



**Katholieke Universiteit Leuven**  
Faculteit Letteren  
Departement Linguïstiek

## **Het naakt(e) adjectief**

*Kwantitatief-empirisch onderzoek naar de adjectivische  
buigingsalternantie bij neutra*

Promotoren:  
Prof. dr. Dirk Geeraerts  
Prof. dr. Dirk Speelman

Proefschrift aangeboden tot  
het behalen van de graad  
van Doctor in de Taalkunde

door  
**José TUMMERS**

2005

*Voor Leen*

# Inhoud

INLEIDING	1
DEEL I LITERAATUURSTUDIE: VOORSTELLING EN SITUERING VAN HET ONDERZOEKSOBJECT EN DE ONDERZOEKSMETHODOLOGIE	7
<b>1 Taalkundige traditie op het gebied van de buigingsalternantie bij neutra</b>	<b>9</b>
1.1 Condities op de keuze van het adjectivische buigingsmorfeem	10
1.1.1 Morfosyntactische kenmerken van de elementen in de NP	10
1.1.2 Fonologische kenmerken van het adjectief en het substantief	13
1.1.2.1 Segmentele fonologie	13
1.1.2.2 Prosodische fonologie	13
1.1.3 Lexicale en semantische kenmerken van het adjectief-nomenpaar	16
1.1.3.1 Idiomaticiteit van het adjectief-nomenpaar	16
1.1.3.2 Een systematisch betekenisverschil?	18
1.1.4 Nationale variëteit van het Nederlands	20
1.1.5 Stilistische waarde van de onverbogen vorm	21
1.1.6 Syntactische factoren	21
1.2 Bespreking en evaluatie	22
1.2.1 Hypothesen met betrekking tot de conditionering van de adjectivische buigingsalternantie bij neutra	22
1.2.2 Situering en verantwoording van het onderzoek naar de adjectivische buigingsalternantie bij neutra	25
1.3 Conclusie	29
Noten	31
<b>2 Historische situering van de adjectivische buigingsalternantie</b>	<b>35</b>
2.1 De adnominale buiging in het Indo-Europees en het Oer-Germaans	36
2.2 De adnominale buiging in het Oud- en Middel-Nederlands	37
2.3 De adnominale buiging in het Nieuw-Nederlands	41
2.3.1 Het Nederlands in de zestiende eeuw en de divergentie tussen de zuidelijke en noordelijke dialecten	42
2.3.2 Het Nederlands in de zeventiende eeuw	45
2.3.3 Het Nederlands in de achttiende eeuw	46
2.3.4 Het Nederlands in de negentiende eeuw	47
2.4 Conclusie	47
Noten	49

<b>3 Verbuiging van het attributieve adjectief in het Nederlands</b>	<b>51</b>
3.1 Basismechanisme voor de verbuiging van het geanteponeerde attributieve adjectief	51
3.2 Systematische uitzonderingen op het basismechanisme	52
3.3 Contexten met buigingsvariatie	55
Noten	59
<b>4 Congruentie</b>	<b>61</b>
4.1 Congruentie als morfologisch fenomeen	61
4.2 Definitie en functie van congruentie	64
4.2.1 Definitie congruentie	64
4.2.2 Functies van congruentie	67
4.3 Conclusie	71
Noten	73
<b>5 Methodologische facetten</b>	<b>75</b>
5.1 Motivatie corpusonderzoek	75
5.2 Methodologische fasering	78
5.2.1 Selectie van het materiaal	78
5.2.2 Analyse van het materiaal	78
5.2.3 Samenvatting en interpretatie van de gegevens	79
5.3 Situering van het onderzoek in de corpuslinguïstische traditie	81
5.4 Conclusie	82
Noten	85
<b>DEEL II   CONDITIONERING VAN DE BUIGINGSALTERNANTIE:           OPERATIONALISERING EN KWANTIFICERING</b>	<b>87</b>
<b>6 Samenstelling en analyse van het <i>Corpus Gesproken Nederlands</i></b>	<b>91</b>
6.1 Samenstelling van het <i>Corpus Gesproken Nederlands</i>	91
6.2 Analyse van het <i>Corpus Gesproken Nederlands</i>	93
6.2.1 Zoekopdracht	93
6.2.2 Controle gegevensbank	96
6.3 Resulterende materiaalverzameling	100
Noten	101
<b>7 Bivariate statistiek en stratumanalyse</b>	<b>103</b>
7.1 Bivariate statistiek	103
7.1.1 Technieken voor het bivariaat onderzoek: $\chi^2$ -statistiek, Fisher Exact-test, relatief risico en odds ratio	103

7.1.2	Een fictief voorbeeld: de invloed van de sekse van de spreker op de buigingsalternantie	107
7.2	Stratumanalyse	109
7.2.1	Technieken voor stratumanalyses: <i>Cochran-Mantel-Haenszel</i> -statistiek, <i>Breslow-Day</i> -test en stratumtabellen	109
7.2.2	Het fictieve voorbeeld herbekeken	111
	Noten	114
<b>8</b>	<b>Morfosyntactische conditionering</b>	<b>117</b>
8.1	Woordsoort van de determinator	117
8.1.1	Hypotheses en operationalisering	117
8.1.2	Resultaten en bespreking	118
8.2	Gradatie van het adjectief	121
8.2.1	Hypothese en operationalisering	121
8.2.2	Resultaten en bespreking	122
8.3	Diminuering van het substantief	123
8.3.1	Hypothese en operationalisering	123
8.3.2	Resultaten en bespreking	123
8.4	Woordsoortelijke subclassificatie van het element in de kernpositie	125
8.4.1	Hypotheses en operationalisering	125
8.4.2	Resultaten en bespreking	125
8.5	Bigenericiteit van het substantief	127
8.5.1	Hypothese en operationalisering	127
8.5.2	Resultaten en bespreking	127
8.6	Conclusie	128
	Noten	130
<b>9</b>	<b>Fonologische conditionering I – segmentele fonologie</b>	<b>133</b>
9.1	Hypotheses en operationalisering	133
9.2	Resultaten en bespreking	134
9.2.1	Substantivisch beginsegment	134
9.2.2	Adjectivisch eindsegment	135
9.3	Conclusie	136
	Noten	137
<b>10</b>	<b>Fonologische conditionering II – prosodische fonologie</b>	<b>139</b>
10.1	Hypotheses en operationalisering	139
10.1.1	Onderzoeksdoelstellingen en onderzoekskader	139
10.1.2	Operationalisering: codering van de prosodische patronen	146
10.2	Resultaten en bespreking	153
10.2.1	Falsificatie van de ritmische en eufonische hypothesen	153
10.2.2	Prosodische conditionering van de buigingsalternantie	159

INHOUD

10.2.3 Partiële conditionering van prosodische factoren	169
10.3 Conclusie	173
Noten	176
<b>11 Lexicale/semantische conditionering</b>	<b>181</b>
11.1 Hypothesen en operationalisering	182
11.1.1 Lexicale/semantische kenmerken van het adjectief-nomenpaar	182
11.1.1.1 Problemen bij de identificatie van lexicale idiomen	183
11.1.1.2 Een alternatieve benadering: de lexicale collocabiliteit van het adjectief-nomenpaar	185
11.1.1.3 Lexicale collocaties als operationalisering van de lexicale/semantische kenmerken van het adjectief-nomenpaar	193
11.1.2 Relationaliteit van het adjectief	194
11.2 Resultaten en bespreking	197
11.2.1 Lexicale/semantische conditionering van het adjectief-nomenpaar	197
11.2.2 Relationaliteit van het adjectief	204
11.2.3 Stratumanalyse: lexicale/semantische conditionering van prosodische patronen	209
11.3 Conclusie	213
Noten	217
<b>12 Lectische conditionering</b>	<b>221</b>
12.1 Lectische betekenisaspecten	221
12.2 Hypothesen en operationalisering	223
12.2.1 Hypothesen	223
12.2.2 Operationalisering	227
12.2.2.1 Operationalisering van de geografische differentiëring	227
12.2.2.2 Operationalisering van de stilistische differentiëring	229
12.3 Resultaten en bespreking	233
12.3.1 Geografische differentiëring	233
12.3.2 Stilistische differentiëring	237
12.3.3 Stilistische conditionering van taalinterne factoren	246
12.3.3.1 Woordsoort determinant	247
12.3.3.2 Fonologische factoren	248
12.3.3.3 Lexicale/semantische kenmerken van het adjectief en het adjectief-nomenpaar	250
12.4 Conclusie	252
Noten	255

DEEL III	MODELLERING VAN DE BUIGINGSALTERNANTIE: INTERACTIE EN COMPETITIE	261
<b>13</b>	<b>Multivariate statistiek: logistische regressie</b>	<b>263</b>
13.1	Lineaire regressie	263
13.2	Logistische regressie	265
13.2.1	Constructie logistische-regressiemodel	265
13.2.2	Interpretatie logistische-regressiemodel	269
	Noten	271
<b>14</b>	<b>Multivariate analyse van de adjectivische buigingsalternantie bij neutra</b>	<b>273</b>
14.1	Parametrisering van de verklarende factoren	273
14.2	Impact van de verklarende factoren: <i>forward stepwise</i> -regressie	275
14.3	Logistische-regressiemodel voor de adjectivische buigingsalternantie bij neutra	278
14.3.1	Modelstatistieken	278
14.3.2	Regressiecoëfficiënten: effect van de factorwaarden op de buigingsalternantie	279
14.3.3	Voorspellende kracht van de factorwaardencombinaties	293
14.3.4	Een apart model voor het Belgisch Nederlands?	296
14.4	Algemene bespreking	301
14.5	Conclusie	308
	Noten	312
DEEL IV	LEXICALE EN SEMANTISCHE ANALYSE: SEMANTISCHE CONDITIONERING VAN DE BUIGINGSALTERNANTIE	321
<b>15</b>	<b>Voorstelling en analyse van het <i>De Standaard</i>-corpus</b>	<b>323</b>
15.1	Voorstelling van het <i>De Standaard</i> -corpus	323
15.2	Ontsluiting en analyse van het <i>De Standaard</i> -corpus	324
15.2.1	Compilatie van het deelcorpus met twee- en drieplaatsige NP's	325
15.2.2	Compilatie van de materiaalverzameling met de buigingsalternantie	330
15.3	Resulterende materiaalverzameling	330
	Noten	333
<b>16</b>	<b>Lexicale conditionering</b>	<b>335</b>
16.1	Overzichtsstatistieken voor de lexicale collocabiliteit van de adjectief-nomenparen in het <i>De Standaard</i> -corpus	335
16.2	Invloed van de lexicale collocabiliteit van het adjectief-nomenpaar op de buigingsalternantie	337

INHOUD

16.3	Lexicale collocabiliteit en medium: het buigingsverschil tussen gesproken en geschreven taal als een lexicaal neveneffect	339
16.4	Conclusie	342
	Noten	344
<b>17</b>	<b>Semantische conditionering van de adjectivische buigingsalternantie bij lexicale collocaties</b>	<b>347</b>
17.1	Semantische restricties op de buigingsalternantie in lexicale collocaties	347
17.1.1	Hypothesen en operationalisering van de semantische factoren	348
17.1.2	Resultaten en bespreking	353
17.1.2.1	Bivariate analyses	353
17.1.2.2	Multivariate analyse	357
17.1.2.3	Bespreking van de semantische conditionering van de buigingsalternantie	360
17.2	Relatie tussen de semantische condities en de lexicale collocaties	361
17.2.1	Hypothesen en operationalisering	361
17.2.2	Resultaten en bespreking	365
17.2.2.1	Externe controletest	365
17.2.2.2	Interne controletest	367
17.2.2.3	Bespreking van de semantische tendensen aan de basis van de lexicale collocaties	368
17.3	Algemene bespreking	368
17.4	Conclusie	376
	Noten	379
	CONCLUSIE	385
	Literatuur	393



## INLEIDING

De verbuiging van het attributieve adjectief wordt in de meeste grammatica's en morfologische handboeken van het Nederlands behandeld. Het merendeel van deze beschrijvingen blijft echter beperkt tot de basisregel volgens dewelke het adjectief in attributieve positie altijd verbogen wordt, behalve in een indefiniete NP met een enkelvoudig *het*-woord als hoofd (*een sterk paard*). De talige realiteit is echter complexer dan deze regel laat verstaan: de adjectivische buiging vertoont in verschillende syntactische contexten een alternantie tussen de 'regelmatige' verbogen vorm en de onverbogen vorm van het adjectief. In dit proefschrift concentreren we ons op één welbepaalde context waar beide adjectivische vormen alterneren, met name de definiete naamwoordelijke constituent met een enkelvoudig *het*-woord als hoofd. De ondertitel van dit proefwerk, *de adjectivische buigingsalternantie bij neutra*, specificeren we dan ook tot *de adjectivische buigingsalternantie in een definiete NP met een enkelvoudig het-woord als hoofd*. Deze buigingsalternantie wordt geïllustreerd door de voorbeelden in (1):

- (1) het Openbaar(e) Ministerie; het stoffelijk(e) overschot; het vergelijkend(e) examen; het bruin(e) brood; het vriendelijker(e) kind; het prachtig(e) uniform; het continu(e) schreeuwen; het bruin(e) paard

Deze voorbeelden geven een idee van de complexiteit van de bestudeerde buigingsalternantie. Ten eerste, de kans dat het onverbogen adjectief wordt geselecteerd, daalt naarmate we van links naar rechts opschuiven in het rijtje in (1). Bovendien daalt de aannemelijkheid van de onverbogen variant wellicht sneller voor een Nederlander dan voor een Vlaming. Ten tweede, het is moeilijk – zo niet onmogelijk – om een duidelijke breuklijn te trekken tussen de voorbeelden waar de onverbogen variant gebruikt *moet* worden en de voorbeelden waar de verbogen variant gebruikt *moet* worden. Dit toont aan dat de bestudeerde buigingsalternantie niet discreet maar gradueel van aard is. Ten derde, het is eveneens moeilijk – zo niet onmogelijk – om het gebruik van beide buigingsvarianten door middel van één schematisch principe te beschrijven. Dit blijkt duidelijk uit de analyses die een globaal beeld van de buigingsalternantie willen schetsen (ANS 1997; Booij 1992; Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a). Deze studies vermelden een rist uiteenlopende factoren om het gebruik van beide buigingsvarianten te motiveren. Deze factoren behoren daarenboven tot diverse taalkundige domeinen, die in traditioneel taalkundig onderzoek zelden gecombineerd worden: de morfosyntaxis, de fonologie (zowel segmenteel als suprasegmenteel), de semantiek en de stilistische en geografische variëteit van het Nederlands.

De afwisseling tussen de verbogen en de onverbogen vorm van het adjectief in de voorbeelden in (1) hangt samen met een netwerk van factoren dat complexer is dan sommige (introspectieve) analyses laten uitschijnen. Dat de keuze van het adjectivische buigingsmorfeem voor de taalgebruiker niet altijd evident is, blijkt uit de veelvuldige behandeling van dit onderwerp in werken over taalzuivering (*inter al.* Cockx 2000; Klein & Vischer 1996; Penninckx & Buyse 1997; Permentier & Van Den Eynden 1998; Verhasselt 1974). Nochtans bestaat er – althans voor zover wij weten – geen recent onderzoek dat zich tot taak stelt dit kluwen van factoren te ontwarren. Bestaande analyses van de adjectivische buigingsalternantie bij neutra zijn beperkt tot één factor of bestuderen het afzonderlijke effect van de verschillende potentiële factoren zonder ze in een globaal model te integreren om zo hun cumulatieve impact op de keuze van het buigingsmorfeem te meten. Dit is precies wat wij in dit proefschrift willen doen: het modelleren van de adjectivische buigingsalternantie bij neutra, zoals ze in het standaard (gesproken) Nederlands voorkomt.

Voorliggend proefschrift heeft een dubbele doelstelling: de eerste is descriptief, de tweede is methodologisch.

Vanuit een descriptief standpunt willen we de adjectivische buigingsalternantie bij neutra in kaart brengen, zoals deze in het reële taalgebruik voorkomt. Deze algemene descriptieve doelstelling kunnen we als volgt verfijnen. In de eerste plaats willen we de condities identificeren die de keuze tussen het verbogen en het onverbogen adjectief bij neutra motiveren. De identificatie van deze condities verloopt hoofdzakelijk via het falsificeren van hypothesen die we op basis van bestaand onderzoek formuleren. We beperken ons echter niet tot de verificatie van de hypothesen met betrekking tot de afzonderlijke condities op de keuze van het buigingsmorfeem, zoals dat in bestaand corpusonderzoek het geval is, maar daarnaast willen we het complexe netwerk met deze condities ontwarren.

We zijn van oordeel dat we deze descriptieve doelstelling het best kunnen realiseren middels kwantitatief-empirisch corpusonderzoek. We benadrukken doorheen het proefschrift dan ook meermaals de descriptieve meerwaarde en – naar onze mening – de noodzaak van kwantitatief-empirisch onderzoek om een complex taalkundig verschijnsel, zoals de buigingsalternantie bij neutra, te beschrijven en te verklaren. Bij het empirisch onderzoek streven we naar een maximale objectiveerbaarheid/intersubjectiviteit, waarbij we anekdotische en impressio-nistische bespiegelingen tot een absoluut minimum beperken. Concreet betekent dit dat we voor de analyse van iedere conditie op de buigingsalternantie een repliceerbare procedure voorstellen om de observaties in de materiaalverzameling te analyseren. Als empirische basis voor het onderzoek doen we een beroep op twee corpora, met name het *Corpus Gesproken Nederlands* en de jaargang 2003 van de Vlaamse kwaliteitskrant *De Standaard*.

De combinatie van de descriptieve en methodologische doelstellingen mondt uit in de constructie van een globaal model voor de beschrijving en de verklaring van de adjectivische buigingsalternantie bij neutra in het standaard (gesproken) Nederlands. Dit model presenteert een algemeen beeld van de competitie en de interactie tussen de diverse restricties op de keuze van het adjectivische buigingsmorfeem: het model identificeert immers niet alleen de significante factoren, maar biedt daarnaast ook de mogelijkheid om een rangorde op te stellen voor (de impact van) deze factoren.

Rekening houdend met deze descriptieve en methodologische hoofddoelstelling schrijven we ons *a priori* in geen enkele taalkundige theorie in, maar maken we op eclectische wijze gebruik van de inzichten uit verschillende taaltheoretische modellen bij de analyse van de empirische gegevens en bij de interpretatie van de resultaten. Door de corpuslinguïstische methodologie hebben we uiteraard een grote affiniteit met theoretische modellen die uitgaan van het reële taalgebruik, zoals sommige gebruiksgebaseerde (*usage based*) strekkingen binnen de Constructiegrammatica.

We sluiten de algemene inleiding af met het voorstellen van de globale structuur van onderhavig onderzoek. Dit proefschrift bestaat uit vier delen: het eerste deel is een literatuurstudie; de overige delen zijn de drie opeenvolgende fases van het empirisch onderzoek.

Deel I is een literatuurstudie om het onderzoeksobject en de onderzoeksmethodologie voor te stellen en in een ruimer taalkundig kader te plaatsen. Met betrekking tot het onderzoeksobject stellen we de *status quaestionis* voor waaruit we de hypothesen afleiden die het uitgangspunt van de empirische analyse vormen. Daarnaast situeren we de adjectivische buigingsalternantie bij neutra in een ruimer diachroon en morfosyntactisch kader. Tot slot relateren we de buigingsalternantie aan algemene theoretische inzichten over het fenomeen congruentie. Het laatste hoofdstuk van het eerste deel behandelt de methodologische facetten van het kwantitatief-empirische corpusonderzoek: we verantwoorden het gebruik van corpusmateriaal en belichten de fasering van het empirisch onderzoek en de repercussies ervan op de argumentatie.

De delen II, III en IV stellen het eigenlijke empirisch onderzoek voor. In deel II en deel III, die een methodologisch tweeluik vormen, presenteren we het basisonderzoek naar de modellering van de buigingsalternantie bij neutra in het gesproken standaard Nederlands, zoals deze in het *Corpus Gesproken Nederlands* voorkomt. In deel IV concentreren we ons op één type factor, met name de lexicale en semantische conditionering van de buigingsalternantie. Doordat het *Corpus Gesproken Nederlands* om kwantitatieve (dataschaarste) en kwalitatieve (c.q. de lectische en thematische diversiteit) redenen minder geschikt is voor dit onderdeel van het onderzoek, maken we gebruik van de Vlaamse kwaliteitskrant *De Standaard*.

In deel II identificeren we de afzonderlijke condities op het gebruik van beide adjectivische buigingsvarianten. Deze analyse komt in grote mate neer op het testen van hypothesen die we uit de literatuur hebben afgeleid. In dit deel ligt de nadruk op de operationalisering en de interpretatie van de verschillende verklarende factoren. Dit betekent dat we voor iedere verklarende factor, c.q. hypothese, op zoek gaan naar een repliceerbare procedure en een optimale parametrisering, zowel met het oog op het maximaliseren van het kwantitatieve effect als met het oog op de taalkundige interpretatie. De verschillende hoofdstukken in dit deel groeperen de factoren per taalkundig domein (zie tabel 1).

Het onderzoek van de afzonderlijke condities op de buigingsalternantie in deel II vormt, zoals hierboven is aangegeven, het eerste deel in een methodologisch tweeluik. Na de operationalisering, de kwantificering en de interpretatie van de afzonderlijke verklarende factoren in deel II, gebruiken we deze informatie in deel III om een globaal statistisch model te construeren dat de verschillende factoren integreert. Rekening houdend met de complexiteit van de buigingsalternantie, met een grote diversiteit aan verklarende factoren, en met de ongecontroleerde natuur van het spontane taalgebruik in een corpus, waardoor deze factoren op een ongecontroleerde manier interageren en concurreren, construeren we een statistisch model om het netwerk met de verklarende factoren te ontwarren. Een statistisch model biedt bovendien de mogelijkheid om de descriptieve adequaatheid van het voorgestelde model te evalueren. Anders gezegd: aan het einde van deel III kunnen we ons een beeld vormen van de modelleerbaarheid van de buigingsalternantie en van de beschrijvende en verklarende kwaliteit van het gevoerde onderzoek.

In deel IV gaan we dieper in op het buigingseffect van één type verklarende factor, met name de lexicale en semantische conditionering van de buigingsalternantie. Terwijl deze condities in de introspectieve literatuur uitgebreid aan bod komen, is hun impact tot op heden niet empirisch getoetst – omwille van de problemen die gepaard gaan met de operationalisering van deze condities. In het basisonderzoek in de delen II en III lossen we dit probleem op door het lexicale en/of semantische effect op louter kwantitatieve gronden te analyseren aan de hand van lexicale collocaties. In dit laatste deel stellen we ons de vraag of er semantische regelmatigigheden aan de grondslag liggen van het buigingseffect van de lexicale patronen, c.q. van de lexicale collocaties, die we in de delen II en III kwantitatief geïdentificeerd hebben.

De structuur van dit proefschrift wordt samengevat in tabel 1. Aangezien de argumentatie in dit proefschrift uit de kwantitatieve analyse van empirische gegevens bestaat, besteden we ruim aandacht aan de voorstelling en de analyse van de geraadpleegde corpora, zodat de lezer zich een beeld kan vormen van de empirische basis van dit onderzoek, en aan de gebruikte statistische maten, zodat de lezer die minder vertrouwd is met de gebruikte technieken de kwantitatieve argumentatie kan volgen. Zowel in deel II (hoofdstuk 7) als in deel III (hoofdstuk

13) is een apart hoofdstuk voorzien waarin de gebruikte statistische technieken worden toegelicht. De statistische aspecten die niet essentieel zijn voor een goed begrip van de argumentatie rapporteren we in noten. Zoals uit de overzichtstabel op de volgende pagina kan worden afgeleid, hebben we ervoor geopteerd om ieder aspect in een afgerond hoofdstuk te behandelen.

<b>DEEL I</b> <b>LITERATUURSTUDIE</b>	<b>DEEL II</b> <b>CONDITIONERING VAN DE</b> <b>BUIGINGSALTERNANTIE</b>	<b>DEEL III</b> <b>MODELLERING VAN DE</b> <b>BUIGINGSALTERNANTIE</b>	<b>DEEL IV</b> <b>LEXICALE EN SEMANTISCHE</b> <b>ANALYSE</b>
1. Taalkundige traditie	6. Samenstelling en analyse van het <i>Corpus Gesproken Nederlands</i>	13. Multivariate statistiek: logistische regressie	15. Voorstelling en analyse van het <i>De Standaard</i> -corpus
2. Historische situering	7. Bivariate statistiek en stratumanalyse	14. Multivariate analyse van de buigingsalternantie	16. Lexicale conditionering
3. Verbuiging van het attributieve adjectief	8. Morfosyntactische conditionering		17. Semantische conditionering
4. Congruentie	9. Fonologische conditionering I: Segmentele fonologie		
5. Methodologische facetten	10. Fonologische conditionering II: Prosodische fonologie		
	11. Lexicale/semantische conditionering		
	12. Lectische conditionering		

*Tabel 1: Opbouw van het proefschrift*

# DEEL I

## LITERATUURSTUDIE: VOORSTELLING EN SITUERING VAN HET ONDERZOEKSOBJECT EN DE ONDERZOEKSMETHODOLOGIE

Het eerste deel van dit proefschrift is een algemene inleiding tot het onderzoeksobject en de onderzoeksmethodologie. Dit overzicht mondt uit in het formuleren van een reeks concrete hypotheses met betrekking tot het effect van een aantal factoren die in de literatuur worden aangehaald om de buigingsalternantie te verklaren (zie paragraaf 1.2.1, tabel 1(1)). Daarnaast stellen we een aantal algemene onderzoeksvragen voor die samenhangen met de kwantitatief-empirische methodologie en die lacunes in het bestaande onderzoek opvullen (zie paragraaf 5.4). Deze hypotheses en onderzoeksvragen vormen de algemene leidraad bij het empirisch onderzoek in de delen II, III en IV.

Dit deel is op de volgende wijze gestructureerd: de eerste vier hoofdstukken behandelen de buigingsalternantie; het vijfde hoofdstuk presenteert de kwantitatief-empirische onderzoeksmethodologie.

Het eerste hoofdstuk overloopt de literatuur over de buigingsalternantie bij neutra en is in feite het belangrijkste hoofdstuk van dit inleidende deel. Dit verkennende literatuuroverzicht heeft een dubbele doelstelling. In eerste instantie overlopen we de verschillende factoren die in de literatuur worden aangehaald ter verklaring van het gebruik van beide buigingsvarianten. Dit resulteert in een reeks concrete hypotheses die in het empirisch onderzoek gefalsificeerd worden. Daarnaast presenteren we een kritische bespreking van de bestaande literatuur waarbij we verschillende attitudes ten aanzien van de buigingsalternantie identificeren. Deze attitudes worden later met de empirische realiteit van het corpus geconfronteerd (hoofdstuk 14).

De twee volgende hoofdstukken situeren de buigingsalternantie in een ruimer diachroon en taalstructureel kader. In hoofdstuk 2 presenteren we een kort historisch overzicht van de buigingsalternantie bij neutra. Uit deze bespreking blijkt enerzijds dat het attributieve adjectief bij neutra vanaf de vroegste stadia van het Nederlands buigingsvariatie vertoont en dat verscheidene factoren die in het hedendaags

Nederlands het gebruik van de gemarkeerde onverbogen variant motiveren al in vroegere stadia van het Nederlands aanwezig zijn. Hoofdstuk 3 plaatst de buigingsalternantie in het ruimere structurele kader van de buiging van het attributieve adjectief in het Nederlands. De centrale doelstelling van dit hoofdstuk is de identificatie van de contexten waar het adjectief consistent voor één van beide buigingsvarianten kiest en bijgevolg – naar alle waarschijnlijkheid – variatie uitsluit. Voor neerlandici bevatten deze twee hoofdstukken wellicht weinig nieuwe informatie.

Het volgende hoofdstuk verruimt het perspectief van de adjectivische buiging in het Nederlands naar het algemene fenomeen congruentie (hoofdstuk 4). Hierbij willen we vanuit een theorieonafhankelijk gezichtspunt een algemene definitie presenteren van congruentie en van de gangbare visies op congruentie. Een bijzondere aandacht gaat hierbij uit naar de modellen die het voorkomen van variatie bij congruentie als uitgangspunt nemen.

Het laatste hoofdstuk van dit eerste inleidende deel presenteert de gebruikte kwantitatief-empirische onderzoeksmethodologie en situeert deze in het ruimere kader van het corpuslinguïstische onderzoek (hoofdstuk 5). Nadat we de keuze voor corpuslinguïstisch onderzoek gemotiveerd hebben, belichten we de methodologische fasering van het empirisch onderzoek. Concreet betogen we dat de traditionele afzonderlijke analyses van verklarende factoren ontoereikend zijn om een complex taalkundig fenomeen, zoals de adjectivische buigingsalternantie bij neutra, in kaart te brengen. We pleiten dan ook voor een methodologisch tweeluik, waarbij de afzonderlijke analyses van de verklarende factoren (deel II) de eerste stap vormen naar de constructie van een geïntegreerd (statistisch) model dat het effect van de verschillende condities op de buigingsalternantie in hun onderlinge samenhang in kaart brengt (deel III). De constructie van een geïntegreerd model voor het netwerk van condities op de buigingsalternantie biedt tot slot de mogelijkheid om een aantal algemene onderzoeksvragen te formuleren die in traditioneel (introspectief) onderzoek niet gesteld kunnen worden.



# **Hoofdstuk 1:**

## **Taalkundige traditie op het gebied van de buigingsalternantie bij neutra**

De adjectivische buigingsvariatie staat bij neerlandici en professionele taalgebruikers algemeen bekend als een complex en problematisch verschijnsel. Een blik op grammatica's, morfologische handboeken en taalzuiveringswerken volstaat om zich te vergewissen van de complexiteit van de buigingsalternantie, vooral bij neutra. De meeste beschrijvingen raken echter niet verder dan een algemene probleemschets aangevuld met de vermelding van enkele tendenties die de keuze tussen beide vormvarianten beïnvloeden. Typerend is het veelvuldige gebruik van modaliserende elementen zoals 'vaak', 'min of meer' of 'algemeen', wat duidelijk het graduele karakter van de alternantie aantoont en de moeilijkheid om ze in enkele abstracte regels samen te vatten. Ondanks de vaststelling dat het een complex en problematisch verschijnsel betreft en ondanks de aandacht die de buigingsalternantie in taalzuiveringswerken geniet, zijn diepgaande analyses eerder zeldzaam, zoals in dit hoofdstuk duidelijk wordt.

Dit verkennende literatuuroverzicht heeft een dubbele doelstelling. In de eerste plaats willen we aan de hand van bestaand onderzoek de restricties definiëren die het gebruik van beide buigingsvarianten bepalen. We zijn meer bepaald geïnteresseerd in de factoren die de selectie van de gemarkeerde onverbogen vorm conditioneren. Het overzicht van deze condities resulteert in het formuleren van een reeks hypothesen met betrekking tot de selectie van beide buigingsvarianten. De tweede doelstelling van dit literatuuroverzicht is het algemene beeld schetsen dat ten overstaan van de buigingsalternantie uit de vakliteratuur oprijst. Op basis van een kritische bespreking van het bestaande onderzoek plaatsen we tot slot de eigen analyse in het ruimere kader van het neerlandistische onderzoek naar de adjectivische buiging(s-alternantie).

Dit hoofdstuk bestaat uit twee delen. Het eerste deel biedt een overzicht aan van de verschillende factoren die in de literatuur worden aangehaald om de adjectivische buigingsalternantie bij neutra te beschrijven (1.1). Om een schier eindeloze herhaling van analoge visies te vermijden, hebben we ervoor geopteerd om het literatuuroverzicht in eerste instantie rond de verklarende factoren te organiseren. Deze indeling biedt de mogelijkheid om de verschillende visies op éénzelfde factor met elkaar te confronteren. Het tweede deel is een kritische evaluatie waar enerzijds de hypothesen worden geformuleerd die de basis vormen van het empirisch onderzoek en waar anderzijds de verschillende attitudes in de neerlandistiek ten aanzien van de buigingsalternantie worden besproken (1.2).

## 1.1 Conditie op de keuze van het adjectivische buigingsmorfeem

Het overzicht van de verklarende factoren die in de literatuur worden aangereikt, is ingedeeld in functie van de verschillende domeinen die traditioneel in de taalkunde onderscheiden worden. We bespreken achtereenvolgens de volgende types factoren: de morfosyntactische kenmerken van de elementen in de NP (1.1.1), de fonologische kenmerken van het adjectief en het substantief (1.1.2), de lexicale en semantische kenmerken van het adjectief-nomenpaar (1.1.3), de nationale variëteit van het Nederlands (1.1.4), het register (1.1.5) en de syntactische structuur van de NP (1.1.6).

Als uitgangspunt voor het literatuuroverzicht raadplegen we de schaarse globale beschrijvingen van de adjectivische buigingsalternantie in een definiëte NP met een enkelvoudig onzijdig substantief als hoofd. Naast de ANS (1997: 401-412) betreft het enkele descriptieve (Raidt 1968; Royen 1953) en corpuslinguïstische studies (Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a). Deze analyses worden aangevuld en verfijnd met gevalstudies van afzonderlijke factoren en beschrijvingen in grammatica's en morfologische handboeken.

De doelstelling van deze paragraaf is het schetsen van een algemeen overzicht van de factoren die in de literatuur naar voor worden geschoven om de keuze van het adjectivische buigingsmorfeem bij neutra te verklaren. Dit betekent dat we – voorlopig – niet stilstaan bij de operationalisering of de empirische analyse en analyseerbaarheid van deze factoren. Deze problematiek wordt uitgebreid behandeld in het tweede deel van dit proefschrift (zie deel II).

### 1.1.1 Morfosyntactische kenmerken van de elementen in de NP

We becommentariëren achtereenvolgens de invloed op de buigingsalternantie van de morfosyntactische realisatie van de elementen die de verschillende posities binnen de NP bezetten, met name de determinatorpositie, de adjectivische positie en de nominale kernpositie.

We kijken eerst naar de elementen in de determinatorpositie. Bij een onzijdig substantief als kern wordt na een definiëte determinator in de regel de verbogen vorm van het adjectief geselecteerd (zie paragraaf 3.1, figuur 3(1)). Na de volgende determinatoren kan er buigingsvariatie optreden:<sup>1</sup>

- het bepaald lidwoord *het*
- de aanwijzende voornaamwoorden *dit* en *dat*
- de bezittelijke voornaamwoorden

Synchroon worden deze elementen gewoonlijk als een homogene groep behandeld: sommige analyses construeren uitsluitend voorbeelden met het bepaald lidwoord in determinatorpositie (*inter al.* Booij 1992; Booij & van Santen 1998; Broekhuis 1999); andere analyses plaatsen de drie groepen op één lijn zonder na te gaan of ze een gedifferentieerd effect op de buigingsalternantie hebben (*inter al.* de Haas & Trommelen 1993; Booij 2002c; Schultink 1962).

In het historisch overzicht van de adjectivische buiging zullen we echter een afwijkend congruentiegedrag vaststellen na een bezittelijk voornaamwoord (zie paragraaf 2.2): vanaf het Middel-Nederlands komt na een bezittelijk voornaamwoord het onverbogen adjectief frequent voor. Het gebruik van de onverbogen vorm in deze context is waarschijnlijk ontstaan als gevolg van de assimilatie tussen de adjectivische buigingsvorm en het voorafgaande onverbogen bezittelijk voornaamwoord.<sup>2</sup> Als gevolg van de geleidelijke verspreiding van de verbogen vorm in de omgangstaal is de keuze van de buigingsvorm na een bezittelijk voornaamwoord tijdens de achttiende en de negentiende eeuw het onderwerp geweest van een intens grammaticaal debat (zie paragrafen 2.3.3 en 2.3.4). Tot halverwege de twintigste eeuw blijven grammatici het gebruik van de onverbogen vorm voorschrijven.<sup>3</sup>

Verschillende analyses van de adjectivische verbuiging merken op dat de onverbogen vorm vaker voorkomt na een bezittelijk voornaamwoord dan na een bepaald lidwoord of een aanwijzend voornaamwoord (ANS 1997: 411; Degryse 1959: 73-84; Klein & Visscher 1996: 71; Lebrun & Schurmans-Swillen 1966: 183; Raidt 1968: 100; de Rooij 1980a: 14, 16-17, 20-21; Royen 1953: 103-104). Dat de laatste twee groepen samen horen, is niet verrassend, aangezien het bepaald lidwoord tijdens het Oud-Nederlands uit het aanwijzend voornaamwoord ontwikkeld is (zie paragraaf 2.2). Over de concrete invloed van het bezittelijk voornaamwoord bestaat er geen eenstemmigheid: sommige onderzoekers oordelen dat de onverbogen vorm de dominante vorm is na een bezittelijk voornaamwoord (Raidt 1968; Royen 1953: 103), terwijl anderen zich beperken tot de opmerking dat er in deze context enkel sprake is van een stijging van het aandeel onverbogen vormen (ANS 1997; Klein & Visscher 1996; de Rooij 1980a). Bestaand empirisch onderzoek geeft geen uitsluitsel over het al dan niet dominante statuut van de onverbogen vorm na een bezittelijk voornaamwoord. In de Belgische en Nederlandse literaire teksten van omstreeks 1950 die Degryse (1959: 82) geëxcerpeerd heeft, komt de onverbogen variant respectievelijk in gemiddeld 90% en 70% van de gevallen voor (tegen 47% en 9.5% na een andere determinator).<sup>4</sup> Haar cijfers voor het Belgische materiaal worden bevestigd door de analyse van Lebrun & Schurmans-Swillen (1966: 183): zij tellen in Belgische dagbladen uit 1964 89.9% onverbogen vormen na een bezittelijk voornaamwoord (tegen 42.7% na een andere determinator). Uit het meest recente empirisch onderzoek blijkt daarentegen enkel een significant hogere frequentie van de onverbogen variant na een bezittelijk voornaamwoord: De Rooij (1980a: 17) telt in een corpus dat geschreven en gesproken taalmateriaal<sup>5</sup> uit 1974 en 1979 omvat slechts 34.7% onverbogen vormen na een bezittelijk voornaamwoord (tegen 16.8% na een andere determinator).

Merk op dat het verschil tussen een dominant gebruik van de onverbogen vorm en een grotere frequentie van de onverbogen vorm samenvalt met de datering van het geraadpleegde materiaal: enkel de minst recente studies wijzen op een absolute voorkeur voor de onverbogen vorm na een bezittelijk voornaamwoord. Een analoge temporele breuk vinden we in het overzicht van twintigste-eeuwse grammatica's bij

De Rooij (1980a: 6-7), waaruit blijkt dat vooral grammatica's in de eerste helft van de twintigste eeuw het onverbogen adjectief voorschrijven in navolging van de traditie in de achttiende en negentiende eeuw.<sup>6</sup> Latere grammatica's spreken slechts over een hogere frequentie van de onverbogen vorm na een bezittelijk voornaamwoord. We kunnen dus aannemen dat het literaire canon in 1950 en de dagbladders in 1964 wellicht nog sterk beïnvloed werden door de 'oudere' norm, terwijl de materiaalverzameling van De Rooij, die uit 1974 en 1979 dateert en naast geschreven materiaal ook spreektaalig materiaal bevat, 'moderner' is en aansluit bij de hedendaagse grammatica's die alle definiëte determinatoren als een homogene groep beschouwen met betrekking tot de keuze van het adjectivische buigingsmorfeem.

Het bezittelijk voornaamwoord interfereert tot slot met andere factoren. In de eerste plaats is het gebruik van de onverbogen vorm na een bezittelijk voornaamwoord volgens sommige auteurs beperkt tot bepaalde registers. De meningen hierover zijn echter niet eensluidend: Raidt (1968: 100) oordeelt dat de selectie van de onverbogen buigingsvariant na het bezittelijk voornaamwoord vooral in de informele omgangstaal voorkomt, terwijl De Rooij (1980a: 6-7) na een bespreking van twintigste-eeuwse grammatica's concludeert dat het gebruik van de onverbogen vorm na een bezittelijk voornaamwoord kenmerkend is voor de ambtelijke en literaire taal. De ANS (1997: 411) wijst op een andere interferentie, meer bepaald met de prosodische structuur van het adjectief: het gebruik van de onverbogen variant na een bezittelijk voornaamwoord is een versterking van het prosodische effect dat zich voordoet bij adjectieven die op *-ig* of *-lijk* eindigen (zie paragraaf 1.1.2.2).<sup>7</sup> Raidt (1968: 100) stelt deze interferentie vast in de gevallen waar het adjectief een comparatief is.

Na de bespreking van het element in de determinatorpositie richten we nu onze aandacht op de inherente inflectie van het adjectief. Bij de vormen van de vergrotende trap wordt vaker het onverbogen buigingsmorfeem geselecteerd dan bij de vormen van de stellende trap. De reden waarom de onverbogen variant aannemelijker is bij de vormen van de stellende trap in (1a) dan bij de overeenkomstige vormen van de stellende trap in (1b) is prosodisch. We komen hier dan ook op terug bij de discussie van de prosodische structuur van het adjectief (zie paragraaf 1.1.2.2).

- (1) a. het groter(e) boek; het vriendelijker(e) kind  
b. het grot(e) boek; het vriendelijk(e) kind

Bij het overzicht van de condities die, ongeacht de syntactische configuratie van de NP, consistent één adjectivische buigingsvariant selecteren, zullen we vaststellen dat de vormen van de overtreffende trap in attributieve positie altijd verbogen worden (zie paragrafen 3.2 en 6.2.2).

Tot slot vermelden we de stelling van Raidt (1968: 101) dat een verkleinwoord in de nominale kernpositie het gebruik van de onverbogen adjectivische vorm vergemakkelijkt.

### ***1.1.2 Fonologische kenmerken van het adjectief en het substantief***

De fonologische condities op de buigingsalternantie vallen uiteen in twee groepen, namelijk de restrictie die tot de segmentele fonologie (1.1.2.1) behoort en de condities die tot de suprasegmentele of prosodische fonologie behoren (1.1.2.2).

#### *1.1.2.1 Segmentele fonologie*

De invloed van de beginklank van het substantief is voornamelijk aan bod gekomen in het empirisch onderzoek naar de buigingsalternantie (Degryse 1959: 74; Lebrun & Schurmans-Swillen 1966: 178-180; de Rooij 1980a: 14, 16-17, 20-21). De veronderstelling is dat de buigings-*e* vaker wordt weggelaten vóór een substantief dat met een klinker begint dan vóór een substantief dat met een medeklinker begint. Dit is een instantie van *ease of articulation*: het gebruik van de onverbogen buigingsvariant is articulatorisch gemotiveerd omdat op deze wijze een hiaat op de woordgrens tussen het adjectief en het substantief vermeden wordt.

- (2) a. het traditionel(e) echtpaar; het beschadigd(e) ecoduct
- b. het traditionel(e) koppel; het beschadigd(e) viaduct

Het gebruik van het onverbogen adjectief is aannemelijker in (2a) dan in (2b) omdat de buigings-*e* in deze voorbeelden tot een opeenvolging van twee vocalische elementen zou leiden.

De twee empirische studies die de significantie van deze factor testen, geven een tegenstrijdig resultaat: in de analyse van Lebrun & Schurmans-Swillen (1966) heeft deze factor geen significant effect; in het corpus dat De Rooij (1980a) heeft geraadpleegd, definieert deze factor wel een significante impact op de keuze van de adjectivische buigingsvariant.<sup>8</sup> De verklaring voor deze discrepantie is wellicht stilistisch van aard: Booij (1995: 150-151) kenmerkt de deletie van de prevocalische sjwa boven het woordniveau als informeel. De materiaalverzameling van De Rooij (1980a) bevat naast geschreven materiaal ook spreektaalig materiaal met een groter aandeel prevocalische sjwadeleties in het gesproken deelcorpus.<sup>9</sup> Het empirische materiaal van Lebrun & Schurmans-Swillen (1966) bestaat daarentegen uitsluitend uit krantenmateriaal, wat de afwezigheid van een significant effect kan verklaren.

#### *1.1.2.2 Prosodische fonologie*

De prosodische conditionering van de adjectivische buiging staat in de literatuur bekend als de ‘eufonische’ of ‘ritmische’ kenmerken van het adjectief. Dit wil zeggen dat het weglaten van de buigings-*e* omwille van de welluidendheid gebeurt.

Er worden twee soorten welluidendheidsfactoren onderscheiden: de lengte van het adjectief en de accentuering van de adjectivische eindlettergreep.

We kijken eerst naar het effect van de lengtegraad van het adjectief. Een vaak geciteerd argument stelt dat bij meerlettergrepige adjectieven de buigings-*e* kan worden weggelaten (ANS 1997: 411; Booij 1992: 129-130; Blom 1994: 82; Broekhuis 1999: 216; Degryse 1959: 73-84; Lebrun & Schurmans-Swillen 1966: 178-179, 184-186; Raidt 1968: 99-101; de Rooij 1980a: 14, 16-17, 20-21; Royen 1953: 97-113). Hoe *meerlettergrepig* precies moet worden ingevuld, is niet duidelijk. De ANS (1997) beschouwt drielettergrepige adjectieven als benedengrens voor de swja-apocope, terwijl de voorbeelden geciteerd in Rijpma & Schuringa (1978) tweelettergrepige onverbogen comparatieven bevat. Bestaand empirisch onderzoek (Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a) toont aan dat eenlettergrepige adjectieven bijna zonder uitzondering verbogen worden, zeker in Nederland; vanaf tweelettergrepige adjectieven is er een progressieve stijging waarneembaar van het percentage onverbogen vormen.

De lengte van het adjectief wordt vaak gepreciseerd door het formuleren van bijkomende condities op de beklemtoning van de adjectivische eindlettergreep. Zo zijn we bij het tweede prosodische aspect aanbeland: de accentuering van de adjectivische eindlettergreep. Als de stam van het adjectief op één of meerdere onbeklemtoonde lettergrepen eindigt, wordt de onverbogen vorm gebruikt om een opeenvolging van onbeklemtoonde en doffe lettergrepen te vermijden (ANS 1997; De Vooy 1967: 63; Lebrun & Schurmans-Swillen 1966; Raidt 1968; de Rooij 1980a). Het aantal onbeklemtoonde lettergrepen aan de rechterwoordrand van het adjectief dat nodig is om de onverbogen variant te selecteren, is echter onduidelijk. Uit de voorbeelden aangehaald in *inter al.* Booij (1992: 129), Royen (1953: 111) en Cockx (2000: 30) leiden we af dat het aandeel van de onverbogen buigingsvariant stijgt zodra de adjectivische stam op een sequentie van twee onbeklemtoonde lettergrepen eindigt. Permentier & Van Den Eynden (1998) stellen een striktere drempel voor: pas als de buigings-*e* de vierde opeenvolgende doffe lettergreep aan de rechterwoordgrens is, schrijven zij het gebruik van de onverbogen vorm voor. De voorbeelden in (3) tonen aan dat de aanvaardbaarheid van de onverbogen vorm toeneemt naarmate er meer opeenvolgende onbeklemtoonde en doffe lettergrepen aan de rechterwoordgrens van het adjectief zijn.<sup>10</sup>

- (3) het ánder(e) huis; het vólgend(e) werk; het dágelijks(e) woordgebruik; het vertéderend(e) uiterlijk; het overwéldigend(e) succes; het vriéndelijker(e) kind

Omgekeerd is het gebruik van de onverbogen vorm zeldzaam bij adjectieven met een beklemtoonde eindlettergreep (ANS 1997; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a). In de volgende voorbeelden wordt de sjwaloze vorm als gemarkeerd beschouwd ondanks de meerlettergrepigheid van de adjectieven.

- (4) het gewón(e) leven; het enthousiást(e) onthaal; het rancunéuz(e) kind; het ónvervreemdbàr(e) recht

De lengte van het adjectief lijkt voor een deel een neveneffect te zijn van de accentuering van de eindlettergreep. Het is duidelijk dat (sequenties van) onbeklemtoonde lettergrepen vooral in meerlettergrepige adjectieven voorkomen, terwijl eenlettergrepige adjectieven per definitie op een beklemtoonde lettergreep eindigen.

De beklemtoning van de adjectivische eindlettergreep wordt meebepaald door de morfologische structuur van het adjectief. Deze interactie tussen de prosodische en de morfologische structuur wordt in de literatuur weerspiegeld. De meeste grammatica's spreken niet in termen van accentuering, maar laten de onbeklemtoonde eindlettergreep samenvallen met de doffe affixen *-ig*, *-(e)lijk* en *-isch* (5a). Daarnaast vindt men in nagenoeg alle analyses een speciale vermelding voor de vormen van de vergrotende trap (zie ook paragraaf 1.1.1), die per definitie meerlettergrepig zijn en op minstens één doffe lettergreep eindigen (5b).<sup>11</sup>

- (5) a. het scháttig(e) kind; het vegetárisch(e) menu; het aandóenlijk(e) schouwspel; het afstótelijk(e) gezicht  
b. het slúwer(e) beleid; het uitvóeriger(e) verslag; het gemóedelijker(e) gesprek

Bij de adjectieven met een beklemtoonde eindlettergreep is er eveneens interactie met de morfologische structuur. Lebrun & Schurmans-Swillen (1966) en De Rooij (1980a) splitsen de adjectieven met een beklemtoonde eindlettergreep op in de adjectieven met een uitheems klemtoondragend suffix (6a) en de andere adjectieven met een beklemtoonde eindlettergreep (6b).

- (6) a. het actuél(e) debat; het gevariéerd(e) aanbod; het gecomplicéerd(e) probleem; het monogám(e) koppel  
b. het gewón(e) leven; het onbegrénsd(e) vertrouwen

De gerapporteerde resultaten zijn diametraal aan elkaar tegengesteld: terwijl Lebrun & Schurmans-Swillen (1966) een toename van de onverbogen buigingsvariant vaststellen bij adjectieven afgeleid aan de hand van een Frans suffix, vindt De Rooij (1980a) meer verbogen vormen bij de adjectieven afgeleid door middel van een uitheems suffix. Doordat beide analyses de groep uitheemse suffixen verschillend definiëren (Franse suffixen vs. uitheemse suffixen) en vaag zijn over de complementaire groep is het niet duidelijk welk effect precies gemeten wordt: het onderscheid adjectief met uitheems/Frans suffix vs. andere adjectieven, het onderscheid adjectief met uitheems/Frans suffix vs. andere adjectieven met een beklemtoonde eindlettergreep of het onderscheid adjectief met uitheems/Frans suffix vs. adjectieven met een onbeklemtoonde eindlettergreep.

Overdiep (1937: 259) verruimt het prosodische domein van het woord, c.q. het adjectief, naar de fonologische constituent gevormd door de combinatie van het

adjectief en het substantief. Hij beschouwt een substantief dat met een onbeklemtoonde lettergreep begint als een katalysator voor het gebruik van de onverbogen buigingsvariant in meerlettergrepige adjectieven en comparatieven. In (7) zou de selectie van de onverbogen variant leiden tot een opeenvolging van vier doffe lettergrepen tussen de adjectivische en de substantivische hoofdklemtonen.

- (7) het uitvoeriger(e) bericht

Royen (1953) beschouwt de eufonische en ritmische kenmerken van het adjectief als secundair ten opzichte van de semantische conditionering van de buigingsalternantie (zie paragraaf 1.1.3). Daartegenover staat dan weer de enigszins verrassende recurrentie van deze in wezen spreektaalige parameter in de taalzuiveringsliteratuur (*inter al.* Cockx 2000; Klein & Visscher 1996: 71; Penninckx & Buyse 1997: 40; Permentier & Van Den Eynden 1998; Verhasselt 1974).

### ***1.1.3 Lexicale en semantische kenmerken van het adjectief-nomenpaar***

De lexicale en semantische kenmerken van het adjectief-nomenpaar zijn in de literatuur het grondigst bestudeerd. We groeperen de verschillende condities onder twee hoofdingen: de invloed van het idiomatische gehalte van het adjectief-nomenpaar (1.1.3.1) en de vraag naar een systematisch betekenisverschil tussen beide buigingsvarianten (1.1.3.2).

#### ***1.1.3.1 Idiomaticiteit van het adjectief-nomenpaar***

De meest geciteerde factor ter verklaring van het gebruik van de onverbogen buigingsvorm is het argument dat het adjectief en het substantief in zulke gevallen een “min of meer vaste uitdrukking” of een “semantische eenheid” vormen (*inter al.* ANS 1997: 408; Broekhuis 1999: 210; Booij 1992: 128-129, 2002a: 313-319, 2002c: 48-49; Booij & van Santen 1998: 100; De Vooys 1967: 63; Odijk 1992: 198, 201-204; Penninckx & Buyse 1997: 40). Het adjectief-nomenpaar fungeert in zulke gevallen als een naam, waarbij het – doorgaans onverbogen – adjectief een betekenispecialisatie van het substantief teweeg brengt.

- (8) het stoffelijk overschot; het zelfstandig naamwoord; het bijvoeglijk naamwoord; het lijdend voorwerp; het meewerkend voorwerp; het medisch dossier; het dagelijks bestuur; het uitvoerend comité; het O/openbaar M/ministerie; het koninklijk besluit; het Nationaal Theater

De meervoudsvorm van de adjectief-nomenparen in (8) neemt een buigings-*e* (ANS 1997: 408; Broekhuis 1999: 210; Booij 2002c: 48; Odijk 1992: 198):

- (9) de stoffelijke overschotten; de zelfstandige naamwoorden; de bijvoeglijke naamwoorden; de lijdende voorwerpen; de meewerkende voorwerpen; de



medische dossiers; de dagelijkse besturen; de uitvoerende comités; de O/openbare M/ministeries; de koninklijke besluiten

In sommige gevallen resulteert de buigingsalternantie in een minimaal paar. De onverbogen vorm in (10a) merkt de combinatie van het adjectief en het substantief als een lexicale eenheid die een type wetgevende tekst identificeert die door de koning in zijn hoedanigheid van staatshoofd ondertekend moet worden. De verbogen vorm in (10b) verwijst naar een beslissing van de koning in het privéleven.

- (10) a. Infusiepompen, reservoirs en medische hulpmiddelen moeten beantwoorden aan de bepalingen van *het koninklijk besluit* van 18 maart 1999 betreffende de medische hulpmiddelen.
- b. Het *koninklijke besluit* om de prinsjes naar een strenge kostschool te sturen, kon niet op de onvoorwaardelijke steun van alle huisgenoten rekenen.

De betekenis van de adjectief-nomenparen in de voorbeelden in (8) kan niet compositioneel worden afgeleid: de betekenis van het geheel is niet gelijk aan de som van de betekenis van de samenstellende delen (Broekhuis 1999; Odijk 1992). Het samenvoegen van de betekenissen van *stoffelijk* en van *overschot* volstaat niet om de betekenis ‘lijk’ te construeren. Doordat deze elementen als naam fungeren, is herhaaldelijk op de gelijkenissen met samenstellingen gewezen (Booij 2002a: 313-319; De Vooy 1967: 63; Honselaar 1980: 198).

De lijstjes met voorbeelden die in nagenoeg alle analyses worden aangehaald, staven het lexicale en idiosyncratische statuut dat aan de onverbogen variant wordt toegekend. Van der Horst & van der Horst (1999: 314) merken op dat er voor dit type “‘uitzonderingen’ [...] niet zo makkelijk een precieze regel te geven valt.” Dit idiosyncratische statuut verklaart de diverse pogingen tot thematische classificatie. Hier volgt een uittreksel van de klassen die in de literatuur worden voorgesteld:

- “officiële benamingen”: *het hoger onderwijs; het openbaar ministerie; het Nationaal Fonds voor Wetenschappelijk Onderzoek; het koninklijk besluit; het Europees Parlement; het zelfstandig naamwoord; het lijdend voorwerp* (ANS 1997: 408; Cockx 2000: 30; Klein & Visscher 1996: 69; Lebrun & Schurmans-Swillen 1966: 177; de Rooij 1980a: 12-13)
- “namen van instellingen”: *het openbaar ministerie; het Amsterdams Toneel* (Broekhuis 1999: 210)
- “plaatsnamen”: *het Verre Oosten; het Witte Huis* (Lebrun & Schurmans-Swillen 1966: 177; de Rooij 1980a: 12-13)
- “namen van kranten”: *het Utrechts Nieuwsblad; het Algemeen Dagblad* (Broekhuis 1999: 210)
- “namen van firma’s”: *het Antwerps Centrum voor Productiviteitscontrole* (Lebrun & Schurmans-Swillen 1966: 177; de Rooij 1980a: 12-13)

Let wel dat de onderscheiden categorieën niet (duidelijk) gedefinieerd worden en dat de verschillende voorstellen tot alternatieve classificaties leiden.

### 1.1.3.2 Een systematisch betekenisverschil?

Sinds grammatici de buigingsalternantie bestuderen, zijn zij op zoek naar een systematisch betekenisverschil tussen de verbogen en de onverbogen vormvarianten van het adjectief. Lambert Ten Kate formuleert al in de achttiende eeuw een semantisch onderbouwde regel voor de distributie van *een groot man* en *een grote man* (zie paragraaf 2.3.3). Deze semantische traditie zet zich voort tot halfweg de twintigste eeuw. Uit De Rooijs (1980a: 9) grammaticaoverzicht blijkt dat de semantische analyse van Den Hertog (1897) lange tijd is overgenomen: de verbogen vorm “leent zich meer om uitdrukkelijk te onderscheiden”, zoals in “ik ga mijn oude huis verlaten en mijn nieuwe zelf bewonen”, terwijl de onverbogen vorm “het meest geschikt [is] om te karakteriseeren”, zoals in “Hij zou zijn oud kasteel niet licht voor een modern buitengoed ruilen”. Royen (1953: 105) stelt een connotatieve herziening van dit onderscheid voor:

“Naar mijn taalgevoel (al van jaren her) komt na *het, dit, dat* en de possessieven, bij verbuiging een duidelijke waardering, een sprekender affectief element tot uitdrukking dan bij niet-verbuiging. Vergelijk bijv. *Het oude klavier* tegenover *het oud klavier, mijn oude huis* tegenover *mijn oud huis*. [...] Als ik dus zeg: ‘ik ga eens naar mijn zieke paard kijken’, spreekt daar meer affect uit dan wanneer ik zeg: ‘ik ga eens naar mijn ziek paard kijken’, terwijl toch zowel het verbogen als het onverbogen adj. ‘karakteriseert’.”

Als we de door Den Hertog en Royen geciteerde voorbeelden bekijken, kunnen we niet anders dan samen met De Rooij (1980a: 9) besluiten dat “evenmin als de betekenisonderscheiding van Den Hertog zegt die van Royen me iets, althans in verband met het gebruik van de verbogen en onverbogen adjectiefvormen”. Op basis van de afwijzing van bovenstaande semantische analyses wordt de mogelijkheid van een systematische betekenisonderscheiding expliciet verworpen (Raidt 1968; Schultink 1962).

Na meer dan een kwarteeuw semantische windstilte duikt de ambitie om de adjectivische buigingsalternantie met een systematisch betekenisverschil te associëren opnieuw op in de jaren ‘80 en ‘90 van de vorige eeuw (Blom 1994, 1995; Booij 2002a, 2002c; Honselaar 1980; van der Horst 1992, 1995), vooral als een reactie tegen de ongestructureerde opsommingen van adjectief-nomenidiomen of adjectief-nomenparen met naamstatus (zie paragraaf 1.3.1).

Volgens Booij (2002a) hebben de idiomatische adjectief-nomenparen vaak een classificerende functie. De combinatie in zijn geheel identificeert een categorie: *een koninklijk besluit* is een bepaald type besluit (naast bijvoorbeeld *een ministerieel besluit*). De stelling dat het onverbogen adjectief een categoriserende functie heeft,

wordt ook verwoord in de semantische analyses van Honselaar (1980) en Blom (1994, 1995). Daarnaast merken Booij (2002a) en Heynderickx (2001) op dat de adjectivische positie in adjectief-nomenparen met naamstatus frequent door een relationeel adjectief wordt bezet, zoals in *het koninklijk besluit*, *het medisch dossier*, *het Haarlems Dagblad*, *het Europees Parlement* of *het Nederlands elftal*.

Honselaar (1980) associeert het gebruik van het verbogen en het onverbogen adjectief met het onderscheid tussen directe en indirecte karakterisering, dat gebaseerd is op Bolingers (1967) distinctie tussen “referent modification” en “reference modification”. Bij directe karakterisering heeft de eigenschap aangeduid door het adjectief betrekking op de referent van het substantief (“referent modifier”); bij indirecte karakterisering slaat de adjectivische eigenschap op de referentie (“reference modifier” of “feature modifier”). De twee betekenissen van *een rode pen* illustreren dit onderscheid. Als met deze uiting een pen die rood is, wordt aangeduid, modificeert het adjectief de referent; als daarentegen een pen die rood schrijft, wordt bedoeld, modificeert het adjectief niet de referent maar de wijze waarop de referentie moet worden geconstrueerd. Dit onderscheid past Honselaar (1980: 195-202) toe op de buigingsalternantie. Hij associeert de verbogen vorm van het adjectief met *referent modification* en de gemarkeerde onverbogen vorm met *feature modification*. Het onverbogen adjectief modificeert met andere woorden niet de referent maar een kenmerk of kwaliteit van de referent.

- (11) a. het Nederlands elftal; het Haarlems dagblad  
 b. het Nederlandse elftal; het Haarlemse dagblad  
 (voorbeelden overgenomen uit Honselaar 1980: 199-200)

De onverbogen adjectieven in (11a) karakteriseren een specifiek kenmerk van het elftal of het dagblad, zoals blijkt uit Honselaars respectievelijke parafrazen: “the national soccer team of Holland” en “the newspaper of Haarlem”. De verbogen vorm sluit omwille van zijn ongemarkeerde status de *feature modification*-lezing niet *a priori* uit, maar kan daarnaast ook andere betekenissen uitdrukken. Honselaar vermeldt de volgende mogelijke lezingen voor de verbogen vormen in (11b).

- (11) b'. “‘the soccer team which comes from Holland’, ‘... consists of Dutch players’, etc.”  
 b''. “‘a newspaper that is made in Haarlem’, ‘a newspaper that was found in Haarlem’, ‘a newspaper that has a character typical of Haarlem’, etc.”

Een vergelijkbare, maar afwijkende visie vinden we bij Blom (1994, 1995). Haar uitgangspunt is Van der Horsts (1992, 1995: 127-133) analyse van het onderscheid tussen *vele* en *veel*:

“*Veel* individualiseert, onderstreept het stuk-voor-stuk karakter [...] *Veel* doet geen uitspraak over individualiteit of collectiviteit, en kan dus naar beide zijden opgevat worden” (van der Horst 1995: 130)

In navolging van deze analyse kent Blom (1994, 1995) aan de verbogen vorm een individualiserende betekenis toe. Dit stemt overeen met Honselaars *referent modification*. De onverbogen vorm kan uit hoofde van zijn niet-individualiserende hoedanigheid verschillende functies hebben: het modificeren van een categorie (12a), het aanduiden van een formele status (12b) of het verduidelijken van een specifieke notie of een specifiek aspect van het substantief (12c):

- (12) a. het Europees parlement  
 b. het koninklijk ontbijt (‘een royaal ontbijt’)  
 c. het tientallig stelsel  
 (voorbeelden overgenomen uit Blom 1994)

Merk op dat Blom in tegenstelling tot Honselaar (1980) niet de verbogen maar de onverbogen vorm als ongemarkeerd beschouwt. Dit is een gevolg van haar compositionele visie op de verbogen vorm als ‘onverbogen adjectief + *e*’, waarbij de buigings-*e* voor het individualiserende betekenisaspect zorgt.

#### **1.1.4 Nationale variëteit van het Nederlands**

De adjectivische buiging bij de neutra kent in de zuidelijke en de noordelijke dialecten een divergerende evolutie: vanaf het einde van de Middel-Nederlandse periode vindt er in de zuidelijke dialecten een geleidelijke apocopering plaats van de buigings-*e* bij enkelvoudige onzijdige substantieven in een definiëte NP.<sup>12</sup> Als resultaat is het onderscheid tussen de sterke en de zwakke buiging in de zuidelijke dialecten volledig verdwenen: zowel in definiëte als in indefiniëte NP’s is het adjectief voor een onzijdig substantief onverbogen (Daan 1969; Peters 1937; de Rooij 1980b; Taeldeman 1980).<sup>13</sup> Op deze divergerende evolutie tussen Nederland en Vlaanderen komen we uitgebreid terug in het historisch overzicht (zie paragraaf 2.3.1).

Dit systemische onderscheid tussen de noordelijke dialecten en de zuidelijke dialecten beïnvloedt de algemene omgangstaal: de onverbogen vorm kent een bredere verspreiding in het Belgisch Nederlands dan in het Nederlands Nederlands<sup>14</sup> (ANS 1997: 412; van Marle 1995a: 291; Raidt 1968: 101; de Rooij 1980a: 15-18). Ter illustratie worden doorgaans onverbogen eenlettergrepige adjectieven geciteerd die geen lexicale of semantische eenheid met het substantief vormen, zoals de voorbeelden in (13): terwijl deze vormen vreemd klinken in de oren van een Nederlander, ziet de doorsnee Vlaming er geen graten in.

- (13) het bruin paard; het groot feest

De invloed van de nationale variëteit van het Nederlands wordt empirisch aangetoond door De Rooij (1980a: 20-21): zelfs in geschreven bronnen (c.q. dagbladen) telt het Belgisch Nederlands significant meer onverbogen vormen bij neutra dan het Nederlands Nederlands.

### ***1.1.5 Stilistische waarde van de onverbogen vorm***

Over de stilistische waarde van beide buigingsvarianten bestaan twee lijnrecht tegenovergestelde standpunten: de ene visie beschouwt de onverbogen variant als uitermate formeel, de andere merkt de onverbogen variant net als informeel aan. Deze paradoxale situatie is het gevolg van de tegengestelde status van de buigingsvarianten in beide nationale variëteiten van het Nederlands (zie paragraaf 1.1.4). Eerst nemen we de situatie in het Nederlands Nederlands – en dus ook de standaardtaal – onder de loep; daarna kijken we wat de situatie voor het Belgisch Nederlands is.

In navolging van Overdiep (1937: 260) kenmerkt De Rooij (1980a: 19, 29) de onverbogen vorm als formeel en zelfs archaïsch. De Rooij (1980a: 29) spreekt van een “bijzondere stijl”, die hij meer bepaald met de ambtelijke en literaire taal in verband brengt. Deze visie sluit aan bij de historische evolutie van de adjectivische buiging in het noorden van het taalgebied. Vanaf de zeventiende eeuw is het verbogen adjectief de basisvariant in de zich ontwikkelende – noordelijke – standaardtaal en krijgt de onverbogen vorm een ambtelijke en literaire connotatie. De verbogen vorm is bovendien de algemeen verspreide variant in de noordelijke dialecten (zie paragraaf 2.3.1). Aangezien het Algemeen Nederlands in het noorden ontwikkeld is, verwoordt ook de ANS (1997: 412) dit standpunt.

In het Belgisch Nederlands is de situatie omgekeerd. Onderzoek naar de informele omgangstaal in Vlaanderen identificeert de veralgemening van het gebruik van de onverbogen vorm bij enkelvoudige *het*-woorden als een algemeen en prominent kenmerk van deze variëteit (Goossens 2000: 8-10; Haeseryn 1996: 118; de Rooij 1972: 7-8, Kaart III). Dit is niet verwonderlijk, rekening houdend met de exogene status van de ‘regelmatige’ verbogen vorm en de endogene status van de onverbogen vorm in Vlaanderen (zie paragrafen 1.1.4 en 2.3.1).

Bestaand empirisch onderzoek heeft echter geen significante stilistische differentiatie tussen beide buigingsvarianten aangetoond (de Rooij 1980a: 20-21).

In het empirisch onderzoek zullen we de nationale variëteit van het Nederlands en het register, c.q. de formaliteit van de gesprekscontext, onder de gezamenlijke noemer lectische condities behandelen. We gebruiken de term *lectisch* om te verwijzen naar alle types van talige variëteiten of *lecten*: dialecten, regiolecten, nationale variëteiten, registers, enz.

### ***1.1.6 Syntactische factoren***

In deze laatste paragraaf bespreken we twee syntactische factoren. Lebrun & Schurmans-Swillen (1966: 181-182) tellen het aantal woorden tussen de

determinator en het adjectief en het aantal woorden tussen het adjectief en het substantief. In geen van beide gevallen stellen ze een significant effect vast.

Verhagen (1993) bestudeert de buiging van het adjectief bij een genominaliseerde infinitief. Vertrekkend van de impliciete assumptie dat de buigings-*e* een dependentierelatie aanduidt, stelt hij dat het verbogen adjectief de genominaliseerde infinitief modificeert en dat de sequentie van het adjectief en de infinitief voldoende specifiek is om onafhankelijk geconceptualiseerd te worden (14a). In geval van een onverbogen adjectief, zoals in (14b), is er sprake van een vage semantische invulling van de infinitief, zodat er – bij voorkeur interne – complementen nodig zijn om de sequentie van het adjectief en de infinitief te conceptualiseren.

- (14) a. dat nichterige dansen van hem  
 b. het ongemotiveerd afbreken van de zwangerschap  
 (voorbeelden overgenomen uit Verhagen 1993: 105-106)

Deze stelling wordt door empirische gegevens gestaafd: de onverbogen variant komt significant vaker voor bij infinitieven die gevolgd worden door een intern complement dan bij infinitieven die niet door een complement gevolgd worden (Verhagen 1993: 104).

## 1.2 Bespreking en evaluatie

Na het verkennende overzicht van de verschillende factoren die de buigingsalternantie bepalen, zetten we deze informatie nu om in hypothesen (1.2.1). Daarna zullen we aan de hand van een kritische bespreking van de bestaande analyses het eigen onderzoek situeren en verantwoorden binnen het ruimere domein van de neerlandistiek (1.2.2).

### 1.2.1 *Hypothesen met betrekking tot de conditionering van de adjectivische buigingsalternantie bij neutra*

Tabel 1(1) toont een overzicht van de verschillende factoren die in de literatuur besproken worden en hun effect op de buigingsalternantie. Voor iedere factor wordt de volgende informatie samengebracht: de naam van de factor, het type factor (dit is het taalkundige domein waaronder de factor traditioneel geklasseerd wordt), de waarde(n) van de factor die het gebruik van de verbogen vorm bevordert, en de waarde(n) van de factor die het gebruik van de onverbogen vorm bevordert.

Deze tabel is een schematische voorstelling van de hypothesen die het uitgangspunt van het empirisch onderzoek vormen. De concrete operationalisering van de hypothesen in tabel 1(1) zal worden uitgewerkt bij de empirische analyse van de betreffende factor (zie deel II). Merk op dat we twee verschillende stilistische factoren onderscheiden, met name *stijl<sub>BEL</sub>* en *stijl<sub>NED</sub>*. Deze tweedeling lijkt ons gefundeerd, rekening houdend met de tegengestelde stilistische waarde van de

onverbogen vorm in beide nationale variëteiten (zie paragraaf 1.1.5). Het effect van de vergrotende trap hebben we bij de morfosyntactische factoren ingedeeld, hoewel dit effect vermoedelijk ook bij de fonologische factoren geplaatst kan worden. De syntactische factoren (zie paragraaf 1.1.6) zijn niet in dit overzicht opgenomen: de samenstelling van onze materiaalverzameling op basis van het *Corpus Gesproken Nederlands* bevat alleen instanties van de sequentie [DET<sub>[DEF]</sub> ADJ N<sub>[ONZ,ENK]</sub>]<sub>NP</sub> (zie paragraaf 6.2) en sluit bijgevolg de analyse van het aantal woorden tussen de determinator en het adjectief en het aantal woorden tussen het adjectief en het substantief bij voorbaat uit; de conceptualisatie van de genominaliseerde infinitief is een vage en moeilijk operationaliseerbare parameter, die bovendien slechts op een kleine fractie van de materiaalverzameling van toepassing is.

Type	Factorwaarde verbogen vorm	Factor	Factorwaarde onverbogen vorm
morfosyntactisch	bepaald lidwoord, aanwijzend vnv, genitief	woordsoort determinator	bezittelijk vnv
	stellende trap	gradatie A	vergroten trap
	neen	diminuering N	ja
	medeklinker	beginklank N	klinker
fonologisch	een- of tweelettergreep	lengte A	meerlettergreep
	beklemtoond	accentuering eindlettergreep A	onbeklemtoond
	beklemtoond	accentuering beginlettergreep N	onbeklemtoond
lexicaal semantisch	neen	semantische eenheid AN	ja
	kwalificerend	semantische categorie A	relatoneel
	niet-categoriserend	semantische categorie A	categoriserend
lectisch	Nederlands Nederlands	nationale variëteit	Belgisch Nederlands
	formeel, ongemarkeerd	stijl <sub>BEL</sub>	informeel
	ongemarkeerd	stijl <sub>NED</sub>	(zeer) formeel

*Tabel 1(1): Overzicht van de verklarende factoren en hun effect op de keuze van het adjectivische buigingsmorfeem*

Het samenvattende schema in tabel 1(1) illustreert de complexiteit van de adjectivische buigingsalternantie bij neutra. De selectie van het buigingsmorfeem is niet alleen het resultaat van een samenspel van verschillende factoren, maar deze factoren zijn bovendien zowel vormelijk als inhoudelijk van aard, zowel taalintern als taalextern (i.e. lectisch).

De verkenning van de literatuur heeft aangetoond dat het niet volstaat om het effect van de verschillende factoren afzonderlijk te meten. Daarnaast moet er ook rekening worden gehouden met mogelijke interacties tussen factoren. Het effect van een verklarende factor kan immers geconditioneerd worden door andere factoren. Tabel 1(2) schematiseert de potentiële interacties die in het empirisch onderzoek

getest worden. De linker kolom bevat de factor waarvan het effect getest wordt, de rechter kolom bevat de conditionerende factor die het effect van de bestudeerde factor kan beïnvloeden.<sup>15</sup>

<b>Factor</b>	<b>Conditionerende factor</b>
woordsoort determinator	stijl
beginklank N	stijl
lengte A	accentuering eindlettergreep A
klankkwaliteit eindlettergreep A	woordsoort determinator
semantiek AN	stijl
stijl	nationale variëteit

*Tabel 1(2): Potentiële interacties tussen verklarende factoren*

We geven een korte toelichting bij de mogelijke partiële interacties tussen verklarende factoren die in tabel 1(2) vermeld worden. Raidt (1968) is van oordeel dat het deflecterende effect van het bezittelijk voornaamwoord vooral in de informele omgangstaal tot uiting komt; De Rooij (1980a) is de tegengestelde mening toegedaan: hij beschouwt dit effect als kenmerkend voor de formele registers (zie paragraaf 1.1.1). Van een *ease of articulation*-effect, zoals het gebruik van een onverbogen adjectief om een hiaat te vermijden vóór een substantief dat met een klinker begint (zie paragraaf 1.1.2.1), kunnen we aannemen dat het groter is in minder gecontroleerde informele contexten dan in formele contexten. Deze stelling wordt ondersteund door Booij (1995: 150-151), die de deletie van een eindsjwa om de hiaat boven het woordniveau te vermijden als “characteristic of casual speech” typeert. De relatie tussen de lengte van het adjectief en de accentuering van de eindlettergreep hebben we al in paragraaf 1.1.2.2 aangekaart: een eenlettergrepig adjectief eindigt per definitie op een beklemtoonde lettergreep, bij meerlettergrepige adjectieven neemt de kans op een (opeenvolging van) doffe lettergreep(en) toe. De ANS (1997) wijst erop dat het deflecterende effect bij een adjectief met een onbeklemtoonde eindlettergreep versterkt wordt wanneer een bezittelijk voornaamwoord de determinatorpositie bezet (zie paragraaf 1.1.1). De voorbeelden die in de literatuur worden aangehaald om de selectie van het onverbogen buigingsmorfem in idiomatische adjectief-nomenparen te illustreren, duiden voor het merendeel officiële instanties aan. Dit blijkt ook uit de voorgestelde classificaties (zie paragraaf 1.1.3.1). Het lijkt dan ook aannemelijk dat de selectie van de onverbogen vorm om idiomatische adjectief-nomenparen te markeren, vooral in formele(re) registers zichtbaar is. Tot slot moet er ook bij de lectische factoren rekening worden gehouden met een wederzijdse beïnvloeding. Als gevolg van de divergerende evoluties sinds het einde van het Middel-Nederlands (zie paragrafen 1.1.4, 1.1.5 en 2.3.1), vermoeden we een tegengestelde stilistische betekenis voor de onverbogen buigingsvariant in beide nationale variëteiten van het Nederlands: (zeer) formeel en ambtelijk in het Nederlands Nederlands tegenover informeel in het



Belgisch Nederlands. We zullen de buigingsverschillen tussen beide nationale variëteiten in een ruimer perspectief plaatsen door na te gaan of de adjectivische buigingsalternantie bij neutra twee afzonderlijke modellen vereist, een eerste voor het Belgisch Nederlands en een tweede voor het Nederlands Nederlands (zie paragraaf 14.3.4).

De methodologische en technische aspecten van het empirisch onderzoek zullen tijdens de presentatie van de gebruikte methodologie uitgebreid aan bod komen (voor de algemene methodologie verwijzen we naar hoofdstuk 5 en voor de gebruikte statistische technieken naar de hoofdstukken 7 en 13).

### ***1.2.2 Situering en verantwoording van het onderzoek naar de adjectivische buigingsalternantie bij neutra***

Na een overzicht te hebben gegeven van de verschillende verklarende factoren van de buigingsalternantie bij neutra wisselen we nu van invalshoek. We benaderen de buigingsalternantie niet langer vanuit het perspectief van de verklarende factoren, maar vanuit het perspectief van de analyses van de adjectivische buiging(s)-alternantie). Deze benadering wil de verschillende attitudes binnen de neerlandistiek ten aanzien van de buigingsalternantie in kaart brengen. Tot slot situeren we voorliggend onderzoek binnen dit neerlandistische kader.

De geraadpleegde analyses vertonen één van de volgende drie attitudes ten overstaan van de buigingsalternantie:

- *marginaliserende attitude*: analyses die de buigingsalternantie niet erkennen of marginaliseren
- *reducerende attitude*: analyses die de buigingsalternantie tot één of twee factoren herleiden
- *globaliserende attitude*: analyses die de complexiteit van de buigingsalternantie in beeld proberen te brengen

Deze driedeling is uiteraard een simplificatie waarin een aantal analyses moeilijk kan worden ingepast.

De marginaliserende attitude groepeert in eerste instantie algemene analyses van de buiging van het attributieve adjectief (De Schutter 1994, 1997; Kester 1996; Van Eynde 2003). De primaire doelstelling van dit type onderzoek is de ontwikkeling van een formeel model voor de beschrijving van de adjectivische buiging. Het buigingsmechanisme dat het object van deze beschrijvingen vormt, is het model in figuur 3(2). Het bestaan van variatie wordt genegeerd of tot een idiosyncratisch (Kester 1996)<sup>16</sup> of dialectisch (De Schutter 1994: 29-30 noten 5 en 22, 1997: 242 noot 36; Van Eynde 2003: 116 noot 9, 119) fenomeen gereduceerd dat niet tot de kern van de grammatica behoort. Deze attitude vinden we ook terug in enkele algemene grammatica's (*inter al.* Klooster 2001: 82-83<sup>17</sup>; Rijpma & Schuringa 1978: 115-116; van den Toorn 1982: 165). Het is geen toeval dat deze analyses uitgaan van een autonome grammatica die beschreven kan worden door het raadplegen van de intuïties van een geïdealiseerde taalgebruiker in een homogene

taalgemeenschap – of gebruik maken van een theoretisch kader dat een autonome grammatica veronderstelt.

Het model voorgesteld door Odijk (1992) is een twijfelgeval tussen de marginaliserende en de reducerende attitude. Vertrekkend van dezelfde theoretische assumpties als bovenstaande analyses, kent hij de adjectief-nomenparen die een semantische eenheid vormen een positie toe in een formeel congruentiemodel in de trant van *Principles and Parameters*. Odijk (1992: 201) stelt een “Formal Lexicalisation”-regel voor om deze gevallen correct af te leiden. Deze regel houdt in dat er een lexicale knoop N boven de syntagmatische knoop N' wordt geplaatst, zodat de sequentie van het adjectief en het substantief ‘onzichtbaar’ is voor het toekennen van het buigingsmorfeem (op het niveau van N' of N''). Met andere woorden, de sequentie van adjectief en substantief wordt als een lexicale eenheid beschouwd. Omdat de lexicale eenheden doorgaans bij de idiosyncratische uitzonderingen worden ondergebracht, delen we Odijks analyse bij de marginaliserende groep in.

De analyses die de tweede, reducerende attitude aanhangen, relateren het optreden van de buigingsalternantie bij neutra aan de lexicale en semantische factoren en/of aan de prosodische factoren. Het merendeel van de morfologische analyses behoort tot deze groep: sommige beschrijvingen vermelden alleen de lexicale of semantische eenheid van het adjectief-nomenpaar – al dan niet aangevuld met de systematische uitzonderingen (zie paragraaf 3.2) (Booij 2002c: 45-48; Booij & van Santen 1998: 99-100; de Haas & Trommelen 1993: 285-286). Andere studies voegen aan deze semantische factor de eufonische kenmerken van het adjectief toe (Booij 1992: 128-130). Dit is ook de algemene houding in de taalzuiveringsliteratuur (Cockx 2000: 29-30; Klein & Visscher 1996: 69, 71; Penninckx & Buyse 1997: 39-40; Permentier & Van Den Eynden 1998: 182-183). Tot de reducerende attitude rekenen we ook de analyses van Blom (1994, 1995), Booij (2002a) en Honselaar (1980) die een systematisch betekenisverschil voor de buigingsalternantie voorstellen.

De derde en laatste attitude groepeerde de analyses die de complexiteit van de buigingsalternantie in kaart willen brengen. Naast de ANS (1997) zijn dit vooral descriptieve en variationistische analyses (Raidt 1968; Royen 1953). Enkele studies maken gebruik van corpusgegevens (Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a, 1980b). De rapportering neemt in deze analyses de vorm aan van een opsomming van de (significante) condities op de keuze van het buigingsmorfeem.

Hoewel de adjectivische buiging en de buigingsalternantie een recurrent onderwerp zijn in grammatica's en morfologische analyses, raken de meeste beschrijvingen niet verder dan het citeren van dezelfde factoren die ze aan de hand van dezelfde geconstrueerde voorbeelden illustreren. In dit verband valt het verschil in aandacht op die de verschillende types factoren genieten. De lexicale of semantische eenheid van het adjectief-nomenpaar is verreweg de meest geciteerde

factor. Daarna volgen de prosodische kenmerken, niet zelden in een morfologisch kleedje gestoken (cf. het gebruik van de onverbogen variant na bepaalde onbeklemtoonbare suffixen; zie paragraaf 1.1.2.2). De morfosyntactische en lectische factoren worden daarentegen nauwelijks onderkend. Een plausibele verklaring voor de beperkte aandacht die deze factoren in de literatuur genieten, is het introspectieve karakter van het gros van het onderzoek dat lectische en historisch bepaalde verschillen nivelleert in een geïdealiseerd taalmodel.<sup>18</sup> Het citeren van hetzelfde beperkte aantal geconstrueerde voorbeelden leidt bovendien tot de circulaire situatie waar de talige ‘data’ die de grondslag van de beschrijving vormen op basis van het ‘resulterende’ model geselecteerd worden.

Bij variationele fenomenen veroorzaakt introspectief onderzoek twee bijkomende problemen. In eerste instantie sluit introspectief onderzoek de mogelijkheid uit om de impact van een parameter te meten. Illustratief is het standpunt van Odijk (1992: 198) dat bij “Type-I exceptions [combinaties van een adjectief en een *het*-woord met een niet-compositionele betekenis; JT] [...] the adjective does not get the *e*-suffix in the singular, not even in a definite NP”, terwijl er in de literatuur een ruime consensus bestaat over de graduele aard van de alternantie tussen beide buigingsvarianten. De formalisering van lexicale ‘uitzonderingen’ in een symbolisch model dwingt Odijk tot het gebruik van binaire criteria die homogene groepen creëren, dit wil zeggen een groep voor de (‘regelmatige’) verbogen vormen tegenover een groep voor de (‘onregelmatige’) onverbogen vormen. In tweede instantie resulteert het ontbreken van informatie over de impact van de verschillende parameters en eventuele interacties in een schijnbaar onontwarbaar kluwen van factoren dat onder de hoofding “speciale regels en twijfelgevallen” wordt ingedeeld (ANS 1997: 408-412) of ervoor zorgt dat “beide vormen (de verbogen en de onverbogen) dikwijls naar eigen goeddunken gebruikt mogen worden” (Verhasselt 1974: 140-141). Deze visie op de buigingsalternantie als een aleatoir fenomeen is kenmerkend voor het introspectieve onderzoek dat de globaliserende attitude is toegedaan. Dit onderzoek vormt de antipode van het marginaliserende standpunt, dat de adjectivische buigingsalternantie hoogstens als een marginaal randverschijnsel erkent.

De adjectivische buigingsalternantie bij neutra is in het verleden al het onderwerp geweest van kwantitatief-empirisch onderzoek (Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a). Hoewel dit onderzoek een belangrijke stap voorwaarts is ten overstaan van het introspectieve onderzoek, getuige de uitgebreide aandacht voor verklarende factoren, zijn er een aantal methodologische en inhoudelijke punten van kritiek. Ten eerste, deze analyses berekenen enkel de betrouwbaarheid waarmee de nulhypothese – er is geen invloed van de bestudeerde verklarende factor op de buigingsalternantie – kan worden weerlegd.<sup>19</sup> De effectieve impact van de afzonderlijke parameters wordt met andere woorden niet gemeten, zodat de rapportering beperkt is tot een opsomming van de factoren en hun significantie. Ten tweede, het testen van de significantie gebeurt voor iedere factor

afzonderlijk in bivariate analyses (zie paragraaf 5.2.3). Er wordt dus geen rekening gehouden met de mogelijke interacties die het overzicht van de verklarende factoren aan het licht heeft gebracht (zie paragraaf 1.2.1, tabel 1(2)). Bovendien worden de significante factoren enkel opgesomd, zonder ze te integreren in een globaal model dat vervolgens taalkundig geïnterpreteerd kan worden (zie paragraaf 5.2.3 en deel III). Ten derde, het onderzoek van Degryse (1959) en Lebrun & Schurmans-Swillen (1966) is beperkt tot één stilistisch stratum, respectievelijk literair proza en dagbladen. Daarnaast omvat het empirische materiaal van Lebrun & Schurmans-Swillen (1966) slechts één nationale variëteit, met name het Belgisch Nederlands.<sup>20</sup> De Rooij (1980a) gebruikt wel een stilistisch (geschreven materiaal vs. gesproken materiaal) en regionaal (Belgisch Nederlands vs. Nederlands Nederlands) gedifferentieerde materiaalverzameling, maar hij kampt met dataschaarste in het deelcorpus met het gesproken (c.q. informele) materiaal. Ten vierde, de factor die het vaakst in de (introspectieve) literatuur wordt aangehaald, met name de lexicale of semantische eenheid gevormd door het adjectief-nomenpaar, wordt expliciet uit het bestaande empirisch onderzoek geweerd, enerzijds omdat deze uitdrukkingen geen variatie zouden vertonen, anderzijds omdat de lexicale of semantische eenheid van het adjectief-nomenpaar moeilijk op een objectieve manier geoperationaliseerd kan worden. In het licht van het tweede argument mag het dan ook verwonderlijk heten dat het verwijderen van deze uitdrukkingen op introspectieve basis noch voor Lebrun & Schurmans-Swillen (1966) noch voor De Rooij (1980a) een probleem lijkt te vormen.<sup>21</sup>

Op basis van bovenstaande recensie situeren we tot slot het eigen onderzoek in het ruimere kader van het neerlandistische onderzoek naar de adjectivische buiging(salternantie). Vooreerst is het na 25 jaar nodig om de adjectivische buigingsalternantie bij neutra opnieuw aan een diepgravend kwantitatief-empirisch onderzoek te onderwerpen. Hierbij laten we ons door geen enkel theoretisch apriori leiden. Dit wil zeggen dat we geen potentiële verklarende factoren uitsluiten op basis van motieven eigen aan een bepaald theoretisch taalkundig kader. Bijgevolg bestuderen we ook de lectische conditionering van de buigingsalternantie. De doelstelling van de kwantitatief-empirische analyse is de constructie van een globaal model dat de verschillende verklarende factoren integreert. Op deze wijze kunnen we het buigingseffect van de verschillende condities kwantificeren en een rangorde tussen deze condities aanbrengen met betrekking tot hun impact op de keuze van het adjectivische buigingsmorfeem. Daarnaast evalueren we de empirische adequaatheid van de verschillende attitudes die we ten aanzien van de buigingsalternantie onderscheiden hebben door ze met het globale model voor de buigingsalternantie te confronteren.

We hopen dat deze werkwijze ons in staat stelt om de algemene tendenties die de buigingsalternantie in het reële taalgebruik bepalen niet alleen in kaart te brengen maar ook te verklaren om zo het kluwen van concurrerende en interagerende verklarende factoren te ontwarren. De lexicale en semantische conditionering van de

buigingsalternantie, een factor waarvoor tot heden een empirische analyse ontbreekt, vormt het onderwerp van deel IV. In dat vierde en laatste deel van het proefschrift trachten we de vraag te beantwoorden of het gebruik van het onverbogen adjectief bij de – zogenaamde – idiomatische adjectief-nomenparen een louter lexicaal effect is, dan wel of er semantische patronen geïdentificeerd kunnen worden aan de basis van deze deflecterende buigingstendens.

### 1.3 Conclusie

Dit hoofdstuk presenteert een verkennend overzicht van de taalkundige traditie op het gebied van de adjectivische buigingsalternantie bij neutra. Het eerste deel van dit hoofdstuk geeft een overzicht van de verschillende factoren die in de literatuur worden aangehaald om de selectie van het onverbogen adjectief te motiveren (1.1). Drie vaststellingen dringen zich hierbij op. Ten eerste, het onverbogen adjectief is taalkundig de gemarkeerde vorm: niet alleen zijn de contexten waar deze variant voorkomt, beperkter in aantal, maar deze contexten kunnen bovendien vaak als taalkundig gemarkeerd beschouwd worden. Ten tweede, de factoren die het gebruik van de onverbogen variant motiveren, zijn zeer divers en behoren tot verschillende taalkundige domeinen, die zowel vormelijk (bv. de woordsoort van het element in de determinatorpositie en de kernpositie; de prosodische kenmerken van het adjectief en het substantief) als inhoudelijk (bv. de lexicale en semantische kenmerken van het adjectief-nomenpaar) van aard zijn, en die zowel taalintern als lectisch zijn (de nationale variëteit; de formaliteit van het register). Deze grote verscheidenheid aan factoren sluit een overkoepelende hypothese over de schematische betekenis/functie van de buigingsalternantie nagenoeg bij voorbaat uit. Ten derde, de literatuur vermeldt verschillende mogelijke interacties tussen verklarende factoren. Het is niet uitgesloten dat het effect van verschillende taalstructurele factoren (onder andere de fonologische factoren) groter is in de informele(re) registers dan in de formele(re) registers. De meest ingrijpende conditionerende factor is wellicht de nationale variëteit van het Nederlands. De ‘regelmatige’ verbogen vorm heeft immers een tegengestelde status in beide nationale variëteiten: terwijl de verbogen vorm boven de Moerdijk – en dus ook in de noordelijke en zuidelijke standaardtaal – de basisvariant is, kennen de zuidelijke dialecten uitsluitend de onverbogen vorm bij een onzijdig substantief (zie ook paragraaf 2.3.1). De verbogen variant is met andere woorden endogeen in het Nederlands Nederlands en exogeen in het Belgisch Nederlands. We zullen bijgevolg de noodzaak onderzoeken om een afzonderlijk model te construeren voor de buigingsalternantie in beide nationale variëteiten (zie paragraaf 14.3.4).

Het tweede deel van de literatuurstudie vat eerst de verklarende factoren samen en somt verscheidene mogelijke interacties tussen deze factoren op (1.2.1). Daarnaast bespreken we de verschillende attitudes in de neerlandistische literatuur ten aanzien van de buigingsvariatie (1.2.2). We hebben drie attitudes onderscheiden: de

## DEEL I

marginaliserende attitude, de reducerende attitude en de globaliserende attitude. Vervolgens hebben we getoond dat in geen enkele van deze attitudes een globaal model voor de beschrijving van de buigingsalternantie wordt voorgesteld. De constructie van een globaal model vormt dan ook de hoofddoelstelling van dit onderzoek: we willen op basis van het reële taalgebruik de tendenties modelleren die de adjectivische buigingsalternantie bij neutra in het standaard gesproken Nederlands bepalen. In hoofdstuk 5 bespreken we de methodologische benadering die we het meest geschikt achten om deze doelstelling te verwezenlijken.

## Noten

1. Aan deze lijst moeten we de geanteponeerde genitieven toevoegen die in de literatuur vaak over het hoofd worden gezien (bv. *Jans mooie boek*). In deze paragraaf komen de geanteponeerde genitieven niet meer aan bod, maar ze worden niet *a priori* uit het empirisch onderzoek geweerd (zie paragraaf 8.1). Bij de beschrijving van de materiaalverzameling op basis van het *Corpus Gesproken Nederlands* kijken we ook naar het buigingseffect van andere definiete determinatoren, zoals de aanwijzende voornaamwoorden *zo'n* en *zulk* of de collectiverende onbepaalde voornaamwoorden (zie paragraaf 6.2.2).
2. Aan het einde van het Middel-Nederlands leidt het proclitische gebruik tot het verlies van de buigingsuitgang van het bezittelijk voornaamwoord (zie paragraaf 2.2).
3. Van der Horst & van der Horst (1999: 315) merken op dat het echter onduidelijk is wat in deze periode de werkelijke adjectivische buigingsvorm na een bezittelijk voornaamwoord was.
4. Doordat Degryse (1959) enkel de relatieve frequenties vermeldt, hebben we het ongewogen gemiddelde berekend voor het aandeel onverbogen vormen bij de eenlettergrepige, de tweelettergrepige en de meerlettergrepige adjectieven.
5. De Rooij (1980a: 10-11) heeft het materiaal gebruikt van de “Werkgroep Frequentieonderzoek van het Nederlands”.
6. Voor de eerste helft van de 20<sup>e</sup> eeuw wordt deze inferentie bevestigd door het literatuuroverzicht in Degryse (1959) en Royen (1953).
7. Uit de voorbeelden geciteerd door Klein & Visscher (1996: 71) leiden we af dat zij eveneens deze mening zijn toegedaan.
8. Bij het onderscheid tussen een klinker ( $\frac{\text{verbogen} = 194}{\text{onverbogen} = 70}$ ) en een medeklinker ( $\frac{\text{verbogen} = 1,512}{\text{onverbogen} = 323}$ ) als substantivisch beginsegment, moeten we opmerken dat De Rooij (1980a: 17, 21) op basis van de trendtoets van Van Eeden rapporteert dat deze oppositie geen significant effect op de keuze van het buigingsmorfeem heeft ( $p = 0.084$ ). Wanneer we echter De Rooijs frequentieverdeling aan de hand van de  $\chi^2$ -statistiek analyseren, is het resultaat wel significant ( $\chi^2 = 11.4697$ ,  $df = 1$ ,  $p = 0.0007074$ ) (zie paragraaf 7.1 voor meer informatie over het gebruik van de  $\chi^2$ -statistiek).
9. Wij hebben de significantietest ook voor het spreektaalige materiaal overgedaan en bekomen opnieuw – in tegenstelling tot De Rooij (1980a) – een significant effect. De frequentieverdeling in het gesproken materiaal ziet er als volgt uit: substantief begint met medeklinker: 159 verbogen vormen tegen 21 onverbogen vormen; substantief begint met klinker: 15 verbogen tegen 9 onverbogen vormen (de Rooij 1980a: 17). De Rooij (1980a: 21) rapporteert voor de trendtoets van Van Eeden een  $p$ -waarde gelijk aan 0.53. Op basis van de *Fisher Exact*-test bekomen wij voor deze frequentieverdeling een  $p$ -waarde gelijk

aan 0.002764 (zie paragraaf 7.1 voor meer informatie over het gebruik van de *Fisher Exact-test*).

10. Het klemtoonpatroon wordt in de voorbeelden (3) tot en met (9) aan de hand van accenten aangeduid: op de klinker in de lettergreep met de hoofdklemtoon plaatsen we een *accent aigu* (´), op de klinker in de lettergreep(en) met de nevenklemtoon een *accent grave* (`), de klinker in een onbeklemtoonde lettergreep blijft accentloos.
11. De vormen van de vergrotende trap kunnen ook bij *de*-woorden onverbogen blijven, vooral als de stellende trap van het adjectief op *-ig* of *-(e)lijk* eindigt (Broekhuis 1999; Rijpma & Schuringa 1978; Royen 1953). Terwijl Broekhuis (1999) deze mogelijkheid tot indefiniete NP's met een enkelvoudig substantief als hoofd (ia) beperkt, vinden we bij Royen (1953) en Rijpma & Schuringa (1978) ook voorbeelden met een definiete determinator (ib) of een meervoudsvorm (ic).
  - (i) a. een slúwer(e) diplomaat; een uitvoériger(e) beschrijving; een gemakkelijker(e) oplossing
  - b. de uitvoériger(e) behandeling  
(voorbeeld overgenomen uit Royen 1953: 112)
  - c. Bspóttelijker mensen zijn er niet.  
(voorbeeld overgenomen uit Rijpma & Schuringa 1978: 118)
12. De volledigheid gebiedt te zeggen dat de eindsjwa in het zuidwesten (West-Vlaanderen en het westelijke deel van Oost-Vlaanderen) van het taalgebied in vele contexten niet apocopeert (Taeldeman 1978; Weijnen 1991).
13. De kaart van De Rooij (1980b) voor *het bruin(e) paard* bevat slechts vijf meetpunten met een verbogen adjectief in Vlaanderen: voor drie meetpunten in West-Vlaanderen en twee meetpunten in Oost-Vlaanderen hebben de respondenten de verbogen vorm als basisvorm opgegeven.
14. In navolging van Geeraerts, Grondelaers & Speelman (1999) spreken we van Belgisch Nederlands en Nederlands Nederlands wanneer het over de nationale variëteiten van het Nederlands gaat. De scheiding tussen zuidelijke en noordelijke dialecten valt immers niet samen met de rijksgrens (zie paragraaf 2.3.1).
15. Technisch gezien kunnen de kolom met de (verklarende) factoren en de kolom met de conditionerende factoren van plaats gewisseld worden. Bij het bepalen van de conditionerende factor hebben we ons hoofdzakelijk laten leiden door de taalkundige interpretatie. Het is in de taalkunde bijvoorbeeld de gewoonte om lectisch gedefinieerde variëteiten als aparte talige systemen met een eigen interne structuur te beschouwen. Hoewel we deze strikte scheiding tussen taalinterne en taalexterne factoren verwerpen (zie paragraaf 12.1), is het interpretatief makkelijker taalinterne factoren in functie van de taalvariëteiten te differentiëren dan omgekeerd. Als een lectische factor meerdere taalinterne factoren conditioneert, moet uiteraard getest worden in welke mate een eventueel lectisch effect een neveneffect is van de verschillende taalinterne effecten.
16. Alternanties van het type *een knappe schilder/een knap schilder* worden wel verklaard in het model voorgesteld door Kester (1996: 99-104).



17. Hoewel Klooster (2001: 83) op de mogelijkheid van buigingsvariatie bij *het*-woorden wijst, gaat hij er niet dieper op in, in tegenstelling tot de buigingsvariatie die bij *de*-woorden optreedt.
18. Relevant voor deze discussie is de relatie tussen introspectief onderzoek aan de ene kant en het aanhangen van een geïdealiseerd 'taalmodel' aan de andere kant. De exacte oorzakelijke relatie tussen introspectie en het gebruik van een geïdealiseerd 'taalmodel' is van ondergeschikt belang.
19. Degryse (1959) rapporteert alleen de relatieve frequenties (zie ook noot 4), zonder melding van de absolute frequenties of de significantie van de vastgestelde frequentieverdelingen.
20. Degryse (1959) excerpert romans van zowel Belgische als Nederlandse auteurs, maar behandelt het onderscheid tussen beide nationale variëteiten niet als een verklarende factor.
21. Het criterium om te bepalen of een adjectief-nomenpaar in het onderzoek wordt opgenomen, is het al dan niet voorkomen van beide buigingsvarianten bij de attestaties van het adjectief-nomenpaar in kwestie: indien de realisaties van een adjectief-nomensequentie consistent één buigingsvariant kiezen, wordt deze sequentie uit het onderzoek geweerd. We hebben echter de stellige indruk dat De Rooij (1980a) enkel de observaties van de adjectief-nomenparen waarvan hij *a priori* aanneemt dat ze een lexicale of semantische eenheid (kunnen) vormen, gecontroleerd heeft op het samen voorkomen van de verbogen én de onverbogen vorm. Bijgevolg heeft hij *hapax legomena* en adjectief-nomenparen die om andere redenen door slechts één buigingsvariant gerealiseerd worden niet uit zijn materiaalverzameling verwijderd. Illustratief is de vaststelling dat De Rooij (1980a: 13) onder andere *het openbaar vervoer* en *het lijdend voorwerp* niet uitsluit van zijn empirische analyse omdat zijn materiaalverzameling één attestatie van *het openbare vervoer* en één attestatie van *het zelfstandige naamwoord* (sic) bevat.

## Hoofdstuk 2:

# Historische situering van de adjectivische buigingsalternantie

In het kader van het onderzoek naar de adjectivische buigingsalternantie in een definiëte NP met een enkelvoudig *het*-woord als hoofd is het nuttig de grote lijnen van de historische ontwikkeling van de adjectivische buiging(salternantie) te beschouwen. We beperken ons hierbij veelal tot de nominatief- en de accusatief-casus: Van Haeringen (1937: 247) en Geerts (1966: 127-128) voeren aan dat de vormen van de accusatief de basis van de huidige taal zijn; bij onzijdige en vrouwelijke substantieven verdwijnt het verschil tussen nominatief en accusatief al in een vroeg stadium, zoals de tabellen 2(1) tot en met 2(5) voor de vormen bij *het*-woorden aantonen.

Het historische overzicht wordt in drie fases opgesplitst: de periode van het Indo-Europees en het Oer-Germaans (2.1), de periode van het Oud- en Middel-Nederlands (2.2), en tot slot de periode van het Nieuw-Nederlands (2.3). De relevante ontwikkelingen met betrekking tot de buiging van de andere adnominale elementen komen eveneens aan bod.

Onderstaande beschrijving is schatplichtig aan de volgende analyses van de ontwikkeling van het Nederlands. De basisreferentie voor dit historische overzicht is Raidt (1968). Zij beschrijft de ontwikkeling van de adjectivische buiging van het Indo-Europees tot het hedendaags Nederlands. Haar beschrijving stoelt op teksten en grammatica's uit de betreffende periodes. We vullen Raidts beschrijving aan met algemene naslagwerken over de geschiedenis van het Nederlands en specifieke studies voor de afzonderlijke periodes.<sup>1</sup> Naast deze diachrone analyses hebben we ook de resultaten van dialectologisch onderzoek geraadpleegd. Hier stelt zich het probleem dat globale dialectologische analyses van de adjectivische buiging zeldzaam zijn (Daan 1969; Van Haeringen 1937; Peters 1937; de Rooij 1980b; Taeldeman 1980).<sup>2</sup>

Het is niet de bedoeling om een exhaustief overzicht van de adjectivische buiging(svariatie) te presenteren. We beperken ons tot de grote lijnen om de adjectivische buigingsalternantie bij neutra in een ruimer diachroon kader te situeren. De relevante bevindingen worden in de conclusie op een rijtje gezet (zie paragraaf 2.4).

## 2.1 De adnominale buiging in het Indo-Europees en het Oer-Germaans

Algemeen wordt aangenomen dat de adjectieven oorspronkelijk onverbogen bleven (van Bree 1980; Van Loey 1970). In het Indo-Europees is de adjectivische buiging die van het substantief: “Morphologically, adjectives are nouns, and in Indo-European their declension does not differ from that of the corresponding noun stems” (Prokosch 1939: 259). De evolutie van het Indo-Europees naar het Oer-Germaans wordt gekenmerkt door een differentiëring van de adjectieven ten overstaan van de substantieven door het overnemen van de flexie-uitgangen van de voornaamwoorden en door de ontwikkeling van een syntactisch gemotiveerde oppositie tussen een sterke en een zwakke verbuiging.

Naast de twee syntactisch gemotiveerde flexietypes onderscheidt het Oer-Germaans verschillende lexicale adjectiefklassen. Op basis van het stamvormende element worden de volgende klassen onderscheiden: de *o/a*-stammen, die veruit de grootste groep vormen, de *jo/ja*-stammen, de *i*-stammen en de *u*-stammen. Het verschil tussen deze groepen verdwijnt progressief vanaf de Germaanse periode.<sup>3</sup> Niettemin zal het vervagende onderscheid tussen *jo/ja*-stammen en *o/a*-stammen in het Oud- en Middel-Nederlands bijdragen tot de verzwakking van het adjectivische flexiesysteem (zie paragraaf 2.2).

In het Oer-Germaans worden de traditionele substantivische uitgangen van het adjectief ten dele vervangen door toevoeging van pronominale elementen. Dit resulteert in een gemengde pronominaal-substantivische buiging die bekend staat als de sterke buiging. Tabel 2(1) geeft een overzicht van de sterke verbuiging voor de *o/a*-stammen bij een enkelvoudig onzijdig substantief in het Gotisch:

<b>Nom.</b>	<i>blind, blindata</i>
<b>Gen.</b>	<i>blindis</i>
<b>Dat.</b>	<i>blindamma</i>
<b>Acc.</b>	<i>blind, blindata</i>

Tabel 2(1): Sterke adjectivische buiging van *blind* in het Gotisch  
(overgenomen uit Krahe & Meid 1969: 77)

We stellen vast dat de pronominale (niet onderlijnd) en nominale (onderlijnd) uitgangen door elkaar lopen. In de nominatief en de accusatief is er zelfs concurrentie tussen beide varianten.

De pronominale buigingsvormen zijn de adjectivische flexie binnengedrongen via pronominale adjectieven zoals *ander* (Krahe & Meid 1969; Van Loey 1970). De buigingsmorfemen zijn deze van het aanwijzend voornaamwoord dat volgens de sterke buiging verbogen wordt (cf. *infra*): de vorm van *blind* voor de nominatief enkelvoud onzijdig is in het Gotisch *blindata* naar analogie met het aanwijzend voornaamwoord *ata* (Prokosch 1939: 276).

Naast de sterke verbuiging ontwikkelt zich ook een zwakke verbuiging. Deze zwakke verbuiging ontstaat uit het substantivische gebruik van het adjectief dat de verbuiging van de nominale *n*-stammen overneemt. Het onderscheid tussen sterke en zwakke flexie is syntactisch bepaald: de sterke flexie komt in onbepaalde contexten voor; de zwakke flexie wordt gebruikt in definiete contexten, i.e. na voornaamwoorden, bij genominaliseerde adjectieven en bij pronominale adjectieven. Dit syntactische onderscheid is semantisch van oorsprong: “The ‘indefinite’ function of the ‘strong’ and the ‘definite’ function of the ‘weak’ declension is closely related to the individualizing sense of the former and the generalizing sense of the latter” (Prokosch 1939: 260-261). De overige adnominale elementen, i.e. de adjectivisch gebruikte voornaamwoorden, volgen de sterke adjectivische buiging (Krahe & Meid 1969: 59).

## 2.2 De adnominale buiging in het Oud- en Middel-Nederlands

Het Oud-Nederlands onderscheidt nog duidelijk de zwakke en de sterke flexie. Het inflectionele paradigma van het attributieve adjectief bij een enkelvoudig onzijdig substantief ziet er volgens Quak & van der Horst (2002: 41-42) als volgt uit:

	Sterke flexie	Zwakke flexie
<b>Nom.</b>	<i>blint</i>	<i>blinda</i>
<b>Gen.</b>	<i>blindes</i>	<i>blindin</i>
<b>Dat.</b>	<i>blindin</i>	<i>blindin</i>
<b>Acc.</b>	<i>blint</i>	<i>blinda</i>

Tabel 2(2): Adjectivische buiging van *blint* voor het onzijdige genus in het Oud-Nederlands

In de *Wachtendonckse Psalmen*<sup>4</sup>, het materiaal waaruit de gegevens in tabel 2(2) geëxtraheerd zijn, is de competitie tussen de pronominale en de nominale vorm in de nominatief en de accusatief verdwenen. Quak & van der Horst (2002: 42) melden dat de zwakke buiging “vrij zeldzaam” is door het ontbreken van lidwoorden in het materiaal dat zij geraadpleegd hebben. Nochtans vindt de ontwikkeling van het bepaald lidwoord uit het aanwijzend voornaamwoord plaats in het Oud-Nederlands (van den Toorn et al. 1997: 60-61). De bezittelijke voornaamwoorden zijn in het Oud-Nederlands ontstaan uit de genitiefvormen van het persoonlijk voornaamwoord: *mîn*, *thîn*, *sîn*<sup>5</sup>, *unser*, *iuwer* en *sîn* (van den Toorn et al. 1997: 52-53).

Tijdens de drie eeuwen die het Middel-Nederlands overspant<sup>6</sup>, ondergaat de adjectivische buiging ingrijpende veranderingen die uitmonden in de desintegratie van het Germaanse buigingssysteem. Belangrijk voor onze doeleinden is de vaststelling dat de zwakke buiging aan het eind van het Middel-Nederlands grotendeels met de sterke buiging samenvalt. Aan het begin van deze periode kunnen beide flexiesystemen nog duidelijk onderscheiden worden (Raidt 1968). Tabel 2(3) toont

het paradigma van het attributieve adjectief bij een enkelvoudig *het*-woord bij aanvang van het Middel-Nederlands: deze vormen vertonen een sterke gelijkenis met de vormen voor het Oud-Nederlands in tabel 2(2). Het paradigma is gereconstrueerd aan de hand van Raidt (1968: 4, 7), Te Winkel (1901: 133)<sup>7</sup> en Van Helten (1973: 391, 400).

	<b>Sterke flexie</b>	<b>Zwakke flexie</b>
<b>Nom.</b>	<i>blint</i>	<i>blinde</i>
<b>Gen.</b>	<i>blind(e)s</i>	<i>blinden</i>
<b>Dat.</b>	<i>blinden</i>	<i>blinden</i>
<b>Acc.</b>	<i>blint</i>	<i>blinde</i>

Tabel 2(3): *Adjectivische buiging van blint voor het onzijdige genus bij aanvang van het Middel-Nederlands*

In het Laat-Oud-Nederlands treedt een klankwettige verdoffing van de onbeklemtoonde lettergrepen op (van Loon 1986, 1988). Aangezien de adjectiefuitgangen onbeklemtoond zijn, treedt ook hier verdoffing op (Te Winkel 1901), zoals blijkt uit de vormen in tabel 2(3). Een samenspel van uiteenlopende factoren leidt tot de atrofïering van dit systeem tijdens de Middel-Nederlandse periode. Zoals aangegeven in de inleiding spitsen wij ons toe op de processen die de onzijdige nominatief en accusatief beïnvloed hebben.

Het Oer-Germaanse onderscheid tussen *o/a*-stammen met korte klinker en *jo/ja*-stammen met lange klinker<sup>8</sup> resulteert door het toonloos worden van de uitgangen in een distinctie tussen adjectieven waarvan de stam uitgangsluys is en adjectieven waarvan de stam op *-e* eindigt (Te Winkel 1901; Van Helten 1973). In de volgende voorbeelden, die overgenomen zijn uit Van Helten (1973: 396-397), is de *-e* aan het eind van het adjectief geen buigingsuitgang maar een deel van de stam:

- (1) scone weder, een vreemde dier, een woeste land, een niwe spel

Doordat het onderscheid tussen beide lexicale klassen vormelijk identiek is aan het onderscheid tussen beide flexiesystemen voor de masculiene en onzijdige nominatief en de onzijdige accusatief, wordt de stam-sjwa regelmatig als buigingsmorfem geïnterpreteerd en bijgevolg weggelaten in een onbepaalde context:

- (2) een scoen wesen, scoen gelaet, vreemt oly, een woest eilant  
(voorbeelden overgenomen uit Van Helten 1889: 398)

Deze reïnterpretatie werkt ook andersom: het predicatief gebruikte adjectief, dat vanaf het Middel-Nederlands onverbogen is, wordt in (3) naar analogie met de oude *jo/ja*-stammen verbogen:

- (3) wonderlike es sijn doen  
(voorbeeld overgenomen uit Van Loey 1970: 151)

De lexicale adjectiefklassen verdwijnen aan het eind van het Middel-Nederlands door de apocope van de auslautende sjwa (van Loon 1986: 97-99; Marynissen 2004). Door de dubbele werking van analogie vervaagt niet alleen het onderscheid tussen beide adjectiefklassen, maar ook het onderscheid tussen de vormen van de zwakke en de sterke flexie in de onzijdige nominatief en accusatief en in de masculiene nominatief. Dit werkt de vervlakking van het flexiesysteem in de hand door het voorkomen van sterke vormen in definitieve contexten (4a) en het voorkomen van zwakke vormen in indefinitieve contexten (4b):

- (4) a. dat half jaer, dat arm wijf, tvalsc venijn, dat swart sabelijn  
(voorbeelden overgenomen uit Van Helten 1973: 402)  
b. een ganse lant, een engelsche leven, een hovesche wijf  
(voorbeelden overgenomen uit Van Helten 1973: 395)

Daarnaast is er een algemene neiging om de buigings-*e* te apocoperen in adjectieven eindigend op *-el*, *-en* en *-er*, zowel in de sterke als in de zwakke buiging. Tot deze groep behoren ook de deelwoorden op *-en* en de stofnamen op *-en* of *-ijn*:

- (5) dat edel diet, dit wandel aerdsche goet, in een doncker stat, beloken  
duren, die drunken liede, die steinen pilare  
(Raidt 1968: 10-11; Van Helten 1973: 392-393; Van Loey 1949: 34)

De klankwettige verdoffing van het flectionele systeem wordt gecompenseerd door een toenemend gebruik van het aanwijzend voornaamwoord en het lidwoord. Onder invloed van het lidwoord en het aanwijzend voornaamwoord die het adjectief voorafgaan en volgens de sterke flexie verbogen worden, is er een verloop van de zwakke naar de sterke flexie in contexten waar oorspronkelijk de zwakke flexie de regel is (van Bree 1980: 247; Peters 1937: 357; Van Helten 1973: 400; Van Loey 1949: 33, 1970: 149). Deze tendens ontstaat aan het eind van het Oud-Nederlands (van Loon 1988; van den Toorn et al. 1997: 51). Dit proces vervangt bijvoorbeeld in de masculiene en feminiene genitief enkelvoud progressief de ‘regelmatige’ zwakke vormen in (6a) door de overeenkomstige sterke vormen in (6b):

- (6) a. des goeten mans, der goeden vrouwen  
b. des goets mans, der goeder vrouwen  
(voorbeelden overgenomen uit Van Loey 1970: 149)

De onderstreepte morfemen in (6b) illustreren de assimilatie van de adjectivische uitgang met de uitgang van de determinator. De onzijdige nominatief en accusatief enkelvoud evenals de mannelijke nominatief enkelvoud vormen uitzonderingen op dit proces (Van Helten 1973: 400). Het paradigma in tabel 2(4) toont het resultaat van het verloop van de zwakke naar de sterke buiging voor het attributief gebruikte adjectief bij een enkelvoudig onzijdig substantief:

<b>Nom.</b>	<i>blinde, blint</i>
<b>Gen.</b>	<i>blind(e)s</i>
<b>Dat.</b>	<i>blinden</i>
<b>Acc.</b>	<i>blinde, blint</i>

Tabel 2(4): Adjectivische buiging van blint voor het onzijdige genus aan het eind van het Middel-Nederlands

De belangrijkste vaststelling voor onderhavig onderzoek is het verdwijnen van het onderscheid tussen sterke en zwakke buiging aan het einde van het Middel-Nederlands, behalve in de onzijdige nominatief en accusatief en de masculiene nominatief, die uitzonderingen vormen op de algemene tendens tot assimilatie van de adjectivische zwakke flexie met de sterke flexie van de voorafgaande determinator. In deze contexten blijft het onderscheid tussen de sterke/onverbogen en de zwakke/verbogen vormen gehandhaafd: de verbogen vorm wordt in definitie contexten geselecteerd en de onverbogen vorm in indefiniete contexten. Dit onderscheid is echter niet absoluut. De hierboven aangehaalde variatie tussen verbogen en onverbogen vormen blijft in deze contexten optreden: in definitie contexten komt naast de verbogen ook de onverbogen vorm voor zoals in (4a), en in indefiniete contexten komt naast de onverbogen ook de verbogen vorm voor zoals in (4b). Na een bezittelijk voornaamwoord volgt het adjectief oorspronkelijk de sterke verbuiging, zoals in *u daghelics broot* of *ons vorseit godshuus* (Van Helten 1973: 406). Naar analogie met het gebruik van de zwakke vormen na het aanwijzend voornaamwoord en het hiervan afgeleide bepaald lidwoord, vindt men naar het einde van het Middel-Nederlands toe ook zwak verbogen adjectieven na het bezittelijk voornaamwoord, zoals in *syn goede swaert* en *syn grise haer* (Van Helten 1973: 406).<sup>9</sup> De vormen van de trappen van vergelijking worden oorspronkelijk uitsluitend zwak maar later ook sterk verbogen (Raidt 1968; Van Helten 1973).

Tot slot kijken we naar de andere adnominale elementen, met name het bezittelijk voornaamwoord, het aanwijzend voornaamwoord en het lidwoord. Deze woordsoorten worden, op enkele uitzonderingen na, in adnominale positie sterk verbogen. Onbepaalde voornaamwoorden en lidwoorden zijn oorspronkelijk sterk verbogen maar later komen ook zwak verbogen vormen voor (Raidt 1968: 13). De verbuiging van het bezittelijk voornaamwoord is die van het sterke adjectief (Te Winkel 1901), maar als gevolg van het proclitische gebruik komen al vroeg flexie-loze vormen voor in alle genera: *ons vader*, *naar ons zin* (Van Loey 1970: 144).<sup>10</sup> In het hedendaags Nederlands komt enkel nog de onverbogen vorm van het bezittelijk voornaamwoord voor, met uitzondering van de dyade *onze/ons* die een genus-onderscheidende functie heeft: *onze* wordt bij *de*-woorden en meervoudsvormen gebruikt, *ons* bij enkelvoudige *het*-woorden (zie ook paragraaf 8.1.2).

Als gevolg van de desintegratie van het flexiesysteem vertoont het Middel-Nederlands progressief meer analytische kenmerken, waaronder een striktere woordorde en het gebruik van functiewoorden, zoals de veralgemening van het

gebruik van het lidwoord ter onderscheiding van de sterke en de zwakke verbuiging (Raidt 1968).<sup>11</sup>

Voor de periode van het Middel-Nederlands besluiten we dat het adjectivische flexiesysteem sterk gereduceerd is door het vervagen van de grenzen tussen de casussen, door het verdwijnen van de lexicale klassen, en door de uitholling van het onderscheid tussen de sterke en de zwakke verbuiging. Alleen in de masculiene en onzijdige nominatief en in de onzijdige accusatief is er nog sprake van een onderscheid tussen zwakke en sterke uitgangen, maar dit onderscheid is niet vrij van variatie als gevolg van de wederzijdse beïnvloeding tussen beide flexiesystemen.

De transitie van het Middel-Nederlands naar het Nieuw-Nederlands wordt gekenmerkt door de apocope van de auslautende sjwa (van Loon 1986). Aan het eind van het Middel-Nederlands ontstaat bijgevolg een gunstige situatie voor het verdwijnen van de restanten van het onderscheid tussen sterke en zwakke buiging in de nominatief en de accusatief. Van Loon (1986) en Van Haeringen (1937) melden dat de sjwa met morfeemstatus behouden blijft, zoals de sjwa in het verledentijdsuffix *-de/-te*. Deze functionele verklaring is echter niet van toepassing op de adjectivische buiging, die strikt genomen redundant geworden is door de veralgemening van het gebruik van het lidwoord (van Bree 1980; van Haeringen 1937; van Loon 1988). Dat de buigings-*e* toch niet geapocopeerd is, wordt verklaard door het inroepen van de masculiene accusatiefvorm op *-en* die tot *-e* gereduceerd is door de apocope van de eind-*n* (van Bree 1980; van Haeringen 1937; Van Loey 1970). Met andere woorden: *-en* is tot *-e* geapocopeerd, en *-e* is tot zero geapocopeerd. Deze verklaring wordt gestaafd door de algemene tendens tot apocopering van de buigings-*e* bij vrouwelijke substantieven in de zuidelijke dialecten (van Haeringen 1937; Peters 1937; Raidt 1968). Volgens deze verklaring zou dan ook het algemene gebruik van de onverbogen vorm bij onzijdige substantieven in de zuidelijke dialecten verklaard kunnen worden als een gevolg van de apocope van de eindsjwa.<sup>12</sup> In paragraaf 2.3.1 komen we terug op het veralgemeende gebruik van de sterke vorm in de zuidelijke dialecten.

### 2.3 De adnominale buiging in het Nieuw-Nederlands

Voor de periode na het Middel-Nederlands, die meestal als het Nieuw-Nederlands wordt aangeduid, overlopen we de evolutie van de adnominale flexie eeuw per eeuw. Naast het onderzoek van Raidt (1968), maken we gebruik van Van der Horst & van der Horst (1999), De Rooij (1980b), Te Winkel (1901) en Van Loey (1970). Door het talrijker wordende materiaal kan het taalgebruik, zoals bij Raidt (1968), in functie van geografische en stilistische factoren gedifferentieerd worden.



### 2.3.1 *Het Nederlands in de zestiende eeuw en de divergentie tussen de zuidelijke en noordelijke dialecten*

Tijdens de zestiende eeuw gaat de uitholling van het flexiesysteem voort, vooral op het vlak van de casussen. Zo worden bij feminiene en onzijdige substantieven de adjectivische vormen van de nominatief en de accusatief vermengd. Van den Toorn et al. (1997: 219) melden dat in het Nieuw-Nederlands vaak alleen het lidwoord nog casusmarkering draagt.

Als we het onderscheid tussen de zwakke en de sterke buiging bij onzijdige substantieven bekijken, stellen we een veralgemening vast van de onverbogen/sterke vorm in indefiniete contexten. In haar schematische overzicht maakt Raidt (1968: 33-34) geen gewag van afwijkende buigingsvormen in deze context. In definitie contexten is de verbogen/zwakke vorm duidelijk de basisvorm, maar ondervindt hij nog steeds concurrentie van het onverbogen/sterke alternatief, zoals aangetoond door *het schoon uer-werek* of *dat vroom ende cloeck feyt* (Raidt 1968: 23). Daarnaast zijn er uiteenlopende factoren die de selectie van de onverbogen vorm na een definitie determinator vergemakkelijken. Na een bezittelijk voornaamwoord is de onverbogen vorm frequent, zoals in *u zot bestier* (Raidt 1968: 23). Fonologische factoren spelen eveneens een rol: meerlettergrepige adjectieven op *-lijk* en *-igh* verkiezen bijna altijd de onverbogen vorm, zoals in *het namelyk woord* en *dat aerdich juweelken* (Raidt 1968: 23), net als adjectieven waarvan de stam op een diftong eindigt, zoals in *mieu casteel* of *dit luij maeijken* (Raidt 1968: 24).

Raidt (1968: 33-34) stelt eveneens lectische effecten vast. Geografisch onderscheidt ze Hollandse en Brabantse bronnen: Hollandse documenten vertonen een duidelijke voorkeur voor de verbogen variant in definitie contexten, terwijl in de Brabantse documenten de onverbogen variant naast de verbogen variant voorkomt. De Brabantse bronnen vertonen stilistische verschillen: in Brabantse archiefstukken is de onverbogen vorm de norm; literaire teksten beschouwen de onverbogen vorm veeleer als afwijkend; kluchten vertonen een onduidelijke situatie met een voorkeur voor de verbogen vorm in de nominatief en een voorkeur voor de onverbogen vorm in de accusatief.

Bij de vormen van de comparatief is de onverbogen variant de meest frequente. Het beeld voor de deelwoorden en de adjectieven op *-en* is net als voor de stofnamen enigszins chaotisch met een voorkeur voor de onverbogen variant (Raidt 1968: 29-31). Voor de overige adnominale elementen vermeldt Raidt (1968: 30-31) het wijdverspreide gebruik van de onverbogen vorm *ons*, die in het Middel-Nederlands uit het proclitische gebruik van de bezittelijke voornaamwoorden is ontstaan.

De geografische differentiatie tussen de Brabantse en de Hollandse bronnen die we hierboven hebben vastgesteld, wordt nog steeds in de hedendaagse dialecten weerspiegeld. Vanaf de Middeleeuwen kent het gebruik van de onverbogen vorm van het adjectief na het bepaald lidwoord en na een bepaald voornaamwoord uitbreiding in het zuiden van het taalgebied en geleidelijk wordt dit de algemene

zuidelijke variant (de Rooij 1980b). Op basis van de toestand van het Brabants zoals hoger in deze paragraaf beschreven, wordt aangenomen dat deze evolutie in de zestiende eeuw haar eindpunt nog niet bereikt heeft.

Taeldeman (1980: 224-225) stelt dat het flexiesysteem in de zuidelijke dialecten gesimplificeerd is, terwijl het behoud van het onderscheid tussen verbogen en onverbogen vormen in de noordelijke dialecten een historisch relict is. Als verklaring haalt hij de invloed van het voorafgaande voornaamwoord aan. Een analoge verklaring wordt door Daan (1969) aangereikt. Op basis van de analyse van de geografische verspreiding van beide adjectivische buigingsvarianten in de sequentie *ons oud(e) paard* concludeert zij:

“het lijkt zo te zijn dat in de zuidelijke [Vlaanderen, Zeeland, Noord-Brabant, Limburg en Gelderland tot de Waal en de Linge; JT] en noordelijke [Friesland, Groningen en Noord-Drenthe; JT] delen van ons taalgebied de assimilatie in verbuiging van het adjectief aan het pronomen zich heeft voltrokken, maar dat het onderscheid tussen de zwakke en de sterke flexie in het midden [Noord-Holland, Zuid-Holland, Utrecht, het noorden van de Betuwe, de Veluwe, de Lijmers, de Graafschap en Twente; JT] bewaard is gebleven” (Daan 1969: 114)

Daans bevindingen worden ondersteund door het onderzoek van De Rooij (1980b) naar de adjectivische buiging in *het bruin(e) paard*, *ons bruin(e) paard* en *een bruin(e) paard*. Hij plaatst evenwel enkele kanttekeningen bij de conclusies van Daan (1969), waardoor een genuanceerd beeld ontstaat. Enerzijds stelt hij dat “het woord *oud* [...] niet zo geschikt is voor het onderkennen van de oppositie verbogen-onverbogen” (de Rooij 1980b: 110), door het frequent wegvallen van de eindmedeklinker (zie ook de Rooij 1972 en Taeldeman 1980). Om die reden vervangt hij het adjectief *oud* door het adjectief *bruin*. Anderzijds analyseert hij de buigingsalternantie niet enkel na een bezittelijk voornaamwoord, een context die de selectie van de onverbogen variant vergemakkelijkt (cf. *supra*), maar ook na een bepaald en een onbepaald lidwoord. We becommentariëren nu de resultaten van De Rooijs onderzoek.

De dialectkaart voor *een bruin(e) paard* bevat slechts vijf meetpunten met een verbogen adjectief. De Rooij stelt dat het hier hoogstwaarschijnlijk om vergissingen gaat.<sup>13</sup> Na een bepaald lidwoord of een bezittelijk voornaamwoord is er een homogeen onverbogen zone in het zuiden van het taalgebied die Vlaanderen, Noord-Brabant en Nederlands-Limburg omvat. De vergelijking van de buiging na een bezittelijk voornaamwoord en na een bepaald lidwoord leert ons dat er enkele menggebieden zijn in het noordoosten van Nederland (Friesland, Groningen, Oost-Drenthe), in het rivierengebied en in Zeeuws-Vlaanderen. In deze gebieden is de verbogen variant dominant na het bepaald lidwoord. Na een bezittelijk voornaamwoord, meer bepaald na *ons*, is er in de noordoostelijke gebieden een voorkeur voor de onverbogen variant, terwijl in de andere menggebieden beide varianten een vergelijkbare verspreiding hebben.

De Rooij (1980b: 113-119) toetst deze bevindingen aan de beschrijvingen in dialectgrammatica's. Na een onbepaald lidwoord vermelden alle dialectgrammatica's uitsluitend het gebruik van het onverbogen adjectief. Na een bepaald lidwoord bevestigen de dialectgrammatica's in grote lijnen het beeld van de dialectkaarten. Het zuiden van het taalgebied selecteert, enkele uitzonderingen niet te na gesproken, consistent de onverbogen variant in een definiete context. In Friesland, Groningen en de rest van het noordoosten is de verbogen variant de norm na een definiete determinator. Voor Groningen en Friesland rapporteren meerdere dialectgrammatica's het gebruik van de onverbogen variant na een bezittelijk voornaamwoord. In dialectgrammatica's voor de provincie Noord-Holland en het westen van Nederland is de verbogen vorm de norm, ook na een bezittelijk voornaamwoord.

In de zuidelijke dialecten is het verschil tussen sterke en zwakke flexie dus volledig verdwenen, ook bij *het*-woorden. Peters (1937: 358) legt een verband met het behoud van het driegerasysteem in de zuidelijke dialecten<sup>14</sup>: "Wij hebben gezien, dat in het Zuiden de genusaanduiding zeer sterk is, en hier hebben wij, juist door het verschil tusschen den onzijdigen en den manneliken of vrouwelijken vorm zulk een genusaanduiding gekregen". Taeldeman (1980: 224) presenteert de volgende onderliggende buigingsvormen voor het attributieve adjectief in de zuidelijke dialecten:

	Enkelvoud	Meervoud
<b>Mannelijk</b>	-★n	
<b>Vrouwelijk</b>	-★	-★
<b>Onzijdig</b>	-∅	

Tabel 2(5): Onderliggende buigingsvormen van het attributieve adjectief in de zuidelijke dialecten

Dit zijn de bevindingen van het overzicht van het dialectologische onderzoek naar de adjectivische buiging bij neutra:

- Na een onbepaald lidwoord selecteren alle dialecten in het Nederlandse taalgebied de onverbogen variant.
- Na een bepaald lidwoord en een bezittelijk voornaamwoord selecteren de dialecten in kern-Nederland het verbogen adjectief. Het zuiden van het taalgebied gebruikt consistent het onverbogen adjectief bij een onzijdig *het*-woord. Dit is in mindere mate het geval bij de dialecten in het noordoosten van Nederland en Zeeland, waar de onverbogen vorm van het adjectief na een bezittelijk voornaamwoord geselecteerd kan worden. Op deze geografische verschillen komen we terug bij de analyse van de lectische conditionering van de buigingsalternantie (zie paragrafen 12.2.2.1 en 12.3.1).

In de rest van dit historische overzicht concentreren we ons op de noordelijke contreien waar de standaardtaal ontwikkeld is.<sup>15</sup>

### 2.3.2 *Het Nederlands in de zeventiende eeuw*

Na de val van Antwerpen in 1585 vindt een migratie plaats van de zuidelijke naar de noordelijke Nederlanden. Dit heeft – vooral tijdens de eerste helft van de zeventiende eeuw – een weerslag op het taalgebruik: tot ongeveer 1650 overheerst de zuidelijke – Brabantse – variëteit, terwijl na 1650 de noordelijke – Hollandse – variëteit dominant wordt. Kenmerkend voor de zeventiende en de achttiende eeuw zijn de pogingen van grammatici tot instandhouding en reconstructie van het casussysteem (van der Horst & van der Horst 1999: 311), terwijl in de niet-literaire taal het casussysteem steeds verder erodeert.

In indefiniëte contexten is het onverbogen adjectief de regel bij onzijdige substantieven en is de verbogen variant hoogst zeldzaam. In definiëte contexten daarentegen heerst onzekerheid en hebben beide buigingsvarianten nagenoeg dezelfde frequentie (Raidt 1968: 67). Deze onzekere situatie wordt weerspiegeld door de grammatici die de variatie erkennen maar het niet eens raken over een norm: sommigen beschouwen de onverbogen vorm als de basis, anderen de verbogen variant (Raidt 1968: 67-70).

De populariteit van de onverbogen vorm in de zeventiende eeuw is toe te schrijven aan de invloed van het Brabants, waar de onverbogen variant de norm is, zoals kan worden afgeleid uit de hedendaagse Vlaamse dialecten (zie paragraaf 2.3.1). Dit verklaart wellicht de voorkeur voor de onverbogen vorm in de literaire teksten, terwijl de verbogen pendant dominant is in niet-literaire teksten. Van Marle (1995a) stelt dat de situatie in de zeventiende eeuw gemarkeerd is door de aanwezigheid in het noorden van zuidelijke migranten wier taal als prestigieus ervaren wordt, althans in de eerste helft van de eeuw. Zoals aangegeven in de inleidende alinea kan deze eeuw in twee gesplitst worden: tijdens de eerste helft van de eeuw is het onverbogen adjectief de dominante vorm onder invloed van de migranten uit het zuiden; na 1650 wordt het Hollands de dominante variëteit en wordt het verbogen adjectief de basisvariant. Raidt (1968: 78-79) besluit dat het buigingssysteem aan het eind van de zeventiende eeuw relatief stabiel is en niet noemenswaardig afwijkt van het hedendaagse gesproken Hollands.

Na de bespreking van de algemene tendensen bij de alternantie tussen verbogen en onverbogen adjectieven, sluiten we de bespreking van het zeventiende-eeuwse Nederlands af met het aanstippen van enkele bijzonderheden (Raidt 1968: 73-78). Bij de vormen op *-en* duurt de chaotische situatie van de zestiende eeuw voort: naast *-en* (*open*) komen ook *-ene* (*opene*) en *-e* (*ope*) voor. Bij de stoffelijke adjectieven wordt *-en* soms als een flexie-uitgang geïnterpreteerd, zodat men ook *aerde* en *kopere* aantreft. De vormen van de trappen van vergelijking selecteren tegengestelde buigingsvormen: de superlatief eindigt altijd op *-ste*; de comparatief verkiest overwegend de onverbogen variant.

### 2.3.3 *Het Nederlands in de achttiende eeuw*

In de achttiende eeuw is de adnominale flexie relatief stabiel, behalve bij neutra. Grammatici als Lambert Ten Kate bekritisieren enerzijds het gebruik van de verbogen vorm na *een*, *geen* en een nul-determinator, en anderzijds het gebruik van de onverbogen vorm na een bepaald lidwoord en een aanwijzend of bezittelijk voornaamwoord (Raidt 1968: 83). David van Hoogstraten is dan weer van mening dat in een definiete context beide buigingsvarianten aanvaardbaar zijn: naast *het heerlyke werk* aanvaardt hij ook *het heerlyk werk*. Met betrekking tot de verbuiging van het adjectief na een bezittelijk voornaamwoord is er een interessante polemiek tussen Ten Kate en Huydecoper: terwijl Ten Kate de verbogen variant *myn soete kint* voorschrijft, oordeelt de conservatievere – aldus Van der Horst & van der Horst – Huydecoper dat de onverbogen vorm wenselijk is (van der Horst & van der Horst 1999: 315; Raidt 1968: 83-84). Uit de beoordeling van Ten Kate, die veeleer ‘descriptief’ dan ‘prescriptief’ is, kunnen we afleiden dat de verbogen vorm in de achttiende eeuw opgeld doet na een bezittelijk voornaamwoord naar analogie met het gebruik van de verbogen vorm in andere definiete contexten.

In tegenstelling tot de zeventiende eeuw, waar de selectie van de adjectivische buigingsvariant bij neutra in definiete contexten sterk geografisch bepaald is, is de buigingsalternantie in de achttiende eeuw veeleer stilistisch bepaald: Raidts tekst-analyse suggereert dat de onverbogen variant vooral het emphatische en affectieve taalgebruik kenmerkt (Raidt 1968: 85-90).

Voltooid deelwoorden op *-en* worden verbogen, wellicht in een poging tot systematiek (van der Horst & van der Horst 1999: 315-316). Dezelfde trend tot systematiek verklaart eveneens dat de grammatici adjectieven op *-er* verbuigen. Het effect van deze voorschriften is duidelijk zichtbaar in de literaire taal, maar veel minder in niet-literaire teksten, zoals reisjournalen, die in aansluiting met de zeventiende eeuw de onverbogen vorm blijven hanteren: *het ander land*, *het helder licht* (Raidt 1968: 94). De vormen van de vergrotende trap blijven onverbogen, zoals tijdens de vorige eeuwen: *het hooger einde* (Raidt 1968: 94). De *-en* aan het eind van stoffelijke adjectieven wordt al in de zeventiende eeuw als een flexiomorfeem geïnterpreteerd, zodat ook in de achttiende eeuw nog vaak de *-e* voorkomt met verlies van de *-n*, zoals in *het stale wapen*, en het nul-morfeem na *een* of *geen*, zoals in *een stael wapen* (Raidt 1968: 93).<sup>16</sup>

Bij een mannelijke persoonsnaam kent Ten Kate na een indefiniete determinator een semantische motivatie toe aan het onderscheid tussen beide buigingsvarianten:

- (8) a. *een groot man* : “ijmand van groot bedrijf of bestier”  
 b. *een groote man* : “een die hoog en breed van statuer is”  
 (voorbeelden en citaten overgenomen uit Raidt 1968: 91)

Deze semantische distinctie vinden we in hedendaagse grammatica’s terug.

### 2.3.4 *Het Nederlands in de negentiende eeuw*

Het stilistische onderscheid tussen de ‘lage’ verbogen en de ‘hoge’ onverbogen vorm bij een *het*-woord na een bepaald lidwoord of een aanwijzend voornaamwoord wordt overgedragen van de achttiende naar de negentiende eeuw. Na een bepaalde determinator is het gebruik van de onverbogen variant een kenmerk van deftig taalgebruik; de verbogen variant is typisch voor het gewone taalgebruik. Na een bezittelijk voornaamwoord blijft het gebruik van het verbogen adjectief veeleer uitzonderlijk (Raidt 1968: 95).

Uit het overzicht van een aantal grammatica's in De Rooij (1980b: 123-124) blijkt dat de onverbogen vorm de norm is na onbepaalde determinatoren en dat de verbogen vorm na het bepaald lidwoord en het aanwijzend voornaamwoord gebruikt wordt. Er bestaat geen eenduidigheid over de keuze van de buigingsvariant na een bezittelijk voornaamwoord: nu eens wordt de verbogen vorm voorgeschreven, dan weer de onverbogen vorm. Eén grammatica vermeldt het gebruik van de onverbogen vorm in tweelettergrepige en langere adjectieven: *het liefderijk aanbod* of *het vaardig antwoord* (de Rooij 1980b: 124).

Terwey relateert de selectie van beide buigingsvormen na een definiete determinator voor het eerst aan een betekenisverschil: de onverbogen vorm wordt gebezigd wanneer “men de hoedanigheid noemt om de hoedanigheid”, terwijl de verbogen vorm “de hoedanigheid noemt als een middel tot onderscheiding eener zelfstandigheid van andere, welke die hoedanigheid niet bezitten” (de Rooij 1980b: 124):

- (9) a. *mijn oud huis* : “mijn huis dat oud is”  
 b. *mijn oude huis* : “mijn vroegere huis”  
 (voorbeelden overgenomen uit De Rooij 1980b: 124)

Net als in de achttiende eeuw worden voltooid deelwoorden op *-en* omwille van de systematiek verbogen (van der Horst & van der Horst 1999: 315-316). De stoffelijke adjectieven vertonen nog steeds variatie: vormen als *zijn goud horloge* en *'t staal rapier* komen voor naast de vormen op *-en* (Raidt 1968: 95).

De schrijftaal is sterk beïnvloed door de regels van Huydecoper, waardoor een opwaardering van oude vormen kan worden vastgesteld. Het gebruik van de casusvormen vinden we terug in grammatica's van het begin van de twintigste eeuw: Van der Horst & van der Horst (1999: 310) merken echter op dat deze vormen beperkt zijn tot de schrijftaal en in de gesproken taal in onbruik zijn geraakt.

## 2.4 Conclusie

Het historische overzicht van de adjectivische buiging toont aan dat het globale proces er een van deflexie is. Ten eerste, het casussysteem is verdwenen, enkele idiomatische tournures zoals *in koelen bloede* of *van ganser harte* niet te na gesproken. Ten tweede, het onderscheid tussen de verschillende adjectivische

stammen is verdwenen. Ten derde, het onderscheid tussen de zwakke en de sterke verbuiging is verdwenen, behalve bij *het*-woorden waar het onderscheid tussen het verbogen en het onverbogen adjectief syntactisch bepaald is: de onverbogen vorm komt enkel voor in indefiniete NP's. Bij masculiene *de*-woorden voorziet de grammatica een semantisch geconditioneerde alternantie tussen verbogen en onverbogen vormen (zie ook hoofdstuk 3).

De contexten waar in het hedendaags Nederlands variatie optreedt, zijn al in vroegere stadia variatiegevoelig, met name de nominatief- en accusatiefvormen van het attributieve adjectief bij *het*-woorden en de nominatiefvormen bij masculiene substantieven, meer bepaald bij indefiniete persoonsbepalingen.

Wanneer we de evolutie van de adjectivische buiging(salternantie) bij *het*-woorden beschouwen, komen we tot deze conclusies:

- In indefiniete contexten vertoont het adjectief al vanaf de zestiende eeuw een zeer sterke voorkeur voor de sterke/onverbogen variant.
- In definiete contexten is de buiging van het adjectief bij een *het*-woord doorheen de evolutie van het Nederlands aan variatie onderhevig geweest, met nu eens de sterke/onverbogen variant en dan weer de zwakke/verbogen variant als dominante vorm. De zestiende eeuw vormt een breekpunt (zie paragraaf 2.3.1): terwijl de zuidelijke dialecten in alle contexten met een enkelvoudig *het*-woord het gebruik van de onverbogen vorm beginnen te generaliseren, stellen we in de noordelijke dialecten – en de ontluikende standaardtaal – een toenemende voorkeur voor de verbogen variant vast. Deze divergente evoluties worden nog steeds in de hedendaagse dialecten en – zij het in mindere mate – in de nationale variëteiten weerspiegeld (zie paragraaf 1.1.4).
- De adjectivische buiging na een bezittelijk voornaamwoord neemt een speciale positie in doordat de adjectieven in deze context vaker dan bij andere definiete determinatoren de onverbogen variant selecteren. Dit is wellicht een instantie van buigingsassimilatie tussen het onverbogen voornaamwoord en het daaropvolgende adjectief.
- De factoren die de keuze tussen beide buigingsvarianten bepalen, zijn zeer divers en vertonen gelijkenissen met de factoren die in het literatuur-overzicht besproken zijn (zie paragraaf 1.1): de woordsoort van de determinator, de lengte van het adjectief, de prosodische structuur van het adjectief (in de vorm van de suffigering), de semantische relatie tussen het adjectief en het substantief, de regionale variëteit en het register.

Dit korte – en allesbehalve exhaustieve – historische overzicht toont aan dat de adjectivische buigingsalternantie die we in het hedendaags Nederlands bij neutra aantreffen al sinds het Middel-Nederlands voorkomt. Daarnaast zijn verscheidene verklarende factoren die in het empirisch onderzoek in deel II aan bod komen al meerdere eeuwen actief.

## Noten

1. Historische grammatica's van het Nederlands, zoals Te Winkel (1901), Van Bree (1980) en Van Loey (1970 [1959]), besteden – althans wat de adjectivische buiging betreft – hoofdzakelijk aandacht aan de periode van het Middel-Nederlands. Van den Toorn et al. (1997) bekijken ook het Oud-Nederlands en het Nieuw-Nederlands. Naast deze algemene naslagwerken hebben we de volgende referenties geraadpleegd voor de specifieke historische stadia. De beschrijving van het Indo-Europees en het Oer-Germaans is hoofdzakelijk gebaseerd op Prokosch (1939) en Krahe & Meid (1969). De gegevens over het Oud-Nederlands zijn afkomstig van het onderzoek van Van Loon (1988) naar de sterke en zwakke flexie in het Oud-Nederlands en van de grammatica van Quak & van der Horst (2002). Voor het Middel-Nederlands hebben we naast de rijk gedocumenteerde analyse van Van Helten (1973 [1889]), de grammatica van Van Loey (1949) en het inleidend overzicht van Van Kerckvoorde (1993) geraadpleegd. Van der Horst & van der Horst (1999) bespreken de ontwikkeling van de adjectivische buiging vanaf de achttiende eeuw tot en met de hedendaagse toestand.
2. Voor een uitgebreide bespreking van het dialectologische onderzoek naar de adjectivische buigingsalternantie bij neutra verwijzen we naar paragraaf 2.3.1. De eerste aflevering van de *Morfologische Atlas der Nederlandse Dialecten* die in het voorjaar van 2005 verschenen is, behandelt inherente inflectionele verschijnselen (de nominale meervoudsvorming en de diminuering).
3. In het Gotisch zijn de *i-* en *u-*stammen volledig door de klasse van de *jo/ja*-stammen geabsorbeerd (Prokosch 1939). In het Oud-Hoog-Duits en het Oud-Saksisch worden de *jo/ja*-stammen verbogen zoals de *o/a*-stammen (Krahe & Meid 1969).
4. De *Wachtendonckse Psalmen* zijn een Oud-Nederlandse bewerking van een Hoog-Duits origineel. De bewerking wordt omstreeks 950 gedateerd.
5. *Sîn* wordt oorspronkelijk enkel reflexief gebruikt. Daarnaast zijn er ook attestaties van de vorm *haro* (van den Toorn et al. 1997: 52-53).
6. Het Middel-Nederlands wordt meestal gedateerd als de periode van 1200 tot 1500.
7. Te Winkel (1901) maakt als enige gewag van een alternantie bij de nominatief en de accusatief: naast *blint* vermeldt hij ook *blinde*. Het is echter onduidelijk of hij hiermee de alternantie tussen de sterke en de zwakke buiging identificeert die voorkomt bij de vormen van de nominatief en de accusatief, of dat dit de voorzetting is van de oorspronkelijke nominale en pronominale vormen uit het (Oer-)Germaans (zie tabel 2(1)).
8. Deze klasse had de *i-* en *u-*stammen opgeslokt (zie paragraaf 2.1; Te Winkel 1901; Van Helten 1973).
9. In het Middel-Nederlands wordt het attributieve adjectief ook gepostponeerd ten overstaan van het nominale hoofd. In deze positie wordt het adjectief doorgaans niet meer verbogen, zoals in *eenre mouwen wit* of *in den brant scaerp* (Van Helten 1973: 413). Al in het Oud-



Nederlands is het gebruik van het gepostponeerde adjectief en het gepostponeerde bezittelijk voornaamwoord zeldzaam (van den Toorn et al. 1997: 60-61).

10. Van Kerckvoorde (1993: 67) verklaart de verspreiding van *ons* en *u* ten nadele van de respectievelijke klankwettige vormen *onse* en *uwe* als een instantie van paradigmatische nivellering met de andere bezittelijke voornaamwoorden *mijn*, *dijn* en *sijn*.
11. Van Loon (1986) is de mening toegedaan dat de veralgemening van het lidwoord noodzakelijk was om syncretismen te voorkomen, veeleer dan om het onderscheid tussen definitieve en indefinitieve NP's in stand te houden.
12. Deze verklaring heeft – naar onze mening – echter met drie problemen te kampen: ten eerste, het – door ons gekende – dialectologische onderzoek behandelt in deze context de apocope van de buigings-*e* bij vrouwelijke substantieven en niet bij onzijdige substantieven; ten tweede, dialectologen verklaren het gebruik van de sterke vorm in de zuidelijke dialecten van het taalgebied als een gevolg van de assimilatie met de buiging van de determinator (Daan 1969; Taeldeman 1980); tot slot blijft de vraag, ook voor het vrouwelijk, waarom de apocope van het buigingsmorfeem enkel de zuidelijke en niet de noordelijke dialecten heeft getroffen waar paradoxaal de oorsprong van dit proces gesitueerd is (Marynissen 2004; Van Loey 1970).
13. Oorspronkelijk zijn er 12 meetpunten met een verbogen vorm na een onbepaald lidwoord. Voor zeven van deze meetpunten kunnen de respondenten getraceerd worden. In alle zeven gevallen is er sprake van een vergissing. Op basis van deze vaststelling veronderstelt De Rooij dat de vijf niet-traceerbare verbogen varianten hoogstwaarschijnlijk ook vergissingen zijn.
14. Het driegerasysteem wordt vormelijk niet alleen in de adjectivische buiging gerealiseerd (Peters 1937; Taeldeman 1980), maar ook in de vormen van het bepaald lidwoord (van Ginneken 1934) en het onbepaald lidwoord (van Ginneken 1937).
15. Op basis van dit geografische onderscheid beperkt Raidt haar analyse van het Nieuw-Nederlands (vanaf de zeventiende eeuw) tot het Nederlands boven de rivieren. Een andere reden die deze keuze motiveert, is dat Raidts (1968) historische overzicht de vergelijkingsbasis vormt voor het Kaapse Nederlands en het Afrikaans.
16. Huydecoper schrijft in alle contexten het gebruik van de etymologische vorm op *-en* voor, zoals dit nog steeds in het hedendaags Nederlands het geval is.

## Hoofdstuk 3:

# Verbuiging van het attributieve adjectief in het Nederlands

In dit hoofdstuk stellen we bondig de adjectivische buiging in het Nederlands voor, zoals deze in ruimere of mindere mate in iedere grammatica van het Nederlands beschreven wordt. Deze beschrijving is hoofdzakelijk op de ANS (1997: 400-412) gebaseerd.

De beschrijving van de adjectivische buiging omvat drie delen. We vertrekken van het basismechanisme van de verbuiging van het geanteponeerde attributieve adjectief in het Nederlands (3.1) dat we vervolgens op twee manieren verfijnen: eerst identificeren we een reeks fonologische en morfosyntactische factoren die consistent één buigingsmorfeem selecteren, ongeacht wat de kenmerken van de syntactische context zijn (3.2); daarna richten we onze aandacht op de contexten waar het adjectief buigingsvariatie vertoont (3.3). Deze trapsgewijze verfijning van het buigingsmechanisme van het attributieve adjectief eindigt bij de buigingsalternantie die het onderwerp van dit onderzoek is, met name de alternantie tussen het verbogen en het onverbogen adjectief in een definiete NP met een enkelvoudig *het*-woord als hoofd.

De factoren die systematisch één buigingsvariant selecteren in paragraaf 3.2 worden uit de materiaalverzamelingen gefilterd die de empirische basis van dit onderzoek vormen (zie paragraaf 6.2 voor het *Corpus Gesproken Nederlands* en paragraaf 15.2 voor het *De Standaard*-corpus).

### 3.1 Basismechanisme voor de verbuiging van het geanteponeerde attributieve adjectief

In het hedendaags Nederlands heeft het geanteponeerde attributieve adjectief twee verschillende realisatievormen, de zogenaamde buigingsvormen: de onverbogen vorm die gelijk is aan de stam, en de verbogen vorm die ontstaat door toevoeging van het morfeem *-e* aan de stam. Hoewel de buiging van het adjectief in het Nederlands op het eerste zicht eenvoudig lijkt – er zijn slechts twee morfemen –, wordt de distributie van beide buigingsmorfemen door de combinatie van drie morfosyntactische parameters bepaald:

- het genus van het substantief: *de*-woord tegenover *het*-woord
- het getal van het substantief: enkelvoud tegenover meervoud
- de definietheid van de determinator: definiet tegenover indefiniet

Figuur 3(1) is een schematische voorstelling van het ‘regelmatige’ gebruik van de verbogen en de onverbogen vormen van het attributieve adjectief, zoals we dat in de meeste grammatica’s (*inter al.* ANS 1997: 404-408; Broekhuis 1999: 205) en morfologische handboeken (*inter al.* Booij 2002c: 43; Booij & van Santen 1998: 99; de Haas & Trommelen 1993: 285) van het Nederlands aantreffen.<sup>1</sup>

	enkelvoud		meervoud
	definitief	indefinitief	
<b>de-woord</b>	-e	-e	-e
<b>het-woord</b>	-e	-∅	-e

Figuur 3(1): Verdeling van de buigingsvormen van het attributieve adjectief

Het buigingssysteem in figuur 3(1) wordt door de volgende vuistregel samengevat: een adjectief in attributieve positie wordt altijd verbogen (1a-d, 1f), behalve in een indefiniete NP met een enkelvoudig *het*-woord als hoofd (1e).

- (1)
- a. de mooie / \*mooi man ; de mooie / \*mooi vrouw
  - b. een mooie / \*mooi man ; een mooie / \*mooi vrouw
  - c. de mooie / \*mooi mannen ; de mooie / \*mooi vrouwen
  - d. het mooie / \*mooi kind
  - e. een \*mooie / mooi kind
  - f. de mooie / \*mooi kinderen

### 3.2 Systematische uitzonderingen op het basismechanisme

Het buigingssysteem zoals voorgesteld in figuur 3(1) en geïllustreerd in (1) is echter een idealisatie die geen rekening houdt met variatie. Alvorens de contexten te analyseren waar er sprake is van ‘echte’ variatie, i.e. de contexten waar de verbogen basisvorm met de onverbogen variant alterneert (zie paragraaf 3.3, figuur 3(2)), bekijken we eerst een aantal fonologische en morfosyntactische restricties die systematisch een verbogen of een onverbogen vorm vereisen, ongeacht wat de configuratie is van de morfosyntactische kenmerken binnen de NP, c.q. genus, getal en definitie. We nemen eerst de fonologische factoren onder de loep.

Adjectieven eindigend op een sjwa zijn onverbuigbaar (ANS 1997; Booij 1992; de Haas & Trommelen 1993; Royen 1953; Schultink 1962). Dit betekent dat er geen sjwa kan worden toegevoegd en dat de eindsjwa niet gedeleerd kan worden:

- (2) een roze / \*rozeë / \*roos jasje; een timide / \*timideë / \*timied kind

Wanneer een adjectief op een klinker eindigt, wordt het evenmin verbogen (ANS 1997; Booij 1992; Broekhuis 1999; de Haas & Trommelen 1993; de Rooij 1980a):<sup>2</sup>

- (3) de lila / \*lilae jurk; de indigo / \*indigoë japon; de kaki / \*kakië jekker;  
de flashy / \*flashy'e jas

Deze conditie moet verder verfijnd worden: wanneer de eindklinker een /u/, een /e/, een /y/ (4a) of een diftong (4b) is, neemt het adjectief wel een buigings-*e* volgens het mechanisme beschreven in figuur 3(1) (ANS 1997; Broekhuis 1999; de Haas & Trommelen 1993; Schultink 1962).

- (4) a. het moeë kind; het weeë gevoel; het ambigue voorbeeld  
b. het blije kind, het veie koren

Van Oostendorp (2000: 100) reikt een prosodische verklaring aan voor het verschillende buigingsgedrag van de voorbeelden in (3) en (4): wanneer de eindlettergreep onbeklemtoond is, wordt de onverbogen vorm gebruikt (3<sup>˘</sup>); wanneer de eindlettergreep beklemtoond is, de verbogen vorm (4<sup>˘</sup>).<sup>3</sup> Het resultaat is telkens een tweelettergrepige trocheïsche metrische voet aan de rechterwoordrand van het adjectief (zie paragraaf 10.1.1).

- (3<sup>˘</sup>) de líla jurk; de índigo japon; de káki jekker; de fláshy jas  
(4<sup>˘</sup>) a. het móeë kind; het wéeë gevoel; het ambigúe voorbeeld  
b. het blíje kind; het véie koren

Deze prosodische argumentatie verklaart bovendien het verplichte gebruik van de verbogen vorm in *een náë bloedverwant*.

Adjectieven op *-en* [★n] nemen geen buigings-*e* (5a)<sup>4</sup>, evenmin als de voltooid deelwoorden op *-en* (5b) (ANS 1997; Booij 1992, 1998, 2002b, 2002c; Broekhuis 1999; de Haas & Trommelen 1993: 315; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a; Schultink 1962):

- (5) a. de effen / \*effene baan; het eigen / \*eigene belang  
b. het vermeden / \*vermedene ongeval

Stofnamen worden niet verbogen, ongeacht of de origine inheems (6a) (c.q. de stofadjectieven op *-en*) of uitheems is (6b) (ANS 1997; Booij 1992, 2002c; Broekhuis 1999; Royen 1953). Hetzelfde geldt voor de van voorwerpen afgeleide adjectieven op *-en* in (6c) (Royen 1953):

- (6) a. de houten / \*houtene tafel; het gouden / \*goudene servies  
b. de plastic / \*plastique rugzak; de aluminium / \*aluminiume tank  
c. de planken / \*plankene vloer; de platelen / \*platelene vaas

Drie types adjectieven waarvan de stam op *-er* eindigt, worden nooit verbogen. De eerste groep wordt gevormd door de adjectieven op *-er* die van geografische namen zijn afgeleid (ANS 1997; Booij 1992, 2002c; Broekhuis 1999; de Haas & Trommelen 1993; de Rooij 1980a; Royen 1953):

DEEL I

- (7) de Maastrichter / \*Maastrichere ster; de Edammer / \*Edammere kaas;  
de Haarlemer / \*Haarlemere olie

De van religieuze ordes afgeleide adjectieven vormen de tweede groep onverbuigbare adjectieven op *-er* (de Haas & Trommelen 1993: 319; van Haeringen 1962; Royen 1953: 89-95):

- (8) de franciscaner / \*franciscanere monnik; de cisterciënzer /  
\*cisterciënzere abdij

De derde groep onverbuigbare adjectieven op *-er* omvat de afleidingen van hoofdtelwoorden op *-er* in de sequentie *de X-er jaren* (9) (Royen 1953: 86-88):

- (9) de tachtiger / \*tachtigere jaren

Na het overzicht van de fonologisch geconditioneerde uitzonderingen richten we nu onze aandacht op de morfosyntactisch geconditioneerde uitzonderingen. De adjectivische positie kan bezet worden door morfosyntactische vormen die consistent de verbogen of de onverbogen vorm selecteren. De vormen van de overtreffende trap worden in attributieve positie altijd verbogen (10a), zelfs in de zeldzame gevallen waar de superlatief bij een onzijdig substantief en een onbepaald lidwoord voorkomt (10b) (de Haas & Trommelen 1993: 287; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a: 12). Tot deze groep behoren ook de gelexicaliseerde superlatieven, zoals *voorst* en *achterst* (11):

- (10) a. het mooiste / \*mooist paard  
b. een sterkste / \*sterkst paard bestaat niet  
(11) een achterste / \*achterst kamertje is er niet in dit huis

Een uitzondering hierop zijn de vormen van de overtreffende trap die met *aller-*geprefigeerd zijn (Booij 2002c):

- (12) een allerliefst / \* allerliefste meisje; het ?allermooist / allermooiste meisje

Als het element in adjectivische positie een rangtelwoord is, eindigt het altijd op een *sjwa* (13) (Booij 1992: 127; Lebrun & Schurmans-Swillen 1966). Een hoofdtelwoord is daarentegen altijd onverbogen (14a), met uitzondering van *een* dat in attributieve positie altijd een buigings-*e* neemt (14b).

- (13) het derde / \*derd jaar; het tweede / \*tweed kind  
(14) a. zes / \*zesse kinderen; acht / \*achte mensen  
b. het ene / \*een kind

Kwantificerende adjectieven, zoals *beide* en *verschillende*, krijgen altijd een buigings-*e* omdat ze naar betekenis meervoudig zijn (15) (Booij 1992: 127). Een

bijzonder geval is *veel*, dat na een determinator altijd verbogen wordt, maar zonder voorafgaande determinator zowel verbogen als onverbogen kan zijn (16).<sup>5</sup> Aan deze reeks kunnen we ook *zelfde* toevoegen.

- (15) de beide / \*beid kinderen, \* het beid(e) kind; de verschillende / \*verschillend boeken, \* het verschillend(e) boek  
 (16) veel/vele boeken, de vele / \*veel boeken van Jan, \*het vel(e) boek

Tot slot is er nog een reeks idiosyncratische gevallen, zoals *gratis*, *privé*, *oblong*, *volbloed*, *halfbloed*, *rechter* en *linker*, die nooit de buigings-*e* selecteren (ANS 1997; Booij 1992; Broekhuis 1999; de Rooij 1980a; Royen 1953).

- (17) de gratis / \*gratisse krant; de oblong / \*oblonge figuur; de halfbloed / \*halfbloede merrie; de rechter / \*rechtere hand; de linker / \*linkere hand

### 3.3 Contexten met buigingsvariatie

Naast de fonologische en morfosyntactische factoren die systematisch één van beide buigingsvormen selecteren, zijn er tot slot een aantal contexten waar de verbogen basisvorm in mindere of meerdere mate met de onverbogen vorm alterneert. Figuur 3(2) herneemt het basissysteem uit figuur 3(1), maar vult het aan met variationele informatie (*inter al.* ANS 1997: 408-412; Booij 2002c: 47-48; Booij & van Santen 1998: 100; Broekhuis 1999: 210-215; de Haas & Trommelen 1993: 285; Odijk 1992; Royen 1953: 34-75):

	enkelvoud		meervoud
	definit	indefinit	
<b>de-woord</b>	-e/-∅ [a]	-e/-∅ [a][b]	-e/-∅ [a]
<b>het-woord</b>	-e/-∅ [c]	-∅	-e

Figuur 3(2): Verdeling van de buigingsmorfemen van het attributieve adjectief met inclusie van variatie

Zowel bij *de*-woorden als bij *het*-woorden zijn er contexten waar de ‘regelmatige’ verbogen vorm alterneert met de gemarkeerde onverbogen variant. Algemeen wordt aangenomen dat de conditionering van deze alternantie veel strikter is bij de *de*-woorden dan bij de *het*-woorden. De twee alternantietypes bij *de*-woorden, type [a] en type [b] in figuur 3(2), zijn – althans volgens de (introspectieve) literatuur – strikt semantisch geconditioneerd. Het derde alternantietype, type [c] in figuur 3(2), is complex(er).

Het eerste alternantietype, type [a], treedt op wanneer het adjectief-nomenpaar refereert aan “officieel vastgelegde titels die in hun geheel een bepaalde functie of bevoegdheid of een bepaald beroep aanduiden” (ANS 1997: 408-412). In zulke

gevallen blijft het adjectief zowel in het enkelvoud, na een definiete (18a) en een indefiniete determinator (18b), als in het meervoud (18c) onverbogen:

- (18) a. de bouwkundig ingenieur; de geheim agent; de eerstaanwend expert-meteoroloog; de waarnemend burgemeester  
 b. een bouwkundig ingenieur; een geheim agent; een eerstaanwend expert-meteoroloog; een waarnemend burgemeester  
 c. de bouwkundig ingenieurs; de geheim agenten; de eerstaanwend expert-meteorologen; de waarnemend burgemeesters

In sommige gevallen wordt er zelfs een minimaal paar gevormd op basis van de verbogen-onverbogen distinctie. Het verbogen adjectief in (19a) wijst op een toevallige relatie tussen het adjectief en het substantief: het is een arts die toevallig op de plaats van het ongeval aanwezig was. Het gebruik van de gemarkeerde onverbogen vorm in (19b) identificeert het adjectief-nomenpaar als een semantische eenheid en verwijst naar de arts bij wie iemand in behandeling is.

- (19) a. Het verkeersslachtoffer was er zo erg aan toe dat zelfs de *behandelende arts* flauw viel.  
 b. Kan ik de *behandelend arts* even spreken?  
 (voorbeelden overgenomen uit de ANS 1997: 409)

Het volgende voorbeeld toont een analoge oppositie: het verbogen adjectief in (20a) identificeert een melioratieve eigenschap van de hoogleraar in kwestie; de onverbogen variant in (20b) signaleert dat het adjectief-nomenpaar een semantische eenheid vormt, met name een academische titel.

- (20) a. Professor Janssen is een *buitengewone hoogleraar*. Hij is niet alleen een eminent onderzoeker maar is bovendien begaan met het lot van zijn studenten.  
 b. De *buitengewoon hoogleraar* kon niet kandideren voor het rectoraat.

Het tweede type buigingsalternantie dat bij *de*-woorden optreedt, type [b] in figuur 3(2), komt voor in indefiniete NP's (ANS 1997: 408-412; Kester 1996: 99-104; Stuurman 1989). We onderscheiden hier twee soorten restricties. In het eerste geval is het nominale hoofd het substantief *man* of *persoon*, of duidt het nominale hoofd een beroep of een functie aan (21a); in het tweede geval duidt het adjectief een eigenschap aan die afgeleid is van een geografische naam (21b).

- (21) a. een groot(e) schilder; een beroemd(e) dichter; een slecht(e) schaker  
 b. een Frans(e) filosoof; een Amerikaans(e) zakenman

Hoewel het substantief doorgaans een mannelijke persoonsnaam is, vinden we ook onverbogen adjectieven bij vrouwelijke persoonsnamen:

- (22) zij is een voortreffelijk(e) spreekster; zij is een groot(e) pianiste

Dit type buigingsalternantie is beperkt tot indefiniete NP's met een enkelvoudig substantief als hoofd. Indien aan één van deze condities niet voldaan wordt, is het gebruik van de onverbogen vorm uitgesloten: bij een definiete determinator of bij een meervoudsvorm selecteren de adjectieven uit de voorbeelden in (21) en (22) altijd de verbogen vorm.

- (21') a. de grote / \*groot schilder(s); de beroemde / \* beroemd dichter(s); de slechte / \*slecht schaker(s)  
 b. de Franse / \*Frans filosoof(en); de Amerikaanse / \*Amerikaans zakenman, de Amerikaanse / \*Amerikaans zakenlui
- (22') de voortreffelijke / \*voortreffelijk spreekster(s); de grote / \*groot pianiste(s)

De literatuur suggereert dat dit alternantietype minder dwingend is dan het vorige. Dit is de reden waarom we de verbogen variant niet principieel hebben uitgesloten in de voorbeelden in (21) en (22) door de buigings-*e* tussen haakjes te plaatsen. De ANS (1997) stelt bovendien dat de verbogen vorm vooral in de gesproken taal wordt gebruikt. Royen (1953: 56) haalt voorbeelden aan die een lexicaal effect suggereren: "Waarom wel een *braaf man* naast een *brave man*, maar alleen een *brave kerel*?". Nochtans kan ook bij dit alternantietype de oppositie verbogen-onverbogen aan de basis liggen van minimale paren:

- (23) a. Napoleon is geen *grote man* maar wel een *groot man*.  
 b. Jan is een *knap schilder* maar geen *knappe schilder/man*.

Dit semantische onderscheid is al in de achttiende eeuw door Lambert Ten Kate behandeld (zie paragraaf 2.3.3). De onverbogen vorm signaleert in het [b]-type dat de eigenschap van het adjectief niet op de referent van het substantief betrekking heeft, maar op de hoedanigheid die het substantief identificeert: in (23b) signaleert de onverbogen vorm dat de eigenschap aangeduid door het adjectief *knap* enkel op Jans hoedanigheid als schilder betrekking heeft en niet op de rest van zijn persoon. Deze betekenispecialisatie vinden we ook terug bij onverbogen vormen van het [a]-type: in de academische titel *buitengewoon hoogleraar* in (20b) heeft het kenmerk aangeduid door het adjectief enkel betrekking op het hoogleraarschap; voor het overige kan het om een zeer middelmatig persoon gaan.

Ondanks de strikte semantische condities die in de literatuur geformuleerd worden met betrekking tot het gebruik van de onverbogen adjectivische vorm bij een *de*-woord, is het gebruik niet afdwingbaar bij ontstentenis van een constant betekenisverschil. Schultink (1962: 65-66) onderschrijft de mogelijkheid tot betekenisverschil bij de buigingsalternantie van het type [a] en het type [b], maar stelt de algemeenheid en de systematiciteit ervan in vraag. Op basis van voorbeelden verzameld door Royen (1953) concludeert hij dat er geen sprake is van een constant



betekenisverschil bij alle taalgebruikers. Daarnaast bestaat in de literatuur de neiging om instanties waar de onverbogen vorm gebruikt wordt als idiosyncratieën te beschouwen (de Haas & Trommelen 1993; Odijk 1992; Van Eynde 2003). Tot slot is er een complicerende factor aanwezig bij het [b]-type, met name de stilistische variatie: de ANS (1997) suggereert dat het gebruik van de onverbogen vorm vooral in de formele registers voorkomt.

Het derde type buigingsalternantie treffen we aan in definiëte NP's met een enkelvoudig *het*-woord als hoofd. Dit is type [c] in figuur 3(2) en omvat gevallen als:

- (24) het openbaar(e) ministerie; het stoffelijk(e) overschot; het vergelijkend(e) examen; het bruin(e) brood<sup>6</sup>; het vriendelijker(e) kind; het prachtig(e) uniform; het bruin(e) paard

Deze buigingsalternantie vormt het onderwerp van voorliggend proefschrift. Het volstaat een blik te werpen op de voorbeelden in (24) en de aanvaardbaarheid ervan te beoordelen om vast stellen dat de keuze tussen beide buigingsmorfemen niet absoluut is: de aanvaardbaarheid van de onverbogen vorm neemt af naarmate we van links naar rechts opschuiven in de lijst; de aanvaardbaarheid neemt bovendien sterker af voor een Nederlander dan voor een Vlaming. Tot slot lijkt het onmogelijk om de verschillende alternanties tussen het verbogen en het onverbogen adjectief in (24) met één conditie samen te vatten. De doelstelling van dit onderzoek is het identificeren en modelleren van de restricties (zie paragraaf 1.1) op het gebruik van beide adjectivische buigingsvarianten in definiëte NP's met een enkelvoudig *het*-woord als hoofd.

Tot slot wensen we te onderstrepen dat de driedeling van de adjectivische buigingsalternantie in een [a]-type, een [b]-type en een [c]-type niet impliceert dat het drie verschillende en ongerelateerde types zijn. We hebben hierboven al aangevoerd dat in voorbeelden als *de buitengewoon hoogleraar* de betekenisaspecten van het [a]- en het [b]-type nauw verweven zijn. Daarnaast zijn de officiële titels bij het [a]-type verwant met de officiële titels bij het [c]-type, hoewel het in het laatste geval typisch over instellingen (*het dagelijks bestuur, het uitvoerend comité, het openbaar ministerie*; zie ook paragraaf 1.1.3.1) en niet over personen gaat. Tot slot lijkt het niet uitgesloten dat in voorbeelden zoals *het groot geheim* (type [c]) de onverbogen buigingsvariant gebruikt wordt om de figuurlijke betekenis van *groot* in de verf te zetten op een manier die verwant is aan *een groot schilder* in (21a) (type [b]). Deze opmerkingen tonen aan dat voorliggend onderzoek naar de adjectivische buigingsalternantie bij *het*-woorden in een volgend stadium moet worden uitgebreid naar de buigingsalternantie bij *de*-woorden. Het valt trouwens nog te bezien of de buigingsalternantie bij *de*-woorden in het reële taalgebruik zo strikt (semantisch) geconditioneerd is als doorgaans in de literatuur gesuggereerd wordt (cf. *supra*).

## Noten

1. We gebruiken het symbool ‘-ø’ in figuur 3(1) om de onverbogen vorm te identificeren. We spreken ons niet uit over de theoretische status van het zogenaamde nulmorfeem. Deze discussie overschrijdt ruimschoots de descriptieve en methodologische doelstellingen van dit onderzoek.
2. Het adjectief *na* wordt wel verbogen: *een naë bloedverwant* (ANS 1997).
3. De klinker in de adjectivische lettergreep die de hoofdklemtoon draagt, markeren we door middel van een *accent aigu* (´).
4. Royen (1953: 82-84) citeert voorbeelden van verbogen vormen op *-en*: *onze eigene belangen*.
5. Voor een uitgebreide analyse van het betekenisverschil dat de distinctie tussen het gebruik van *veel* en *vele* bepaalt, verwijzen we naar Van der Horst (1992, 1995: 127-133) (zie ook paragraaf 1.1.3.2).
6. De *Elektronische Grote Van Dale* (Geeraerts 2000) twijfelt hier tussen een samenstelling en een constituent: naast het lemma *bruinbrood* vinden we onder het lemma *brood* de constituent *bruin brood*.

## Hoofdstuk 4:

# Congruentie

Dit hoofdstuk beschrijft congruentie als taalkundig fenomeen. Het is de bedoeling een algemeen en inleidend overzicht te presenteren van de aard en de functie van congruentie, zonder dat we hierbij aanspraak willen maken op enige exhaustiviteit. Deze beschrijving wil bovendien theorieonafhankelijk zijn; alleen indien nodig wordt naar specifieke theoretische modellen verwezen. Vertrekkend van de wijdverspreide tweeledige visie op morfologie situeren we congruentiefenomenen in het ruimere kader van de morfologie (4.1). Vervolgens richten we onze aandacht op de afzonderlijke congruentiefenomenen (4.2). Deze paragraaf geeft een kort overzicht van de verschillende congruentiemechanismen en van de functies die congruentiefenomenen binnen de grammatica (kunnen) vervullen. Hierbij evolueren we van een puur formalistische en syntactische visie op congruentie naar een meer pragmatisch en discursief standpunt.

De relevante vragen met betrekking tot het empirisch onderzoek naar de adjectivische buigingsalternantie worden samengevat in paragraaf 4.3.

### 4.1 Congruentie als morfologisch fenomeen

Traditioneel wordt de morfologie in twee deelcomponenten verdeeld, met name de derivationele morfologie en de inflectionele morfologie. De band die beide deelcomponenten verbindt, is formeel: de derivationele en de inflectionele morfologie delen dezelfde bouwstenen en mechanismen. Het onderscheid tussen beide is functioneel. De derivationele morfologie omvat de elementen en mechanismen die gebruikt worden bij de vorming van nieuwe woorden, vanwaar de alternatieve benaming lexicale morfologie. In het Nederlands wordt deze functie hoofdzakelijk door de afleiding en de samenstelling vervuld.<sup>1</sup> De inflectionele morfologie bestudeert de restricties op het gebruik van de verschillende morfologische realisatievormen van een woord. Dit is het syntactische deel van de morfologie: “inflectional morphology is what is relevant to the syntax” (Anderson 1982: 587). Anderson (1988) en Stump (1998: 21-26) onderscheiden de volgende fenomenen binnen de inflectionele morfologie op basis van de syntactische relaties die relevant zijn voor de selectie van de morfosyntactische vorm:<sup>2</sup>

- “inherent properties”: Dit zijn de morfosyntactische eigenschappen die met een woordsoort geassocieerd worden. Met de categorie van de substantieven worden onder andere getal en genus geassocieerd, met de categorie van de adjectieven gradatie (c.q. de trappen van vergelijking).

- “governed properties”: Deze eigenschappen worden aan een element toegerekend op basis van zijn positie of functie in een ruimere syntactische structuur. De grammaticale casus drukt bijvoorbeeld de grammaticale functie van een element binnen een ruimere syntactische structuur uit.
- “agreement properties”: Dit zijn de eigenschappen die aan een element worden opgelegd door een ander element in dezelfde syntactische structuur. De adjectivische verbuiging drukt in prototypische gevallen onder andere het genus en het getal van het substantief uit.

In de traditionele tweedelige visie op de morfologie behoort congruentie (*agreement* of *concord*) dus tot het deel dat de realisatie van syntactische informatie op woordniveau bestudeert.

De derivationale en inflectionele morfologie worden op basis van een reeks criteria gedifferentieerd (Booij 2000; Stump 1998). Deze criteria kennen lexicale eigenschappen toe aan de derivationale morfologie en syntactische eigenschappen aan de inflectionele morfologie: derivationale morfemen hebben een lexicaal-semantic inhoud en zijn niet vrij van vormelijke en inhoudelijke idiosyncrasieën, allemaal lexicale kenmerken; de inflectionele morfologie wordt daarentegen gekenmerkt door een grote regelmaat, voorspelbaarheid en semantische transparantie, typisch syntactische eigenschappen. Terwijl inflectionele eigenschappen door hun regelmaat en abstracte inhoud tot de regelcomponent van de grammatica behoren, worden derivationale eigenschappen – sneller – in het lexicon opgenomen.<sup>3</sup> Omdat inflectionele kenmerken – idealiter – bij alle leden van een woordsoort voorkomen en omdat alle leden van een woordsoort – idealiter – alle waarden van de inflectionele kenmerken kunnen uitdrukken, worden ze in paradigmata georganiseerd (Booij 2000; Carstairs-McCarthy 1998). Een paradigma bevat de waarden van alle relevante combinaties van de morfosyntactische kenmerken en de inflectionele vormen die ze realiseren. Het inflectionele paradigma van een Nederlands substantief omvat vier vormen die de combinaties van de morfosyntactische eigenschappen getal en diminuering uitdrukken. Elk Nederlands substantief realiseert – idealiter – deze vier verschillende morfosyntactische vormen, zoals geïllustreerd in tabel 4(1).

	enkelvoud	meervoud
–diminutief	<i>poot</i>	<i>poot-en</i>
diminutief	<i>poot-je</i>	<i>poot-je-s</i>

Tabel 4(1): Substantivisch paradigma in het Nederlands

Het adjectivische paradigma bestaat uit zes vormen die de morfosyntactische kenmerken gradatie en buiging realiseren. Tabel 4(2) illustreert hoe de inherente kenmerken (trappen van vergelijking) en de congruentiekenmerken (buigingsvormen) samen voorkomen in het adjectivische paradigma.

De abstracte vorm die aan de basis van een paradigma ligt, noemen we het lemma<sup>4</sup>; de verschillende inflectionele realisaties in het paradigma zijn de woordvormen.

	onverbogen	verbogen
<b>stellende trap</b>	<i>klein</i>	<i>klein-e</i>
<b>vergroten trap</b>	<i>klein-er</i>	<i>klein-er-e</i>
<b>overtreffende trap</b>	<i>klein-st</i>	<i>klein-st-e</i>

Tabel 4(2): *Adjectivisch paradigma in het Nederlands*

Deze tweeledige visie op de morfologie staat bekend als *split morphology*, maar wordt het laatste decennium sterk gecontesteerd (Booij 1993, 1996, 2000; van Marle 1995b). De belangrijkste bezwaren kunnen onder twee noemers worden samengebracht. Ten eerste, de typering van inflectie als ‘wat relevant is voor de syntaxis’ (cf. *supra*) is problematisch. De interpretatie van deze definitie wordt volledig bepaald door de visie op syntaxis en de bijhorende syntactische theorie en leidt bijgevolg tot circulariteit.<sup>5</sup> Bauer (2003: 104) is van oordeel dat congruentie het enige inflectionele fenomeen is dat omwille van zijn syntactische basis altijd tot de ‘syntactisch relevante inflectionele morfologie’ behoort. Ten tweede, de inflectionele morfologie vormt geen homogeen domein. Bij de beschrijving van de inflectionele morfologie is al gebleken dat ze verschillende aspecten omvat. Booij (1993, 1996, 2000) groepeert deze aspecten in de inherente inflectie en de contextuele inflectie: de inherente inflectie bevat de waarden in het inflectionele paradigma die inherente eigenschappen van het item of de woordsoort uitdrukken, zoals de meervoudsvorming en de diminuering bij het substantief, en de trappen van vergelijking bij het adjectief; de contextuele inflectie omvat de vormen binnen het inflectionele paradigma die op basis van de kenmerken van de syntactische context geselecteerd worden, zoals het buigingsmorfeem van het adjectief en de casus van de N(P).<sup>6</sup> Booij (1993, 1996, 2000) en Van Marle (1995b) tonen aan dat de inherente inflectie meerdere derivationale kenmerken vertoont.<sup>7</sup> Zij besluiten dat het adequater is het beeld van de morfologische tweedeling te vervangen door een continuum dat zich uitstrekt van het lexicon tot de grammatica (tabel 4(3)). In dit continuum bezet de derivationale morfologie de lexicale pool en de contextuele inflectie de grammaticale pool. De inherente inflectie neemt een overgangspositie in als gevolg van het combineren van lexicale en syntactische kenmerken.

LEXICON		GRAMMATICA
derivationale morfologie	inherente inflectie	contextuele inflectie

Tabel 4(3): *De morfologie als continuum*

Congruentiefenomenen behoren tot de meest syntactische aspecten van de morfologie, waarbij de inflectionele vormen geen inherente eigenschappen van het

element in kwestie uitdrukken maar geselecteerd worden op basis van de morfo-syntactische kenmerken van de syntactische structuur waarin het element in kwestie voorkomt. Het continuüm in tabel 4(3) wordt weerspiegeld in de aanhechtingsvolgorde van de affixen aan de stam: de lexicale morfemen staan het dichtst bij de stam; de grammaticale morfemen zijn het verst van de stam verwijderd.

(1) het vriend-elijk<sub>DER.MOR</sub>er<sub>INH.INFL</sub>e<sub>CONT.INFL</sub> kind

De affixvolgorde in (1) is iconisch: ze weerspiegelt het semantische gehalte van de suffixen, waarbij de contextuele inflectie semantisch het minst en syntactisch het meest relevant is (Bybee 2000).

Het continuüm in tabel 4(3) reflecteert ook de aandacht die de verschillende morfologische aspecten binnen de literatuur genieten: terwijl de derivatieve morfologie veruit het grondigst bestudeerd is, kan de contextuele inflectie gerust het zwakke broertje van de morfologie genoemd worden. Een constante in de analyses over congruentie (c.q. *agreement properties*; *inter al.* Booij 1992; De Schutter 1994; Ferguson & Barlow 1988; Kester 1996) is de opmerking dat descriptieve en theoretische analyses schaars zijn. Een blik op morfologische handboeken leert ons dat de derivatieve morfologie het merendeel van de aandacht krijgt. Binnen de inflectionele morfologie gaat de meeste aandacht uit naar de inherente inflectie. Het aandeel van de contextuele inflectie is nagenoeg beperkt tot een obligate vermelding, met vaak de opmerking dat het een fenomeen van de oppervlaktegrammatica is. Eenzelfde beeld rijst op uit twee recente overzichtsartikelen van de Nederlandse morfologie: de recensies in Booij (1999) en Schultink (2000) tonen aan dat ook binnen de neerlandistiek de derivatieve morfologie een bevoorrechte positie inneemt ten nadele van de inflectionele morfologie, en de contextuele inflectie in het bijzonder.

## 4.2 Definitie en functie van congruentie

Nu we congruentie aan de syntactische pool van het morfologische continuüm gesitueerd hebben, richten we onze aandacht op de kenmerken en de functie van congruentie (4.2.2), nadat we eerst dit morfologische fenomeen gedefinieerd hebben (4.2.1).

### 4.2.1 Definitie congruentie

Corbett (1998) merkt op dat deze intuïtief transparante notie moeilijk definieerbaar is.<sup>8</sup> Na verschillende definities te hebben besproken, besluit hij dat “the essential notion is the covariance or matching of feature specifications between two separate elements” (Corbett 1998: 191). Dit idee ligt aan de grondslag van de twee meest invloedrijke definities: de definitie van Moravcik (1978) en de definitie van

Ferguson & Barlow (1988). De meest gedetailleerde definitie wordt door Moravcik (1978: 333) geformuleerd:

“a grammatical constituent A will be said to agree with a grammatical constituent B in properties C in language L if C is a set of meaning-related properties of A and there is a covariance relationship between C and some phonological properties of a constituent B<sub>1</sub> across some subset of the sentences of language L, where constituent B<sub>1</sub> is adjacent to constituent B and the only meaning-related non-categorical properties of constituent B<sub>1</sub> are the properties C”

Op analoge wijze definiëren Ferguson & Barlow (1988: 1) congruentie als het verschijnsel “by which a grammatical element X matches a grammatical element Y in property Z within some grammatical configuration”. Uit beide definities kan worden afgeleid dat de volgende elementen in een congruentierelatie betrokken zijn (Corbett 1998):

- controlerend element (“controller”): Het element dat de congruentierelatie en -kenmerken bepaalt. [A in Moravciks definitie; Y in de definitie van Ferguson & Barlow]
- doelelement (“target”): Het element waarvan de vorm bepaald wordt door de congruentierelatie. [B in Moravciks definitie; X in de definitie van Ferguson & Barlow]
- congruentiekenmerken (“agreement features”): De (morfosyntactische) kenmerken waarvoor het doelelement met het controlerend element congrueert. De meest relevante kenmerken zijn genus, getal en persoon (Corbett 1998; Moravcik 1978; Stump 1998). Over het statuut van definitie bestaat geen eenduidigheid: sommigen beschouwen het als een kenmerk dat congruentie meebepaalt (Moravcik 1978), terwijl anderen definitie expliciet buiten de congruentie laten (Corbett 1998; Zwicky 1986).<sup>9</sup> [C in Moravciks definitie; Z in de definitie van Ferguson & Barlow]
- exponent (“exponent”): Het vormelijke element, meestal een affix, dat de congruentiekenmerken realiseert in het doelelement. [B<sub>1</sub> in Moravciks definitie]
- domein (“domain”): De syntactische context waarin de congruentie optreedt. [‘some grammatical configuration’ in de definitie van Ferguson & Barlow]

Uit deze definities kan worden afgeleid dat congruentie een asymmetrische relatie is. De asymmetrie vloeit voort uit het feit dat de inflectionele vorm van het doelelement uitsluitend bepaald wordt door kenmerken van het controlerend element. Met andere woorden: de syntaxis definieert het congruentie-domein en de kenmerken betrokken in de congruentierelatie; de rol van de morfologie is beperkt tot het uitspellen van de kenmerkenwaarden in de vorm van een congruentie-exponent. Grammaticale modellen vertalen deze asymmetrie in het kopiëren van kenmerken van het controlerend element naar het doelelement of in het verifiëren

van de compatibiliteit van de kenmerken van het doelelement met de kenmerken van het controlerend element. Deze behandeling van inflectionele morfemen wordt dadelijk verduidelijkt aan de hand van de voorbeelden (2) en (3).

Na het definiëren van het algemene begrip congruentie stellen we nu de verschillende subtypes voor. Tabel 4(4) geeft een overzicht van de congruentietypes die in de literatuur onderscheiden worden (Anderson 1991; Barlow 1991; Lehmann 1988; Moravcik 1978; Pollard & Sag 1994).

<b>Congruentietype Instantie</b>	
zinsextern	congruentie van anafoor met antecedent
zinsintern	
NP-extern	subject-predikaatcongruentie
NP-intern	congruentie van adnominale elementen

Tabel 4(4): Congruentietypes

Dit schema toont aan dat er altijd een substantief of een NP in een congruentierelatie betrokken is, i.e. een refererend element. Dit nominale element is bovendien altijd het controlerend element in de congruentierelatie: het antecedent bij zinsexterne congruentie, het subject bij NP-externe congruentie en het nominale hoofd van de NP bij NP-interne congruentie. Bij zinsinterne congruentie varieert de richting van de asymmetrische relatie in functie van het congruentietype. Bij NP-interne congruentie, ook *modifier-head agreement* genoemd, bepaalt het hoofd de vorm van de modificeerder: de congruentie-exponent van het doelelement, c.q. het adjectief (en de determinator), moet de relevante kenmerken van het controlerend element, c.q. het nominale hoofd, uitspellen. Als voorbeeld kijken we naar de congruentie van het attributief gebruikte adjectief in het Frans. De vier uitgangen van het adjectief *grand* in (2a) worden uitsluitend door het genus en het getal van het nominale hoofd gedefinieerd. Het volstaat bijgevolg om de relevante kenmerken van het substantief naar het adjectief te kopiëren of om hun aanwezigheid in het adjectief te controleren. De agrammaticaliteit van (2b) wordt verklaard door de conflicterende morfo-syntactische kenmerken: de kenmerken [+FEM,+ENK] van de adjectivische uitgang accorderen niet met de kenmerken [+MASC,+ENK] van het substantief.

- (2) a. grand- $\emptyset$  garçon, grand-s garçons, grand-e fille, grand-es filles  
 b. \* grand-e garçon ('groot-FEM.ENK man.MASC.ENK')

Bij subject-predikaatcongruentie, een instantie van NP-externe congruentie, is de congruentierelatie omgekeerd: het syntactisch ondergeschikte element, het subject, bepaalt de vervoeging van het hoofd, het werkwoord. In het Nederlands wordt de vervoegde vorm van de tegenwoordige tijd uitsluitend bepaald door de persoon en het getal van het subject:

- (3) ik drink- $\emptyset$ , jij/hij/zij drink-t, wij/jullie/zij drink-en



Als gevolg van deze syntactische voorspelbaarheid en redundantie is aan congruentiefenomenen lange tijd weinig aandacht besteed: door middel van percolatie worden de relevante kenmerken doorheen de syntactische structuur gepropageerd om uiteindelijk in de post-syntactische component als een congruentie-exponent te worden uitgespeld. Het uitspellen is niets anders dan het vormgeven aan structuren die tot dan enkel morfosyntactisch gespecificeerd zijn (zie o.a. Anderson 1991: 122-135). Toegepast op de adjectivische buiging in het Nederlands betekent dit dat de kenmerkenstructuur [+ONZ,+ENK,-DEF] als  $-\emptyset$  wordt uitgespeld en alle andere kenmerkenstructuren als  $-e$  (zie paragraaf 3.1, figuur 3(1)).<sup>10</sup>

In recente formeel taalkundige stromingen, zoals *Minimalistic Program* (Kester 1996) en *Head-Driven Phrase Structure Grammar* (HPSG; Pollard & Sag 1994), worden de inflectionele vormen in het lexicon gegenereerd en worden de geïnfecteerde vormen in plaats van de lemmata in de syntactische structuur ingevoegd. De syntaxis voorziet dan in een mechanisme om de compatibiliteit van de geïnfecteerde lexicale vormen in de syntactische structuur te testen. Hiervoor wordt onder andere de informatie aanwezig in congruentiemorfemen gebruikt: net omwille van de morfosyntactische redundantie vormen congruentiemorfemen een geschikt testcriterium. Voor theoretische modellen van de verbuiging van het Nederlandse adjectief moeten we terecht bij dezelfde taalkundige stromingen: *Principles and Parameters* (De Schutter 1994, 1997; Odijk 1992), *Minimalistic Program* (Kester 1996) en HPSG (Van Eynde 2003). We herinneren de lezer eraan dat we deze analyses in het literatuuroverzicht bij de marginaliserende attitude hebben ingedeeld, i.e. de analyses van de adjectivische buiging die de buigingsalternantie niet erkennen of als een randverschijnsel afdoen (zie paragraaf 1.1.2).

#### 4.2.2 Functies van congruentie

In deze paragraaf kijken we naar de functie van congruentie. Rekening houdend met voorgaande bespreking kan men geneigd zijn congruentie door haar voorspelbaarheid en redundantie als functioneel en zelfs overbodig te beschouwen. Deze stelling wordt onder andere door Haiman (1985) verwoord. Hij is de mening toegedaan dat “the phenomenon of grammatical agreement seems a clear case of the victory of the indexical aspect of language over its iconic aspect inasmuch as categories such as number and case, properly associated with nouns, are copied onto verbs and adjectives” (Haiman 1985: 162). Hij besluit bijgevolg dat “grammatical agreement is redundant: not only non-iconic, but meaning-less. Rather than have one meaning represent one form, we find one meaning represented in several forms at once, signaled again and again” (Haiman 1985: 164-165). Dat de basisfunctie van congruentie indexicaal is, valt niet te betwisten. Of de interpretatie zo negatief hoeft te zijn als bij Haiman (1985) is voor discussie vatbaar. We zullen stapsgewijs aantonen dat congruentie-exponenten betekenisvol/functioneel (kunnen) zijn.

Bybee (2000) en Anderson (1991) vullen de indexicale basisfunctie van congruentie positief in: een congruentie-exponent verwijst naar de aanwezigheid van bepaalde eigenschappen van het controlerend element (Bybee 2000). Anderson (1991) spreekt bij NP-interne congruentie van “registration”, waarbij de congruentie-exponent de aanwezigheid signaleert van een element met bepaalde morfosyntactische kenmerken:

“‘Registration’ consists in marking an ‘agreeing’ element in some way so as to record the presence of another element with certain properties somewhere in the surrounding structure. [...] Once the appropriate features have been passed to a modifier element, there is no reason to believe that their marking on the modifier corresponds to the establishment of any other relation.” (Anderson 1991: 112-113)<sup>11</sup>

De functie van congruentie is niet meer of minder dan het signaleren dat twee elementen samen horen en informatie over dezelfde referent verstrekken. Dit verklaart waarom congruentie altijd uitgaat van een substantief, zoals we hebben vastgesteld bij de verschillende congruentietypes in tabel 4(4). In synthetische talen is deze functie essentieel, voor een analytische taal als het Nederlands is de vraag naar de functionaliteit van de adjectivische buiging terecht (*inter al.* van Bree 1980; De Groot 1949; van Haeringen 1937; Peters 1937).

Op basis van onderzoek naar discongruenties (i.e. het voorkomen van discordante kenmerken bij het controlerend element en de congruentie-exponent) en probleemgevallen (bv. congruentie waar het controlerend element ontbreekt, congruentie over de zinsgrenzen heen) is de indexicale functie van congruentie uitgebreid (Barlow 1999; Corbett 1979, 1998; Pollard & Sag 1994). Corbett (1979) onderscheidt naast de traditionele formele (morfo)syntactische congruentie ook semantische congruentie.<sup>12</sup> De formele congruentie is de vormelijk bepaalde en dus morfosyntactisch voorspelbare congruentie, zoals we tot nu toe besproken hebben. Semantische congruentie daarentegen is niet louter op morfosyntactische eigenschappen gebaseerd, maar veronderstelt de inbreng van referentiële en semantische informatie. Vanuit een strikt formeel en morfosyntactisch perspectief is er in (4b) sprake van discongruentie: het getal van het onderwerp (enkelvoud) en het getal van de vervoegde persoonsvorm (meervoud) zijn onverenigbaar. Corbetts alternatief is dat het werkwoord niet met de vorm van het subject congrueert, maar met de collectieve – en dus meervoudige – referent.

- (4)     a. The committee has decided.  
           b. The committee have decided.  
           (voorbeelden overgenomen uit Corbett 1979: 203)

De alternatieve congruentievormen in (4a) en (4b) zijn dus functioneel gemotiveerd: het gebruik van de twee congruentie-exponenten gaat gepaard met een verschillende conceptualisatie. Corbett (1979) stelt de zogenaamde congruentiehiërarchie

(*agreement hierarchy*) voor om de verdeling tussen de semantische en de syntactische congruentie te beschrijven. De waarschijnlijkheid van semantisch gemotiveerde congruentie stijgt naarmate we van links naar rechts opschuiven in tabel 4(5).

SYNTAXIS		SEMANTIEK	
attributie	predicatie	betrekkelijk vnw	persoonlijk vnw

Tabel 4(5): *Congruentiehiërarchie (Corbett 1979: 204)*

De hiërarchie in tabel 4(5) beschouwt congruentie als gradueel veeleer dan als discreet, waarbij het voorkomen van semantische congruentie gerelateerd wordt aan de afstand tussen het doelelement en het controlerend element in de congruentierelatie. De buiging van het attributieve adjectief, dat vlakbij het nominale hoofd staat, bevindt zich aan de syntactische pool van het congruentiecontinuum; de semantische pool wordt bezet door de persoonlijke voornaamwoorden die in een andere zin dan hun antecedent (kunnen) staan.

Barlow (1991, 1999) gaat een stap verder en beschouwt congruentie als een discursief fenomeen, waarbij de informatie die nodig is om een referent te interpreteren over verschillende talige elementen verdeeld is. Congruentiemorfemen zijn niet langer markeerders van redundante en voorspelbare morfosyntactische informatie, maar betekenisvolle elementen die een volwaardige functie vervullen in het construeren van de informatie die over een referent verstrekt wordt. De primaire functie van congruentie is dan niet morfosyntactisch maar wel discursief: de congruentie-exponent identificeert geen morfosyntactische kenmerken van het controlerend element, maar wel kenmerken van een referent die in het discours vermeld wordt. Een substantief identificeert een referent waaraan bepaalde interpretatieve eigenschappen verbonden zijn. Sommige van deze eigenschappen worden in de inflectionele vorm gerealiseerd. Een enkelvoudig substantief identificeert bijvoorbeeld een individu, een meervoudig substantief verwijst naar een groep individuen. Analoge kenmerken zijn aan het congruentiemorfeem verbonden. Door het gebruik van unificatie en onderspecificatie wordt de betekenis van de congruentierelatie niet bepaald door de informatie aanwezig in één element, c.q. het controlerend element, maar is ze het resultaat van het cumuleren van de informatie aanwezig in het controlerend element en de informatie aanwezig in de congruentie-exponent. De congruentie-exponent spelt dus niet langer louter redundante en voorspelbare informatie uit, maar participeert op actieve wijze in de conceptualisatie van de referent.

We zullen deze benadering illustreren aan de hand van de congruentievormen in (4). De nominale en verbale inflectionele vormen [+ENK] of [+MV] (rechte haken) worden in (5) geassocieerd met de respectievelijke referentiële eigenschappen <INDIVIDUEEL> en <BESTAAND-UIT-INDIVIDUEN> (vishaken). In tegenstelling tot [+ENK] en [+MV] sluiten <INDIVIDUEEL> en <BESTAAND-UIT-INDIVIDUEN> elkaar niet wederzijds uit.<sup>13</sup>

- (5) a.  $V_{[+ENK]} \rightarrow \langle \text{INDIVIDUEEL} \rangle$   
 b.  $V_{[+MV]} \rightarrow \langle \text{BESTAAND-UIT-INDIVIDUEN} \rangle$   
 c.  $N_{[+ENK]} \rightarrow \langle \text{INDIVIDUEEL} \rangle$   
 d.  $N_{[+MV]} \rightarrow \langle \text{BESTAAND-UIT-INDIVIDUEN} \rangle$

De traditionele morfosyntactische analyse van de congruentierelatie tussen subject en predikaat in (4b) in termen van de overdracht van morfosyntactische kenmerken van het subject naar het predikaat leidt tot discongruentie door het verenigen van incompatibele morfosyntactische kenmerken, met name [+ENK] voor het substantief en [+MV] voor het predikaat. De unificatie van de referentiële kenmerken, daarentegen, creëert een nieuwe informatiestructuur,  $\{\langle \text{INDIVIDUEEL} \rangle, \langle \text{BESTAAND-UIT-INDIVIDUEN} \rangle\}$ , die de sleutel vormt tot de interpretatie van de referent van *committee*. In tegenstelling tot (4a), waar de referent van *committee* als een individu geconceptualiseerd wordt (cf.  $\{\langle \text{INDIVIDUEEL} \rangle\}$  als resulterende kenmerkenstructuur), wijst de resulterende kenmerkenstructuur  $\{\langle \text{INDIVIDUEEL} \rangle, \langle \text{BESTAAND-UIT-INDIVIDUEN} \rangle\}$  in (4b) op de interpretatie van *committee* als een collectivum, c.q. een structuur die in feite een verzameling van individuen is. Samengevat: het conflict tussen de morfosyntactische kenmerken fungeert op iconische wijze als een signaal voor de codering van bijkomende of nieuwe informatie over een referent. De selectie van een congruentievorm is volgens Barlows discursieve benadering een classificatorische daad: de congruentievorm verstrekt informatie over de conceptualisatie van de substantivische referent.

De discursieve en classificatorische functie van congruentie-exponenten kan uitgebreid worden naar pragmatische, stilistische en sociolinguïstische informatie. In deze gevallen heeft de congruentie-informatie niet langer betrekking op de eigenschappen van de betrokken N(P) maar op de eigenschappen van de referent van de gesprekspartners, hun onderlinge relatie en de gesprekssituatie. Kiefer (1998: 273) spreekt van morfopragmatiek of expressieve morfologie (*expressive morphology*) die vooral op de voorgrond treedt “whenever we have to do with competing realizations of morphological rules”. Pollard & Sag (1994) bespreken de rol van contextuele informatie aan de hand van beleefdheidsmarkering door middel van meervoudsvormen in het Duits en *honorifics* in het Koreaans. Corbett (1979) toont aan dat lectische factoren, zoals leeftijd, regio en sociale klasse, in het Russisch voor variatie zorgen bij de congruentie van het attributieve adjectief en de vervoegde werkwoordsvorm. In een discursieve benadering wordt deze variatie als betekenisvol beschouwd, waarbij de selectie van de congruentie-exponent informatie verschaft over de sociale status, woonplaats enz. van de spreker (zie paragraaf 12.1).

Kortom, de traditionele visie op congruentie als een louter syntactisch bepaald verschijnsel moet worden genuanceerd: de congruentie-exponent drukt niet alleen syntactisch voorspelbare en redundante informatie uit, maar neemt op actieve wijze deel aan de constructie van het discursieve interpretatiekader door het verstrekken van informatie voor de conceptualisatie van de referenten, de gesprekspartners en de gesprekssituatie.

### 4.3 Conclusie

Congruentie behoort tot de meest syntactische aspecten van de morfologie (4.1). De selectie van het congruentiemorfeem wordt door contextuele elementen, meer bepaald door de morfosyntactische kenmerken van het controlerend element, gedefinieerd en niet door intrinsieke kenmerken van het doelelement. De syntaxis bepaalt het domein waar congruentie optreedt evenals de morfosyntactische kenmerken die met de congruentie-exponent geassocieerd worden (4.2.1). Door de syntactische conditionering is de distributie van de congruentiemorfemen voorspelbaar en redundant: de inflectionele paradigmata specificeren welke woordvorm bij welke morfosyntactische configuratie past. Congruentie is in deze optiek van elke functionele motivatie ontdaan. Deze louter indexicale functie verklaart wellicht voor een deel de geringe aandacht die aan congruentiefenomenen wordt besteed.

Onderzoek naar discongruenties en naar variatie bij het uitspellen van congruentiekenmerken heeft het traditionele beeld van congruentie als de meest syntactisch geconditioneerde component van de morfologie in twee opzichten bijgesteld (4.2.2). Ten eerste, bij het uitspellen van de congruentiekenmerken spelen naast syntactische argumenten ook semantische, pragmatische, stilistische en sociolinguïstische factoren een rol. Ten tweede, Barlows (1991, 1999) discursieve visie op congruentie beschouwt congruentiemorfemen niet langer als passieve markeerders van elementen die samen voorkomen, maar verleent hen een actieve inbreng in het construeren van de conceptuele inhoud van de referent. De discursieve benadering kan bovendien worden toegepast op de conceptualisatie van de gesprekscontext en de relatie tussen de gesprekspartners, waar de pragmatische, stilistische en sociolinguïstische factoren een prominente rol spelen.

Als een voorbeeld van NP-interne congruentie behoort de adjectivische buiging tot de meest syntactische congruentiefenomenen (zie tabel 4(5)), die sowieso al de meest syntactische/grammaticale fenomenen van de morfologie uitmaken (zie tabel 4(3)). Een blik op de theoretische modellering van de adjectivische buiging in het Nederlands bevestigt deze stelling: de adjectivische buiging wordt gemodelleerd als een (morfo)syntactisch verschijnsel, waarbij de interactie tussen de kenmerken van het substantief en de kenmerken van de determinator het centrale aandachtspunt is.

In dit onderzoek wordt geen centrale positie toegekend aan de (binaire) syntactische conditionering van de adjectivische buiging. De syntactische conditionering wordt enkel gebruikt om het onderzoeksdomein af te bakenen, met name de adjectivische buigingsalternantie in een definiëte NP met een enkelvoudig *het*-woord als hoofd (variatietype [c] in figuur 3(2); we verwijzen naar paragraaf 6.2 en paragraaf 15.2 voor de samenstelling van de materiaalverzamelingen die de empirische basis van het onderzoek vormen). We concentreren ons op de volgende vragen:

- Hoe voorspelbaar/redundant is de keuze van het buigingsmorfeem in een definiëte NP met een enkelvoudig *het*-woord als hoofd? Of: is de keuze van het buigingsmorfeem in deze context volledig syntactisch voorspelbaar?

## DEEL I

- Welke factoren bepalen de selectie van de buigingsmorfemen die in deze context optreden? Wat is de aard van deze factoren?

De verkennende literatuurstudie in paragraaf 1.1 doet vermoeden dat het adjectivische buigingsmorfeem bij neutra niet volledig voorspelbaar is op basis van de syntactische kenmerken van de NP. Bovendien zijn de factoren die in de literatuur worden aangehaald niet-syntactisch van aard: de keuze van de adjectivische buigingsvorm wordt volgens de hypothesen in tabel 1(1) mee bepaald door fonologische, lexicale, semantische en lectische factoren, wat ten dele aansluit bij de discursieve congruentiemodellen.

## Noten

1. In realiteit zijn er ook andere mechanismen, zoals acronymie (*NAVO*) en woordversmelting (*docudrama*). Voor een overzicht van de woordvormingsmechanismen in het Nederlands verwijzen we naar De Haas & Trommelen (1993) en Meesters (2002).
2. Deze driedeling is afkomstig van Stump (1998). Anderson (1988) hanteert een vierdeling waarbij de *governed properties* opgedeeld zijn in de eigenschappen die aan een woord worden toegekend (“configurational properties”) en de eigenschappen die aan een constituent worden toegekend (“phrasal properties”) in een dependentierelatie.
3. In de meeste theoretische modellen worden twee types inflectionele vormen in het lexicon opgeslagen: onregelmatige vormen en vormen met een hoge tokenfrequentie.
4. In de morfologie spreekt men veeleer van het lexeem.
5. Voor een kritische discussie verwijzen we naar Bauer (2003: 103-105).
6. Als we terugblikken op de inflectionele categorieën van Stump (1998) en Anderson (1988) die we hierboven hebben opgesomd, dan zijn deze als volgt verdeeld over de inherente en contextuele inflectie: de *agreement properties* en de *governed properties* behoren tot de contextuele inflectie, want ze worden extern bepaald door andere elementen in de syntactische context; de *inherent properties* vormen de inherente inflectie.
7. Ten eerste, inherente eigenschappen drukken vaak conceptuele inhouden uit die de betekenis van de basisvorm veranderen, zoals dat voor het substantief het geval is bij de diminuering (Bakema 1998) en de meervoudsvorming. Merk op dat het uitdrukken van lexicale inhouden doorgaans aan de lexicale morfologie wordt voorbehouden. Ten tweede, de inherente inflectie kan woordsoorteigenschappen aanpassen: een verkleinwoord is in het Nederlands altijd onzijdig, ongeacht wat het genus van de basisvorm is. Ten derde, de voorspelbaarheid van inflectionele paradigmata is niet zo algemeen als doorgaans wordt aangenomen: niet alle inherente eigenschappen kunnen door alle leden van een woordsoort gerealiseerd worden. Dit zijn de zogenaamde defectieve woorden: *antiek*, *genade* en *verzekerdheid* missen een meervoudsvorm (Booij 1996: 31). Ten vierde, bij sommige inflectionele vormen is er sprake van lexicalisatie, zoals in *beter* ‘niet meer ziek’ of *ontzettend* ‘erg’ (Booij 1996: 34). Tot slot kan de selectie van het morfeem voor het uitdrukken van een inherente eigenschap semantisch gedifferentieerd zijn, zoals geïllustreerd door de meervoudsvorming van *vader*: *vaders* ‘verwekkers’ tegenover *vaderen* ‘voorouders’ (Booij 1996: 32).
8. Ferguson & Barlow (1988) beginnen de inleiding van hun bundel over congruentie met de opmerking dat er geen unanimititeit bestaat over welke fenomenen tot de congruentie behoren. De afbakeningskwestie is irrelevant voor ons onderzoek omdat de buiging van het attributieve adjectief in alle analyses tot de NP-interne congruentie wordt gerekend. De discussie over de domeinafbakening gaat hoofdzakelijk over de inclusie/exclusie van persoonlijke en betrekkelijke voornaamwoorden. Wij zullen bij deze bespreking – in

navolging van de meeste recente analyses – aannemen dat de voornaamwoorden tot het domein van de congruentie behoren.

9. Zwicky (1986) onderscheidt voor het Duits, waar de distinctie tussen de zwakke en sterke buiging behouden is (zie hoofdstuk 2), tussen “agreement” en “government”: de propagatie van de kenmerken *getal*, *genus* en *casus* sorteert hij onder de noemer congruentie, want ze emaneren van het nominale hoofd; de distinctie tussen sterke en zwakke flexie beschouwt hij echter als een instantie van “idiosyncratic government” (Zwicky 1986: 984), want hun distributie wordt bepaald door de categorie van de determinator.
10. Het is duidelijk dat de congruentie van het attributieve adjectief in het Nederlands, zoals beschreven in figuur 3(1), problematisch is voor benaderingen van congruentie die NP-interne congruentie als een instantie van *modifier-head agreement* beschouwen. De verbuiging van het attributieve adjectief in het Nederlands – en andere Germaanse talen – kent een complicerende factor, met name de definietheid van de determinator. Anderson (1991: 108-109) stelt voor om de morfosyntactische kenmerken niet onmiddellijk van het hoofd naar de modifier te kopiëren, maar om dit middels het niveau van de NP te doen waar dan alle relevante kenmerken gecumuleerd worden. Deze mogelijkheid is expliciet aanwezig in grammaticale modellen die gebruik maken van unificatie, zoals HPSG. De bestaande modellen voor de congruentie van het Nederlandse attributieve adjectief passen dit inzicht op een of andere wijze toe (De Schutter 1994, 1997; Kester 1996; Odijk 1992; Van Eynde 2003).
11. NP-externe en zinsexterne congruentie heeft volgens Anderson (1991: 113-114) een co-indexerende functie. De vorm van het vervoegde werkwoord heeft een refererende functie als het subject afwezig is, zoals in het Italiaanse *parl-o* (‘spreken-PERS1.ENK’, ‘ik spreek’), en een corefererende functie als het subject aanwezig is, zoals in (3).
12. In HPSG worden morfosyntactische congruentie en “index agreement” onderscheiden (Pollard & Sag 1994). *Index agreement* is de tegenhanger van semantische congruentie en heeft betrekking op de kenmerken die geen eigenschap van het talige teken maar van de referent uitdrukken.
13. Hier kan de indruk ontstaan dat referentiële kenmerken elkaar nooit uitsluiten. Dit is niet het geval: Barlow verklaart de onmogelijkheid van een NP als *grand-e garçon* (‘groot-FEM.ENK jongen’) in (2b) als een gevolg van de onverenigbare kenmerken <IN-FEMINIENE-KLASSE> bij het adjectivische buigingssuffix en <IN-MASCULIENE-KLASSE> bij het nominale hoofd.



## Hoofdstuk 5:

# Methodologische facetten

De gehanteerde onderzoeksmethodologie kan het best omschreven worden als kwantitatief-empirisch onderzoek en meer in het bijzonder als kwantitatief corpusonderzoek. Deze methode is uitermate geschikt voor het realiseren van de descriptieve basisdoelstelling van deze dissertatie: het in kaart brengen van de buigingsvariatie die optreedt in definiete NP's met een enkelvoudig *het*-woord als hoofd. Deze basisdoelstelling kan verfijnd worden tot deze drie doelstellingen:

- de identificatie van de buigingsvariatie die optreedt in definiete NP's met een enkelvoudig *het*-woord als hoofd
- de identificatie van de contexten waarin beide buigingsvarianten voorkomen
- de identificatie en modellering van de significante tendenties die het gebruik van beide buigingsvarianten bepalen

In de volgende paragrafen beschrijven we de methodologische stappen die we zullen volgen om deze onderzoeksdoelstellingen te realiseren. Eerst motiveren we het gebruik van corpusmateriaal en bespreken we de weerslag hiervan op de argumentatie (5.1). Vervolgens wordt de globale methodologische fasering van het onderzoek voorgesteld (5.2). Hierbij wordt stilgestaan bij de implicaties van de methodologische keuzes voor het gevoerde onderzoek. Tot slot situeren we de gebruikte methode in de ruimere corpuslinguïstische onderzoekstraditie (5.3).

### 5.1 Motivatie corpusonderzoek

In deze paragraaf motiveren we de keuze voor corpusonderzoek aan de hand van de vier verschillende datagaringsmethoden die Labov (1972b) onderscheidt: introspectie, enquêtering, experimenten en corpusonderzoek. Deze vier technieken om talig materiaal te verzamelen, kunnen op basis van drie parameters beschreven worden (Tummers, Heylen & Geeraerts 2005). Deze parameters zijn:

- de epistemologische basis van de gegevens: zijn de gegevens een gerapporteerde intuïtie of betreft het reëel taalgebruik?
- de spontaneïteit van de gegevens: zijn de gegevens spontaan gerealiseerd of geëliciteerd/geconstrueerd?
- de processuele status van de gegevens: zijn de gegevens georiënteerd op het product van taalgebruik (items, constructies, patronen) of op het proces van taalgebruik?

Tabel 5(1) toont de waarden van deze parameters voor de verschillende door Labov onderscheiden datagarings technieken:

	<b>introspectie</b>	<b>enquêtering</b>	<b>experiment</b>	<b>corpus</b>
<b>epistemologische basis</b>	intuïtie	intuïtie	taalgebruik	taalgebruik
<b>spontaneïteit</b>	elicitatie	elicitatie	elicitatie	spontaan
<b>processuele status</b>	product	product	proces/product	product

Tabel 5(1): Datagaringstechnieken

Zoals elke synopsis is ook het schema in tabel 5(1) generaliserend. De breuklijn tussen de verschillende technieken is niet altijd even scherp als het schema suggereert. Zo zijn er geavanceerde vormen van enquêtering, vooral in antropologisch veldonderzoek (Levinson & Meira 2003), die qua opzet en conditionering vergelijkbaar zijn met psycholinguïstische experimenten. Ondanks de inherente simplificatie toont de tabel onweerlegbaar aan waarom corpusonderzoek de meest geschikte methode is voor onze onderzoeksdoelinden: het is de enige manier om spontaan gerealiseerde taaldata, zeg maar het spontane taalgebruik, te verzamelen.

Corpusonderzoek is niet alleen de meest aangewezen methode voor de realisatie van onze onderzoeksdoelstellingen, maar het gebruik van de overige technieken is in verschillende opzichten problematisch. Hudson (1996: 148-150) sluit introspectie uit bij variatieonderzoek omdat het graduele karakter van – lectische en socio-linguïstische – variatie tot kwantitatief onderzoek noopt.

Tegen het gebruik van (traditionele) enquêtes, waar de respondenten op de hoogte zijn van de onderzoeksvraag, kan worden aangevoerd dat de flexibiliteit van het reële taalgebruik vaak het vermogen van de individuele taalgebruikers overstijgt door de discrepantie tussen wat taalgebruikers denken te doen en hun feitelijke taalgebruik (Geeraerts, Grondelaers & Bakema 1994: 17). De sociolinguïstische motivatie voor enquêtering, namelijk de compositie van een gestratificeerde materiaalverzameling, is niet van toepassing door het gebruik van een geografisch en stilistisch gedifferentieerd corpus, met name het *Corpus Gesproken Nederlands*: de macrostructuur van dit corpus bevat zowel een geografische (Nederlands Nederlands vs. Belgisch Nederlands) als een stilistische dimensie (formele vs. informele gesprekscontexten) (zie paragraaf 6.1).

Experimenteel onderzoek en geavanceerde enquêtes stellen dan weer hoge eisen aan de gebruikte items met betrekking tot o.a. de frequentie en de verwerkingscomplexiteit. Een bijkomende conditie op experimenteel onderzoek is dat andere, interfererende factoren gecontroleerd moeten worden. Doordat er geen globaal descriptief model voor de buigingsalternantie bestaat, is geavanceerd experimenteel onderzoek – nog – niet aan de orde. Corpusonderzoek daarentegen biedt de mogelijkheid om de condities op het gebruik van een item/constructie en de mogelijke interdependenties tussen de verschillende condities te identificeren. Daarnaast kan de globale impact van de verschillende condities gemodelleerd worden (zie paragraaf 5.2.3). Experimenteel onderzoek vormt dan de volgende fase

in de onderzoekscyclus waarbij de impact van één of meerdere factoren op het talige productie- of verwerkingsproces wordt getest, rekening houdend met het effect van andere factoren en eventuele interdependenties die in het corpusonderzoek geïdentificeerd zijn. Deze methodologie leidt tot een stapsgewijze verfijning van de onderzoeksvraag. Voorbeelden van zulke graduele verfijning zijn Pander Maat & Sanders (2001) en Grondelaers et al. (2001). Daarnaast worden experimentele gegevens vaak bovenop corpusmateriaal gebruikt als onafhankelijke evidentie voor de cognitieve mechanismen aan de basis van een item/constructie (Bybee te verschijnen; Croft & Cruse 2004).

De onderzoeksmethodologie die we toepassen, kan het best als *corpus-gebaseerd* gekwalificeerd worden.<sup>1</sup> Dit betekent dat het corpusmateriaal de kern van het onderzoek vormt en het uiteindelijke descriptieve model bepaalt. Anekdotische en impressionistische beschrijvingen zijn bijgevolg uitgesloten. De corpus-gebaseerde oriëntatie impliceert dat we niet in individuele gevallen geïnteresseerd zijn maar in de algemene tendenties die de selectie van beide buigingsvarianten in het reële taalgebruik sturen. Deze demarche vereist dat alle observaties in het taalgebruik door middel van een zo objectief mogelijke procedure geanalyseerd en gecategoriseerd worden (zie paragraaf 5.2.2).

De kwantitatief-empirische en corpuslinguïstische methodologie beïnvloedt de gevoerde argumentatie. In corpuslinguïstisch onderzoek verschuift de aandacht van de 'potentialiteit' naar de 'probabiliteit'. Het is de gangbare traditie in (introspectief) taalkundig onderzoek om de 'potentialiteit' te beschrijven. De taalkundige is geïnteresseerd in wat mogelijk, c.q. grammaticaal, is in een bepaalde taal. Hiervoor doet h/zij hoofdzakelijk een beroep op negatieve evidentie: door middel van grammaticaliteitsoordelen gaat de taalkundige het toepassingsgebied van een item/constructie bepalen door in eerste instantie na te gaan wat het item/de constructie niet is (c.q. de agrammaticale items/constructies).<sup>2</sup> Kwantitatief-empirisch onderzoek daarentegen heeft als doelstelling om middels positieve evidentie te beschrijven wat waarschijnlijk is in een taal en hoe waarschijnlijk een item/constructie is gegeven de aan- of afwezigheid van bepaalde condities. Het resulterende model is bijgevolg niet absoluut maar probabilistisch. Daarnaast kan het model niet alle observaties in de materiaalverzameling correct voorspellen. Dit vormt echter geen probleem wanneer men de inherente variabiliteit van het taalgebruik als vertrekpunt van het onderzoek neemt (Paolillo 2001: 2-4). Dit staat in schril contrast met de taalkundige modellen die in de structuralistische en generatieve traditie zijn voorgesteld en waar alle introspectief verzamelde items door middel van (een zo economisch mogelijke verzameling van) symbolische regels beschreven worden.

## 5.2 Methodologische fasering

In corpuslinguïstisch onderzoek kunnen *grosso modo* drie methodologische fases onderscheiden worden: de selectie van het materiaal (5.2.1), de analyse en annotatie van het materiaal (5.2.2), en tot slot het samenvatten van de gegevens en het interpreteren van de resultaten (5.2.3).

### 5.2.1 Selectie van het materiaal

De empirische basis van deze dissertatie bestaat uit twee corpora, met name de 5<sup>e</sup> pre-release van het *Corpus Gesproken Nederlands* en de jaargang 2003 (meer bepaald de edities van januari tot en met september) van de Vlaamse kwaliteitskrant *De Standaard*. Uit beide corpora zijn de attestaties geëxtraheerd van de syntactische context waar de bestudeerde buigingsalternantie optreedt: [DET<sub>[+DEFINIET]</sub> ADJ N<sub>[+ONZIJDIG,+ENKELVOUD]</sub>]<sub>NP</sub>. Beide materiaalverzamelingen en de procedure gebruikt bij het exciperen ervan worden besproken in hoofdstuk 6 (*Corpus Gesproken Nederlands*) en hoofdstuk 15 (*De Standaard*-corpus).

### 5.2.2 Analyse van het materiaal

De analyse van de materiaalverzameling is de meest arbeidsintensieve fase van het empirisch onderzoek en bestaat uit de falsificatie van de hypotheses die op basis van de bestaande literatuur geformuleerd kunnen worden of die op basis van eigen observaties tot stand zijn gekomen. Rekening houdend met de ongelijke status van beide buigingsvarianten, de verbogen vorm als zijnde de basisvariant en de onverbogen vorm als zijnde de gemarkeerde variant, zullen we hoofdzakelijk op zoek gaan naar condities die het gebruik van de gemarkeerde variant motiveren. Deze methode is grotendeels hypothetisch-deductief (*top-down* of *hypothesis-driven*), hoewel we ook hypotheses testen die uit eigen observaties volgen (*bottom-up* of *data-driven*).

Door de hypothese-testende oriëntatie van het onderzoek zullen we een bijzondere aandacht besteden aan de operationalisering van de hypotheses. Hoewel dit een essentieel onderdeel is van empirisch onderzoek, wordt er doorgaans omwille van de vanzelfsprekendheid weinig aandacht aan besteed. Bij de operationalisering en de falsificatie van de onderzoekshypotheses streven we naar:

- een maximale consistentie
- een maximale objectiviteit indien mogelijk, en anders een maximale intersubjectiviteit
- een maximale repliceerbaarheid

Uiteraard zijn de drie genoemde kenmerken onderling nauw verbonden: repliceerbaarheid veronderstelt objectiviteit, die op haar beurt consistentie vereist. Om een maximale consistentie te bereiken, zullen we de inbreng van de intuïtie van de onderzoeker tot een minimum trachten te beperken. Wanneer een beroep op de eigen intuïtie wordt gedaan, zal deze door het formuleren van criteria gestuurd

worden. Deze criteria zijn gericht op een gelijke – of op zijn minst vergelijkbare – behandeling van alle observaties in de materiaalverzameling.

Bij de analyse en de annotatie van het materiaal maken we gebruik van *Abundantia Verborum* (Speelman 1997). Dit programma is ontworpen voor het verzamelen en analyseren van taalkundige gegevens.<sup>3</sup> Het gebruik van een digitale onderzoeksomgeving dwingt de onderzoeker tot een maximale consistentie bij de analyse en de classificatie van de observaties in de materiaalverzameling.

### 5.2.3 *Samenvatting en interpretatie van de gegevens*

Na de selectie van het relevante materiaal en de analyse en annotatie van de observaties in de materiaalverzameling rest de samenvatting en de interpretatie van de gegevens. Dit is het domein van de statistiek: we maken gebruik van technieken uit de descriptieve statistiek voor het beschrijven en het samenvatten van de observaties in de gegevensbank; technieken uit de inferentiële statistiek worden aangewend om na te gaan of de vaststellingen voor de staalname, c.q. het corpus, naar de gehele populatie geëxtrapoleerd kunnen worden.

Globaal gezien bestaat de empirische analyse uit twee delen: een bivariaat deel (deel II) en een multivariaat deel (deel III). De bivariate analyse komt chronologisch het eerst en behelst de analyse van de relatie tussen twee variabelen, waarbij één van beide variabelen de responsvariabele is en de andere de verklarende/onafhankelijke variabele.<sup>4</sup> Deze analyse neemt doorgaans de vorm aan van een  $2 \times k$ -contingencietabel (een binomiale responsvariabele en een binomiale of multinomiale onafhankelijke variabele). Het doel van deze analyse is na te gaan of er een significante interactie of dependentie tussen beide variabelen is en, indien ja, wat de aard van deze interactie is. Deze fase is uitermate geschikt voor het optimaliseren van de parametrisering van een verklarende variabele. Hieronder verstaan we de identificatie van de ideale configuratie van een verklarende variabele, niet alleen voor het maximaliseren van de kwantitatieve impact van de variabele in kwestie, maar ook met het oog op de taalkundige interpretatie en verklaring.

De bivariate analyse is echter een idealisering omdat er geen rekening wordt gehouden met het effect van andere verklarende factoren en mogelijke interferenties tussen verschillende verklarende factoren. Rekening houdend met de complexiteit van het fenomeen natuurlijke taal lijkt het ons zeer onwaarschijnlijk dat het taalgebruik in het algemeen en de buigingsalternantie in het bijzonder aan de hand van één enkele verklarende variabele kan worden verklaard. Bijgevolg worden de bivariate analyses van de afzonderlijke verklarende factoren aangevuld met stratumanalyses en een multivariate analyse. Deze technieken zijn onmisbaar voor de identificatie van mogelijke interacties tussen verschillende verklarende variabelen en voor het modelleren van het cumulatieve effect van de rist verklarende variabelen die in de literatuur worden aangehaald (zie paragraaf 1.1).

De eerste methodologische extensie van het traditionele bivariaat onderzoek is de identificatie van partiële of conditionele associaties door middel van een

stratumanalyse (Agresti 1996: 54-60). Een stratumanalyse wordt uitgevoerd om na te gaan of het effect van een verklarende variabele op significante wijze wordt beïnvloed door een tweede verklarende variabele, de zogenaamde stratumvariabele. We testen niet systematisch alle logische twee-aan-twee combinaties van verklarende factoren, maar we beperken het stratumonderzoek tot de combinaties waarvoor het gegrond is een conditionerend effect te veronderstellen op basis van het literatuuroverzicht<sup>5</sup> of op basis van observaties tijdens het empirisch onderzoek.

De tweede methodologische extensie ten aanzien van het bivariaat onderzoek is het uitvoeren van een multivariate analyse. In een multivariate analyse wordt de impact van iedere verklarende factor gekwantificeerd terwijl het effect van de andere factoren gecontroleerd wordt. Dit resulteert in een globaal statistisch model waarin de verschillende significante factoren geïntegreerd zijn. Het construeren van een multivariaat model biedt de mogelijkheid om vragen te beantwoorden die in bivariaat onderzoek eenvoudigweg niet gesteld (kunnen) worden:

- Welke factoren hebben een significante impact op de responsvariabele? Met andere woorden: welke factoren reduceren op significante wijze de variatie in de materiaalverzameling?
- Wat is de impact op de responsvariabele van de verschillende significante factoren in het model?
- Wat is de verklarende waarde van het geïntegreerde model? Met andere woorden: hoeveel variatie verklaart het geïntegreerde model?

Het testen van meerdere factoren en het gebruik van multivariate technieken vloeit voort uit de ongecontroleerde samenstelling van corpora. Onder ‘ongeccontroleerd’ verstaan we dat corpora spontaan gerealiseerd materiaal bevatten, niet dat corpora lukrake samenraapsels zijn (bij wijze van voorbeeld verwijzen we naar de samenstelling van het *Corpus Gesproken Nederlands*, beschreven in paragraaf 6.1 en tabel 6(1)). In tegenstelling tot onderzoekers die een beroep doen op enquêtes en experimenten, waar gebruik gemaakt wordt van lijsten met test-items die de invloed van een bepaalde factor testen en andere factoren onder controle houden, wordt de corpuslinguïst geconfronteerd met de volledige heterogeniteit en complexiteit van het taalgebruik. Deze vaststelling heeft verschillende analytische implicaties. Ten eerste, het onderzoek kan niet beperkt worden tot één enkele verklarende variabele, maar de talige constructie moet in zijn volledige complexiteit worden benaderd.<sup>6</sup> Ten tweede, de onderlinge samenhang tussen de verschillende verklarende factoren moet onderzocht worden evenals hun gezamenlijke impact op het bestudeerde fenomeen. Ten derde, de heterogeniteit van spontaan gerealiseerd materiaal is niet beperkt tot taalstructurele factoren maar heeft ook betrekking op externe factoren zoals de regionale/nationale variëteit, het register en andere kenmerken van de gesprekscontext. Aangezien een corpus vaak over verschillende lectische variëteiten en gesprekscontexten geaggregeerd is, moet het eventuele effect van deze factoren gecontroleerd worden.

De statistische technieken die gebruikt worden bij het bivariaat en het multivariaat onderzoek worden respectievelijk toegelicht in de hoofdstukken 7 en 13. De statistische analyses zijn uitgevoerd met behulp van het R-softwarepakket (R Development Core Team 2004).<sup>7</sup>

### 5.3 Situering van het onderzoek in de corpuslinguïstische traditie

De situering van voorliggend onderzoek binnen het domein van de corpuslinguïstiek bestaat uit twee delen. In eerste instantie plaatsen we het onderzoek in het ruimere kader van de corpuslinguïstiek, meer bepaald het multivariate corpusonderzoek. Daarna wijzen we aan de hand van de specifieke kenmerken van het variantenonderzoek binnen de onderzoeksgroep *Quantitative Lexicology and Variational Linguistics* (QLVL) aan de KU Leuven op de specificiteit van dit onderzoek ten aanzien van andere multivariate corpusanalyses.

De corpuslinguïstiek is geen taalkundige theorie maar een methodologisch kader voor kwantitatief-empirisch onderzoek. Deze primair methodologische oriëntatie vertaalt zich in onderzoek dat de nadruk legt op de descriptieve aspecten van het beschreven fenomeen en nauwelijks aandacht aan de theoretische interpretatie besteedt. Dit is onder andere duidelijk het geval in de Angelsaksische traditie (*inter al.* Hunston 2001; Stubbs 1995, 2001). Het omgekeerde komt ook voor: taalkundige stromingen met een primair theoretische oriëntatie tonen nauwelijks interesse in empirische onderzoeksmethoden. Het laatste decennium stellen we echter een groeiende interactie vast tussen theorievormende en methodologische strekkingen binnen de taalkunde.<sup>8</sup> Hoewel we in de vorige paragraaf betoogd hebben dat corpusgegevens door hun aard in feite om multivariaat onderzoek vragen, worden deze technieken relatief zelden aangewend in taalkundig onderzoek. Tijdens de laatste jaren is er niettemin een toename van kwantitatief-empirisch onderzoek, waarin ook meer en meer gebruik wordt gemaakt van multivariate technieken.<sup>9</sup> Als voorbeeld kunnen de volgende analyses worden aangehaald: Bergen (te verschijnen) verricht multivariaat onderzoek naar het optreden van de *liaison* in het Frans; Gries (2001, 2003) modelleert de verschillende factoren die de positie van het partikel bij de zogenaamde *phrasal verbs* in het Engels reguleren; Poplack (2001) bestudeert de alternantie tussen verschillende conditionele constructies in het Frans; Szmrecsanyi (2005) analyseert de condities op het voorkomen van *priming*.

Het gebruik van multivariate technieken wordt echter al langer toegepast in de taalkunde. Het sociolinguïstische onderzoek naar *variable rules* in de jaren '60 en '70 van de vorige eeuw maakt gebruik van multivariate analyses (Labov 1969, 1980): technisch gezien zijn de *VarbRul*-analyses (Rousseau & Sankoff 1978) logistische regressies. Dit is de multivariate techniek die in onderhavig onderzoek gebruikt wordt (zie hoofdstukken 13, 14 en 17). Vandaag vormt de *VarbRul*-techniek – of de vernieuwde *GoldVarb*-techniek (Rand & Sankoff 1990) – nog steeds de methodologische ruggengraat van een niet onbelangrijke strekking binnen

de sociolinguïstiek (*inter al.* Clarke 1997; Dubois & Horvath 2003; Guy 1994; Tagliamonte 2003). Ook binnen de Angelsaksische corpuslinguïstische traditie is er sporadisch met multivariaat onderzoek geëxperimenteerd (Leech, Francis & Xu 1994), net als binnen de historische taalkunde (Ebert 1992; Nevilainen & Raumolin-Brunberg 2003).

Binnen QLVL vormen multivariate analyses naar grammaticale alternanties de kern van het variantenonderzoek. Dit is het onderzoek naar de condities op het voorkomen van alternatieve realisatievormen van een item/constructie.<sup>10</sup> Deze onderzoekslijn omvat het onderzoek naar het gebruik van presentatief *er* in het Nederlands (Grondelaers 2000; Grondelaers et al. 2001, 2002), het onderzoek naar de rode en groene woordorde in een Nederlandse completiefzin (De Sutter 2005) en het onderzoek naar de woordvolgordevariatie in het *Mittelfeld* in het Duits (Heylen 2005). Uit deze verschillende onderzoeksprojecten rijst een coherent beeld op: al deze alternanties zijn complexe types van variatie die gereguleerd worden door een samenspel van uiteenlopende factoren die in de traditionele taalkunde zelden gecombineerd worden omdat ze tot verschillende taalkundige domeinen behoren.

Het QLVL-onderzoek onderscheidt zich op de volgende punten van het hierboven vermelde multivariate onderzoek. Ten eerste, taalinterne en taalexterne factoren worden gecombineerd. Naast taalstructurele, semantische en discursieve restricties, wordt expliciet onderzoek verricht naar het effect van lectische factoren en hun impact op enerzijds de bestudeerde alternantie en anderzijds het voorkomen van de andere factoren. Bovendien worden interne en externe factoren niet strikt gescheiden. Lectische factoren worden beschouwd als onderdeel van de conceptueel-semantische structuur van een item/constructie. Taalgebruikers wenden lectische factoren actief aan om bij een talige interactie een bepaald beeld van zichzelf te presenteren of om de relatie met de gesprekspartner te definiëren (Kristiansen 2003). Ten tweede, het QLVL-onderzoek is gericht op de identificatie van de factoren die aan de basis van het taalgebruik liggen en hoe deze factoren het gebruik van talige varianten conditioneren. Dit is een belangrijk verschilpunt met o.a. het onderzoek van Gries (2003): het einddoel van Gries' multivariate analyse is de ontwikkeling van een cognitief beslissingsmechanisme om de keuze tussen twee alternatieve constructies te reguleren.<sup>11</sup>

#### 5.4 Conclusie

In dit hoofdstuk hebben we de methodologische facetten van het onderzoek voorgesteld. Eerst hebben we betoogd dat kwantitatief corpuslinguïstisch onderzoek de meest geschikte keuze is ten aanzien van onze onderzoeksdoelstelling, met name het modelleren van de tendenties die de buigingsalternantie bij neutra bepalen, en ten aanzien van de stand van zaken met betrekking tot het onderzoeksobject, namelijk het ontbreken van een globaal model voor de buigingsvariatie (5.1). Vervolgens hebben we de methodologische fasering voorgesteld (5.2). Deze omvat



drie delen: het verzamelen van het materiaal, het analyseren van het materiaal en het samenvatten en interpreteren van het materiaal. We hebben aangevoerd dat de analyse van ongecontroleerd taalgebruik een verfijndere analyse behoeft dan de traditionele bivariate analyses, waar aan de hand van  $2 \times k$ -contingentietabellen het effect van afzonderlijke verklarende factoren wordt getest. De verfijning bestaat uit twee stappen: primo, mogelijke interacties tussen verklarende factoren worden getest door middel van stratumanalyses; secundo, de verschillende verklarende factoren worden in een globaal statistisch model geïntegreerd aan de hand van een logistische-regressieanalyse. Deze twee verfijningen vloeien voort uit de bevindingen op basis van het literatuuroverzicht, waar enerzijds potentiële interacties tussen verklarende factoren aan het licht zijn gekomen en waar anderzijds op het ontbreken van een globaal descriptief model voor de buigingsalternantie bij neutra is gewezen (zie hoofdstuk 1). Tot slot hebben we het onderzoek binnen de QLVL-onderzoeksgroep in het algemeen en het eigen onderzoek in het bijzonder gesitueerd binnen de corpuslinguïstische traditie (5.3).

Het originele karakter van onderhavig onderzoek ligt in de combinatie van een ruime descriptieve invalshoek en het aanwenden van een doorgedreven kwantitatief-empirische methodologie. De descriptieve basis bestaat niet uit geconstrueerde en immer herkauwde voorbeelden (zie 1.2.2), maar uit reële en spontaan gerealiseerde taaldata. Met het oog op het verklaren van de buigingsalternantie wordt geen enkele verklarende factor *a priori* om theoretische motieven uitgesloten. Tot slot biedt de gebruikte methodologie de mogelijkheid om een geïntegreerd model te construeren waarin de impact van de verklarende factoren in hun onderlinge samenhang wordt berekend.

Op basis van de recensie van de bestaande literatuur en de gehanteerde methodologie kunnen de volgende onderzoeksvragen geformuleerd worden:

- Wat is de verdeling van beide buigingsvarianten? Met andere woorden: hoe marginaal is de onverbogen variant?
- Welke factoren (die in de literatuur worden aangehaald) hebben een significante impact op de buigingsalternantie?
- Wat is de impact van de verschillende significante factoren in het geïntegreerde model?
- Hoeveel variatie kunnen we op basis van het geïntegreerde model verklaren? Met andere woorden: hoe aleatoir is de buigingsalternantie? En: hoe adequaat is het model dat we construeren voor de beschrijving en de verklaring van de buigingsalternantie?
- Wordt het taalgebruik in Nederland en in Vlaanderen door verschillende onderliggende systemen bepaald? Met andere woorden: zijn er factoren met een significant verschillend effect of een significant verschillende impact in het Belgisch Nederlands en het Nederlands Nederlands?

DEEL I

Tot slot hopen we dat de eigenlijke analyse de descriptieve en methodologische meerwaarde van kwantitatief-empirisch onderzoek binnen de taalkunde illustreert.

## Noten

1. In Tummers, Heylen & Geeraerts (2005) onderscheiden we “corpus-based” onderzoek van “corpus-illustrated” onderzoek. In het tweede type onderzoek wordt een corpus beschouwd als een databank voor de selectie van niet-geconstrueerde voorbeelden die introspectieve gegevens aanvullen of vervangen. In tegenstelling tot wat in corpus-gebaseerd onderzoek het geval is, vormen de empirische gegevens in corpus-geïllustreerd onderzoek niet de ruggengraat van de analyse, maar wordt de aanwezigheid van een item/constructie in het corpus als het ware als een positief grammaticaliteitsoordeel geïnterpreteerd.
2. Dit kan beschouwd worden als het doordenken van de Saussures differentieeltheesthese.
3. Omdat we de corpora door middel van *Python*-scripts (*Corpus Gesproken Nederlands*) en *Perl*-scripts (*De Standaard*) geëxcerpeerd hebben, maken we enkel gebruik van de databankmodule binnen *Abundantia Verborum*. In het algemeen streven we bij het empirisch onderzoek een maximale automatisering door middel van *Perl*-scripts na, zonder echter de kwaliteit van de analyses in gevaar te brengen.
4. In voorliggend onderzoek is de buigingsalternantie de responsvariabele en vormen de verschillende verklarende factoren de onafhankelijke variabelen. We zullen de termen onafhankelijke variabele, verklarende variabele, (verklarende) parameter en (verklarende) factor door elkaar gebruiken.
5. We verwijzen naar paragraaf 1.2.1 en tabel 1(2) voor een overzicht van de mogelijke conditionele associaties tussen verklarende variabelen.
6. Zelfs indien men slechts in het effect van één enkele factor geïnteresseerd is, moet men strikt genomen het effect van de andere factoren controleren.
7. We hebben versie 2.0.0 gebruikt (datum release: 4 oktober 2004). R biedt een grotere controle op de processen en de uitvoer alsook de mogelijkheid om zelf functies te definiëren.
8. Een dergelijke toenadering tussen theoretische en kwantitatief-empirische strekkingen binnen de taalkunde vindt onder andere plaats in het kader van de Cognitieve Taalkunde. Voor een kritische bespreking van het kwantitatief-empirisch onderzoek binnen het domein van de Cognitieve Taalkunde verwijzen we naar Tummers, Heylen & Geeraerts (2005).
9. De opkomst van geavanceerd kwantitatief-empirisch onderzoek voor de analyse van (traditionele) grammaticale fenomenen mag niet overroepen worden. Zo bevat de bundel die een overzicht geeft van de *usage-based* benadering in de neerlandistiek geen enkele multivariate analyse (Verhagen & van de Weijer 2003).
10. Naast het variantenonderzoek, met focus op de afzonderlijke talige verschijnselen, is er binnen QLVL een tweede onderzoeklijn met focus op de teksten (c.q. deelcorpora) als

## DEEL I

verzamelingen van talige verschijnselen. Dit is het zogenaamde variëteitenonderzoek, dat o.a. de afstand tussen het Belgisch Nederlandse taalgebruik en het Nederlands Nederlandse taalgebruik meet (Geeraerts, Grondelaers & Speelman 1999) en dat de verschillende registers in het Belgisch Nederlands identificeert en onderling positioneert (Speelman, Grondelaers & Geeraerts 2003).

11. Voor een uitgebreide bespreking van het onderzoek binnen de QLVL-onderzoeksgroep verwijzen we naar Heylen, Tummers & Geeraerts (te verschijnen).

## DEEL II

### CONDITIONERING VAN DE BUIGINGSALTERNANTIE: OPERATIONALISERING EN KWANTIFICERING

Na de algemene voorstelling en situering van het onderzoeksobject en de kwantitatief-empirische onderzoeksmethodologie vatten we in het tweede deel van dit proefschrift de eigenlijke empirische analyse aan. De doelstelling van dit deel is het identificeren en kwantificeren van de condities op de adjectivische buigingsalternantie bij neutra, en meer bepaald van de restricties op het gebruik van het gemarkeerde onverbogen adjectief. De werkwijze die we hierbij toepassen, is het falsificeren van onderzoekshypotheses die we op basis van de bestaande literatuur (zie hoofdstuk 1) of tijdens het empirisch onderzoek formuleren. Dit doen we door middel van bivariate analyses die de impact van de afzonderlijke potentiële verklarende factoren op de buigingsalternantie berekenen.

Het zwaartepunt van deze bivariate analyses ligt bij de operationalisering en de interpretatie van de afzonderlijke condities op de buigingsalternantie. We zullen voor iedere potentiële verklarende factor een procedure voorstellen die een objectiveerbare en repliceerbare analyse van de observaties in de materiaalverzameling mogelijk maakt. Het resultaat van deze onderzoeksfase vormt de empirische basis van de kwantitatieve analyses die in de volgende hoofdstukken gerapporteerd worden. Deze initiële en arbeidsintensieve onderzoeksfase wordt doorgaans als een vanzelfsprekendheid beschouwd en derhalve tot een obligate randbemerking beperkt. Rekening houdend met de methodologische rigueur die we onszelf opleggen, besteden we bijzondere aandacht aan de operationalisering van de verklarende factoren, aangezien deze keuzes voor een (groot) deel de resultaten en de interpretatie van het empirisch onderzoek beïnvloeden.

Dit tweede deel is op de volgende wijze georganiseerd. Eerst presenteren we de samenstelling en de raadpleging van het *Corpus Gesproken Nederlands* (hoofdstuk 6). Daarna lichten we de gebruikte statistische technieken toe (hoofdstuk 7). De bedoeling van dit hoofdstuk is om de lezer een algemeen inzicht te geven in de statistische technieken die in het bivariate onderzoek gebruikt worden: we zullen de interpretatie van de verschillende statistische technieken aan de hand van een becommentarieerd voorbeeld verduidelijken. De overige hoofdstukken behandelen

de potentiële verklarende factoren van de buigingsalternantie, gegroepeerd per taalkundig domein. We analyseren achtereenvolgens de volgende verklarende factoren: de morfosyntactische conditionering (hoofdstuk 8), de fonologische conditionering, zowel segmenteel (hoofdstuk 9) als suprasegmenteel (hoofdstuk 10), de lexicale/semantische conditionering (hoofdstuk 11) en de lectische conditionering (hoofdstuk 12).

De analyse van de verklarende factoren verloopt volgens een vast stramien: eerst herhalen we de hypothese(s) die op basis van het literatuuroverzicht voorgesteld is/zijn (zie hoofdstuk 1) en lichten we de operationalisering van deze hypothese(s) toe; vervolgens worden de resultaten voorgesteld en besproken. Indien nodig worden tijdens de bespreking nieuwe hypothesen geformuleerd en getest.

Alvorens de empirische analyse aan te vatten, introduceren we een nieuwe term: het buigingsprofiel van een adjectivisch lemma. Hieronder verstaan we de frequentieverdeling van beide buigingsvarianten van een adjectivisch lemma.<sup>1</sup> We illustreren deze notie aan de hand van een voorbeeld. Als het adjectivische lemma *groot* in het corpus met 9 verbogen attestaties en 1 onverbogen attestatie voorkomt, is het buigingsprofiel van dit lemma  $\frac{\text{verbogen}=9(90\%)}{\text{onverbogen}=1(10\%)}$ . Deze notie kan uiteraard toegepast worden op alle relevante contexten in het onderzoek. We zullen met andere woorden spreken over het buigingsprofiel van een substantivisch lemma of het buigingsprofiel van een regionale variëteit van het Nederlands: als bij de vormen van het substantivische lemma *kind* bijvoorbeeld 6 keer een verbogen adjectief en 4 keer een onverbogen adjectief geselecteerd wordt, dan vormt de verdeling  $\frac{\text{verbogen}=6(60\%)}{\text{onverbogen}=4(40\%)}$  het buigingsprofiel voor het substantivische lemma *kind*.

Binnen het geheel van dit proefschrift vormen de bivariate analyses van de afzonderlijke verklarende factoren het eerste deel van een methodologisch tweeluik. Na de bivariate analyses, waar de aandacht hoofdzakelijk uitgaat naar de operationalisering en de interpretatie van de afzonderlijke condities op de buigingsalternantie, zullen we in het derde deel van dit proefschrift deze verklarende factoren in een globaal model integreren voor de verklaring van de buigingsalternantie in het standaard gesproken Nederlands, waarvan het *Corpus Gesproken Nederlands* een staalname is.

## Noten

1. Deze profielgebaseerde aanpak is kenmerkend voor het onderzoek binnen de QLVL-onderzoeksgroep. In Geeraerts, Grondelaers & Speelman (1999) vormen onomasio-logische profielen het uitgangspunt voor de berekening van de lexicale afstanden tussen het Belgisch Nederlands en het Nederlands Nederlands. Speelman, Grondelaers & Geeraerts (2003) belichten de meerwaarde van de profielgebaseerde aanpak bij het clusteren van talige variëteiten op basis van heterogene taalkundige variabelen. De responsvariabelen in het variantenonderzoek binnen QLVL definiëren binaire profielen (zie ook paragraaf 5.3): presentatieve zinnen met *er* vs. presentatieve zinnen zonder *er* (Grondelaers 2000), rode woordvolgorde vs. groene woordvolgorde (De Sutter 2005), volgorde subject-object vs. volgorde object-subject (Heylen 2005), verbogen adjectief vs. onverbogen adjectief.

## **Hoofdstuk 6:**

# **Samenstelling en analyse van het *Corpus Gesproken Nederlands***

Het empirisch onderzoek in deel II en deel III maakt gebruik van het *Corpus Gesproken Nederlands* (CGN). We zullen achtereenvolgens de samenstelling van dit corpus bespreken (6.1) en de wijze waarop we het geëxcerpeerd hebben (6.2). Tot slot presenteren we de resulterende materiaalverzameling voor de analyse van de buigingsalternantie in het gesproken Nederlands (6.3).

### **6.1 Samenstelling van het *Corpus Gesproken Nederlands***

De doelstelling van het CGN is “de aanleg van een databank van het hedendaags Standaardnederlands zoals dat wordt gesproken door volwassenen in Nederland en Vlaanderen” (Oostdijk 2000: 281). De samenstelling van het CGN incorporeert zowel een geografische als een stilistische dimensie: de geografische dimensie onderscheidt de twee nationale variëteiten van het Nederlands; de stilistische dimensie is opgedeeld in 14 componenten die variëren van minder gecontroleerde (component 1) tot zeer gecontroleerde (component 14) gesprekssituaties. De formaliteit is gedefinieerd in termen van drie situationele parameters:

- interactietype: monoloog vs. dialoog/multiloog [mono vs. dial]
- tekstuele voorbereiding: (min of meer) voorbereid vs. spontaan [vbr vs. spont]
- publiek: publiek vs. privé [pub vs. priv]

Op het ogenblik dat we de empirische analyse hebben uitgevoerd, beschikten we over de 5<sup>e</sup> pre-release van het CGN (5 april 2002). Deze versie van het corpus omvat iets minder dan 5 miljoen woorden waarvan ongeveer driekwart van woordsoort-annotatie is voorzien. Tabel 6(1) toont de structuur en de omvang van het CGN in de vorm zoals wij het geraadpleegd hebben. De volgende informatie is in deze tabel opgenomen: het nummer en de inhoudelijke omschrijving van de corpuscomponent, de situationele kenmerken van de component en het aantal woorden met woordsoortannotatie in het Belgische en het Nederlandse deelcorpus.<sup>1</sup>

Vanuit het standpunt van onze specifieke onderzoeksdoelstelling biedt het CGN verscheidene voordelen. In eerste instantie is (een deel van de 5<sup>e</sup> pre-release van) het CGN voorzien van woordsoortannotatie. Deze informatie vergemakkelijkt de identificatie van de syntactische context waar de bestudeerde buigingsalternantie optreedt (zie paragraaf 6.2). Ten tweede, de geografisch en stilistisch



gedifferentieerde samenstelling van het CGN maakt het mogelijk om de impact te meten van de lectische factoren die voor ons onderzoek relevant zijn (zie paragrafen 1.1.1.4 en 1.1.1.5). Tot slot, de gesproken taal is de meest natuurlijke realisatievorm van taal. Van Marle (1997) definieert de gesproken standaardtaal als een hybride variëteit: als gevolg van haar positie tussen de zeer formele geschreven standaardtaal en de zeer informele dialecten vormt de gesproken standaardtaal een stilistisch continuüm met zeer formele en zeer informele variëteiten. De formele variëteiten ontstaan door het aanwenden van de geschreven norm in formele gesprekssituaties. Deze variëteiten van de gesproken standaardtaal zijn strikt genormeerd en benaderen de geschreven norm. De stilistische tegenpool omvat de informele variëteiten voor het alledaagse taalgebruik. Deze registers zijn minder strikt genormeerd en vormen de ideale biotoop voor mechanismen die met de talige productie worden geassocieerd: “in dialects [en bij uitbreiding de informele variëteiten van de gesproken standaardtaal; JT] the efficacy of performance mechanisms governing speech production may be more prominent than in standard languages” (van Marle 1997: 27).

CGN-component	Situatiele kenmerken			Belgisch Nederlands	Nederlands Nederlands	$\Sigma$
1 conversaties	dial	priv	spont	238,358	1,024,211	1,262,569
2 interviews	dial	priv	spont	315,604	229,856	545,460
4 zakelijke onderhandelingen	dial	priv	spont	0	110,294	110,294
5 interviews en discussies	dial	pub	vbr	231,199	7,101	238,300
6 discussies, debatten, vergaderingen	dial	pub	spont	130,813	204,956	335,769
7 lessen	dial	pub	spont	102,006	0	102,006
9 spontaan commentaar	mono	pub	spont	74,600	4,209	78,809
10 actualiteitsrubrieken, reportages	mono	pub	vbr	74,906	0	74,906
11 nieuwsbulletins	mono	pub	vbr	87,356	36,143	123,499
12 beschouwingen, commentaren	mono	pub	vbr	65,632	5,210	70,842
13 lezingen, toespraken	mono	pub	vbr	52,558	60,102	112,660
14 voorgelezen tekst	mono	pub	vbr	322,342	327,239	649,581
<b><math>\Sigma</math></b>				<b>1,695,374</b>	<b>2,009,321</b>	<b>3,704,695</b>

*Tabel 6(1): Samenstelling van het CGN  
(deel met woordsoortannotatie in de 5<sup>e</sup> pre-release)*

Daarnaast zijn er ook enkele minpunten aan het CGN verbonden. Ten eerste, het beschikbare materiaal is niet evenredig over de verschillende componenten verdeeld,

met zelfs enkele lege cellen (i.e. component 4 in het Belgische deelcorpus en de componenten 7 en 10 in het Nederlandse deelcorpus) en twee ontbrekende componenten (i.e. de componenten 3 en 8). Dit is uiteraard een gevolg van het gebruik van een pre-release, maar ook de definitieve versie van het CGN kent een onevenwichtige spreiding van het materiaal over de verschillende componenten en de nationaal gedefinieerde deelcorpora. Ten tweede, het Belgisch Nederlandse materiaal en het Nederlands Nederlandse materiaal zijn niet gelijk over de verschillende stilistisch gedifferentieerde componenten verdeeld: het Belgische materiaal is oververtegenwoordigd in de formele(re) componenten en het Nederlandse materiaal in de informele(re) componenten. Tot slot, bij de lexicale analyse zal duidelijk worden dat de gebruikte versie van het CGN kwantitatief te beperkt is.<sup>2</sup>

## 6.2 Analyse van het *Corpus Gesproken Nederlands*

Het extraheren van de relevante observaties uit het corpus vormt de eerste fase van de corpuslinguïstische methodologie (zie paragraaf 5.2). Voor de bevraging van het CGN hebben we de syntactische context waar de adjectivische buiging voorkomt, afgebakend aan de hand van de beschrijvingen in de ANS (1997) en De Schutter (1997). De mogelijke syntactische configuraties van de NP hebben we vervolgens vertaald naar sequenties van woordsoortannotatiecodes in het CGN (6.2.1). De observaties in de resulterende gegevensbank zijn vervolgens individueel gecontroleerd (6.2.2).

### 6.2.1 Zoekopdracht

Voor het opstellen van de zoekopdracht moeten we de verschillende elementen beschrijven die de drie posities kunnen bezetten in een definiëte NP met een enkelvoudig *het*-woord als hoofd. Deze posities worden in (1) voorgesteld:

$$(1) \quad \text{DET}_{[+DEF]} \text{A N}_{[+ONZ,+ENK]}$$

We zullen achtereenvolgens de restricties op de determinatorpositie, de adjectivische positie en de nominale kernpositie doornemen.

#### *Determinatorpositie*

De determinatorpositie in (1) is voorbehouden aan definiëte determinatoren. De ANS (1997) beschouwt de volgende elementen als definiëte determinatoren:

- bepaald lidwoord *het* – bv. *het moeilijk(e) boek*
- aanwijzend voornaamwoord – bv. *dit moeilijk(e) boek*
- bezittelijk voornaamwoord – bv. *jouw moeilijk(e) boek*
- geanteponeerde genitief – bv. *Jans moeilijk(e) boek*
- prenominaal betrekkelijk voornaamwoord – bv. *Jan, wiens moeilijk(?e) boek verdwenen is*

- collectiverend onbepaald voornaamwoord – bv. *ieder moeilijk(?) boek*

De ANS (1997: 404) en Booij (2002c: 44) beperken de buigingsalternantie tot de eerste vier categorieën in deze lijst, met de bijkomende restrictie dat het aanwijzend voornaamwoord *dit* of *dat* is. Voor de andere aanwijzende voornaamwoorden (*zo'n*, *zulk*) en categorieën in bovenstaande lijst gaan we na in welke mate ze variatie toelaten (zie paragraaf 6.2.2).

#### *Adjectivische positie*

Wanneer de adjectivische positie tot één item beperkt is, kunnen de volgende elementen als adjectief fungeren (ANS 1997: 837-844):

- adjectief – bv. *het grot(e) boek*
- voltooid en onvoltooid deelwoord – bv. *het gevreesd(e) boek*; *het glimmend(e) boek*

#### *Nominale kernpositie*

De conditie dat de nominale kernpositie door een enkelvoudig onzijdig element wordt bezet, beperkt deze positie tot de volgende elementen:

- onzijdig substantief – bv. *het moeilijk(e) boek*
- verkleinwoord – bv. *het grappig(e) vertellertje*
- substantivering: Elk “willekeurig woord” kan als een onzijdig substantief gebruikt worden (ANS 1997: 36-37, 153-154). Deze mogelijkheid is in de CGN-annotatieset voorzien. – bv. *het constant(e) snurken*

#### *Nominale constituent*

Tot slot zijn er nog twee situaties die om bijkomende toelichting vragen. Een eerste probleemgeval vormen de naamwoordelijke constituenten met een complexe kern, zoals in (2).

- (2) [het nieuw(e) [[hoofd] buitenlandse betrekkingen]]; [het dagelijks(e) [[traject] Brussel-Antwerpen]]

In zulke gevallen kijken we enkel naar het eerste substantief, i.e. het meest linkse substantief, dat het genus en het getal van de NP bepaalt. In de voorbeelden in (2) selecteren we bijgevolg enkel het onderstreepte deel van de NP: deze sequentie bevat de determinator en het adjectief waarvan de congruentiemorfologie door het linkse substantief wordt bepaald.

Daarnaast zijn er ook complexe NP's met een ingebedde kwantiteitsaanduidende NP in determinatorpositie. In deze gevallen wordt enkel naar de onderstreepte sequentie tussen vierkante haken gekeken. Dit is: de eerste NP die we tegenkomen en aan de syntactische structuur in (1) beantwoordt.

- (3) [het klein(e) stuk] kaas; [het grot(e) glas] melk; [het fijn(e) sneetje] brood

De gevallen onder (2) en (3) zijn te zeldzaam om na te gaan of ze een afwijkend gedrag vertonen met betrekking tot de keuze van het buigingsmorfeem.

### Zoekopdracht

Nu vertalen we bovenstaande restricties op de posities in de NP naar een zoekopdracht die gebruik maakt van de woordsoortannotatie in het CGN.<sup>3</sup> Tabel 6(2) stelt de zoekopdracht voor: de eerste kolom identificeert de drie posities binnen de NP, de tweede kolom somt de woordsoortannotatiecodes op, de derde kolom geeft een korte toelichting bij iedere code in de tweede kolom.<sup>4</sup>

Positie in NP	CGN woordsoortannotatie	Toelichting
DETERMINATOR	LID(bep.*evon)	bepaald onzijdig lidwoord
	LID(bep,dial)	dialectisch lidwoord
	VNW(bez,det.*prenom.*)	bezittelijk voornaamwoord
	VNW(aanw,det.*prenom.*)	aanwijzend voornaamwoord
	VNW(vb,det.*prenom.*)	vragend/betrekkelijk voornaamwoord
	N(. *gen)	nominale genitief
	VNW(onbep,det.*prenom.*)	onbepaald voornaamwoord
	VNW(. *det,dial)	dialectisch voornaamwoord
ADJECTIEF	ADJ(prenom.*)	adjectief
	ADJ(dial)	dialectisch adjectief
	WW(. *prenom.*)	preminale werkwoordsvorm
	WW(dial)	dialectische werkwoordsvorm
SUBSTANTIEF	N(. *ev,basis,onz.*)	onzijdig substantief
	N(. *,ev,basis,genus.*)	bigenerisch substantief
	N(. *ev,dim.*)	diminutief
	N(. *ev.*gen)	naamwoordelijke genitief
	N(. *ev.*dat)	naamwoordelijke datief
	N(. *dial)	dialectisch substantief
	ADJ(nom.*zonder-n)	genominaliseerd adjectief
	ADJ(dial)	dialectisch adjectief
	WW(. *nom.*zonder-n)	genominaliseerde werkwoord
	WW(dial)	dialectische werkwoordvorm

Tabel 6(2): Zoekopdracht voor de extractie van  $DET_{[+DEF]} A N_{[+ONZ,+ENK]}$  uit het CGN

Tot slot moeten bij de zoekopdracht in tabel 6(2) nog enkele algemene opmerkingen worden geformuleerd. Ten eerste, de zoekopdracht is overgeneraliserend om de kans op het missen van een attestatie te minimaliseren. Daarenboven sluit de formele inslag van de CGN-codes bepaalde distincties uit, zoals het onderscheid tussen collectiverende en niet-collectiverende onbepaalde voornaamwoorden, of het onderscheid tussen vragende en betrekkelijke voornaamwoorden. De observaties waarin deze elementen voorkomen, worden manueel gecontroleerd (zie paragraaf 6.2.2). Ten tweede, omdat een algemeen

corpus (zie paragraaf 6.1) met gesproken materiaal – bijna onvermijdelijk – dialectische woorden bevat, voorzien we voor iedere woordsoort op iedere positie ook de dialectische realisatievormen. Ten derde, voor de kernpositie zijn geen genominaliseerde voornaamwoorden opgenomen, zoals *het mijne* of *het welke*. Tijdens een afzonderlijk vooronderzoek hebben we geen enkele attestatie gevonden van een sequentie met een genominaliseerd voornaamwoord in de nominale kernpositie die de structuur in (1) realiseert. Tot slot, we hebben geen telwoorden in de zoekopdracht opgenomen, noch in de adjectivische positie noch in de nominale kernpositie. Door het ontbreken van inflectionele kenmerken zijn deze elementen voor dit onderzoek irrelevant als adjectief. De meervoudige betekenis sluit de genominaliseerde telwoorden bovendien uit van de nominale kernpositie. In (4) toont de determinator *die* dat de NP meervoudig is.

(4) die andere twee<sub>rW(hoofd,nom,zonder-n,basis)</sub>

Door de vormelijke oriëntatie prevaleert bij de CGN-woordsoortannotatie de morfologische ongeleedheid van het telwoord, zodat de constituent in (4) als enkelvoudig gecodeerd is (dit is: het derde argument *zonder-n* codeert het telwoord als enkelvoudig).

### 6.2.2 Controle gegevensbank

De zoekopdracht voorgesteld in tabel 6(2) excerpert 6,711 observaties uit de 5<sup>e</sup> pre-release van het CGN. Ondanks de woordsoortannotatie moeten we de bekomen gegevensbank manueel controleren omwille van de overgeneraliserende zoekopdracht, de vormgebaseerde codes voor de woordsoortannotatie en de onvermijdelijke inconsequenties en fouten in iedere woordsoortannotatie.<sup>5</sup> We zullen achtereenvolgens de verificatie van de elementen in de verschillende posities van de NP bespreken.

#### *Determinatorpositie*

Bij de beschrijving van de definiete determinatoren in de vorige paragraaf hebben we erop gewezen dat niet alle elementen die als definiete determinator geboekstaafd staan de verbogen vorm van het adjectief selecteren en dus de buigingsalternantie mogelijk maken. We zullen deze punten van divergentie bespreken, samen met enkele overgeneralisaties inherent aan de CGN-woordsoortannotatie.

- *aanwijzend voornaamwoord*: De codes van de CGN-woordsoortannotatie maken geen onderscheid tussen de bepaalde (*die, dat*) en de onbepaalde (*zo'n, zulk*) aanwijzende voornaamwoorden. Bijgevolg moeten we de attestaties met een aanwijzend voornaamwoord in determinatorpositie manueel controleren. Bij de onbepaalde vormen van het aanwijzend voornaamwoord komt bijna zonder uitzondering de onverbogen vorm van het adjectief voor: alle 11 observaties met *zulk* in determinatorpositie bevatten een onverbogen adjectief; in de 257 observaties met *zo'n* als determinator

selecteert het element in adjectivische positie slechts één keer de buigings-*e*, zo'n *mooie gordijn*. Merk op dat het element in de kernpositie een bigenerisch substantief is, zodat de buigings-*e* het substantief *gordijn* wellicht als *de*-woord identificeert (cf. *infra*). Alle observaties met een onbepaald aanwijzend voornaamwoord in determinatorpositie worden uit de materiaalverzameling verwijderd, aangezien ze consistent de onverbogen buigingspendant selecteren in het CGN.

- collectiverend onbepaald voornaamwoord: Van de 23 observaties met een collectiverend onbepaald voornaamwoord in determinatorpositie zijn er 2 verbogen. Beide verbogen vormen komen na de determinator *alle* voor, zodat er wellicht sprake is van buigingsassimilatie tussen de 'uitgang' van de determinator en het uitgangsmorfeem van het adjectief: *alle andere werk*, *alle denkbare kattenkwaad*. De andere twee attestaties met *alle* bevatten een onverbogen adjectief: *alle warm water*, *alle wetenschappelijk onderzoek*.<sup>6</sup> Deze 23 observaties zijn uit de gegevensbank met de buigingsalternantie verwijderd omdat ze consequent het onverbogen buigingsmorfeem kiezen.
- vragend en betrekkelijk voornaamwoord: Er is geen enkele attestatie van een betrekkelijk voornaamwoord in determinatorpositie. Bij alle 13 attestaties wordt de determinatorpositie bezet door het vragend voornaamwoord *welk* en selecteert het adjectief de verbogen vorm.
- foutieve annotaties: Het merendeel van de annotatiefouten heeft betrekking op het onderscheid tussen het bijvoeglijk en het zelfstandig gebruik van het voornaamwoord, zoals in *ze vonden die aangenaam gezelschap*, waar de onderstreepte vorm van de volgende adnominale annotatie (cf. *prenom* als vierde argument) voorzien is: *VNW(aanw, det, stan, prenom, zonder, rest)*.

#### *Adjectivische positie*

Bij de algemene beschrijving van de verbuiging van het attributieve adjectief hebben we een aantal systematische uitzonderingen geformuleerd op het algemene buigingssysteem (zie paragraaf 3.2). Het betreft fonologische en morfosyntactische condities die consistent de verbogen of de onverbogen buigingsvariant selecteren ongeacht de syntactische configuratie van de NP.

	Verbogen	Onverbogen
<b>A eindigend op klinker</b>	0	15
<b>A op [★n]<sub>(exclusief stofnamen)</sub></b>	0	390
<b>A afgeleid van stofnaam</b>	0	45
<b>A op -er</b>	0	0
<b>Superlatief<sup>7</sup></b>	667	0
<i>zelfde, verscheidene, ...</i>	0	5
<i>gratis, linker, ...</i>	0	9

Tabel 6(3): *Overzicht van de buiging van de adjectivische vormen die systematisch één van beide buigingsmorfemen selecteren*

Aangezien deze elementen systematisch alle variatie uitsluiten, worden ze niet in het empirisch onderzoek opgenomen. Het overzicht in tabel 6(3) toont de systematiciteit van deze elementen met betrekking tot de keuze van het adjectivische buigingsmorfem.

#### *Nominale kernpositie*

Bij de elementen in de kernpositie moet het genus gecontroleerd worden. Het Nederlands classificeert de substantieven in drie verschillende genusklassen: de *de*-woorden, de *het*-woorden en de bigenerische substantieven die beide genera kunnen aannemen (ANS 1997). Deze driedeling is in de CGN-annotatieset overgenomen door de ternaire distinctie “zijdig” (*de*-woorden), “onzijdig” (*het*-woorden) en “genus” (bigenerische nomina) (Van Eynde 2001: 16). De bigenerische genusklasse valt uiteen in een groep waar de genusdistinctie semantisch gemotiveerd is (5) en een groep waar beide genera vrij alterneren (6).

- (5) a. voetbal: object (*de*) vs. sport (*het*)
- b. koppel: draagriem (*de*) vs. paar (*het*)
- (6) de/het gordijn; de/het matras

Aan de hand van de *Van Dale*-genusannotatie<sup>8</sup> hebben wij de *de*-woorden en de bigenerische substantieven in de materiaalverzameling geïdentificeerd. De uiteindelijke evaluatie van het genus gebeurt op basis van het effectieve taalgebruik in het corpus. Alle substantieven die volgens de *Van Dale*-annotatie onzijdig zijn en waarvoor de context geen indicatie van het tegendeel bevat, worden als *het*-woord gelabeld. Voor de desambiguering van de bigenerische substantieven gelden de volgende criteria. Het genus van de bigenerische substantieven wordt in eerste instantie op basis van de morfosyntactische context, c.q. de determinator en de adjectivische buiging, gedesambigueerd. Wanneer de determinatorpositie door een bepaald lidwoord of een aanwijzend voornaamwoord wordt bezet, kan het genus probleemloos gedesambigueerd worden: *de, deze, die* (7) vs. *het, dit, dat* (8).

- (7) die houten soort : *de*-woord
- (8) het gevreesde silhouet : *het*-woord

Bij een bezittelijk voornaamwoord zijn enkel de vormen van de eerste persoon meervoud gemarkeerd voor het genusonderscheid: *ons* (*het*-woord) vs. *onze* (*de*-woord). De overige bezittelijke voornaamwoorden zijn ambigu met betrekking tot de genusdistinctie. Indien het genusverschil niet op basis van de morfosyntactische context gedesambigueerd kan worden, kijken we of het genusverschil semantisch gemotiveerd is, zoals bij de voorbeelden in (5). Observaties met bigenerische substantieven die we noch op basis van de morfosyntactische context noch op basis

van de betekenis kunnen desambigueren, zoals in (9), worden uit de materiaalverzameling verwijderd.

- (9) haar natte voorschoot

In de zuidelijke dialecten kunnen dialectische buigingsvormen zulke gevallen desambigueren. De *en*-uitgangen in (10) wijzen op een mannelijk substantief.

- (10) a. hare witten boord  
b. den Europese nummer<sup>9</sup>

Voor de nominale hoofden die niet in de *Van Dale*-lijsten voorkomen, is nagegaan of het genus kan worden vastgesteld op basis van morfologische, semantische of contextuele kenmerken. Nominalisaties (11) en verkleinwoorden (12) zijn zonder uitzondering *het*-woorden, net als deverbatieven met het prefix *ge-* (13). Samenstellingen met een *het*-woord als hoofd zijn onzijdig (14). Dit geldt ook voor plaatsnamen (15).

- (11) mijn verkrampte pedaleren; het aandachtig kijken; het nieuwe links  
(12) het lage salontafeltje; het vies apeke; dat kleine ukkie  
(13) het hysterische gehinnik; het voortdurende gebonk; het zachte geknetter  
(14) dat hele computergedoe; het voorgestelde stadsvernieuwingsfonds  
(15) het Franse Beauvais; dat sterke Japan; het traditionele Afrika

Voor de resterende observaties is een beroep gedaan op de syntagmatische context en de eigen encyclopedische kennis. Deze groep bevat acroniemen (16) en eigenamen (17):

- (16) het vorige MAP (Mest Actie Plan; JT)  
(17) het Spaanse Telefonica; dat vijandige Ajax

De toepassing van deze criteria leidt tot de eliminatie van 28 *de*-woorden en 25 twijfelgevallen. Daarnaast hebben we ook vier meervoudsvormen uit de gegevensbank verwijderd.

#### *Naamwoordelijke constituent*

Tot slot zijn de observaties in de gegevensbank getest op het overgeneraliserende karakter van de zoekopdracht en de aanwezigheid van bepaalde labels in de CGN-woordsoortannotatie. De verwijderde observaties behoren tot de volgende groepen:

- opeenvolging van twee determinatorvormen: In deze gevallen brengt de CGN-woordsoortannotatie geen onderscheid aan tussen het gebruik van een voornaamwoord in determinatorpositie en in adjectivische positie, twee prenominales posities vanuit het perspectief van de CGN-codering, zoals in *Japan is het enige*<sub>VNW(onbep,det,stan,prenom,met-e,rest)</sub> *beschaafde land* of in *geen enkel*<sub>VNW(onbep,det,stan,prenom,zonder,evon)</sub> *gezellig hoekje*.



- dialectische woorden: Het CGN-label *dia1* is problematisch en overgeneraliserend bij ontstentenis van bijkomende informatie. Observaties met dialectwoorden, zoals *dat kleine grei*, zijn uit de materiaalverzameling verwijderd. We hebben een uitzondering gemaakt voor de fenomenen die in Vlaanderen als tussentalig worden bestempeld, zoals verkleinwoorden op *-ke*, *het schoon lieke*, en vormen als *dagdagelijks*, ‘*t dagdagelijks spreken*. Nederlandse verkleinwoorden op *-ie*, zoals *dat kleine ukkie*, hebben we evenmin uit de materiaalverzameling verwijderd.
- onbekende woorden: Onbekende en niet-identificeerbare woorden zijn uit de materiaalverzameling verwijderd: *het nachtelijk spoocht*.<sup>10</sup>

### 6.3 Resulterende materiaalverzameling

Tabel 6(4) stelt de resulterende materiaalverzameling en de verdeling van beide buigingsvarianten voor:

	n	%
<b>verbogen</b>	3,810	76.75
<b>onverbogen</b>	1,154	23.25
<b>Σ</b>	<b>4,964</b>	<b>100.00</b>

Tabel 6(4): Verdeling van de adjectivische buigingsvarianten in het CGN

De gegevens in tabel 6(4) tonen duidelijk aan dat de verbogen vorm van het adjectief de basisvariant is in een definiete NP met een enkelvoudig *het*-woord als hoofd. De onverbogen vorm kan echter niet als een marginaal randverschijnsel worden afgedaan: het corpus bevat bijna één onverbogen vorm voor drie verbogen vormen. We kunnen bovendien met 95% betrouwbaarheid stellen dat de proportie van de onverbogen vorm bij neutra in het standaard gesproken Nederlands tussen 22.07% en 24.42% bedraagt. Deze constatering vormt een sterke empirische motivatie voor onderhavig onderzoek.

## Noten

1. De componenten 3 en 8 zijn niet in tabel 6(1) opgenomen omdat ze ontbreken in de 5<sup>e</sup> pre-release.
2. Vanuit het perspectief van onze onderzoeksdoelstelling zal ook de definitieve versie van het CGN dit euvel niet oplossen. De stilistische en geografische stratificatie vormt net een belemmering voor een lexicale analyse (c.q. collocatie-analyse) omdat het materiaal in verschillende (te) kleine lectisch gedefinieerde deelverzamelingen is opgedeeld (zie ook paragraaf 15.1).
3. De codes voor de woordsoortannotatie in het CGN (Van Eynde 2001) bestaan uit een complex label dat de vorm aanneemt van een predicator: de functor identificeert de woordsoort en de eventuele argumenten coderen bijkomende morfosyntactische informatie van de woordvorm in kwestie. De gecodeerde morfosyntactische informatie is het resultaat van het lemmatiseringsproces, zoals de getalwaarde van een substantief, en kenmerken gerelateerd aan “woordsoortbehoudende derivatie [...] zoals de diminutievorming bij substantieven” (Van Eynde 2001: 7). In de sequentie *het stoeltje* wordt het label  $N(\text{soort}, \text{ev}, \text{dim}, \text{onz}, \text{stan})$  aan het substantief *stoeltje* toegekend: de functor  $N$  identificeert het item als een substantief en de argumenten coderen achtereenvolgens het “naamwoordelijk type” (*soort*), het getal (*ev*), de graad (*dim*), het genus (*onz*) en de casus (*stan*). Bij de toekenning van de woordsoortannotatie primeert de vorm op de betekenis, c.q. functie, die aan de hand van een argument in het complexe label geïdentificeerd wordt. Het label van *verwittigd* in *een verwittigd man* is  $WW(\text{od}, \text{prenom}, \text{zonder})$ , waarbij de functor  $WW$  de oorspronkelijke woordsoort aanduidt en het tweede argument *prenom* de syntactische functie identificeert op basis van de distributie van dit element, namelijk de prenominale positie in een NP.
4. Deze zoekopdracht hebben we door middel van een *Python*-script over het corpus laten lopen. Wij danken bij deze Dirk Speelman voor de *Python*-scripts die we gebruikt hebben bij het bevragen van het CGN en het converteren van de zoekresultaten naar een *Abundantia Verborum*-gegevensbank.
5. Voor dit laatste punt verwijzen we naar Rietveld, Van Hout & Ernestus (2004: 344-350).
6. Doordat de categorie der onbepaalde voornaamwoorden in de CGN-annotatieset niet intern gedifferentieerd is, hebben we ook 258 observaties met *geen* in determinatorpositie uit de oorspronkelijke materiaalverzameling verwijderd. In deze context neemt geen enkel adjectief een buigings-*e*.
7. De 30 superlatieven die met *aller-* geprefigeerd zijn, nemen zonder uitzondering een buigings-*e*.
8. Deze gegevens zijn ons via Dirk Geeraerts door *Van Dale Lexicografie* ter beschikking gesteld. Wij wensen hierbij *Van Dale Lexicografie* en Dirk Geeraerts te bedanken.
9. Het afwijkende genus van *nummer* in (10b) is typisch voor de (zeer) informele Belgisch-Nederlandse omgangstaal, de zogenaamde tussentaal (zie paragraaf 12.2). Daarnaast zijn er ook substantieven met een systematisch genusverschil in functie van de nationale

## DEEL II

variëteit, zoals *deken* ('kleed'): terwijl dit substantief in Nederland een *de*-woord is, wordt het in Vlaanderen overwegend als een *het*-woord gebruikt.

10. Deze items zijn in het CGN voorzien van een speciale code: *\*u*.

# Hoofdstuk 7:

## Bivariate statistiek en stratumanalyse

Alvorens de eigenlijke empirische analyse aan te vatten, presenteren we de statistische technieken die we in het bivariaat onderzoek zullen aanwenden (7.1). Daarna bespreken we de technieken voor de stratumanalyse (7.2). Beide discussies volgen hetzelfde stramien: we stellen eerst de technische aspecten van de gebruikte maten voor en vervolgens illustreren we ze aan de hand van een voorbeeld. Hierbij besteden we vooral aandacht aan de interpretatie van de resultaten. De lezer die niet in de technische details geïnteresseerd is, kan de technische paragrafen overslaan en onmiddellijk naar de becommentarieerde voorbeelden gaan.

We wensen te benadrukken dat bij de interpretatie van statistische gegevens uiteindelijk het gezond verstand en de domeinkennis van de onderzoeker doorslaggevend is.

Lezers die vertrouwd zijn met de gebruikte statistieken voor het bivariaat ( $\chi^2$ -statistiek, *Fisher Exact-test*, relatief risico, odds ratio) en het stratumonderzoek (*Cochran-Mantel-Haenszel*-statistiek) kunnen dit hoofdstuk overslaan onmiddellijk naar de resultaten van het empirische onderzoek gaan.

### 7.1 Bivariate statistiek

De bespreking van de statistische technieken voor het bivariaat onderzoek bestaat uit twee delen: we lichten eerst de technische aspecten van de verschillende bivariate tests toe (7.1.1); daarna illustreren we deze tests aan de hand van een fictief voorbeeld (7.1.2).

#### 7.1.1 Technieken voor het bivariaat onderzoek: $\chi^2$ -statistiek, Fisher Exact-test, relatief risico en odds ratio

Bivariate analyses worden uitgevoerd om het verband tussen twee variabelen te testen. Hiervoor gebruiken we  $2 \times k$ -contingencietabellen waarin de binomiale responsvariabele (c.q. verbogen adjectief vs. onverbogen adjectief) tegen een binomiale of multinomiale verklarende/onafhankelijke variabele wordt uitgezet.<sup>1</sup> We zullen bij voorkeur  $2 \times 2$ -tabellen gebruiken omdat hiervoor meer statistische technieken ontwikkeld zijn, zoals het relatief risico en de odds ratio (cf. *infra*).<sup>2</sup>

De  $\chi^2$ -statistiek – of correcter Pearsons  $\chi^2$  – is de basistest om de significantie van het verband tussen de onafhankelijke variabele en de responsvariabele in een  $2 \times k$ -tabel te meten. Deze inferentiële test wordt gebruikt om de waarschijnlijkheid te meten waarmee de nulhypothese ( $H_0$ ), er is geen verband tussen beide variabelen,

verworpen kan worden ten voordele van de onderzoekshypothese ( $H_1$ ), er is een verband tussen beide variabelen.<sup>3</sup> Bij iedere  $\chi^2$ -test wordt de volgende informatie gerapporteerd: de  $\chi^2$ -score, die een cijfer op de associatiesterkte tussen de twee variabelen in de contingentietabel kleeft, het aantal vrijheidsgraden<sup>4</sup>, en de  $p$ -waarde. Indien de  $p$ -waarde kleiner is dan 0.05, beschouwen we de  $H_0$  als verworpen: een  $p$ -waarde kleiner dan 0.05 betekent namelijk dat de kans dat de  $H_0$  correct is minder dan 5% bedraagt – en dat de  $H_0$  bijgevolg verworpen kan worden. Indien relevant zal ook de partiële  $\chi^2$ -statistiek gerapporteerd worden. Deze variant op de  $\chi^2$ -statistiek kent aan iedere cel in een contingentietabel een waarde toe voor (de significantie van) de bijdrage van deze cel tot de globale  $\chi^2$ . Deze test is vooral informatief bij multinomiale verklarende variabelen. Voor een toelichting van de partiële  $\chi^2$ -statistiek verwijzen we naar paragraaf 9.1.2.

De  $\chi^2$ -statistiek moet echter met de nodige omzichtigheid gebruikt worden. Rietveld, van Hout & Ernestus (2004) wijzen erop dat zowel bij lage als bij hoge frequenties problemen kunnen optreden bij de  $\chi^2$ -statistiek. Vooreerst zijn er de restricties op het gebruik van de  $\chi^2$ -statistiek bij kleine steekproeven. Deze restricties worden in de zogenaamde *Cochran*-regel samengevat.<sup>5</sup> Indien niet aan de voorwaarden van de *Cochran*-regel is voldaan, gebruiken we automatisch de *Fisher Exact*-test.<sup>6</sup> Daarnaast tonen Rietveld, van Hout & Ernestus (2004) dat kleine steekproeven alleen grote effecten kunnen identificeren en dat in zulke gevallen niet-significantie vaak zeer informatief is. Rekening houdend met de omvang van de gebruikte materiaalverzamelingen zal dataschaarste niet vaak voor problemen zorgen (zie paragrafen 6.2 en 15.2).

Grote staalnames confronteren onderzoekers met het probleem dat minimale effecten significant kunnen zijn (Rietveld, van Hout & Ernestus 2004). Het volstaat dus niet om naar de significantie te kijken waarmee de  $H_0$  verworpen kan worden. We meten bijgevolg ook de omvang en de zin van de associatie tussen de responsvariabele en de verklarende variabele. Hiervoor maken we gebruik van het relatief risico (RR) en de odds ratio (OR) (Agresti 1996: 21-27). Deze maten kennen een score toe aan het verband tussen de responsvariabele en de verklarende variabele. Deze score varieert tussen 0 en  $+\infty$ , met 1 als scharnierpunt:

- een score van 1 duidt aan dat er geen verband is
- een score tussen 0 en 1 wijst op een negatief verband: de waarde van de responsvariabele neemt af
- een score groter dan 1 wijst op een positief verband: de waarde van de responsvariabele neemt toe

De grootte van deze score kwantificeert het effect van de verklarende variabele op de responsvariabele.

We definiëren het RR en de OR aan de hand van de contingentietabel in tabel 7(1). Deze tabel schematiseert de structuur van de contingentietabellen die we in dit onderzoek zullen rapporteren. De rijvariabele toont de waarden van de respons-

variabele ('-Succes' vs. 'Succes') en de kolomvariabele toont de waarden van de verklarende variabele ('Groep1' vs. 'Groep2').

		Verklarende variabele		
		Groep1	Groep2	$\Sigma$
Respons	-Succes	$a_{11}$	$a_{12}$	$a_{1+}$
	Succes	$a_{21}$	$a_{22}$	$a_{2+}$
	$\Sigma$	$a_{+1}$	$a_{+2}$	$a_{++}$

Tabel 7(1): Schematische voorstelling van een 2x2-contingentietabel

Het RR (relatief risico) vergelijkt de kans op Succes in de twee groepen die door de verklarende variabele gedefinieerd worden. Vertrekkend van tabel 7(1) wordt het RR voor Succes ( $RR_{\text{Succes}}$ ) op basis van deze formule berekend:

$$(1) \quad RR = \frac{P(\text{Succes} | \text{Groep2})}{P(\text{Succes} | \text{Groep1})} = \frac{(a_{22}) / (a_{+2})}{(a_{21}) / (a_{+1})}$$

Het RR vergelijkt de waarschijnlijkheid van één waarde van de responsvariabele, de waarde 'Succes' in (1), voor de twee waarden van de verklarende variabele, c.q. 'Groep 1' en 'Groep 2'.<sup>7</sup> Het RR in bovenstaande formule duidt aan hoeveel vaker er sprake is van Succes in Groep 2 vergeleken met Groep 1. Het RR is intuïtief makkelijk interpreteerbaar: gesteld dat het  $RR_{\text{Succes}}$  (berekend volgens de formule in (1)) gelijk is aan 3, dan betekent dit dat de kans op Succes in Groep 2 het drievoud bedraagt van de kans op Succes in Groep 1 (dit is: 200% meer kans op Succes t.o.v. het scharnierpunt 1). Als het  $RR_{\text{Succes}}$  0.5 bedraagt, dan betekent dit dat het aantal voorkomens van Succes in Groep 2 slechts half zo frequent is als in Groep 1, of nog: in Groep 1 is de kans op Succes dubbel zo groot als in Groep 2 (cf.  $1/0.5 = 2$ ).

De OR (odds ratio) is de ratio van twee odds'en. De odds ( $\Omega$ ) zijn de kans op Succes (bv. het adjectief is onverbogen) gedeeld door de kans op mislukking (c.q. -Succes; bv. het adjectief is *niet* onverbogen, c.q. verbogen). De odds zijn afkomstig uit de wereld van het paardenrennen, waar deze maat gebruikt wordt om de kans op Succes tegen de kans op Mislukking (i.e. -Succes) uit te drukken, bijvoorbeeld '3 tegen 1'. De odds kunnen worden omgezet in een kans: een odds van  $3/1$  voor  $\text{Succes} / \text{-Succes}$  betekent dat Succes in 75% van de gevallen voorkomt, dit is:  $3/(3+1) = 75\%$ . De OR vergelijkt de odds'en in de twee groepen die door de verklarende variabele gedefinieerd worden. Toegepast op tabel 7(1) geeft dit de volgende berekening:

$$(2) \quad OR = \frac{\Omega_{\text{Groep2}}}{\Omega_{\text{Groep1}}} = \frac{P(\text{Succes} | \text{Groep2}) / P(\text{-Succes} | \text{Groep2})}{P(\text{Succes} | \text{Groep1}) / P(\text{-Succes} | \text{Groep1})} = \frac{(a_{22}) / (a_{12})}{(a_{21}) / (a_{11})}$$

Deze maat drukt uit in welke mate de odds  $\text{Sukses}/\text{-Sukses}$  stijgen/dalen in Groep 2 vergeleken met Groep 1. Onder de veronderstelling dat de OR 3 bedraagt, kunnen we niet stellen dat de kans op Succes driemaal groter is in Groep 2 dan in Groep 1, zoals dit het geval is bij het RR. We kunnen enkel stellen dat de odds  $\text{Sukses}/\text{-Sukses}$  driemaal groter zijn in Groep 2 dan in Groep 1. Ondanks de interpretatieve complexiteit van de OR, zullen we naast het RR ook de OR rapporteren. Deze keuze garandeert niet alleen een optimale vergelijkbaarheid met de logistische regressieanalyse, de techniek die we bij de multivariate analyse gebruiken en waarin de OR een belangrijke rol speelt (zie paragraaf 13.2.1 en 14.3.2), maar de OR is bovendien mathematisch superieur ten overstaan van het RR. Vooreerst is het een symmetrische maat die onafhankelijk is van de oriëntatie van de tabel (zie paragraaf 7.1.2 voor een nadere toelichting aan de hand van een voorbeeld). Daarnaast is het resultaat van het  $\text{RR}_{\text{Sukses}}$  sterk afhankelijk van de kans op Succes in de ongemarkeerde context (c.q. ‘Groep 1’ in tabel 7(1)): wanneer de kans op Succes in de ongemarkeerde context 50% bedraagt, kan deze kans maximaal verdubbeld worden (cf. een kans bedraagt maximaal 100%).

Omdat we in eerste instantie geïnteresseerd zijn in het effect van de gemarkeerde waarde van een verklarende variabele, kwantificeren de gerapporteerde scores van het RR en de OR in het empirisch onderzoek het effect van de gemarkeerde waarde van de verklarende variabele in vergelijking met de neutrale waarde of de ijkwaarde van deze variabele. Behalve bij indicatie van het tegendeel is de meest linkse kolomwaarde in de contingentietabel de neutrale ijkwaarde van de verklarende variabele, i.e. de waarde ‘Groep 1’ in tabel 7(1).

Bij de interpretatie van het RR en de OR moeten we rekening houden met het betrouwbaarheidsinterval (BI)<sup>8</sup>: het effect gemeten door het RR of de OR is enkel significant wanneer het betrouwbaarheidsinterval de waarde 1 niet bevat. Dit is de waarde die de afwezigheid van associatie aanduidt: bij een  $\text{RR}_{\text{Sukses}}$  gelijk aan 1 is de kans op Succes gelijk in beide groepen; bij een OR gelijk aan 1 zijn de odds  $\text{Sukses}/\text{-Sukses}$  gelijk in beide groepen (cf. *supra*).

Het gebruik van het RR en de OR in combinatie met de betrouwbaarheidsintervallen biedt enkele descriptieve en evaluatieve voordelen ten overstaan van significantietests, zoals de  $\chi^2$ -statistiek. Ten eerste, het RR en de OR kennen een waarde toe aan een onbekende parameter: ze meten de impact van een parameterwaarde op de responsvariabele, terwijl een significantietest enkel nagaat met welke waarschijnlijkheid de  $H_0$  kan worden verworpen. Met andere woorden: de  $\chi^2$ -statistiek beantwoordt enkel de vraag *of* er een verband is tussen twee variabelen (‘ja’ vs. ‘neen’); het RR en de OR antwoorden daarnaast ook op de vraag *hoe groot* dit verband is. Ten tweede, het BI geeft informatie over de precisie van de berekende waarden: hoe dichter de beneden- en bovengrens van het BI bij elkaar liggen, des te nauwkeuriger de uitspraak die we kunnen doen. Ten derde, BI’en kunnen ook gebruikt worden om verschillende steekproeven te vergelijken: wanneer de 95% BI’en voor twee steekproeven niet overlappen, betekent dit dat in 95% van

de situaties de waarde van de ene steekproef groter/kleiner is dan de waarde van de andere steekproef. Deze toepassingsmogelijkheid van BI'en passen we toe in de stratumanalyse om na te gaan of het effect van een verklarende variabele (berekend door middel van de OR) significant verschilt in de groepen gedefinieerd door de stratumvariabele (zie paragraaf 7.2). De BI'en kunnen tot slot ook als significantietest gebruikt worden: wanneer het BI de scharnierwaarde (1 voor het RR en de OR) bevat, is het effect van de verklarende variabele niet significant.

### 7.1.2 Een fictief voorbeeld: de invloed van de sekse van de spreker op de buigingsalternantie

We illustreren de abstracte definities uit de vorige paragraaf aan de hand van een fictief voorbeeld. Veronderstel dat we het effect van de sekse van de taalgebruiker op de adjectivische buigingsalternantie willen testen. Hiervoor gebruiken we de fictieve materiaalverzameling voorgesteld in tabel 7(2). In aansluiting met de sociolinguïstiek beschouwen we het taalgebruik van vrouwen als normatiever dan dat van mannen en bijgevolg als de 'neutrale' ijkwaarde, vanwaar de positie van de waarde 'vrouw' als de meest linkse kolomvariabele. De rijvariabelen definiëren de twee waarden van de responsvariabele: verbogen adjectief ('-Succes' in tabel 7(1)) vs. onverbogen adjectief ('Succes' in tabel 7(1)). We herinneren de lezer eraan dat, behalve bij vermelding van het tegendeel, de rijvariabele (verbuiging adjectief) de responsvariabele voorstelt en de kolomvariabele (sekse) de verklarende variabele. De meest linkse waarde van de kolomvariabele ('vrouw') is de 'neutrale' waarde of de ijkwaarde die als vergelijkingspunt dient om het effect van de andere waarde(n) van de verklarende variabele te berekenen ('man').

Voor iedere cel in de contingentietabel wordt niet alleen de absolute frequentie maar ook de relatieve frequentie gerapporteerd. Behalve bij bericht van het tegendeel drukken de relatieve frequenties het aandeel verbogen/onverbogen vormen uit in de deelverzameling gedefinieerd door een bepaalde waarde van de verklarende variabele. Toegepast op tabel 7(2) betekent dit dat de percentages het aandeel verbogen en onverbogen vormen in het taalgebruik van vrouwen en het taalgebruik van mannen uitdrukken. Indien de omvang van de tabel het toelaat, rapporteren we ook de marginaal ( $\Sigma$ ).

	vrouw		man		$\Sigma$	
<b>verbogen</b>	900	81.82	700	66.67	1,600	74.42
<b>onverbogen</b>	200	18.18	350	33.33	550	25.58
<b><math>\Sigma</math></b>	<b>1,100</b>	<b>100.0</b>	<b>1,050</b>	<b>100.0</b>	<b>2,150</b>	<b>100.0</b>

Tabel 7(2): Fictieve materiaalverzameling met het effect van de sekse van de spreker op de adjectivische buigingsalternantie

De frequentieverdeling in tabel 7(2) is significant. Wanneer we naar de  $p$ -waarde van de  $\chi^2$ -statistiek kijken, stellen we vast dat deze kleiner is dan 0.05:  $\chi^2 = 63.9879$ ,



$df = 1$ ,  $p = 1.252e-15$ .<sup>9</sup> Met andere woorden: we kunnen de  $H_0$ , er is geen verband tussen het geslacht van de spreker de selectie van het buigingsmorfeem, verwerpen omdat de kans dat ze correct is minder dan 5% bedraagt. Hoewel het strikt gesproken overbodig is – de frequentieverdeling in tabel 7(2) voldoet aan alle voorwaarden van de *Cochran*-regel –, rapporteren we ook de *Fisher Exact*-test. Het resultaat van deze test bestaat uitsluitend uit een  $p$ -waarde (die doorgaans een tikkeltje afwijkt van de  $p$ -waarde voor de  $\chi^2$ -statistiek):  $p = 8.066e-16$ .

Nu we weten dat de sekse van de spreker een significante invloed op de buigingsalternantie definieert, willen we de zin en de omvang van dit effect kennen. Hiervoor berekenen we het RR (relatief risico) en de OR (odds ratio). De toepassing van de definitie in (1) levert een  $RR_{\text{onverbogen}}$  gelijk aan 1.8333 op. Dit cijfer wordt als volgt bekomen:

$$(1) \quad RR = \frac{P(\text{onverbogen.adjectief} | \text{man})}{P(\text{onverbogen.adjectief} | \text{vrouw})} = \frac{(350)/(1,050)}{(200)/(1,100)} = 1.8333$$

Een  $RR_{\text{onverbogen}}$  gelijk aan 1.8333 betekent dat de kans op een onverbogen vorm 83.33% groter is in het taalgebruik van een man dan in het taalgebruik van een vrouw. Het 95% BI (95% betrouwbaarheidsinterval) bedraagt [1.5752, 2.1338]. Dit BI bevestigt het resultaat van de  $\chi^2$ -statistiek: aangezien de scharnierwaarde 1 buiten het BI valt, definieert de sekse van de spreker een significant effect ten aanzien van de buigingsalternantie (zie paragraaf 7.1.1). Als gevolg van het relatief enge BI kunnen we een redelijk nauwkeurige uitspraak doen met betrekking tot het RR: na afronding is in 95% van de gevallen de kans op een onverbogen adjectief tussen de 1.5 en 2 keer groter in het taalgebruik van mannen dan in het taalgebruik van vrouwen.

Op basis van de gegevens in tabel 7(2) is het ook makkelijk aan te tonen dat het RR geen symmetrische maat is. De score voor het  $RR_{\text{onverbogen}}$  is verschillend van de score voor het  $RR_{\text{verbogen}}$ :  $RR_{\text{verbogen}} = 0.8148$  [0.7743, 0.8575], wat betekent dat de kans op een verbogen adjectief bijna 23% ( $1/0.8148 \approx 1.2273$ ) kleiner is in het taalgebruik van mannen dan in het taalgebruik van vrouwen. Zoals hierboven reeds is aangegeven, rapporteren we het  $RR_{\text{onverbogen}}$  omdat we geïnteresseerd zijn in het effect van de verklarende variabele op de gemarkeerde onverbogen variant. Indien relevant wordt uiteraard ook het  $RR_{\text{verbogen}}$  gerapporteerd.

Wanneer we de OR voor het fictieve voorbeeld in tabel 7(1) berekenen, zijn we geïnteresseerd in welke mate de odds  $^{\text{onverbogen}}_{\text{verbogen}}$  stijgen/dalen in het taalgebruik van mannen vergeleken met het taalgebruik van vrouwen (c.q. de ijkwaarde van de verklarende variabele). Hiervoor vullen we de formule in (2) als volgt in:

$$(2) \quad OR = \frac{\Omega_{\text{man}}}{\Omega_{\text{vrouw}}} = \frac{350/700}{200/900} = 2.25$$

Deze berekening levert een OR gelijk aan 2.25 op, met [1.8424, 2.7477] als 95% BI: de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  zijn met andere woorden 2.25 maal groter in het taalgebruik van mannen dan in het taalgebruik van vrouwen. Het BI bevestigt de resultaten van de  $\chi^2$ -test en het  $RR_{\text{onverbogen}}$ : de verdeling is significant en we kunnen met 95% betrouwbaarheid stellen dat de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  tussen 1.8 en 2.7 maal groter zijn in het taalgebruik van mannen dan in het taalgebruik van vrouwen.

Zoals aangegeven in de vorige paragraaf is de OR – in tegenstelling tot het RR – wel een symmetrische maat: voor de grootte van de OR maakt het niet uit of we de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  ( $OR = 2.25 = \frac{1}{0.4444}$ ) of de odds  $\frac{\text{verbogen}}{\text{onverbogen}}$  ( $OR = 0.4444 = \frac{1}{2.25}$ ) vergelijken in het taalgebruik van vrouwen en mannen. Aangezien we in eerste instantie geïnteresseerd zijn in de conditionering van de gemarkeerde onverbogen buigingsvariant, rapporteren we standaard de OR die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vergelijkt.

We sluiten deze paragraaf af met een notationale bemerking: we gebruiken de punt voor de identificatie van de decimalen en de komma als identificatie van de duizendtallen (bv. 1,000 = duizend).

## 7.2 Stratumanalyse

We belichten eerst de technische aspecten van de stratumanalyse, met name de technieken die in de verschillende stappen gebruikt worden (7.2.1). Daarna passen we deze technieken toe op het fictieve voorbeeld dat we in de vorige paragraaf geïntroduceerd hebben (7.2.2). We zullen meer bepaald illustreren dat het buigings-effect van de sekse genuanceerd moet worden omdat deze factor deels door de leeftijd van de spreker geconditioneerd wordt.

### 7.2.1 Technieken voor stratumanalyses: Cochran-Mantel-Haenszel-statistiek, Breslow-Day-test en stratumtabellen

Het doel van een stratumanalyse is, zoals reeds aangegeven in het methodologische hoofdstuk (zie paragraaf 5.2.3), testen of er sprake is van een conditionele of partiële interactie tussen twee verklarende variabelen met betrekking tot het verklaren van het gedrag van de responsvariabele. Hiervoor worden de gegevens als een  $2 \times 2 \times k$ -datamatrix geformatteerd. De verklarende variabele die het effect van een andere verklarende variabele conditioneert, noemen we de stratumvariabele.

Een stratumanalyse kan op de volgende manier gevisualiseerd worden. Een traditionele contingentietabel met een verklarende variabele ( $x$ ) en een responsvariabele ( $y$ ) definieert een curve in een tweedimensionale ruimte bestaande uit een X-as en een Y-as (i.e. een vlak). In een stratumanalyse wordt hieraan een derde dimensie toegevoegd voor de stratumvariabele ( $z$ ) die het effect van de verklarende variabele (mogelijkerwijs) beïnvloedt: de Z-as zorgt voor de ‘diepte’-dimensie. Wat er in een stratumanalyse gebeurt, is niets anders dan het achter elkaar plaatsen van  $k$  verschillende tweedimensionale curves – of contingentietabellen zo men wilt –

gedefinieerd door de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ) voor de  $k$  verschillende waarden op de  $Z$ -as. Deze verschillende curves worden vervolgens vergeleken op (significante) verschillen.

Concreet verloopt een stratumanalyse in drie fases. In de eerste fase testen we of er over de verschillende strata ( $z$ ) heen een significant verband is tussen de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ). Hiervoor gebruiken we de *Cochran-Mantel-Haenszel*-statistiek, of voortaan kortweg de CMH-statistiek (Agresti 1996: 60-66). Deze statistiek bekijkt wat er gemiddeld over de verschillende waarden van de stratumvariabele gebeurt en test de  $H_0$  dat er geen verband is tussen de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ) wanneer we de verschillende waarden van de stratumvariabele controleren ( $z$ ). Bij een  $2 \times 2 \times k$ -matrix worden de volgende resultaten gerapporteerd:

- *Mantel-Haenszel- $\chi^2$*  (MH- $\chi^2$ ): Deze variant van de  $\chi^2$ -statistiek berekent of het verband tussen de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ) significant is over de verschillende strata ( $z$ ).<sup>10</sup> De  $p$ -waarde voor deze statistiek geeft een antwoord op de volgende vraag: is er een significant verband tussen de verklarende variabele ( $x$ ) en de responsvariabele ( $y$ ) wanneer we het effect van de stratumvariabele controleren ( $z$ )? Het is immers mogelijk dat Pearsons  $\chi^2$  op basis van de marginale tabel, i.e. de tabel die enkel de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ) in rekening neemt, geen significante associatie identificeert als gevolg van tegengestelde onderliggende stratumeffecten. Dit is: het is best mogelijk dat de verklarende variabele ( $x$ ) een gedifferentieerd effect heeft bij de verschillende waarden van de stratumvariabele ( $z$ ).
- *Cochran-Mantel-Haenszel-OR* (CMH-OR): De CMH-OR meet de impact van de verklarende variabele ( $x$ ) op de responsvariabele ( $y$ ) terwijl het effect van de stratumvariabele ( $z$ ) gecontroleerd wordt. Deze variant van de OR beantwoordt de volgende vraag: wat is de zin en de omvang van het effect van de verklarende variabele ( $x$ ) op de responsvariabele ( $y$ ) wanneer we het effect van de stratumvariabele ( $z$ ) controleren?
- *Breslow-Day*-test: Bij de interpretatie van de CMH-OR moeten we rekening houden met de homogeniteit van de OR's in de verschillende strata. Dit berekenen we door middel van de *Breslow-Day*-statistiek (Agresti 1996: 63), die de  $H_0$  test dat de OR's in de verschillende strata homogeen zijn, i.e. niet significant van elkaar verschillen. Een  $p$ -waarde groter dan 0.05 betekent dat de OR's in de verschillende strata homogeen zijn en dat er geen sprake is van een partiële conditionering van de verklarende variabele ( $x$ ) door de stratumvariabele ( $z$ ). Bij een  $p$ -waarde kleiner dan of gelijk aan 0.05 wordt het verband tussen de verklarende variabele ( $x$ ) en de responsvariabele ( $y$ ) op significante wijze door de stratumvariabele ( $z$ ) beïnvloed.

Bij multinomiale verklarende variabelen kunnen deze tests niet worden toegepast en gaan we onmiddellijk over naar de tweede fase.

De tweede fase van de stratumanalyse bestaat uit de analyse van de stratumtabellen. Een stratumtabel is een gewone contingentietabel met een responsvariabele ( $y$ ) en een verklarende variabele ( $x$ ), met dit verschil dat ze het effect van de verklarende factor ( $x$ ) kwantificeert voor één waarde  $k_i$  van de stratumvariabele ( $z$ ). Een stratumanalyse op basis van een  $2 \times 2 \times k$ -matrix bevat  $k$  stratumtabellen, dit is: één afzonderlijke stratumtabel voor iedere waarde van de stratumvariabele. De stratumtabellen worden als een contingentietabel geanalyseerd: de  $\chi^2$ -statistiek wordt gebruikt om de significantie van de contingentierelatie tussen de responsvariabele ( $y$ ) en de verklarende variabele ( $x$ ) te bepalen; de OR meet het effect van de gemarkeerde waarde van de verklarende factor ( $x$ ) op de responsvariabele ( $y$ ).

De derde en laatste fase van de stratumanalyse is het samenvatten en interpreteren van de resultaten voor de afzonderlijke stratumtabellen. Deze resumptie bestaat uit de vergelijking van de OR's voor de afzonderlijke stratumtabellen en hun respectievelijke 95% BI'en. Deze OR's vullen de *Breslow-Day*-statistiek aan: terwijl de laatste maat in de eerste fase van de stratumanalyse een globale score voor de partiële associatie tussen de stratumvariabele en de verklarende variabele berekent, identificeren de OR's en hun respectievelijke 95% BI'en de effectieve impact van de partiële associatie. Uit de OR's kunnen we afleiden of het effect van (de gemarkeerde waarde van) een verklarende factor groter/kleiner is voor deze of gene waarde van de stratumvariabele en bijgevolg of er sprake is van een partiële conditionering van de verklarende variabele ( $x$ ) door de stratumvariabele ( $z$ ). De vergelijking van de 95% BI'en leert ons of de waargenomen verschillen voor de  $k$  waarden van de stratumvariabele significant zijn: als de 95% BI'en voor twee waarden van de stratumvariabele ( $k_i$  en  $k_j$ ) niet overlappen, betekent dit dat in 95% van de gevallen de OR in stratum  $k_i$  groter/kleiner is dan de overeenkomstige waarde in stratum  $k_j$ .

Indien de *Breslow-Day*-test in de eerste fase van de stratumanalyse niet significant is, is het vervolg van de stratumanalyse overbodig, want dan is er geen partiële conditionering van de verklarende variabele ( $x$ ) door de stratumvariabele ( $z$ ).<sup>11</sup>

### 7.2.2 *Het fictieve voorbeeld herbekeken*

We illustreren de stratumanalyse aan de hand van hetzelfde fictieve voorbeeld dat we in paragraaf 7.1.2 gebruikt hebben. Op basis van de bivariate analyse van de gegevens in tabel 7(2) hebben we besloten dat mannen significant vaker de onverbogen buigingsvariant selecteren dan vrouwen. Wanneer we naar de leeftijd van de sprekers in deze fictieve materiaalverzameling kijken, met de distinctie 'jong' vs. 'oud', moeten we dit beeld echter nuanceren. We definiëren de verschillende variabelen in de stratumanalyse op de volgende wijze:

- responsvariabele (y): de buigingsalternantie ('verbogen' vs. 'onverbogen')
- verklarende variabele (x): de sekse van de spreker ('vrouw' vs. 'man')
- stratumvariabele (z): de leeftijd van de spreker ('jong' vs. 'oud')

In de eerste fase van het stratumonderzoek kijken we naar de globale stratumstatistieken. Deze tests meten het verband tussen de buigingsalternantie en de sekse van de spreker terwijl de leeftijdsklasse gecontroleerd wordt. Het is bijvoorbeeld niet *a priori* uitgesloten dat er bij de jongere en de oudere sprekers verschillende sekse-effecten zijn. De CMH-statistieken staven de resultaten van de bivariate analyse in paragraaf 7.1.2: ten eerste, de MH- $\chi^2$  wijst op een significant verband tussen de sekse van de spreker en de keuze van het buigingsmorfeem wanneer we de leeftijdsgroep controleren (MH- $\chi^2 = 55.6887$ ,  $df = 1$ ,  $p < 2.2e-16$ ); ten tweede, de CMH-OR identificeert een significante stijging van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in het taalgebruik van de mannelijke sprekers vergeleken met het taalgebruik van de vrouwelijke sprekers wanneer we de leeftijdsgroep van de spreker controleren (CMH-OR = 2.1888 [1.7759,2.6976]).

De significante *p*-waarde voor de *Breslow-Day*-test ( $p < 0.0001$ ) signaleert echter dat de OR die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bij de mannelijke en de vrouwelijke sprekers vergelijkt niet homogeen over de leeftijdsgroepen verdeeld is. Met andere woorden: er is sprake van een partiële conditionering van het buigingseffect van de sekse van de spreker door de leeftijdsklasse van de spreker. Indien de *p*-waarde voor de *Breslow-Day*-test groter is dan 0.5 is de stratumanalyse beëindigd, want dit betekent dat het sekse-effect in beide leeftijdsgroepen niet significant verschilt, zodat de bivariate analyse volstaat. Dit is in het fictieve voorbeeld niet het geval, zodat we nu het buigingseffect van de sekse van spreker in de verschillende strata, c.q. leeftijdsgroepen, onder de loep nemen om de exacte aard van de partiële conditionering van de sekse door de leeftijdsgroep bloot te leggen (zoals: is de zin van het buigingseffect van de sekse gelijk in beide strata?; in welk stratum is het sekse-effect het grootst?). Deze stratumtabellen vatten voor ieder stratum, c.q. voor iedere leeftijdsgroep ('oud' vs. 'jong'), het effect samen van de verklarende variabele, c.q. de sekse van de spreker ('man' vs. 'vrouw'), op de responsvariabele, c.q. de buigingsalternantie ('verbogen' vs. 'onverbogen').

We kijken eerst naar de stratumtabel voor de oudere leeftijdsgroep: bij de oudere sprekers stellen we geen significant verschil vast tussen de buigingsprofielen voor de mannelijke en de vrouwelijke sprekers ( $\chi^2 = 1.0349$ ,  $df = 1$ ,  $p = 0.309$ ):

	vrouw		man		$\Sigma$	
<b>verbogen</b>	650	86.67	550	84.62	1,200	85.71
<b>onverbogen</b>	100	13.33	100	15.38	200	14.29
<b><math>\Sigma</math></b>	<b>750</b>	<b>100.0</b>	<b>650</b>	<b>100.0</b>	<b>1,400</b>	<b>100.0</b>

Tabel 7(3a): Fictieve materiaalverzameling met het effect van de sekse van de spreker op de adjectivische buigingsalternantie in het stratum gedefinieerd door de oudere sprekers

Tabel 7(3b) is de stratumtabel voor de jonge leeftijdsgroep. In dit stratum verschilt het buigingsgedrag wel significant bij beide seksen ( $\chi^2 = 84.9783$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	vrouw		man		$\Sigma$	
<b>verbogen</b>	250	71.43	150	37.50	400	53.33
<b>onverbogen</b>	100	28.57	250	62.50	350	46.67
<b><math>\Sigma</math></b>	<b>350</b>	<b>100.0</b>	<b>400</b>	<b>100.0</b>	<b>750</b>	<b>100.0</b>

*Tabel 7(3b): Fictieve materiaalverzameling met het effect van de sekse van de spreker op de adjectivische buigingsalternantie in het stratum gedefinieerd door de jongere sprekers*

Als tussentijdse conclusie stellen we dat de sekse van de spreker enkel voor de jongere leeftijdsgroep een significant buigingsverschil definieert.

In de derde en laatste fase van de stratumanalyse vatten we de resultaten samen. In tabel 7(3c) rapporteren we de OR's en de bijbehorende 95% BI'en die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  ( $y$ ) bij de mannelijke en de vrouwelijk sprekers ( $x$ ) vergelijken voor de oudere en de jongere leeftijdsgroep ( $z$ ):

Stratum	OR	95% BI OR
Oude sprekers	1.1818	[0.8759,1.5946]
Jonge sprekers	4.1667	[3.0627,5.6686]
CMH-OR	2.1888	[1.7759,2.6976]

*Tabel 7(3c): Samenvatting van de stratumanalyse op basis van het fictieve voorbeeld*

Op basis van deze fictieve gegevens moeten we de deflecterende tendentie in het taalgebruik van mannen, zoals vastgesteld op basis van de bivariate analyse in paragraaf 7.1.2, nuanceren: enkel in de jongere leeftijdsgroep selecteren de mannelijke sprekers significant vaker dan de vrouwelijke sprekers het onverbogen adjectief bij neutra; in de oudere leeftijdsgroep is er geen significant verschil tussen het buigingsgedrag van beide seksen. Het verschil tussen de jongere en de oudere sprekers is significant, zoals blijkt uit de niet-overlappende 95% BI'en. De niet-overlappende 95% BI'en voor de jonge en de oude sprekers moeten we als volgt interpreteren: in 95% van de gevallen is de OR die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in het taalgebruik van mannelijke en vrouwelijke sprekers vergelijkt significant groter voor de jongere leeftijdsgroep ([3.0627,5.6686]) dan voor de oudere leeftijdsgroep ([0.8759,1.5946]). Deze niet-overlappende BI'en concretiseren de significante  $p$ -waarde voor de *Breslow-Day*-test die enkel aangeeft dat de OR's voor beide strata, c.q. leeftijdsgroepen, significant verschillen.

Na deze algemene voorstelling van de statistische technieken die we bij het bivariaat onderzoek gebruiken, starten we in het volgende hoofdstuk met de eigenlijke empirische analyse van de buigingsalternantie.

## Noten

1. Bij nominale variabelen – zowat alle variabelen in dit onderzoek – is het statistisch irrelevant welke variabele de responsvariabele is. De statistische significantie- en associatiematen berekenen in eerste instantie de significantie en de sterkte van de associatie. De onderzoeksvraag bepaalt welke variabele als de responsvariabele en welke variabele als de verklarende variabele geïnterpreteerd wordt. De specifieke analytische problemen die samenhangen met (multinomiale) nominale variabelen komen uitgebreid aan bod in paragraaf 13.2.1.
2. Merk op dat een  $2 \times k$ -tabel in  $k-1$   $2 \times 2$ -tabellen kan worden opgesplitst. Eén waarde van de verklarende variabele moet dan wel als ijkwaarde in alle  $k-1$  tabellen worden opgenomen.
3. We spreken van *verband* om de situatie aan te duiden waar de waarde van de verklarende variabele de waarde van de responsvariabele (op significante wijze) beïnvloedt.
4. Gegeven een tabel met  $r$  rijen en  $k$  kolommen is het aantal vrijheidsgraden gelijk aan  $(r-1) \cdot (k-1)$ .
5. De *Cochran*-regel formuleert de volgende voorwaarden op het gebruik van de  $\chi^2$ -test:
  - bij  $2 \times 2$ -tabellen: alle verwachte celwaarden zijn gelijk aan of groter dan 5
  - bij  $r \times c$ -tabellen: de minimum verwachte frequentie van 1 is toelaatbaar als de verwachte frequentie in niet meer dan 20% van de cellen kleiner dan 5 is.
6. Dit is een exacte test die altijd kan worden uitgevoerd, ongeacht wat de geobserveerde of verwachte frequenties zijn. Deze test berekent de exacte kans op een extremere distributie dan de geobserveerde distributie wanneer de marginalen constant worden gehouden (Agresti 1996: 39-44).
7. Voor beide waarden van de responsvariabele kan een apart RR berekend worden. Doorgaans is men alleen in de kans op Succes geïnteresseerd. Denk bijvoorbeeld aan medisch onderzoek, waar men in eerste instantie in de kans op genezing geïnteresseerd is. Vertaald naar ons onderzoek betekent dit dat we in het RR van de gemarkeerde onverbogen adjectivische vorm geïnteresseerd zijn ( $RR_{\text{onverbogen}}$  of kortweg  $RR_{\text{onv}}$ ).
8. In onderzoek dat gebruik maakt van steekproeven, zoals alle corpuslinguïstisch onderzoek, is het nuttig het betrouwbaarheidsinterval (BI) te berekenen. Het BI geeft het gebied van waarden aan waarbinnen onder een bepaalde waarschijnlijkheid de werkelijke waarde van een variabele in de populatie ligt. Het 95% BI identificeert de zone waarvoor 95% kans bestaat dat de werkelijke waarde erbinnen ligt. Alle BI'en in deze studie zijn 95% BI'en. Dit is onzes inziens de optimale afweging tussen betrouwbaarheid en precisie, aangezien het interval groter wordt naarmate de restricties op de betrouwbaarheid strenger worden (met andere woorden: de limieten van een 99% BI zijn verder van elkaar verwijderd dan de limieten van een 95% BI). Het 95% BI is trouwens de pendant van het significantieniveau gelijk aan 0.05 dat we bij de statistische significantietests hanteren (cf.  $95\% = 100\% - 5\%$ ). Het 95% BI voor de OR wordt als volgt berekend:  $OR \pm 1.96 \cdot \text{Asymptotische Standaardfout}$ .

9. Bij de interpretatie van deze  $p$ -waarde moeten we er rekening mee houden dat 'e' de decimaalwaarde aanduidt:  $1e-03$  is bijvoorbeeld gelijk aan 0.001.
10. Omdat zowel de responsvariabele als de verklarende variabelen in dit onderzoek nominale variabelen zijn, gebruiken we de type-3 *Mantel-Haenszel*-correlatiestatistiek, de zogenaamde *Mantel-Haenszel- $\chi^2$* .
11. Wanneer de *Brelow-Day*-test de  $H_0$  – de OR's voor de verschillende waarden van de stratumvariabele zijn homogeen – niet kan verwerpen, wordt de analyse van de stratumtabellen niet gerapporteerd. Dit impliceert echter niet dat deze tests niet zijn uitgevoerd. Hetzelfde geldt voor de situaties waar de  $MH-\chi^2$  de  $H_0$  niet kan verwerpen. Dit betekent immers niet dat er *a priori* geen verband tussen de verklarende variabele en de responsvariabele kan zijn. Het is – vooral bij kleine materiaalverzamelingen – best mogelijk dat tegengestelde verbanden in verschillende strata elkaar opheffen. Enkel de vergelijking van de OR's voor de afzonderlijke stratumtabellen kan deze tegengestelde tendenties exact in kaart brengen.



## Hoofdstuk 8:

# Morfosyntactische conditionering

We starten het empirisch onderzoek met de analyse van de invloed van de morfosyntactische factoren. Deze factoren voorspellen een gedifferentieerd effect van de morfosyntactische kenmerken van de elementen die de posities in de NP bezetten op de keuze van het adjectivische buigingsmorfeem. We bestuderen achtereenvolgens de woordsoort van de determinator (8.1), de gradatie van het adjectief (8.2), de diminuering van het substantief (8.3), de woordsoortelijke subclassificatie van het element in de nominale kernpositie (8.4) en de bigenericiteit van het substantief (8.5). De eerste drie factoren zijn in het literatuuroverzicht voorgesteld (zie paragraaf 1.1.1), de laatste twee factoren worden op inductieve wijze geïdentificeerd tijdens de empirische analyse.

### 8.1 De woordsoort van de determinator

#### 8.1.1 Hypotheses en operationalisering

De hypothese voorspelt dat de onverbogen vorm van het adjectief vaker geselecteerd wordt wanneer de determinator een bezittelijk voornaamwoord is (ANS 1997: 411; Degryse 1959: 73-84; Klein & Visscher 1996: 71; Lebrun & Schurmans-Swillen 1966: 183; Raidt 1968: 100; de Rooij 1980a: 14, 16-17, 20-21; Royen 1953: 103-104). Deze hypothese sluit aan bij de historische evolutie van de adjectivische buiging: vanaf het Middel-Nederlands tot halverwege de twintigste eeuw is de onverbogen variant dominant na een bezittelijk voornaamwoord (zie paragrafen 2.2 en 2.3).<sup>1</sup> Op basis van de historische verwantschap tussen het bepaald lidwoord en de vormen van het aanwijzend voornaamwoord kan bovendien worden aangenomen dat deze woordsoorten geen differentiërend effect op de buigingsalternantie hebben.

De operationalisering van deze hypothese stelt geen enkel probleem: aan de hand van de woordsoortannotatie in het CGN (zie paragraaf 6.2.1) onderscheiden we de volgende woordklassen in de determinatorpositie:

- het bepaald lidwoord *het*
- het aanwijzend voornaamwoord, met name *dit* en *dat* (we verwijzen naar paragraaf 6.2.2 voor de discussie van het adjectivische buigingsgedrag bij de onbepaalde aanwijzende voornaamwoorden *zo'n* en *zulk*)
- het bezittelijk voornaamwoord
- de geanteponeerde genitief

Vertrekkend van bovengenoemde hypothese beschouwen we het bepaald lidwoord als de ongemarkeerde context die als referentiepunt fungeert voor de evaluatie van de impact van de andere woordsoorten.

### 8.1.2 Resultaten en bespreking

Tabel 8(1) vat het buigingseffect samen van de woordklasse van het element dat de determinatorpositie bezet. Deze frequentieverdeling is significant (*Fisher Exact*-test:  $p = 0.006704$ ).<sup>2</sup>

	<i>het</i>		<i>aanw. vnw.</i>		<i>bez. vnw.</i>		<i>genitief</i>		$\Sigma$	
<b>verbogen</b>	2,897	75.74	460	81.99	438	78.35	15	78.95	3,810	76.75
<b>onverbogen</b>	928	24.26	101	18.01	121	21.65	4	21.05	1,154	23.25
<b><math>\Sigma</math></b>	<b>3,825</b>	<b>100.0</b>	<b>561</b>	<b>100.0</b>	<b>559</b>	<b>100.0</b>	<b>19</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 8(1): Invloed van de woordsoort van de determinator op de adjectivische buigingsalternantie

Deze gegevens tonen aan dat de hypothese op twee punten moet worden bijgesteld. Ten eerste, het bezittelijk voornaamwoord definieert niet de syntagmatische context waar het onverbogen adjectief het vaakst voorkomt. Het aandeel onverbogen adjectieven na een bezittelijk voornaamwoord is zelfs kleiner dan het algemeen gemiddelde, zodat er zeker geen sprake is van een absolute voorkeur voor de onverbogen buigingsvariant na een bezittelijk voornaamwoord. Deze gegevens contrasteren met de resultaten van het empirisch onderzoek van Degryse (1959) en Lebrun & Schurmans-Swillen (1966) en staven de verklaring die we in het literatuuroverzicht hebben aangereikt, met name dat de absolute voorkeur voor het onverbogen adjectief na een bezittelijk voornaamwoord in deze studies bepaald wordt door de datering van het gebruikte materiaal. Aangezien beide studies geschreven materiaal van respectievelijk 1950 en 1964 gebruiken, is het plausibel dat de grammaticale traditie, die tot halverwege de twintigste eeuw het gebruik van het onverbogen adjectief na een bezittelijk voornaamwoord voorschrijft, de absolute voorkeur voor de onverbogen vorm verklaart.

Ten tweede, er is een significant verschil tussen de buigingsprofielen van het bepaald lidwoord en het aanwijzend voornaamwoord: de verbogen vorm van het adjectief wordt significant vaker gebruikt na het aanwijzend voornaamwoord dan na het bepaald lidwoord ( $\chi^2 = 10.324$ ,  $df = 1$ ,  $p = 0.001313$ ;  $RR_{\text{verbogen}} = 1.0826$  [1.0373, 1.1298];  $OR = 0.6854$  [0.5458, 0.8607]). Wanneer we naar de celstatistieken in tabel 8(2) kijken, stellen we zelfs vast dat de significantie van de frequentieverdeling in tabel 8(1) een gevolg is van het zeer lage aandeel onverbogen vormen wanneer een aanwijzend voornaamwoord de determinatorpositie invult. Voor de berekening van de celstatistieken gebruiken we de partiële  $\chi^2$ -statistiek. Deze maat berekent de bijdrage van de verschillende cellen in een contingencietabel tot de globale  $\chi^2$ -statistiek (Wonnacott & Wonnacott 1990: 552-554). We drukken

de bijdrage van de verschillende cellen tot de globale  $\chi^2$ -statistiek uit aan de hand van de 95% betrouwbaarheidsintervallen (BI). Een cel levert een significante bijdrage tot de globale  $\chi^2$ -statistiek wanneer het BI voor deze cel de scharnierwaarde 1 niet bevat. De BI'en worden, net als bij het RR en de OR, op de volgende manier geïnterpreteerd:

- een BI waarvan beide limieten kleiner zijn dan 1, duidt aan dat de geobserveerde frequentie in deze cel significant lager is dan verwacht
- een BI waarvan beide limieten groter zijn dan 1, duidt aan dat de geobserveerde frequentie in deze cel significant groter is dan verwacht
- een BI dat de scharnierwaarde 1 bevat, duidt aan dat de geobserveerde frequentie in deze cel niet significant afwijkt van de verwachte frequentie.

Tabel 8(2) vat de resultaten van de partiële  $\chi^2$ -statistiek samen.

	<i>het</i>	<i>aanw. vnw.</i>	<i>bez. vnw.</i>	<i>genitief</i>
<b>verbogen</b>	[0.9636,1.0050]	[0.9753,1.1613]	[0.9296,1.1122]	[0.5088,1.5483]
<b>onverbogen</b>	[0.9831,1.1041]	[0.6249,0.9239]	[0.7672,1.0950]	[0.0185,1.7927]

Tabel 8(2): 95% BI'en partiële  $\chi^2$ -statistiek voor de invloed van de woordsoort van de determinator op de adjectivische buigingsalternantie

We constateren dat voor slechts één cel een significant resultaat bekomen wordt: bij een aanwijzend voornaamwoord in determinatorpositie is het aantal onverbogen vormen significant kleiner dan verwacht.<sup>3</sup> De BI'en in de andere cellen in tabel 8(2) bevatten allemaal de scharnierwaarde 1. Dit wijst erop dat deze cellen geen significante bijdrage tot de globale  $\chi^2$ -statistiek leveren. Merk op dat de daling van het aantal verbogen vormen bij het bepaald lidwoord net niet significant is.

Het bezittelijk voornaamwoord is intern een zeer heterogene woordklasse: iedere combinatie van de waarden voor persoon en getal (en genus in de derde persoon enkelvoud) wordt door een andere vorm uitgedrukt. Wanneer we de buigingsprofielen van de bezittelijke voornaamwoorden bekijken, valt één vorm op, met name *ons*. Na deze vorm wordt het onverbogen adjectief significant vaker geselecteerd dan na ieder ander bezittelijk voornaamwoord. Tabel 8(3) toont de buigingsprofielen voor *ons* en voor de andere bezittelijke voornaamwoorden ('*bez.vnw-<sub>ons</sub>*'). Het verschil tussen beide buigingsprofielen is significant ( $\chi^2 = 24.2257$ ,  $df = 1$ ,  $p = 8.568e-07$ ).

	<i>bez.vnw-<sub>ons</sub></i>		<i>ons</i>	$\Sigma_{\text{bez.vnw}}$		
<b>verbogen</b>	413	81.14	25	50.00	438	78.35
<b>onverbogen</b>	96	18.86	25	50.00	121	21.65
<b><math>\Sigma</math></b>	<b>509</b>	<b>100.0</b>	<b>50</b>	<b>100.0</b>	<b>559</b>	<b>100.0</b>

Tabel 8(3): Invloed van *ons* op de adjectivische buigingsalternantie

Na *ons* is de kans op een onverbogen adjectief 2.65 [1.9047,3.6897] maal groter dan na een ander bezittelijk voornaamwoord en 2.06 [1.5533,2.7344] maal groter dan na een bepaald lidwoord. De OR's voor de odds <sup>onverbogen</sup>/<sub>verbogen</sub> bedragen respectievelijk 4.302 [2.3677,7.817] ten overstaan van de andere bezittelijke voornaamwoorden en 3.1218 [1.7845,5.4613] ten overstaan van het bepaald lidwoord.<sup>4</sup> Bij de andere bezittelijke voornaamwoorden is het aandeel verbogen vormen vergelijkbaar met het aandeel bij het aanwijzend voornaamwoord.

Het frequente gebruik van het onverbogen adjectief na *ons* is wellicht een instantie van syntagmatische nivellering: de inflectionele oppositie tussen *ons* en *onze* beïnvloedt de verbuiging van het daaropvolgende adjectief. Terwijl de andere bezittelijke voornaamwoorden, net zoals het aanwijzend voornaamwoord en het bepaald lidwoord, geen buigingsalternantie vertonen<sup>5</sup>, hoort de dyade *ons/onze* – vanuit een morfosyntactisch perspectief – thuis in de categorie met de vragende en de onbepaalde voornaamwoorden. Het inflectionele paradigma van deze elementen bevat een verbogen en een onverbogen vorm. Deze vormelijke onderscheiding valt samen met het genusonderscheid tussen *de*-woorden en *het*-woorden, zoals de vormen in tabel 8(4) aantonen.

	<i>de</i> -woord	<i>het</i> -woord
<b>Vragend/betrekkelijk vnw</b>	<i>welk-e</i>	<i>welk</i>
<b>Onbepaald vnw</b>	<i>elk-e</i>	<i>elk</i>
	<i>ieder-e</i>	<i>ieder</i>
	<i>een of ander-e</i>	<i>een of ander</i>
<b>Bezittelijk vnw</b> <sub>[1e PERS,MV]</sub>	<i>onz-e</i>	<i>ons</i>
<b>Adjectief</b>	<i>-e</i>	<i>-∅</i>

Tabel 8(4): Inflectioneel paradigma van het vragend voornaamwoord, het onbepaald voornaamwoord en *ons*

Daarnaast congrueert het volgende adjectief met de vorm van de determinator: is de determinator onverbogen, dan is het adjectief ook onverbogen; is de determinator verbogen, dan wordt het adjectief ook verbogen. De voorbeelden in (1) illustreren deze analogie: bij het *de*-woord *man* nemen zowel de determinator als het adjectief een buigings-*e*; bij het *het*-woord *paard* blijven de elementen in beide adnominale posities onverbogen.

- (1)
- |    |                           |                           |
|----|---------------------------|---------------------------|
| a. | welk-e mooi-e man         | ; welk mooi paard         |
| b. | elk-e mooi-e man          | ; elk mooi paard          |
| c. | ieder-e mooi-e man        | ; ieder mooi paard        |
| d. | een of ander-e mooi-e man | ; een of ander mooi paard |
| e. | onz-e mooi-e man          | ; ons mooi(-e) paard      |

In tegenstelling tot de andere elementen in tabel 8(4) is de syntagmatische nivellering na *ons* geen categorische regelmaat (1a-d) maar een (sterke) tendentie (1e), zoals de cijfers in tabel 8(3) aantonen.

Een voorbeeld van syntagmatische nivellering dat het gebruik van de buigings-*e* lijkt te vergemakkelijken, is de tendentie om een verbogen adjectief te gebruiken na het collectiverend onbepaald lidwoord *alle* (dit is trouwens de reden waarom *alle* niet in tabel 8(4) is opgenomen). Zoals aangestipt bij de bespreking van de materiaalverzameling op basis van het CGN, verschijnt bij *alle*, in tegenstelling tot bij de andere collectiverende onbepaalde voornaamwoorden *ieder* en *elk*, de verbogen variant in 50% van de gevallen (zie paragraaf 6.2.2). Wellicht wordt *alle* in zulke gevallen als *al(l)-e* geanalyseerd, een interpretatie die op haar beurt het gebruik van het verbogen adjectief verklaart.<sup>6</sup>

Tot slot signaleren we de (niet-significante, wegens dataschaarste) toename van het overbogen adjectief na *uw*, wat mogelijk een tussentaligheidseffect is in het Belgisch Nederlandse deelcorpus (zie ook paragraaf 12.3.1).

De analyse van het buigingseffect van de woordsoort van de determinator heeft beide hypothesen weerlegd die we op basis van het literatuuroverzicht geformuleerd hebben. In de eerste plaats stellen we vast dat na het bepaald lidwoord de onverbogen vorm vaker voorkomt dan na een andere determinator, inclusief het bezittelijk voornaamwoord. Wanneer het bezittelijk voornaamwoord *ons* de determinatorpositie invult, wordt het onverbogen adjectief wel significant vaker geselecteerd als gevolg van de syntagmatische nivellering tussen het adjectivische buigingsmorfeem en de onverbogen vorm *ons*. Daarnaast wordt de hypothese dat de buigingsprofielen van het bepaald lidwoord en het aanwijzend voornaamwoord niet significant verschillen evenmin door de empirische gegevens onderbouwd.

In paragraaf 10.2.3 zullen we de interactie tussen de woordsoort van de determinator en de prosodische structuur van het adjectief onderzoeken: in de literatuur wordt immers – om redenen die ons niet duidelijk zijn – aangenomen dat een bezittelijk voornaamwoord het deflecterende effect van een onbeklemtoonde adjectivische eindlettergreep versterkt (ANS 1997; Raidt 1968). In paragraaf 12.3.3.1 testen we tot slot of het buigingseffect van de woordsoort van de determinator stilistisch geconditioneerd is (Raidt 1968, de Rooij 1980a), en zo ja, wat de richting van deze conditionering is.

## 8.2 Gradatie van het adjectief

### 8.2.1 Hypothese en operationalisering

De hypothese beschouwt de comparatiefvorm van het adjectief als een faciliterende context voor het gebruik van de onverbogen buigingsvariant. De reden hiervoor is prosodisch: door de selectie van de onverbogen vorm vermijdt de spreker de opeenvolging van twee (of meer) doffe lettergrepen aan de rechterwoordrand van het adjectief (*inter al.* ANS 1997: 411; de Haas & Trommelen 1993: 286; Klein &

Vischer 1996: 70; Penninckx & Buyse 1997: 70; de Rooij 1980a; Verhasselt 1974: 140).

Voor de identificatie van de vormen van de vergrotende trap kijken we naar de inflectionele structuur van de woordvorm die de adjectivische positie in de NP bezet. De aanwezigheid van het inflectionele suffix *-er(e)* is een noodzakelijke voorwaarde om het element in adjectivische positie als een comparatief te beschouwen. Dit algemene selectie criterium wordt met een bijkomende conditie aangevuld: de stellende trap van de vorm in kwestie moet synchroon duidelijk identificeerbaar zijn. De toepassing van dit criterium leidt tot de exclusie van vormen zoals *nader* en *eerder* waarvoor de stellende trap niet eenduidig kan worden vastgesteld.<sup>7</sup> Vormen zoals *vroeger* en *verder* worden daarentegen wel als analytisch transparante vormen beschouwd: *vroeg-er* en *ver-der*. Suppletievormen, zoals *beter*, hebben we eveneens als comparatieven gecodeerd. In deze analyse kijken we enkel naar de morfosyntactische structuur; de prosodische structuur komt in paragraaf 10.2 aan bod.

Het spreekt voor zich dat de vergrotende trap de gemarkeerde context is en dat de stellende trap het vergelijkingspunt is om de invloed van de vergrotende trap te meten.

### 8.2.2 Resultaten en bespreking

In tegenstelling tot wat algemeen in de literatuur wordt aangenomen, selecteert de vorm van de vergrotende trap niet vaker de onverbogen buigingsvariant dan de vorm van de stellende trap. Er is nauwelijks enig verschil tussen beide contexten ( $\chi^2 = 0.0021$ ,  $df = 1$ ,  $p = 0.9633$ ;  $RR_{\text{onverbogen}} = 0.9773$  [0.8826,1.1487];  $OR = 0.9707$  [0.5437,1.7328]).

	stellende trap		vergroten de trap		$\Sigma$	
<b>verbogen</b>	3,759	76.75	51	77.27	3,810	76.75
<b>onverbogen</b>	1,139	23.25	15	22.73	1,154	23.25
<b><math>\Sigma</math></b>	<b>4,898</b>	<b>100.0</b>	<b>66</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 8(5): Invloed van de gradatie van het adjectief op de adjectivische buigingsalternantie

We komen hier later op terug tijdens de analyse van de prosodische structuur van het adjectief waar we zullen nagaan of deze vaststelling veralgemeend moet worden voor alle adjectieven die op één of meerdere doffe lettergrepen eindigen (zie paragraaf 10.2).

### 8.3 Diminuering van het substantief

#### 8.3.1 *Hypothese en operationalisering*

Volgens de hypothese komt de onverbogen vorm vaker voor wanneer het nominale hoofd een diminutievorm is dan wanneer het de niet-gediminueerde basisvorm van het substantief is. Deze hypothese vinden we terug bij Raidt (1968), zonder dat deze auteur echter een motivatie voor het eventuele deflecterende effect van het verkleinwoord aanreikt.

Voor het testen van deze hypothese hebben we de verkleinwoorden op basis van hun inflectionele vorm geïdentificeerd. We beschouwen substantieven die eindigen op de allomorfen van *-je* (*kamertje*, *boompje*, *gezinnetje*), op *-ke* (*lieke*) en op *-ie* (*ukkie*) als verkleinwoorden. Gelexicaliseerde diminutieven vormen hierop een uitzondering. Dit zijn de verkleinwoorden die synchroon niet meer geanalyseerd kunnen worden, zoals *meisje* of *zootje*, en de verkleinwoorden met een gespecialiseerde betekenis, zoals *jasje*. De niet-gediminueerde basisvorm van het substantief (‘-diminutief’ in tabel 8(6)) dient als referentiepunt voor het meten van de impact van de gemarkeerde diminutievorm op de adjectivische buiging.

#### 8.3.2 *Resultaten en bespreking*

In tabel 8(6) wordt de impact van de diminuering op de buigingsalternantie samengevat.<sup>8</sup>

	-diminutief		diminutief		Σ	
<b>verbogen</b>	3,520	75.93	290	88.45	3,810	76.75
<b>onverbogen</b>	1,116	24.07	38	11.55	1,154	23.25
<b>Σ</b>	<b>4,636</b>	<b>100.0</b>	<b>328</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 8(6): *Invloed van de diminuering van het substantief op de adjectivische buigingsalternantie*

De diminuering van het substantief heeft een significante invloed op de selectie van de adjectivische buigingsvorm ( $\chi^2 = 26.0744$ ,  $df = 1$ ,  $p = 3.285e-07$ ). De zin van de onderliggende tendens is echter tegengesteld aan de hypothese: het onverbogen adjectief wordt significant minder vaak gebruikt vóór een diminutief dan vóór een niet-gediminueerd *het*-woord. Volgens het  $RR_{\text{onverbogen}}$  en de OR voor de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bedraagt het aandeel van de onverbogen pendant bij een verkleinwoord minder dan de helft van het aandeel bij een substantivische basisvorm:  $RR_{\text{onverbogen}} = 0.4813$  [0.3554,0.6518];  $OR = 0.4133$  [0.2928,0.5834].

De sterkere voorkeur voor het gebruik van het verbogen adjectief in NP's waar de kernpositie door een verkleinwoord wordt bezet, kan als een frequentie-effect worden verklaard: als we ervan uitgaan dat de selectie van de adjectivische buigingsvariant bij een substantief mee bepaald wordt door het buigingsprofiel van (het lemma van) dit substantief, dan kunnen we de frequentieverdeling in tabel 8(6)

als een structureel bepaald frequentie-effect verklaren. De verkleinwoorden zijn in het Nederlands altijd neutra, ongeacht of de basisvorm een *het*-woord of een *de*-woord is: doordat dus ook *de*-woorden, waar de onverbogen buigingsvariant bij de niet-gediminueerde basisvorm taalstructureel is uitgesloten (enkele uitzonderingen niet te na gesproken), tot de verkleinwoorden behoren, is het plausibel dat het verbogen adjectief vaker bij een verkleinwoord dan bij de basisvorm van een onzijdig substantief voorkomt (waar de onverbogen variant structureel kan voorkomen). Anders gezegd: het grotere aandeel verbogen vormen bij de verkleinwoorden hangt samen met de gediminueerde *de*-woorden waarvan het buigingsprofiel minder onverbogen vormen bevat dan dat van de *het*-woorden. We testen deze verklaring door de verkleinwoorden, die in tabel 8(6) als een homogene groep worden behandeld, op te splitsen in functie van het genus van de basisvorm: indien de aangereikte verklaring empirisch correct is, verwachten we een significant groter aandeel verbogen vormen bij gediminueerde *de*-woorden dan bij gediminueerde *het*-woorden. De frequentieverdeling in tabel 8(7) staft onze verklaring: het aandeel onverbogen vormen is groter bij de diminutiefvormen van *het*-woorden dan bij de diminutiefvormen van *de*-woorden.

	stam <sub>het-woord</sub>		stam <sub>de-woord</sub>		$\Sigma_{dim}$	
<b>verbogen</b>	84	85.71	205	89.52	289	88.45
<b>onverbogen</b>	14	14.29	24	10.48	38	11.55
<b><math>\Sigma</math></b>	<b>98</b>	<b>100.0</b>	<b>229</b>	<b>100.0</b>	<b>328</b>	<b>100.0</b>

Tabel 8(7): Invloed van het genus van de niet-gediminueerde substantivische basisvorm op de adjectivische buigingsalternantie bij verkleinwoorden

De tendentie in tabel 8(7) is echter niet significant ( $\chi^2 = 0.6326$ ,  $df = 1$ ,  $p = 0.4264$ ).<sup>9</sup> De voorgestelde verklaring stemt bovendien niet overeen met de vaststelling dat het onderscheid tussen de vormen van de stellende trap en de diminutiefvormen van *het*-woorden significant is ( $\chi^2 = 4.5342$ ,  $df = 1$ ,  $p = 0.03322$ ): bij een gediminueerd *het*-woord wordt de buigings-*e* significant vaker geselecteerd dan bij een niet-gediminueerd *het*-woord, zoals kan worden afgeleid uit de cijfers voor het RR ( $RR_{verbogen} = 1.1289$  [1.0396,1.2259]) en de OR ( $OR = 0.5256$  [0.2974,0.9294]). Voor dit verschil kunnen we geen verklaring aanreiken.

We besluiten dat, in tegenspraak met de stelling van Raidt (1968), het onverbogen adjectief significant minder vaak geselecteerd wordt bij een verkleinwoord dan bij een basisvorm van een *het*-woord. Als mogelijke verklaring denken we aan een frequentie-effect, waarbij het buigingsprofiel voor het gehele substantivische lemma de keuze van de buigingsvariant beïnvloedt. Deze verklaring wordt echter niet door de empirische gegevens onderbouwd.



## 8.4 Woordsoortelijke subclassificatie van het element in de kernpositie

### 8.4.1 Hypotheses en operationalisering

De hypothese met betrekking tot het buigingseffect van de woordsoortelijke subclassificatie van het element in de kernpositie is via inductieve weg tot stand gekomen: tijdens de empirische analyse rees het vermoeden dat bij een gesubstantiveerde infinitief het aandeel onverbogen vormen groter is dan bij een substantief. Voor de falsificatie van deze hypothese coderen we het onderscheid tussen genominaliseerde infinitieven en de andere vormen die de kernpositie bezetten.<sup>10</sup>

Het buigingseffect van andere nominalisaties kan omwille van dataschaarste niet empirisch bestudeerd worden: de materiaalverzameling bevat slechts drie genominaliseerde adjectieven, die tweemaal een verbogen adjectief (*het doorgelooide glimbruin*) en eenmaal een onverbogen adjectief (*het Europees gemiddelde*) selecteren.

Naast deze morfosyntactische stelling testen we ook de syntactische hypothese die stelt dat – wanneer de kernpositie door een genominaliseerde infinitief wordt bezet – de buigings-*e* de adjectief-nomensequentie als een semantisch geheel identificeert dat autonoom geconceptualiseerd kan worden, terwijl de selectie van de onverbogen vorm op een semantisch vage constructie wijst die nood heeft aan bijkomende informatie in de vorm van een werkwoordelijk complement (Verhagen 1993; zie paragraaf 1.1.6). Voor de operationalisering van deze stelling repliceren we het onderzoek van Verhagen (1993): we onderscheiden de genominaliseerde infinitieven die gemodificeerd worden door complementen (2a) en de genominaliseerde infinitieven die niet gemodificeerd worden door complementen (2b).

- (2)
- a. om de achtergronden van *het hedendaagse denken over pijn* beter te begrijpen is het nodig dat
  - b. voetje voor voetje schuifelt ze vanaf de auto naar de voordeur een broer voorop en een broer achter haar voor *het eventuele vallen*

Onder complement verstaan we een constituent die tot het valentiepatroon van het werkwoord behoort. Dit zijn: het onderwerp, het lijdend voorwerp, het meewerkend voorwerp en de voorzetselcomplementen die door de werkwoordsvorm geselecteerd worden, zoals de selectie van het voorzetsel *over* door het werkwoord *denken* in (2a).

### 8.4.2 Resultaten en bespreking

De gegevens in tabel 8(8) bevestigen de morfosyntactische hypothese: het aandeel onverbogen vormen is significant groter wanneer een infinitief de nominale kern-

positie bezet dan wanneer deze positie door een substantief wordt ingenomen ( $\chi^2 = 20.3621$ ,  $df = 1$ ,  $p = 6.409e-06$ ).

	substantief		infinitief		$\Sigma$	
<b>verbogen</b>	3,758	77.13	52	56.52	3,810	76.75
<b>onverbogen</b>	1,114	22.87	40	43.48	1,154	23.25
<b><math>\Sigma</math></b>	<b>4,872</b>	<b>100.0</b>	<b>92</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 8(8): Invloed op de adjectivische buigingsalternantie van de woordsoortelijke subclassificatie van het element in de nominale kernpositie

De kans op een onverbogen vorm verdubbelt bijna wanneer een genominaliseerde infinitief de kernpositie bezet:  $RR_{\text{onverbogen}} = 1.901$  [1.4978,2.4140]. De OR die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bij de genominaliseerde infinitieven en de substantieven vergelijkt, bedraagt zelfs meer dan het tweevoud:  $OR = 2.595$  [1.7091,3.9399].

De stijging van het aandeel onverbogen vormen bij een genominaliseerde infinitief wordt wellicht voor een deel bepaald door de adverbiale waarde van het adjectief in deze context. In (3a) modificeert het adjectief *geduldig* de activiteit uitgedrukt door de genominaliseerde infinitief *herstellen*, net zoals het bijwoord *geduldig* dat in (3b) doet bij de vervoegde vorm *herstelt*. Met andere woorden: naar analogie met het (onverbogen) bijwoord in (3b) wordt het adjectief in (3a) niet verbogen.

- (3) a. [hij was] systematisch bezig met het ontwarren van zijn drijfnet of  
*het geduldig herstellen* van een scheur  
 b. hij herstelt *geduldig*/\*geduldige een scheur

We falsificeren nu de tweede, syntactische hypothese. Deze hypothese voorspelt dat een genominaliseerde infinitief die vergezeld is van een complement vaker de onverbogen vorm selecteert dan een 'naakte' genominaliseerde infinitief. De relatieve frequenties in tabel 8(9) zijn in overeenstemming met deze hypothese. Wat de significantie van dit effect betreft, spreken de verschillende statistische maten elkaar tegen: de  $p$ -waarde voor de  $\chi^2$ -statistiek is net niet significant ( $\chi^2 = 3.4417$ ,  $df = 1$ ,  $p = 0.06357$ ), terwijl de 95% BI'en voor het RR en de OR de scharnierwaarde 1 net niet bevatten en dus op een significant verschil wijzen ( $RR_{\text{onverbogen}} = 1.6285$  [1.0332,2.5668];  $OR = 2.4666$  [1.0410,5.8443]). De ruime BI'en voor het RR en de OR wijzen op dataschaarste, wat de randwaarden voor de significantie verklaart.

	Inf <sub>zonder.complement</sub>		Inf <sub>met.complement</sub>		$\Sigma_{\text{Infinitief}}$	
<b>verbogen</b>	37	64.91	15	42.86	52	56.52
<b>onverbogen</b>	20	35.09	20	57.14	40	43.48
<b><math>\Sigma</math></b>	<b>57</b>	<b>100.0</b>	<b>35</b>	<b>100.0</b>	<b>92</b>	<b>100.0</b>

Tabel 8(9): Invloed van de aanwezigheid van een werkwoordelijk complement op de adjectivische buigingsalternantie bij een genominaliseerde infinitief

Dit onderscheid wordt niet in de multivariate analyse opgenomen omdat het slechts op een zeer beperkt deel van de materiaalverzameling van toepassing is.

## 8.5 Bigenericiteit van het substantief

### 8.5.1 Hypothese en operationalisering

Deze hypothese is langs inductieve weg geformuleerd, net zoals de hypothese met betrekking tot de invloed van de woordsoortelijke subclassificatie van het element in de kernpositie in de vorige paragraaf. Bij de controle van het genus van de elementen die de kernpositie bezetten (zie paragraaf 6.2.2), rees het vermoeden dat de onverbogen buigingsvariant vaker voorkomt bij *het*-woorden dan bij bigenerische substantieven, omdat het buigingsprofiel van de laatste groep mogelijkwijs beïnvloed wordt door de niet-onzijdige realisaties.

Voor de operationalisering van deze hypothese hebben we een beroep gedaan op de genusannotatie in de *Grote Van Dale* (zie paragraaf 6.2). De substantieven in de materiaalverzameling zijn als volgt gecodeerd:

- *het*-woord – bv. *het kind*/\**de kind*
- bigenerisch substantief met semantisch gemotiveerde genusonderscheiding – bv. *de punt* (leesteken; scherpe spits) vs. *het punt* (spikkel; aantekening)
- bigenerisch substantief zonder semantisch gemotiveerde genusonderscheiding – bv. *de/het matras*

De *het*-woorden vormen het referentiepunt bij het kwantificeren van het buigings-effect van de gemarkeerde bigenerische substantieven.

### 8.5.2 Resultaten en bespreking

De frequentieverdeling in tabel 8(10) bevestigt de hypothese: de onverbogen vorm van het adjectief komt significant vaker voor bij een onzijdig substantief dan bij een bigenerisch substantief. Bij een bigenerisch substantief bedraagt het aandeel van de onverbogen variant slechts de helft van het aandeel bij een onzijdig substantief:  $RR_{\text{onverbogen}} = 0.4947$  [0.3435,0.7123];  $OR = 0.4273$  [0.2823,0.6469]. De significantie die we op basis van de BI'en kunnen vaststellen, wordt bevestigd door de  $\chi^2$ -statistiek ( $\chi^2 = 16.425$ ,  $df = 1$ ,  $p = 5.061e-05$ ).<sup>11</sup>

	<i>het</i> -woord		bigenerisch		N	$\Sigma$
<b>verbogen</b>	3,615	76.22	195	88.24	3,810	76.75
<b>onverbogen</b>	1,128	23.78	26	11.76	1,154	23.25
<b><math>\Sigma</math></b>	<b>4,743</b>	<b>100.0</b>	<b>221</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 8(10): Invloed van de bigenericiteit van het substantief op de adjectivische buigingsalternantie

Het waargenomen genuseffect is waarschijnlijk een frequentie-effect. De instanties waar een bigenerisch substantief als een *de*-woord gebruikt wordt, sluiten immers om taalstructurele redenen de selectie van de onverbogen vorm uit (zie figuur 3(1) en paragraaf 8.3.2). Het is bijgevolg plausibel dat de selectie van de onverbogen buigingsvariant bij de onzijdige instanties van een bigenerisch substantief op meer weerstand botst doordat het buigingsprofiel van deze klasse substantieven meer verbogen vormen bevat dan het buigingsprofiel van de *het*-woorden (die per definitie nooit als *de*-woord gebruikt worden). Met andere woorden: de aard zelf van de bigenerische substantieven verkleint de kans op de selectie van een onverbogen vorm.<sup>12</sup>

## 8.6 Conclusie

In het onderzoek naar de morfosyntactische conditionering van de buigingsalternantie hebben we de invloed van de volgende factoren getest:

- de woordsoort van de determinator: In tegenstelling tot wat vroeger empirisch onderzoek heeft aangetoond, wordt de onverbogen vorm niet vaker gekozen wanneer de determinator een bezittelijk voornaamwoord is. Het tegendeel is waar: na een bezittelijk voornaamwoord is het aandeel van de onverbogen buigingsvariant lager dan gemiddeld, terwijl het grootste aandeel onverbogen vormen na een bepaald lidwoord voorkomt. Daarnaast hebben we vastgesteld dat bij de helft van de observaties met *ons* als determinator de onverbogen buigingsvariant geselecteerd wordt. Dit is een instantie van syntagmatische nivellering tussen het adjectivische buigingsmorfeem en het onverbogen bezittelijk voornaamwoord *ons*. Bij de multivariate analyse komen we uitgebreid terug op het buigingseffect van het element in de determinatorpositie (zie paragraaf 14.2.3.2).
- de gradatie van het adjectief: De corpusgegevens weerleggen de door veel auteurs aangehangen hypothese dat een adjectivische vorm van de vergrotende trap vaker onverbogen blijft dan een vorm van stellende trap.
- de diminuering van het substantief: De empirische analyse heeft de hypothese ontkracht dat het verkleinwoord een faciliterende factor is voor het gebruik van de onverbogen buigingsvariant. Uit het materiaal blijkt dat bij het verkleinwoord net de verbogen variant significant vaker voorkomt.
- de woordsoortelijke subclassificatie van het substantief: Wanneer het element in de kernpositie een genominaliseerde infinitief is, komt de onverbogen vorm significant vaker voor dan wanneer een ander element, c.q. een substantief, deze positie invult. De stelling dat een genominaliseerde infinitief die vergezeld is van een (werkwoordelijk) complement vaker het verbogen adjectief selecteert, wordt – wellicht als gevolg van dataschaarste – niet eenduidig door de empirische gegevens ondersteund.

- de bigenericiteit van het substantief: Bij een bigenerisch substantief in de nominale kernpositie komt de buigings-*e* significant vaker voor dan bij een *het*-woord.

We ronden dit hoofdstuk af met een algemene kwantitatieve vaststelling: het is opmerkelijk dat geen enkele morfosyntactische factor een parameterwaarde met een absolute voorkeur voor de onverbogen variant definieert. Dit doet vermoeden dat de morfosyntactische factoren bij de multivariate analyse een secundaire status krijgen toegewezen. Bovendien zijn de gemarkeerde waarden van de morfosyntactische factoren laagfrequent in onze materiaalverzameling.

## Noten

1. Pas in de achttiende eeuw, met onder andere de polemieken tussen Huydecoper en Ten Kate, doet de verbogen buigingsvariant opgeld na een bezittelijk voornaamwoord als determinator (zie paragraaf 2.3.3).
2. We herinneren de lezer eraan dat we automatisch de *Fisher Exact*-test gebruiken wanneer de frequentieverdeling niet aan de voorwaarden van de *Cochran*-regel voldoet (zie paragraaf 7.1.1 noot 5); in tabel 8(1) is de verwachte frequentie voor de onverbogen adjectieven bij een genitief kleiner dan 5 (4.417002). Wanneer we de genitief terzijde laten, is de frequentieverdeling eveneens significant ( $\chi^2 = 11.6504$ ,  $df = 2$ ,  $p = 0.002952$ ).
3. We verkrijgen een vergelijkbaar resultaat voor de partiële  $\chi^2$ -statistiek wanneer we de geanteponeerde genitief niet in de berekening opnemen. Ook dan levert slechts één cel een significante bijdrage tot de globale  $\chi^2$ -statistiek: het aantal onverbogen vormen na een aanwijzend voornaamwoord is opnieuw significant lager dan verwacht. Het 95% BI voor deze cel bedraagt na uitsluiting van de genitieven [0.6247, 0.9236].
4. De  $RR_{\text{onverbogen}}$ 's en de OR's die de buigingsalternantie bij *ons* en de andere bezittelijke voornaamwoorden vergelijken, zijn berekend op basis van de gegevens in tabel 8(3). De  $RR_{\text{onverbogen}}$ 's en de OR's die de buigingsalternantie bij *ons* en het bepaald lidwoord vergelijken, zijn berekend op basis van de tweede kolom in tabel 8(3) en de eerste kolom in tabel 8(1), die als referentiepunt fungeert.
5. Bij het aanwijzend voornaamwoord en het bepaald lidwoord wordt het genusonderscheid niet in het uitgangsmorfeem maar in het stammorfeem gemarkeerd: *deze/die* vs. *dit/dat*; *de* vs. *het*. Merk op dat de genusonderscheiding verdwijnt als het substantief een meervoudsvorm is of als de determinator onbepaald of afwezig is. De vormen van het bezittelijk voornaamwoord, met uitzondering van *ons/onze*, drukken noch lexicaal noch inflectieel het genusonderscheid uit. We herinneren de lezer eraan dat lidwoorden, aanwijzende voornaamwoorden en bezittelijke voornaamwoorden in vroegere stadia van het Nederlands wel verbogen werden (zie paragrafen 2.2 en 2.3).
6. De sequenties *al de* en *al het* kunnen als aanknopingspunt voor deze interpretatie fungeren. We zijn er ons terdege van bewust dat de vergelijking ietwat mank loopt (zie ook ANS 1997: 350-351). Een complicerende factor is de vaststelling dat *alle* met een meervoudig telbaar substantief gecombineerd wordt, zoals in (i) en (ii), of met een niet-telbaar enkelvoudig substantief, zoals in (iii) en (iv):
  - (i) a. alle/al de mooi-e mannen  
b. \*alle/al de mooi-e man
  - (ii) a. alle/al de mooi-e paarden  
b. \*alle/al de mooi(-e) paard
  - (iii) alle/al de lekker-e melk
  - (iv) alle/al het mooi(-e) vee

Rekening houdend met deze structurele dominantie van de verbogen vorm kan een paradigmatische nivellering binnen de verzameling van de elementen die op *alle* volgen niet bij voorbaat als verklaring worden uitgesloten. Het aantal attestaties van NP's met *alle* als determinator is helaas te schaars ( $n = 4$ ) om beide nivelleringshypotheses voor de adjectivische buiging bij *alle* aan een empirische toets te onderwerpen.

7. De stellende trap van *nader*, *na*, wordt enkel voor de kwalificatie van verwantschappen gebruikt, zoals in *een naë bloedverwant*. Onze materiaalverzameling bevat geen enkele instantie van deze toepassing. De stellende trap van *eerder*, *eer*, wordt in het hedendaags Nederlands niet meer als adjectief gebruikt.
8. Bij de gelexicaliseerde verkleinwoorden treffen we één onverbogen adjectief aan tegen 36 verbogen vormen.
9. Het totale aantal verbogen adjectieven in tabel 8(7) ( $n = 289$ ) is niet gelijk aan het aantal verbogen vormen bij een diminutief in tabel 8(6) ( $n = 290$ ). Dit komt doordat het lemma van één diminutiefvorm in de materiaalverzameling een bigenerisch substantief is waarvoor het verschil tussen beide genera niet semantisch gemotiveerd is: *het rodekoolkleurige colbertje*.
10. Hoewel ook het werkwoord *leven* bestaat, hebben we de voorkomens van *leven* in de kernpositie als substantieven gecodeerd.
11. In tabel 8(10) brengen we geen onderscheid aan tussen de semantisch gemotiveerde en de niet-gemotiveerde bigenerische substantieven. De reden is dat beide subtypes geen significant verschillende invloed op de responsvariabele hebben ( $\chi^2 = 0.0015$ ,  $df = 1$ ,  $p = 0.9694$ ).
12. We kunnen deze verklaring verifiëren door het percentage onverbogen adjectieven in het buigingsprofiel van een bigenerisch substantief te correleren met het percentage onzijdige realisaties van het bigenerische substantief in kwestie. Onze stelling voorspelt dan een positieve correlatie tussen beide percentages: hoe vaker een bigenerisch substantief als *het*-woord wordt gebruikt, des te groter de structurele kans op het gebruik van een onverbogen adjectief, des te groter de kans op onverbogen adjectieven in het buigingsprofiel van het bigenerische substantief, en des te groter de kans dat de onverbogen vorm effectief geselecteerd wordt. De frequenties in het CGN zijn helaas ontoereikend om deze stelling te falsificeren.

## **Hoofdstuk 9:**

# **Fonologische conditionering I – segmentele fonologie**

Het gebruik van gesproken materiaal biedt een uitgelezen kans om de fonologische conditionering van de adjectivische buigingsalternantie bij neutra in kaart te brengen. Bestaand empirisch onderzoek maakt namelijk hoofdzakelijk gebruik van geschreven bronnen: Degryse (1959) gebruikt literaire bronnen; Lebrun & Schurmans-Swillen (1966) raadplegen krantenmateriaal; het door De Rooij (1980a) geraadpleegde corpus bevat tot slot slechts een klein deel gesproken materiaal.

We verdelen de fonologische factoren in twee groepen: in dit hoofdstuk behandelen we de factoren die tot de segmentele fonologie behoren; in het volgende hoofdstuk komen de suprasegmentele of de prosodische restricties op het gebruik van de adjectivische buigingsvarianten aan bod.

We belichten eerst de operationalisering van de hypothesen met betrekking tot de segmentele kenmerken van het adjectief en het substantief (9.1). Daarna confronteren we deze hypothesen met de empirische realiteit van het corpus (9.2).

### **9.1 Hypothesen en operationalisering**

Bij de analyse van de segmenteel-fonologische condities kijken we naar de segmentele structuur aan de linkerwoordrand van het substantief, met name de aard van het beginsegment, en de segmentele structuur aan de rechterwoordrand van het adjectief, meer bepaald de sonoriteit van de eindmedeklinker van de onverbogen adjectivische basisvorm.

Op basis van het bestaande empirisch onderzoek (Degryse 1959; Lebrun & Schurmans-Swillen 1966; de Rooij 1980a) hebben we de hypothese geformuleerd dat het onverbogen adjectief vaker geselecteerd wordt vóór een substantief met een klinker als beginfoneem dan vóór een substantief met een medeklinker als beginfoneem (zie paragraaf 1.1.2.1).

We breiden het fonologische onderzoek uit met de segmentele kenmerken van de adjectivische eindmedeklinker, meer bepaald de sonoriteit van de adjectivische eindmedeklinker.<sup>1</sup> We onderzoeken of de sonoriteit van de adjectivische eindmedeklinker de selectie van de buigings-*e* bevordert. Een medeklinker met een grotere sonoriteit kan de (articulatorische) realisatie van de buigings-*e* vergemakkelijken door ze voor te bereiden of zelfs ‘voor een deel’ te realiseren.<sup>2</sup> Om de empirische realiteit van dit vermoeden te testen, maken we gebruik van de sonoriteitshiërarchie.



Dit is een rangorde van de medeklinkers in functie van hun sonoriteit (Booij 1995: 25; Kooij & van Oostendorp 2003: 48-51). Tabel 9(1) stelt de sonoriteitshiërarchie voor en berekent de sonoriteitsindex van iedere medeklinker in functie van zijn kenmerkenstructuur: de sonoriteitsindex van een medeklinker is gelijk aan het aantal plusjes bij de sonoriteitskenmerken.

	Steml. obstruent	Stemh. obstruent	Nasaal	Glijklank	Vloeiklank
[vocoid]	-	-	-	-	+
[approximant]	-	-	-	+	+
[sonorant]	-	-	+	+	+
[stemhebbend]	-	+	+	+	+
<b>Sonoriteitsindex</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

*Tabel 9(1): Sonoriteitshiërarchie van de medeklinkers  
(aanvulling bij de tabel overgenomen uit Booij 1995: 25)*

De berekening van de sonoriteitsindex in tabel 9(1) is gebaseerd op het model van Booij (1995: tabel 3.1) dat gebruik maakt van drie parameters: [vocoid], [approximant] en [sonorant].<sup>3</sup> Om de stemhebbende en de stemloze obstruenten te onderscheiden, voegen we hier het kenmerk [stemhebbend] aan toe, dat enkel een negatieve waarde heeft bij de stemloze obstruenten. De onderscheiding tussen stemhebbende en stemloze eindobstruenten stoelt op het eindfoneem van de adjectivische basisvorm (*rond* /rʔnd/ vs. *bont* /bʔnt/). Deze aanvulling op Booij's model resulteert in tabel 9(1) in een rangorde met vijf sonoriteitsgraden variërend van '0' (afwezigheid van sonoriteit) tot en met '4' (maximale sonoriteit).

## 9.2 Resultaten en bespreking

We rapporteren achtereenvolgens het buigingseffect van de substantivische beginklank (9.2.1) en de sonoriteit van de adjectivische eindconsonant (9.2.2).

### 9.2.1 Substantivisch beginsegment

Tabel 9(2) vat het buigingseffect van de beginklank van het substantief samen. De frequenties zijn conform met de hypothese: vóór een substantief dat met een klinker begint, wordt de onverbogen buigingsvariant significant vaker geselecteerd ( $\chi^2 = 50.5109$ ,  $df = 1$ ,  $p = 1.185e-12$ ). Vóór een substantief met een klinker als beginklank bedragen de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bijna het dubbele van de odds vóór een substantief dat met een medeklinker begint (OR = 1.9193 [1.6019,2.2996]). De kans op een onverbogen vorm groeit met 60% (RR<sub>onverbogen</sub> = 1.61 [1.4168,1.8091]). Gegeven het gebruik van gesproken materiaal is het significante effect van de hiaatvermijding niet verwonderlijk: *ease-of-articulation*-effecten, zoals het vermijden van een hiaat, manifesteren zich bij uitstek in de gesproken taal.

	medeklinker		klinker		$\Sigma$	
<b>verbogen</b>	3,406	78.37	404	65.37	3,810	76.75
<b>onverbogen</b>	940	21.63	214	34.63	1,154	23.25
<b><math>\Sigma</math></b>	<b>4,346</b>	<b>100.0</b>	<b>618</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 9(2): Invloed van de beginklank van het substantief op de adjectivische buigingsalternantie

De deletie van de prevocalische sjwa is één van de twee strategieën die in het Nederlands worden aangewend om een hiaat te vermijden (Booij 1995).<sup>4</sup> Deze fonologische regel opereert niet alleen op woordniveau maar ook op het niveau van de prosodische constituent, zoals blijkt uit de volgende voorbeelden waar de eindsjwa van het bepaald lidwoord (1a) en de adjectivische buigingsuitgang (1b) wegvallen vóór een substantief dat met een klinker begint (Booij 1995: 150-151).

- (1) a. d(e) avond [dav|ɛnt]  
 b. rod(e) aardbeien [roda|ɛrbɛij★]  
 (voorbeelden overgenomen uit Booij 1995: 150)

Booij (1995: 151) typeert de sjwa-deletie in deze voorbeelden als instanties van “casual speech”. Zoals aangekondigd in het literatuuroverzicht (zie paragrafen 1.1.2.1 en 1.2.1), zullen we nagaan of de hiaatvermijding zwaarder doorweegt in informele dan in formele contexten (zie paragraaf 12.3.3.2).

### 9.2.2 Adjectivisch eindsegment

We confronteren nu het vermoeden dat de sonoriteit van de adjectivische eindmedeklinker de selectie van de buigings-*e* bevordert met de corpusgegevens. De gegevens in tabel 9(3) ondersteunen de hypothese: het aandeel onverbogen vormen vermindert gradueel naarmate de sonoriteitsindex van de eindmedeklinker stijgt ( $\chi^2 = 129.6125$ ,  $df = 4$ ,  $p < 2.2e-16$ ), met uitzondering van een niet-significante knik tussen de sonoriteitsindexen ‘2’ en ‘3’ ( $\chi^2 = 0.4053$ ,  $df = 1$ ,  $p = 0.5244$ ).

	sonor=0		sonor=1		sonor=2		sonor=3		sonor=4	
<b>verb</b>	1,310	68.95	1,014	79.65	264	82.50	968	80.73	254	93.38
<b>onverb</b>	590	31.05	259	20.35	56	17.50	231	19.27	18	6.62
<b><math>\Sigma</math></b>	<b>1,900</b>	<b>100.0</b>	<b>1,273</b>	<b>100.0</b>	<b>320</b>	<b>100.0</b>	<b>1,199</b>	<b>100.0</b>	<b>272</b>	<b>100.0</b>

Tabel 9(3): Invloed van de sonoriteitsindex van de adjectivische eindmedeklinker op de adjectivische buigingsalternantie

De resultaten van de partiële  $\chi^2$  in tabel 9(4) tonen aan dat de sonoriteitschaal gereduceerd kan worden tot een ternaire oppositie, stemloze obstruent (sonoriteitsindex = ‘0’) vs. sonore medeklinker (‘0’ < sonoriteitsindex < ‘4’) vs. semivocaal (sonoriteitsindex = ‘4’), of zelfs tot een binaire oppositie, stemloze

obstruent (sonoriteitsindex = ‘0’) vs. sonore medeklinker (sonoriteitsindex  $\geq$  ‘1’). In vergelijking met een stemloze eindobstruent, halveren de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bij de adjectieven waar de eindmedeklinker een minimale sonoriteit vertoont (sonoriteitsindex  $\geq$  ‘1’ vs. sonoriteitsindex = ‘0’): OR = 0.5009 [0.4384,0.5724]. De kans op een verbogen vorm stijgt bij een sonore eindmedeklinker met 18% in vergelijking met een niet-sonore eindmedeklinker, c.q. een stemloze eindobstruent:  $RR_{\text{verbogen}} = 1.1834$  [1.1432,1.225].

	sonor=0	sonor=1	sonor=2	sonor=3	sonor=4
<b>verbogen</b>	[0.86,0.95]	[0.99,1.10]	[0.95,1.21]	[0.98,1.10]	[1.06,1.37]
<b>onverbogen</b>	[1.21,1.42]	[0.76,0.96]	[0.55,0.93]	[0.77,0.99]	[0.16,0.45]

Tabel 9(4): 95% BI'en partiële  $\chi^2$  voor de invloed van de sonoriteitshierarchie op de buigingsalternantie

We zullen deze factor in de multivariate analyse dan ook als een binaire factor operationaliseren: niet-sonore eindmedeklinker tegenover sonore eindmedeklinker (zie paragraaf 14.1, tabel 14(1)).

### 9.3 Conclusie

In dit hoofdstuk hebben we de conditionering van de buigingsalternantie door de segmentele en fonotactische eigenschappen van het substantivische beginfoneem en het adjectivische eindfoneem bestudeerd. Wat het beginsegment van het substantief betreft, ondersteunen de empirische gegevens de hypothese dat de buigings-*e* sneller wordt weggelaten vóór een substantief met een beginklinker dan vóór een substantief met een beginmedeklinker om een hiaat te vermijden.

Wanneer we de sonoriteit van de adjectivische eindmedeklinker beschouwen, stellen we vast dat adjectieven die niet op een stemloze obstruent eindigen een inhiberende context vormen voor het gebruik van de onverbogen vorm. We hebben geopperd dat dit mogelijk het gevolg is van articulatorische anticipatie: bij eindmedeklinkers die ten minste één sonoor kenmerk vertonen, bereidt het sonore karakter van de medeklinker de realisatie van de buigings-*e* voor. Uiteraard moet deze aanzet tot verklaring aan de hand van experimenteel onderzoek getest worden.

## Noten

1. De adjectieven die op een klinker eindigen, zijn niet in dit onderzoek opgenomen omwille van de onverbuigbaarheid van deze adjectieven (zie paragraaf 3.2 en paragraaf 6.2.2, tabel 6(3)).
2. Het Frans, waar in tegenstelling tot het Nederlands de eindmedeklinkers niet verdoft worden, kent de zogenaamde *détente consonantique*. Dit is de realisatie van een facultatieve niet-syllabische sjwa [ʎ] aan de rechterwoordrand als gevolg van de sonore uitspraak van de eindmedeklinker, zoals in *bague* [bagʎ], *aime* [aɛmʎ], *quatre* [katrʎ] of *bille* [bijʎ] (Debrock & Mertens 1993: 48).
3. Booij (1995: 25) beschrijft de eerste twee kenmerken als volgt: “The feature [vocoid] characterizes vowels and glides, and the feature [approximant] stands for ‘an articulation in which one articulator is close to another, but without the vocal tract being narrowed to such an extent that a turbulent airstream is produced’”. Het kenmerk [sonorant] onderscheidt de obstruenten van de andere medeklinkers.
4. Het alternatief, de insertie van een homorganische glijklank, is enkel mogelijk indien het linkse segment een hoge vocaal is, zoals de [e] in *zeearend* [zejarʎnt] (Booij 1995: 67).

# Hoofdstuk 10:

## Fonologische conditionering II – prosodische fonologie

Zoals we al in het vorige hoofdstuk hebben aangegeven, vormt het gebruik van gesproken materiaal een uitgelezen mogelijkheid om de prosodische conditionering van de adjectivische buigingsalternantie te bestuderen. We beperken ons dus niet tot de falsificatie van de hypothesen die we op basis van de bestaande literatuur geformuleerd hebben, maar we gaan daarnaast ook op zoek naar algemene prosodische restricties op het voorkomen van beide buigingsvarianten.

We moeten steeds voor ogen houden dat vanuit een prosodisch standpunt het verbuigen van een adjectief neerkomt op het toevoegen van een doffe lettergreep aan de rechterwoordrand van het adjectief.

### 10.1 Hypothesen en operationalisering

We zullen eerst de hypothesen uit de verkennende literatuurstudie hernemen en het algemene analytische en interpretatieve kader voor het prosodische onderzoek schetsen (10.1.1). Daarna bespreken we de operationalisering van de hypothesen en de onderzoeksdoelstellingen samen met de prosodische codering van de observaties in de materiaalverzameling (10.1.2).

#### 10.1.1 *Onderzoeksdoelstellingen en onderzoekskader*

Vertrekkend van de ‘ritmische’ of ‘eufonische’ hypothesen in de literatuur (zie paragraaf 1.1.2.2) bespreken we de uitbreiding van het analytische en interpretatieve kader van het prosodische onderzoek. Op basis van de bestaande literatuur schuiven we de volgende drie hypothesen naar voor:

- de lengte van het adjectief: De keuze van de buigings-*e* wordt minder waarschijnlijk naarmate het adjectief langer wordt. Omdat de literatuur vaag is over de concrete lengte die de selectie van de onverbogen vorm faciliteert, hebben we in het literatuuroverzicht als voorlopige vuistregel voorgesteld om de adjectieven op te splitsen in de een- en tweelettergrepige adjectieven enerzijds en de meerlettergrepige adjectieven anderzijds (ANS 1997: 411; Booij 1992: 129-130; Blom 1994: 82; Broekhuis 1999: 216; Degryse 1959: 73-84; Lebrun & Schurmans-Swillen 1966: 178-179, 184-186; Raidt 1968: 99-101; de Rooij 1980a: 14, 16-17, 20-21; Royen 1953: 97-113).

- de accentuering van de adjectivische eindlettergreep: Wanneer het adjectief op een doffe of een onbeklemtoonde lettergreep eindigt, neemt de kans op een onverbogen vorm toe omdat zo de opeenvolging van meerdere doffe of onbeklemtoonde lettergrepen kan worden vermeden. Bij adjectieven met een beklemtoonde eindlettergreep komt de verbogen variant vaker voor dan gemiddeld (ANS 1997; Booij 1992: 129; De Vooys 1967: 63; Lebrun & Schurmans-Swillen 1966; Raidt 1968; de Rooij 1980a).
- de accentuering van de substantivische beginlettergreep: De invloed van de beklemtoning van de adjectivische eindlettergreep wordt in de literatuur – sporadisch – geprojecteerd op de beginlettergreep van het substantief: een substantief met een onbeklemtoonde beginlettergreep vergemakkelijkt het gebruik van de onverbogen buigingsvariant (Overdiep 1937: 259).

Tot slot wordt er in de literatuur ook op een mogelijke interactie tussen de accentuering van de adjectivische eindlettergreep en de woordsoort van de determinator gewezen (ANS 1997: 411; Raidt 1968: 100): het faciliterende effect van de onbeklemtoonde adjectivische eindlettergreep ten aanzien van de selectie van de onverbogen vorm neemt toe wanneer een bezittelijk voornaamwoord de determinatorpositie bezet (zie ook tabel 1(2)). Het bezittelijk voornaamwoord fungeert met andere woorden als een katalysator voor de selectie van het onverbogen buigingsmorfeem bij een adjectief met een onbeklemtoonde eindlettergreep.

Zoals aangekondigd in de inleiding van dit hoofdstuk is de prosodische analyse niet beperkt tot de falsificatie van bovenvermelde hypothesen. We stellen twee extensies voor ten aanzien van het bestaande onderzoek, zoals samengevat in het literatuuroverzicht (zie paragraaf 1.1.2.2): een uitbreiding van het onderzoeksdomein en een uitbreiding van het interpretatieve kader.

<b>Prosodisch domein</b>	<b>Taalkundig niveau</b>
Lettergreep	Lexicon
Voet	
Prosodisch woord	
Clitische groep	Syntaxis
Fonologische constituent	
Intonatieve constituent	
Uiting	

*Tabel 10(1): Prosodische hiërarchie  
(overgenomen en vertaald uit Nespor & Vogel 1986: 16)*

De eerste uitbreiding ten aanzien van het bestaande onderzoek betreft het prosodische domein dat de basis van de analyse vormt. Alvorens deze uitbreiding toe te lichten, overlopen we de verschillende prosodische domeinen die doorgaans in

de literatuur onderscheiden worden. Deze domeinen worden geordend in de prosodische hiërarchie, zoals geschematiseerd in tabel 10(1) (Booij 1995: 143-145; Nespor & Vogel 1986): de lettergreep vormt de kleinste prosodische eenheid; de uiting (“utterance”) is de grootste eenheid. De eerste drie eenheden in tabel 10(1), met name de lettergreep, de metrische voet en het prosodische woord, behoren tot het lexicon; de overige eenheden worden gecreëerd door de toepassing van syntactische regels.

We bespreken nu de domeinen in tabel 10(1) die relevant zijn voor onderhavig onderzoek naar de prosodische restricties op de buigingsalternantie:

- *lettergreep* ( $\sigma$ ): De lettergreep is de kleinste eenheid in de prosodische hiërarchie (Nespor & Vogel 1986: 61-83). Het woord *vegetarisch* kunnen we in vier lettergrepen splitsen: [ve] <sub>$\sigma$</sub> [ge] <sub>$\sigma$</sub> [ta] <sub>$\sigma$</sub> [risch] <sub>$\sigma$</sub> . Voor een beschrijving van de structuur van de lettergreep in het Nederlands verwijzen we naar Booij (1995: 22-52) en Kooij & van Oostendorp (2003: 35-54).
- *voet* ( $v$ ): De voet is een aaneenschakeling van lettergrepen en vormt het domein van de klemtoontoekenning (Nespor & Vogel 1986: 83-103). Het Nederlands heeft een sterke voorkeur voor tweelettergrepige voeten, waarbij de eerste lettergreep de klemtoon draagt: [[vè] <sub>$\sigma$</sub> [ge] <sub>$\sigma$</sub> ] <sub>$v$</sub> [[tá] <sub>$\sigma$</sub> [risch] <sub>$\sigma$</sub> ] <sub>$v$</sub>  (cf. *infra*).<sup>1</sup>
- *prosodisch woord* ( $\omega$ ): Het samenvoegen van voeten vindt plaats op het niveau van het prosodische woord. Alle voeten en lettergrepen worden in deze structuur opgenomen: [[[vè] <sub>$\sigma$</sub> [ge] <sub>$\sigma$</sub> ] <sub>$v$</sub> [[tá] <sub>$\sigma$</sub> [risch] <sub>$\sigma$</sub> ] <sub>$v$</sub> ] <sub>$\omega$</sub>  (Nespor & Vogel 1986: 109-144). Vanaf dit niveau in de prosodische hiërarchie is er een lexicale inhoud aanwezig.
- *fonologische constituent* ( $\phi$ ): Voor de definitie van een “phonological phrase” stelt Booij (1995: 143-144) dat “it is usually assumed that each word of a lexical category forms a phonological phrase with either the preceding or the following function words”. Bij Nespor & Vogel (1986: 165-186) valt de fonologische constituent grofweg samen met de syntactische constituentstructuur. Relevant voor onderhavig onderzoek is de fonologische constituent gevormd door de opeenvolging van het adjectief en het substantief, zoals *vegetarisch menu*: [[vegetarisch] <sub>$\omega$</sub> [menu] <sub>$\omega$</sub> ] <sub>$\phi$</sub> .

Een blik op bovenvermelde ‘eufonische’ of ‘ritmische’ hypothesen leert ons dat de prosodische conditionering van de buigingsalternantie in het bestaande onderzoek tot het syllabische niveau beperkt is. Enkel de laatste lettergreep van het adjectief en – zij het sporadisch – de eerste lettergreep van het substantief worden afzonderlijk in beschouwing genomen. Wij breiden het prosodische domein in twee stappen uit: eerst naar het niveau van de metrische voet, vervolgens naar het niveau van de fonologische constituent.

De eerste extensie vergroot het prosodische domein van de lettergreep tot de voetstructuur. We kijken meer bepaald naar de adjectivische voet die gevormd wordt door de lettergrepen vanaf de meest rechtse adjectivische klemtoon tot en met de

laatste lettergreep van het adjectief. We maken hier geen onderscheid tussen de hoofd- en de nevenklemtoon. De tweede extensie van het prosodische domein brengt ons naar het niveau van de fonologische constituent die gevormd wordt door de adjectief-nomensequentie. We nemen niet de structuur van de volledige fonologische constituent in beschouwing, maar enkel het patroon dat gevormd wordt door de lettergrepen vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon.

We verduidelijken deze dubbele extensie van het prosodische domein aan de hand van een voorbeeld. In het bestaande onderzoek, en dan vooral het corpus-linguïstische onderzoek (Degryse 1959: 73-84; Lebrun & Schurmans-Swillen 1966: 178-179, 184-186; de Rooij 1980a: 14, 16-17), kijkt men alleen naar de accentuering van de eindlettergreep van de adjectivische basisvorm, aangeduid door de onderstreepte lettergrepen in (1) en (2). Dit heeft als gevolg dat er geen verschil is tussen de voorbeelden in (1) en (2), hoewel *vegetárisch* in (1) op één onbeklemtoonde lettergreep eindigt en *dágelijks* in (2) op een sequentie van twee onbeklemtoonde lettergrepen.

- (1) het vègetárisch(e) menu : [vè]<sub>σ</sub>[ge]<sub>σ</sub>[tá]<sub>σ</sub>[risch]<sub>σ</sub>  
 (2) het dágelijks(e) woordgebruik : [dá]<sub>σ</sub>[ge]<sub>σ</sub>[lijks]<sub>σ</sub>

De uitbreiding van het prosodische domein naar de meest rechtse voet van het adjectief maakt het mogelijk om het prosodische onderscheid tussen (1) en (2) te analyseren. De prosodische structuur vanaf de meest rechts gesitueerde klemtoon bestaat voor *vègetárisch* uit één beklemtoonde lettergreep gevolgd door één onbeklemtoonde lettergreep en voor *dágelijks* uit de opeenvolging van één beklemtoonde en twee onbeklemtoonde lettergrepen, zoals de onderstreepte lettergrepen in (1') en (2') aantonen. Bij *dágelijks* is de laatste lettergreep 'voetloos', dit wil zeggen dat hij in geen enkele metrische voet is opgenomen. Dit komt door de sterke voorkeur van het Nederlands voor tweelettergrepige voeten (cf. *infra*).

- (1') het vègetárisch(e) menu : [[vè]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[tá]<sub>σ</sub>[risch]<sub>σ</sub>]<sub>v</sub>  
 (2') het dágelijks(e) woordgebruik : [[dá]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[lijks]<sub>σ</sub>]<sub>σ</sub>

De tweede extensie van het prosodische domein leidt ons naar het niveau van de fonologische constituent. We nemen de fonologische constituent niet in zijn geheel in de analyse op, maar we bestuderen enkel het patroon dat gevormd wordt door de lettergrepen vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon. De onderstreepte lettergrepen in (1'') en (2'') stellen dit patroon voor.

- (1'') het vègetárisch(e) menú: [[[vè]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[tá]<sub>σ</sub>[risch]<sub>σ</sub>]<sub>v</sub>[[me]<sub>σ</sub>[nú]<sub>σ</sub>]<sub>v</sub>]<sub>φ</sub>  
 (2'') het dágelijks(e) wóórdgebruik: [[[dá]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[lijks]<sub>σ</sub>]<sub>σ</sub>[[wóórd]<sub>σ</sub>]<sub>v</sub>[ge]<sub>σ</sub>[bruik]<sub>σ</sub>]<sub>φ</sub>



Deze dubbele extensie van het prosodische domein biedt de mogelijkheid om de prosodische patronen in kaart te brengen die de keuze van het buigingsmorfeem bepalen (zie paragraaf 10.2.2).

De tweede uitbreiding ten aanzien van het bestaande onderzoek heeft betrekking op de interpretatie van de waargenomen prosodische patronen. We willen niet alleen de patronen in kaart brengen die het gebruik van beide buigingsvarianten bepalen, maar ook de algemene prosodische restricties identificeren die aan de grondslag van de geobserveerde patronen liggen. Hiervoor maken we gebruik van de inzichten in de prosodische of metrische fonologie.<sup>2</sup> We richten ons hierbij vooral op de bevindingen van het prosodische onderzoek in het kader van *Optimality Theory* (OT). We wensen onmiddellijk een mogelijk misverstand uit de weg te ruimen: we maken enkel gebruik van de principes die in het OT-onderzoek ter verklaring van prosodische patronen worden aangehaald; het is geenszins onze bedoeling een analyse in de trant van OT uit te voeren.

Om de prosodische restricties beter te begrijpen, bespreken we bondig de basiskenmerken van OT. Hiervoor baseren we ons op Booij (1998, 2002b) en Zonneveld (2003). Kenmerkend voor OT is dat niet langer het derivatieproces van een vorm middels een reeks vormelijke regels centraal staat, maar wel het verklaren van de uiteindelijke vorm. Hiervoor reikt OT twee concepten aan: “generator” (GEN) en “evaluator” (EV). GEN genereert alle mogelijke vormen die EV op zijn beurt evalueert aan de hand van een taalspecifieke ordening van universele maar overtreedbare condities. Met andere woorden: GEN genereert een reeks kandidaten waaruit EV de meest geschikte kiest, dit wil zeggen de optimale vorm. Dit is de kandidaat die het geringste aantal prosodische condities schendt.

Booij (1998, 2002b) vermeldt de volgende restricties op de structuur van een prosodisch woord in het Nederlands.

- “Trochee”: De eerste lettergreep in de voet is het hoofd, i.e. de sterke lettergreep. In *vegetarisch* is de eerste lettergreep van iedere voet de sterke lettergreep, zoals aangeduid door de onderstreepte lettergrepen: [[vè]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[tá]<sub>σ</sub>[risch]<sub>σ</sub>]<sub>v</sub>. Deze conditie wordt overtreden door *menu* of *alarm*, waar telkens de tweede lettergreep in de voet de beklemtoonde lettergreep is: [[me]<sub>σ</sub>[nú]<sub>σ</sub>]<sub>v</sub>, [[a]<sub>σ</sub>[lár]<sub>σ</sub>]<sub>v</sub>.
- “FootMin”: Een voet is minimaal tweelettergrepig. Het prosodische woord *vegetarisch* bestaat uit twee voeten, die ieder twee lettergrepen bevatten: [[vè]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[tá]<sub>σ</sub>[risch]<sub>σ</sub>]<sub>v</sub>. Deze conditie wordt geschonden door *bier*, dat slechts uit één eenlettergrepige voet bestaat: [[bíer]<sub>σ</sub>]<sub>v</sub>.
- “ParseSyllable” (*ParseSyl*): Lettergrepen worden in de voetstructuur van een woord opgenomen. Dit criterium illustreren we aan de hand van het onderscheid tussen *vegetarisch* en *dagelijks*. Terwijl in *vegetarisch* alle lettergrepen tot een voet behoren, [[vè]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[tá]<sub>σ</sub>[risch]<sub>σ</sub>]<sub>v</sub>, is de eindlettergreep *lijks* van *dagelijks* voetloos, i.e. hij behoort tot geen enkele voet:

[[dá]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[lijks]<sub>σ</sub>]. Anders gezegd: *dagelijks* vormt een inbreuk op de conditie *ParseSyl*.

- “FootMax”: Een voet is maximaal tweelettergrep. In de hiërarchie van de prosodische restricties is deze conditie hoger gerangschikt dan *ParseSyl* (zie (3) *infra*). Deze classificatie verklaart waarom *dagelijks* niet [[dá]<sub>σ</sub>[ge]<sub>σ</sub>[[lijks]<sub>σ</sub>]<sub>v</sub>, met een schending van *FootMax*, maar [[dá]<sub>σ</sub>[ge]<sub>σ</sub>]<sub>v</sub>[[lijks]<sub>σ</sub>, met een schending van het lager gerangschikte *ParseSyl*, als prosodische structuur heeft.
- “SuperH”: Superzware (*SuperHeavy*) lettergrepen (i.e. lettergrepen met drie posities in het rijm) vormen het hoofd van een voet. Een woord als *alarm* vormt geen goede trochee omdat de superzware lettergreep *larm* (*arm* is een rijm met drie posities) de klemtoon draagt: [[a]<sub>σ</sub>[[lár<sub>m</sub>]<sub>σ</sub>]<sub>v</sub>.
- “Morphological Identity” (*MorphI*): Een morfeem moet identificeerbaar zijn in de uiteindelijke vorm. Deze conditie heeft betrekking op Booij’s gevalstudie die we dadelijk bespreken.

De aangehaalde voorbeelden tonen aan dat deze condities geschonden kunnen worden. Doordat de condities niet absoluut geldend zijn, leent het OT-model zich voor de beschrijving van variatie. Hoe zwaar de schending van een conditie doorweegt bij de uiteindelijke selectie van de optimale vorm hangt af van de taalspecifieke hiërarchie. Booij (1998, 2002b) stelt voor het Nederlands de hiërarchie in (3) voor. Een hogere positie in de hiërarchie wordt door het symbool ‘>>’ aangeduid.

(3) MorphI >> SuperH >> Trochee >> FootMax >> ParseSyl >> FootMin

We illustreren het prosodische onderzoek in de trant van OT aan de hand van Booij’s analyse van de prosodische conditionering bij de verbuiging van de adjectieven op *-en*, zoals *open* (Booij 1998, 2002b). Booij verklaart de onmogelijkheid om adjectieven op *-en* te verbuigen (zie ook paragraaf 3.2) als de realisatie van een optimale prosodische vorm. In OT gebruikt men een zogenaamd tableau om de rangorde van de condities (kolommen) uit te zetten tegen de verschillende vormen (rijen) die GEN genereert. Het symbool ‘ $\emptyset$ ’ selecteert de optimale vorm, ‘\*’ identificeert een (lichte) schending, en ‘\*!’ duidt een zware schending aan die de vorm in kwestie definitief als optimale vorm uitsluit (en verdere controle van de andere condities overbodig maakt).

	MorphI	SuperH	FootMax	ParseSyl	FootMin
(o:p★) <sub>v</sub> (n-★) <sub>σ</sub>				*!	
$\emptyset$ (o:p★n- $\emptyset$ ) <sub>v</sub>					

Tabel 10(2): OT-tableau voor open- $\emptyset$   
(overgenomen uit Booij 1998: 154)

Tabel 10(2) toont de meerwaarde van de OT-benadering aan. Uit het tableau kan worden afgeleid waarom de onverbogen vorm *open-∅* en niet de verbogen pendant *open-e* de optimale vorm is: in *open-e* is de eindlettergreep voetloos (cf. ‘\*!’ bij *ParseSyl*), terwijl *open-∅* geen enkele conditie overtreedt. Merk wel op dat een onderliggend nul-morfeem (∅) geponeerd moet worden.<sup>3</sup> Tabel 10(3) verklaart waarom *groen-e* en niet *groen-∅* de verbogen vorm van *groen* is: *groen-∅* vormt geen binaire voet (cf. ‘\*!’ bij ‘FootMin’), terwijl *groen-e* geen enkele conditie schendt.

	MorphI	SuperH	FootMax	ParseSyl	FootMin
$\emptyset$ (( $\uparrow$ ru:) $_{\sigma}$ (n- ★) $_{\sigma}$ ) $_V$					
(( $\uparrow$ ru:n-∅) $_{\sigma}$ ) $_V$					*!

Tabel 10(3): OT-tableau voor groen-e  
(overgenomen uit Booij 1998: 154)

De condities die tot nu toe aan bod zijn gekomen, opereren op het niveau van het prosodische woord. Hieraan moeten we een bijkomende restrictie toevoegen voor het niveau van de fonologische constituent:

- “StressClash”: Een opeenvolging van twee beklemtoonde lettergrepen en de daaruit voortvloeiende klemtoonbotsing of “stress clash” moet worden vermeden (Booij 1995: 161-164; Kager & Visch 1988; Kooij & van Oostendorp 2003: 126-127; Schultink 1977, 1980; Trommelen & Zonneveld 1989: 237-254). Om de opeenvolging van twee (primaire) beklemtoonde lettergrepen te vermijden, kan de hoofdklemtoon van *feodáál* in *feodáál stélsel* naar de eerste lettergreep verschuiven, *féodaal stélsel*.

We zullen de algemene prosodische condities uit OT aanwenden om de geobserveerde prosodische patronen te interpreteren en te verklaren. Van de hiërarchisering van de condities (zoals in (3)) en de daarmee samenhangende modelbouw zullen we echter om de volgende redenen geen gebruik maken. Ten eerste, het is geenszins onze bedoeling een model in de trant van OT te construeren voor de modellering van de prosodische factoren die een rol spelen bij de buigingsalternantie. We willen enkel de geobserveerde prosodische tendenties verklaren, een doelstelling waarvoor het interpretatieve kader van OT geschikt is gezien zijn primair verklarende oriëntatie. Ten tweede, de OT-benadering werkt voor regulariseerbare uitzonderingen, zoals de onverbuigbaarheid van de adjectieven op *-en* (Booij 1998, 2002b). De adjectieven op *-en* vormen immers een homogene deelverzameling of subregelmatigheid: de attributieve adjectieven op *-en* worden onder geen beding verbogen, *\*de open-e deur* (zie ook paragraaf 3.2). Anders gezegd: de (niet-)verbuiging van de adjectieven op *-en* vormt een discreet fenomeen. De basisregel stelt dat het attributieve adjectief verbogen wordt, maar voor (alle) adjectieven op *-en* geldt dat deze vormen altijd onverbogen blijven (om een optimaal

prosodisch patroon te realiseren). Dit is niet het geval bij de adjectivische buigingsalternantie bij neutra, waar het aandeel van beide buigingsvarianten toeneemt of afneemt in een bepaalde prosodische configuratie. In onderhavig onderzoek zijn de vastgestelde effecten dan ook niet discreet maar gradueel van aard, waarbij de uiteindelijke vorm niet altijd de optimale vorm is in functie van een tableau, zoals in tabel 10(2) en tabel 10(3).<sup>4</sup>

In de neerlandistiek is de prosodische conditionering van morfosyntactische alternanties al meermaals bestudeerd. Van der Hulst & Kooij (1998) formuleren metrische condities op het gebruik van de suffixen *-s* en *-en* bij de meervoudsvorming van inheemse ongelede substantieven. Trommelen (1983) beschrijft de metrische conditionering van de keuze van het diminueringsuffix. Booij (1998, 2002b) verklaart de onverbuigbaarheid van adjectieven op *-en* als een prosodische conditionering van de uiteindelijke vorm (cf. *supra*). Deze alternanties zijn echter allemaal discreet: een bepaald lemma kiest op basis van zijn prosodische configuratie consistent één morfosyntactische variant. Als voorbeeld van kwantitatief variationistisch onderzoek, dat weliswaar niet OT-gebaseerd is, naar de metrische conditionering van een graduele morfosyntactische alternantie verwijzen we naar Mondorfs (2003) analyse van het gebruik van de synthetische en analytische vormen van de vergrotende en de overtreffende trap in het Engels.

### 10.1.2 Operationalisering: codering van de prosodische patronen

De operationalisering van de hypotheses en van de prosodische analyse wordt samengevat door de volgende vragen:

- Welke informatie hebben we nodig om bovenvermelde doelstellingen te realiseren?
- Bestaat er een objectiveerbare procedure om deze informatie toe te kennen aan de elementen die de adjectivische positie en de nominale kernpositie in de materiaalverzameling bezetten?

Het antwoord op de eerste vraag is eenvoudig. Voor ieder adjectief en substantief hebben we de volgende informatie nodig:

- de syllabificatie: De syllabificatie verdeelt het prosodische woord in lettergrepen. Deze informatie is nodig om de lengte van het adjectief te berekenen en vormt daarnaast de structurele basis van het klemtoonpatroon. De syllabische structuur van *vegetarisch* is  $[[ve]_{\sigma}[ge]_{\sigma}[ta]_{\sigma}[risch]_{\sigma}]_{\omega}$ . Op basis van deze structuur kunnen we de lengte meten:  $n(\sigma) = 4$ . Daarnaast vormen deze vier lettergrepen de eenheden waaraan de klankkwaliteit wordt toegekend (cf. *infra*).
- het klemtoonpatroon: Het klemtoonpatroon identificeert de opeenvolging van beklemtoonde en onbeklemtoonde lettergrepen. Het klemtoonpatroon van *vegetarisch* is  $[vè]_{\sigma}[ge]_{\sigma}[tá]_{\sigma}[risch]_{\sigma}$ , dit is: '(secundair) beklemtoond – onbeklemtoond - (primair) beklemtoond - onbeklemtoond'. De belangrijkste functie van de secundaire klemtoon is het garanderen van een afwisseling

tussen beklemtoonde en onbeklemtoonde lettergrepen (“Optimal Grid Principle”; Booij 1995: 99). Dit afwisselende klemtoonpatroon ligt aan de basis van de voetstructuur: door de nevenklemtoon op de beginlettergreep bestaat *vègetárisch* uit twee trocheïsche voeten,  $[[vè]_{\sigma}[ge]_{\sigma}]_v[[tá]_{\sigma}[risch]_{\sigma}]_v$ .

Het antwoord op de tweede vraag ligt minder voor de hand. De identificatie van het klemtoonpatroon van de elementen die de adjectivische en de substantivische positie bezetten, vormt het centrale probleem. De toekenning van de hoofdklemtoon aan een Nederlands woord is een complex samenspel van verschillende principes, zoals de morfologische geleedheid van het woord, de onderliggende lettergreepstructuur en het onderscheid tussen inheemse en uitheemse woorden.<sup>5</sup> Hoewel een aantal generalisaties kunnen worden voorgesteld, besluit Booij (1995: 100) dat “many cases of unpredictable stress remain”.

Omdat uitzonderingen het risico op foutieve en inconsistente annotaties vergroten, hebben we besloten de prosodische informatie te gebruiken die de basis vormt van de uitspraakmodule in de *Elektronische Grote Van Dale* (EGVD, Geeraerts 2000).<sup>6</sup> De prosodische informatie in de EGVD kent aan ieder lemma een prosodische code toe. Dit is een sequentie bestaande uit de cijfers ‘1’, ‘2’ en ‘0’ (Heemskerk 2001). Deze prosodische codes vervullen een dubbele functie. De eerste functie is het syllabificeren van het lemma: ieder cijfer in de code staat voor één lettergreep. De tweede functie is de identificatie van het klemtoonpatroon van het lemma: het cijfer ‘1’ identificeert de lettergreep met de hoofdklemtoon, het cijfer ‘2’ de lettergreep(en) met een nevenklemtoon en het cijfer ‘0’ de onbeklemtoonde lettergrepen, ongeacht of ze al dan niet beklemtoonbaar zijn. De prosodische codes uit de EGVD verstrekken dus de nodige informatie voor het realiseren van de onderzoeksdoelstellingen. De volgende voorbeelden illustreren de dubbele functie van de prosodische codes (die we steeds tussen enkele aanhalingstekens plaatsen).

- (4)
- a. *vègetárisch* : ‘2010’
  - b. *símpel* : ‘10’
  - c. *domeín* : ‘01’
  - d. *ónderwèrp* : ‘102’

De prosodische code die in (4a) aan *vegetarisch* is toegekend, identificeert dit lemma als vierlettergrepig en situeert de hoofdklemtoon op de voorlaatste lettergreep en de nevenklemtoon op de eerste lettergreep. De tweede en de laatste lettergreep zijn als onbeklemtoond gecodeerd. Aan de elementen in de adjectivische positie wordt enkel het klemtoonpatroon van de (onverbogen) basisvorm toegekend omdat we op basis van de prosodische kenmerken van de adjectivische basisvorm de selectie van het buigingsmorfeem willen beschrijven.

Na het toekennen van de klemtoonpatronen uit de EGVD aan de elementen die de adjectivische en de nominale positie in de materiaalverzameling bezetten, resten er zes problemen met betrekking tot de prosodische annotatie:

- de woordvormen waarvoor er geen klemtoonpatroon in de EGVD voorzien is
- de woordvormen waarvan het klemtoonpatroon variatie vertoont in functie van de syntactische context
- de woordvormen waarvoor de EGVD meerdere prosodische patronen voorziet
- de woordvormen met een optionele sjwadeletie
- de differentiëring van de onbeklemtoonde lettergrepen aan de rechterrاند van het adjectief: onbeklemtoond tegenover onbeklemtoonbaar/dof (c.q. een lettergreep met [★] als vocalisch element)
- het aanbrengen van een abstracte prosodische code

In de rest van deze paragraaf reiken we een oplossing aan voor deze annotatieproblemen en op het einde vatten we het prosodische annotatiesysteem samen in tabel 10(4) en tabel 10(5).

In eerste instantie zijn er de vormen waarvoor de EGVD geen klemtoonpatroon voorziet. Tot deze groep behoren enerzijds een aantal gelegenheidssamenstellingen en -afleidingen en anderzijds de geïnfecteerde adjectivische, nominale en verbale vormen. Voor de gelegenheidssamenstellingen voegen we de prosodische informatie van de samenstellende delen samen en vervolgens kennen we de samenstellingsklemtoon toe. De regel voor de samenstellingsklemtoon plaatst de hoofdklemtoon op het linkerlid in een samenstelling als het hoofd van de samenstelling een substantief (5a) of een werkwoord (5b) is, en op het rechterlid als het hoofd een adjectief (5c) of een voorzetsel is (5d) (Kooij & van Oostendorp 2003: 125):

- (5)      a. táfelpòòt  
           b. óverzètten  
           c. hèlbláuw  
           d. hògeróp

In (6) wordt de primaire klemtoon van *proces* tot een secundaire klemtoon herleid omdat in de substantivische samenstelling het eerste lid, *onderwijs*, de samenstellingsklemtoon draagt.

- (6)      onderwijsproces : onderwijs + proces<sup>7</sup>  
                               : ‘102’    + ‘0~~1~~<sub>2</sub>’  
                               : ‘10202’

In een geprefigeerd adjectief, zoals *superkatholiek* in (7), draagt het kernwoord, *katholiek*, de hoofdklemtoon waardoor de primaire klemtoon van het prefix *super-* tot een secundaire klemtoon herleid wordt. Merk op dat prefixen een apart prosodisch woord vormen in het Nederlands, zodat de hoofdklemtoon van een geprefigeerd woord aan de hand van de regel voor de samenstellingsklemtoon wordt toegekend (Booij 1995: 47-52, 120-123; Kooij & van Oostendorp 2003: 120-121).<sup>8</sup>

- (7) superkatholiek : super + katholiek  
 : '1<sub>2</sub>0' + '201'  
 : '20201'

De tweede groep vormen waarvoor de prosodische code in de EGVD ontbreekt, zijn de geïnflecteerde adjectivische (vergroten trap), nominale (verkleinwoord) en verbale vormen (onvoltooid deelwoord, voltooid deelwoord). Op basis van het klemtoonpatroon van het lemma kan het patroon van de geïnflecteerde vorm relatief eenvoudig gegenereerd worden:

- (8) a. vergroten trap : *kleiner* : *klein* + *er*  
 : '1' + '0'  
 : '10'
- b. verkleinwoord: *bedrijffe* : *bedrijf* + *je*  
 : '01' + '0'  
 : '010'
- c. onvoltooid deelwoord : *onderwijzend* : *onderwijz* + *end*  
 : '201' + '0'  
 : '2010'
- d. voltooid deelwoord: *opgehoopt* : *ophopen* + *ge ... t*  
 : '120' + '0'  
 : '1020'

Naast de vormen waarvoor er geen klemtoonpatroon voorhanden is, zijn er in tweede instantie de vormen waarvan het klemtoonpatroon in functie van de syntactische functie kan wijzigen en waarvoor de EGVD slechts één van beide klemtoonpatronen geeft. Dit is het geval bij de geprefigeerde adjectieven, de samengestelde adjectieven en de samenstellende afleidingen. In de prosodische literatuur spreekt men van klemtoonverschuiving ("Stress Retraction"): in predicatieve positie valt de hoofdklemtoon op het rechterlid in een samenstelling en een samenstellende afleiding, en op de stam in een geprefigeerd woord (9a, 10a, 11a), zoals voorzien in de regel voor de samenstellingsklemtoon (cf. *supra*); wanneer samengestelde en geprefigeerde adjectieven in attributieve positie gebruikt worden, dragen respectievelijk het linkerlid en het prefix de klemtoon (om een klemtoonbotsing te vermijden) (9b, 10b, 11b) (ANS 1997: 698; Booij 1995; Gussenhoven 1983a, 1983b; Schultink 1977, 1980; Trommelen & Zonneveld 1989).<sup>9</sup> De prosodische informatie in de EGVD bevat in de meeste gevallen alleen de patronen voor de predicatieve context (9a, 10a, 11a).

- (9) a. Het water is ijskoud.  
 b. Het ijskoude water.
- (10) a. Die behandeling is onheus.  
 b. De onheuse behandeling.
- (11) a. de strijd is ondergronds

## b. de óndergróndse strijd

In de literatuur bestaat geen eenduidigheid over het verplichte karakter van deze klemtoonverschuiving, vooral als het geprefigeerde adjectieven betreft;<sup>10</sup> over het verplichte karakter van de klemtoonverschuiving bij samenstellingen bestaat er meer eensgezindheid.<sup>11</sup> Rekening houdend met de vaagheid waarin de vakliteratuur zich hult, hebben we alle observaties in de materiaalverzameling waar een geprefigeerd adjectief<sup>12</sup>, een samengesteld adjectief of een samenstellende afleiding de adjectivische positie bezet in de geluidsbestanden van het CGN opgezocht om te controleren welke lettergreep de hoofdklemtoon draagt.<sup>13</sup> In sommige gevallen vermeldt de EGVD zowel het predicatieve als het attributieve patroon. Ook voor deze vormen is het klemtoonpatroon in de geluidsbestanden gecontroleerd.

- (12) a. overgroot: ‘201’<sub>predicatief</sub> vs. ‘102’<sub>attributief</sub>  
 b. Oost-Vlaams: ‘21’<sub>predicatief</sub> vs. ‘12’<sub>attributief</sub>

Daarnaast is er nog een tweede type klemtoonverschuiving, die Gussenhoven (1983a, 1983b) “Rhetoric Retraction” noemt en die kan optreden bij alle (andere) adjectieven in attributieve positie waar een klemtoonbotsing optreedt tussen de adjectivische en de substantivische klemtoon, zoals in *feodáál stélsel*. Over het optionele karakter van dit type klemtoonverschuiving bestaat er eensgezindheid in de vakliteratuur (Booij 1995; Gussenhoven 1983a, 1983b; Trommelen & Zonneveld 1989). Deze klemtoonverschuiving wordt geconditioneerd door retorische factoren: Gussenhoven (1983b) heeft experimenteel aangetoond dat dit type klemtoonverschuiving vooral optreedt in sterk persuasieve contexten. We zullen dit type klemtoonverschuiving bijgevolg niet coderen in de gegevensbank, zodat in deze gevallen het ‘predicatieve’ klemtoonpatroon van de EGVD gehandhaafd blijft.

De derde verfijning die we aanbrengen bij de EGVD-klemtoonpatronen is het desambigueren van de lemmata met meerdere klemtoonpatronen. Deze ambigue gevallen kunnen in verschillende subtypes worden ingedeeld: de verschillende klemtoonpatronen zijn regionaal bepaald (13), de verschillende klemtoonpatronen zijn morfosyntactisch bepaald (14), de verschillende klemtoonpatronen vallen samen met het onderscheid tussen een glijklank en een vocaal (15), de verschillende klemtoonpatronen worden bepaald door de optionele deletie van een lettergreep (16) en de restgroep (17):

- (13) a. asfalt: ‘10’<sub>Nederland</sub> vs. ‘01’<sub>België</sub>  
 b. onderwijs: ‘102’<sub>Nederland</sub> vs. ‘201’<sub>België</sub>  
 (14) uniform: ‘100’<sub>substantief</sub> vs. ‘201’<sub>adjectief</sub>  
 (15) a. regionaal: ‘2001’<sub>vocaal</sub> vs. ‘201’<sub>glijklank</sub>  
 b. Rwandees: ‘2001’<sub>vocaal</sub> vs. ‘201’<sub>glijklank</sub>  
 (16) a. regime: ‘010’ vs. ‘01’  
 b. vormeloos: ‘102’ vs. ‘12’  
 (17) a. algemeen: ‘100’ vs. ‘201’



## b. maximaal: ‘100’ vs. ‘102’

Deze gevallen zijn eveneens op basis van de CGN-geluidsbestanden gedesambigüeerd. De instanties in (15) beschouwen we als een glijklank, behalve als er duidelijk een lettergreep hoorbaar is op de positie waar we de glijklank verwachten. Bij twijfel is het woord in kwestie in drie afzonderlijke sessies beluisterd en is de annotatie geselecteerd die minstens tweemaal de voorkeur wegdraagt.

Een vierde instantie waar de EGVD-klemtoonpatronen om manuele controle vragen, zijn de contexten waar de sjwa optioneel gedeleerd kan worden. Booij (1995: 128-130) en Van Oostendorp (2000: 200-202) definiëren deze contexten als volgt:

$$(18) \quad ([+obstruent] \_ )_{\sigma}([glijklank] \star)_{\sigma}$$

Op basis van voorbeelden zoals *vorm(★)l★k* hebben we deze context uitgebreid door de conditie [+obstruent] te generaliseren tot alle consonanten. Hierbij moet worden opgemerkt dat de contexten die in aanmerking komen voor sjwadeletie niet beperkt zijn tot de gevallen waar de tweede sjwa tot de stam behoort, zoals in *vorm(★)l★k* of *dod(★)l★k*. We nemen eveneens de contexten in beschouwing waar het buigingsmorfeem de tweede sjwa is, zoals in *and(★)r-★* of *simp(★)l-★*. Alle adjectieven waar de sjwa volgens deze criteria gedeleerd kan worden, zijn in de CGN-geluidsbestanden beluisterd. Als criterium voor het deleren van een sjwa hebben we de afwezigheid van de lettergreep met de deleerbare sjwa vooropgesteld: enkel indien de sequentie die in bovenvermelde voorbeelden tussen haakjes staat niet als een aparte lettergreep in de geluidsbestanden kan worden geïdentificeerd, beschouwen we ze als gedeleerd. Uiteraard zijn er meerdere twijfelgevallen waar het onduidelijk is of er al dan niet sprake is van een lettergreep. Deze gevallen hebben we tijdens drie afzonderlijke sessies beluisterd en vervolgens hebben we het alternatief gekozen dat minstens tweemaal de voorkeur geniet.

De vijfde aanpassing gaat over de onbeklemtoonde lettergrepen (EGVD-code ‘0’). We hebben de onbeklemtoonde lettergrepen aan de rechterrands van het adjectief opgesplitst in de onbeklemtoonbare of doffe lettergrepen (i.e. de lettergrepen met [★] als vocalisch element) aan de ene kant en de sjwaloze onbeklemtoonde lettergrepen (i.e. de andere onbeklemtoonde lettergrepen) aan de andere kant. Hiervoor hebben we de EGVD-code ‘0’ verfijnd tot de tegenstelling tussen ‘0’ voor onbeklemtoonbaar/dof en ‘3’ voor onbeklemtoond. Voor deze verfijning maken we gebruik van de fonetische transcripties in de EGVD. De prosodische code voor *algemeen* verandert van ‘100’ in ‘103’, terwijl de prosodische code van *mannelijk* ‘100’ blijft. Voor de overige lettergrepen is deze aanpassing overbodig, rekening houdend met de focus op het prosodische patroon aan de rechterwoordsrand van het adjectief (zie paragraaf 10.1.1).

De zesde en laatste aanpassing bestaat uit het aanbrengen van een abstract klemtoonpatroon dat de basis vormt van de prosodische analyse. We transformeren

de concrete vierdeling ‘1’ vs. ‘2’ vs. ‘3’ vs. ‘0’ in een binaire oppositie ‘s’ (beklemtoond) vs. ‘w’ (onbeklemtoond), waarbij het onderscheid tussen de hoofd- en de nevenklemtoon en het onderscheid tussen de onbeklemtoonde en de doffe lettergrepen verdwijnt.<sup>14</sup> De voorbeelden in (4) krijgen bijgevolg de volgende abstracte prosodische structuur toegewezen:

- (4) a. *símpel* : ‘10’ ; ‘sw’  
 b. *vègetárisch* : ‘2010’ ; ‘swsw’  
 c. *domeín* : ‘01’ ; ‘ws’  
 d. *ónderwèrp* : ‘102’ ; ‘sws’

De concrete klemtoonpatronen worden gebruikt om de abstracte patronen intern te differentiëren.

Tabel 10(4) vat het prosodische annotatiesysteem samen:

	<b>abstract</b>	<b>concreet</b>
<b>beklemtoond</b>	‘s’ (beklemtoond)	‘1’ (hoofdklemtoon) ‘2’ (nevenklemtoon)
<b>onbeklemtoond</b>	‘w’ (onbeklemtoond)	‘3’ (beklemtoonbaar) ‘0’ (onbeklemtoonbaar/dof)

*Tabel 10(4): Overzicht van de gebruikte prosodische codes*

We verduidelijken aan de hand van een voorbeeld hoe de prosodische informatie aan de observaties is toegekend, en hoe deze informatie gebruikt wordt bij de verificatie van de ‘eufonische’ of ‘ritmische’ hypothesen (zie paragraaf 10.2.1) en bij de prosodische analyse (zie paragraaf 10.2.2). We hernemen het voorbeeld uit (1):

- (1) het *vègetárisch(e) menú*

Aan de *basisvorm* van het adjectief en het substantief zijn de volgende concrete en abstracte prosodische patronen toegekend:<sup>15</sup>

- *vègetárisch* : ‘2013’ ; ‘swsw’
- *menú* : ‘01’ ; ‘ws’

Deze codes vormen het vertrekpunt van de prosodische analyse. Voor de verificatie van de ‘eufonische’ of ‘ritmische’ hypothesen hebben we de volgende informatie nodig (de concrete klemtoonpatronen staan tussen haakjes):

- de lengte van het adjectief *vegetarisch*, uitgedrukt in lettergrepen:  $n(\sigma) = 4$
- de beklemtoning van de eindlettergreep van het adjectief *vegetarisch*: ‘w’ (‘3’)
- de beklemtoning van de beginlettergreep van het substantief *menu*: ‘w’ (‘0’)

Voor de identificatie van de prosodische patronen komt daar de volgende informatie bij:

- het prosodische patroon vanaf de meest rechtse klemtoon tot en met de eindlettergreep van de (onverbogen) basisvorm van het adjectief *vegetarisch*: ‘sw’ (‘13’)
- het prosodische patroon vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon in de fonologische constituent *vegetarisch menu*, met aanduiding van de woordgrens door middel van een verticale streep ‘|’: ‘sw|ws’ (‘13|01’)

Tabel 10(5) vat de toegekende prosodische codes voor *vegetarisch menu* samen:

Prosodische informatie	<i>vegetarisch menu</i>	
Lengte A	4	-
Accentuering eindlettergreep A / beginlettergreep N	‘w’ (‘3’)	‘w’ (‘0’)
Prosodisch patroon meest rechtse voet A	‘sw’ (‘13’)	-
Prosodisch patroon fonologische constituent AN	‘sw ws’ (‘13 01’)	

*Tabel 10(5): Samenvatting van de prosodische annotatie van het adjectief-nomenpaar vegetarisch menu*

## 10.2 Resultaten en bespreking

De bespreking van de resultaten van de prosodische analyse valt uiteen in drie delen. Eerst confronteren we de drie ‘eufonische’ of ‘ritmische’ hypothesen met de empirische realiteit van het CGN (10.2.1). Vervolgens bestuderen we de algemene prosodische restricties op de selectie van beide buigingsmorfemen (10.2.2). Tot slot wordt de potentiële partiële conditionering van enkele prosodische factoren getest in een stratumanalyse (10.2.3); we beperken de stratumanalyse niet tot de interactie tussen de accentuering van de adjectivische eindlettergreep en de woordsoort van de determinator (zie paragraaf 10.1.1), maar we gaan ook na in welke mate het buigingseffect van de adjectivische lengtegraad een neveneffect is van de aanwezigheid van prosodische configuraties die het gebruik van één van beide buigingsvarianten faciliteren.

### 10.2.1 Falsificatie van de ritmische en eufonische hypothesen

#### *Lengte van het adjectief*

De eerste hypothese die we testen, is de voorspelde stijging van het aandeel onverbogen adjectieven naarmate de adjectivische basisvorm langer is. Tabel 10(6) toont de verdeling van beide buigingsvarianten over de verschillende adjectivische lengtegraden. Bij de onderscheiden adjectivische lengtegraden in tabel 10(6) horen twee inleidende opmerkingen. Ten eerste, de adjectivische lengtegraad in de kolomvariabele is de lengte van de adjectivische basisvorm, dit wil zeggen de vorm van het adjectief zonder het buigingsmorfeem. Ten tweede, alle adjectieven langer dan twee lettergrepen worden onder de noemer ‘meerlettergrepig’ gegroepeerd ( $3 \leq$

$n(\sigma) \leq 7$ ). De verdeling van beide buigingsvarianten vertoont geen significante verschillen voor de lengtegraden in deze groep (*Fisher Exact*-test:  $p = 0.14$ ).

De driedeling zoals voorgesteld in tabel 10(6) heeft een significante impact op de buigingsalternantie ( $\chi^2 = 411.9672$ ,  $df = 2$ ,  $p < 2.2e-16$ ). We stellen vast dat het aandeel van de onverbogen variant progressief toeneemt naarmate het adjectief langer wordt.<sup>16</sup>

	$n(\sigma) = 1$		$n(\sigma) = 2$		$n(\sigma) \geq 3$		$\Sigma$	
<b>verbogen</b>	1,698	91.14	1,076	80.78	1,036	58.56	3,810	76.75
<b>onverbogen</b>	165	8.86	256	19.22	733	41.44	1,154	23.25
<b><math>\Sigma</math></b>	<b>1,863</b>	<b>100.0</b>	<b>1,332</b>	<b>100.0</b>	<b>1,769</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 10(6): Invloed van de lengte van de adjectivische basisvorm op de adjectivische buigingsalternantie

Voor de berekening van het  $RR_{\text{onverbogen}}$  en de OR nemen we de tweelettergrepige adjectieven als vergelijkingspunt: enerzijds bestaat het prototypische Nederlandse woord uit één tweelettergrepige voet (Kooij & van Oostendorp 2003: 81-84); anderzijds benadert de verdeling van beide buigingsvarianten voor de tweelettergrepige adjectieven het meest de marginale verdeling. Dus, het  $RR_{\text{onverbogen}}$  voor ' $n(\sigma) = 1$ ' vergelijkt het aandeel onverbogen vormen voor ' $n(\sigma) = 1$ ' met het referentiepunt ' $n(\sigma) = 2$ '.

	$RR_{\text{onverbogen}}$	95% BI	OR	95% BI
$n(\sigma) = 1$	0.4608	[0.3839,0.5531]	0.4084	[0.3311,0.5039]
$n(\sigma) \geq 3$	2.1559	[1.9059,2.4387]	2.9738	[2.5192,3.5105]

Tabel 10(7):  $RR_{\text{onverbogen}}$  en OR voor de lengte van de adjectivische basisvorm met de tweelettergrepige adjectieven als referentiepunt

De BI'en bij de RR's en de OR's in tabel 10(7) identificeren de respectievelijke verschillen tussen de een- en de tweelettergrepige adjectieven en tussen de twee- en de meerlettergrepige adjectieven als significant. Hoewel het aandeel onverbogen vormen zowel bij de een- als bij de tweelettergrepige adjectieven lager is dan het algemeen gemiddelde, suggereren de gegevens in tabel 10(7) dat het niet opportuun is de driedeling tot een binaire oppositie te herleiden: de eenlettergrepige adjectieven selecteren de buigings-*e* significant vaker dan de tweelettergrepige adjectieven; de meerlettergrepige adjectieven kiezen daarentegen significant vaker de onverbogen vorm.<sup>17</sup>

#### *Accentuering adjectivische eindlettergreep*

Na de analyse van het effect van de adjectivische lengtegraad komen nu de twee accentueringshypotheses aan bod. Deze hypothesen voorspellen een groter aandeel onverbogen vormen als de eindlettergreep van het adjectief en de beginlettergreep van het substantief onbeklemtoond zijn. We vergelijken de frequentieverdeling van

beide buigingsvarianten in de deelverzamelingen gedefinieerd door een beklemtoonde en een onbeklemtoonde adjectivische eindlettergreep (tabel 10(8)) en in de deelverzamelingen gedefinieerd door een beklemtoonde en een onbeklemtoonde substantivische beginlettergreep (tabel 10(11)).

De gegevens in tabel 10(8) tonen aan dat een adjectief significant vaker de buigings-*e* selecteert wanneer de eindlettergreep van de basisvorm beklemtoond is dan wanneer hij onbeklemtoond is ( $\chi^2 = 284.8479$ ,  $df = 1$ ,  $p < 2.2e-16$ ). Vergeleken met de adjectieven eindigend op een beklemtoonde lettergreep zijn de odds  $\frac{\text{onverbogen}}{\text{verbogen}} 3.15 [2.749, 3.6087]$  maal groter bij de adjectieven eindigend op een onbeklemtoonde lettergreep. De kans dat een adjectief de buigings-*e* niet selecteert, bedraagt meer dan het dubbele bij een onbeklemtoonde eindlettergreep dan bij een beklemtoonde eindlettergreep:  $RR_{\text{onverbogen}} = 2.34 [2.115, 2.579]$ . De corpusgegevens zijn dus in overeenstemming met de hypothese die het effect van de beklemtoning van de adjectivische eindlettergreep beschrijft. Dit betekent dat de comparatieven, die de onverbogen vorm niet significant vaker selecteren dan de vorm van de stellende trap (zie paragraaf 8.2), een uitzondering vormen binnen de groep met de adjectieven met een onbeklemtoonde eindlettergreep.

	beklemtoond		onbeklemtoond		$\Sigma$	
<b>verbogen</b>	2,807	83.79	1,003	62.14	3,810	76.75
<b>onverbogen</b>	543	16.21	611	37.86	1,154	23.25
<b><math>\Sigma</math></b>	<b>3,350</b>	<b>100.0</b>	<b>1,614</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 10(8): Invloed van de beklemtoning van de adjectivische eindlettergreep op de adjectivische buigingsalternantie

De voorkeur voor vormen die op een tweevoetige trochee eindigen, verklaart de sterke tendens om een adjectief met een beklemtoonde eindlettergreep te verbuigen (zie paragraaf 10.1.1): door de toevoeging van de doffe buigings-*e* ('w<sub>★</sub>'), wordt een tweelettergrepige trocheïsche voet gevormd aan de rechterraand van het adjectief, '[sw<sub>★</sub>]<sub>v</sub>'. Hetzelfde principe verklaart het grotere aandeel onverbogen vormen bij de adjectieven waarvan de eindlettergreep onbeklemtoond is: door de toevoeging van de buigings-*e* ontstaat een sequentie van minimaal twee onbeklemtoonde lettergrepen ('[sw]<sub>v</sub>w<sub>★</sub>' of '[sw]<sub>v</sub>ww<sub>★</sub>'), wat in strijd is met de prosodische conditie dat alle lettergrepen tot een (tweelettergrepige) metrische voet behoren (*ParseSyl*). Niettemin is er in deze groep een duidelijke voorkeur voor de verbogen vorm. Hierop komen we terug tijdens de prosodische analyse, wanneer het prosodische domein van de lettergreep naar de metrische voet wordt uitgebreid (zie paragraaf 10.2.2).

De informatie in tabel 10(8) kan op twee manieren verfijnd worden: in de eerste plaats kunnen we de beklemtoonde eindlettergrepen opsplitsen naargelang ze de hoofd- of de nevenklemtoon dragen (tabel 10(9)); daarnaast kunnen we de onbeklemtoonde eindlettergrepen in twee groepen indelen in functie van de

beklemtoonbaarheid (tabel 10(10)). We formuleren de volgende hypothesen met betrekking tot de invloed van deze concrete klemtoonpatronen op de buigingsalternantie:

- Wanneer we ons toespitsen op de adjectieven met een beklemtoonde eindlettergreep is de kans dat de buigings-*e* geselecteerd wordt groter bij een primair dan bij een secundair beklemtoonde eindlettergreep.
- Wanneer we ons beperken tot de adjectieven met een onbeklemtoonde eindlettergreep is de kans dat het onverbogen buigingsmorfeem gekozen wordt groter bij een eindlettergreep met een sjwa als klinker dan bij een sjwaloze eindlettergreep.

We kijken eerst naar het buigingseffect van het onderscheid tussen de hoofd- en de nevenklemtoon in tabel 10(9). Deze gegevens bevestigen de hypothese ( $\chi^2 = 27.201$ ,  $df = 1$ ,  $p = 1.834e-07$ ).

	nevenklemtoon		hoofdklemtoon		$\Sigma$	
<b>verbogen</b>	293	74.55	2,514	85.02	2,807	83.79
<b>onverbogen</b>	100	25.45	443	14.98	543	16.21
<b><math>\Sigma</math></b>	<b>393</b>	<b>100.0</b>	<b>2,957</b>	<b>100.0</b>	<b>3,350</b>	<b>100.0</b>

Tabel 10(9): Invloed op de adjectivische buigingsalternantie van het onderscheid tussen primair en secundair beklemtoonde adjectivische eindlettergrepen

De buigings-*e* wordt 14% vaker geselecteerd bij een adjectief met een primair beklemtoonde eindlettergreep dan bij een adjectief met een secundair beklemtoonde eindlettergreep:  $RR_{\text{verbogen}} = 1.1404$  [1.0743,1.2105]. De OR die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vergelijkt bij de primair en de secundair beklemtoonde eindlettergrepen bedraagt 0.5163 [0.487, 0.7118].

Wanneer we naar de concrete realisatie van de onbeklemtoonde adjectivische eindlettergrepen kijken, stellen we niet het verwachte patroon vast: de onverbogen vorm komt significant vaker voor wanneer de onbeklemtoonde eindlettergreep geen sjwa bevat ('onbeklemtoond') dan wanneer ze wel een sjwa bevat ('onbeklemtoonbaar').

	onbeklemtoond		onbeklemtoonbaar		$\Sigma$	
<b>verbogen</b>	194	56.40	809	63.70	1,003	62.14
<b>onverbogen</b>	150	43.60	461	36.30	611	37.86
<b><math>\Sigma</math></b>	<b>344</b>	<b>100.0</b>	<b>1,270</b>	<b>100.0</b>	<b>1,614</b>	<b>100.0</b>

Tabel 10(10): Invloed op de adjectivische buigingsalternantie van het onderscheid tussen onbeklemtoonde en onbeklemtoonbare adjectivische eindlettergrepen

De verbogen vorm wordt 12% vaker geselecteerd na een doffe eindlettergreep (i.e. met de [★] als vocalisch element) dan na een andere onbeklemtoonde eindlettergreep:  $RR_{\text{verbogen}} = 1.1295$  [1.0202,1.2506]. De OR voor de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bedraagt 0.737 [0.5787,0.9386]. Niettemin is het aandeel onverbogen vormen bij de

adjectieven met een doffe eindlettergreep significant groter dan bij de adjectieven met een beklemtoonde eindlettergreep ( $\chi^2 = 217.3446$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

De verklaring voor het onverwacht grote aandeel onverbogen vormen bij de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep, moeten we wellicht bij de adjectieven op *-isch* ([is]) zoeken. Een aanzienlijk aantal van deze adjectieven komt voor in adjectief-nomensequenties met naamstatus, zoals de voorbeelden in (19) illustreren:

- (19) academisch ziekenhuis, Belgisch voorzitterschap (van de EU), biologisch ouderschap, economisch belang, Historisch Nieuwsblad, medisch onderzoek, medisch centrum, organisch materiaal, symfonisch orkest, theoretisch rijexamen, tropisch regenwoud, technisch onderwijs

We herinneren de lezer eraan dat het onverbogen buigingsmorfem algemeen beschouwd wordt als een markeerder van adjectief-nomenparen met naamstatus (zie paragraaf 1.1.3.1). Daarnaast zijn de adjectieven in (19) relationele adjectieven, waarmee eveneens een deflecterend effect geassocieerd wordt. Hierop komen we terug bij de analyse van de lexicale/semantische conditionering van de buigingsalternantie. De intuïtieve verklaring die we op basis van de voorbeelden in (19) hebben opgebouwd, wordt empirisch getoetst in paragraaf 11.2.3. Tot slot signaleren we dat De Rooij (1980a) een sterk deflecterend effect vaststelt bij de adjectieven op het suffix *-isch*.

#### *Accentuering substantivische beginlettergreep*

De tweede accentueringshypothese die we op basis van het verkennende literatuuroverzicht geformuleerd hebben, beschouwt een onbeklemtoonde substantivische beginlettergreep als een faciliterende factor voor de selectie van de onverbogen vorm. De gegevens in tabel 10(11) tonen aan dat de onverbogen vorm significant vaker voorkomt bij een substantief met een beklemtoonde beginlettergreep ( $\chi^2 = 24.7563$ ,  $df = 1$ ,  $p = 6.506e-07$ ).

	beklemtoond		onbeklemtoond		$\Sigma$	
<b>verbogen</b>	2,734	78.72	1,076	72.17	3,810	76.75
<b>onverbogen</b>	739	21.28	415	25.71	1,154	23.25
<b><math>\Sigma</math></b>	<b>3,473</b>	<b>100.0</b>	<b>1,491</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

*Tabel 10(11): Invloed van de beklemtoning van de substantivische beginlettergreep op de adjectivische buigingsalternantie*

Het buigingseffect van de accentuering van de substantivische beginlettergreep is kleiner dan dat van de accentuering van de adjectivische eindlettergreep:  $RR_{\text{onverbogen}} = 1.3081$  [1.1791,1.4511],  $OR = 1.4269$  [1.2412,1.6403]. Deze cijfers bedragen ongeveer de helft van hun tegenhangers op basis van de gegevens in tabel 10(8).

Met andere woorden: de impact van de beklemtoning is dubbel zo groot bij de adjectivische eindlettergreep als bij de substantivische beginlettergreep.

Wanneer we de concrete klemtoon invullen (i.e. ‘hoofdklemtoon’ vs. ‘nevenklemtoon’), verwachten we een sterkere neiging om de buigings-*e* te selecteren bij een primair beklemtoonde substantivische beginlettergreep dan bij een secundair beklemtoonde substantivische beginlettergreep. Tabel 10(12) bevat de buigingsprofielen voor de substantieven met een hoofd- en een nevenklemtoon op de beginlettergreep.

	nevenklemtoon		hoofdklemtoon		$\Sigma$	
<b>verbogen</b>	295	66.29	2,439	80.55	2,734	78.72
<b>onverbogen</b>	150	33.71	589	19.45	739	21.28
<b><math>\Sigma</math></b>	<b>445</b>	<b>100.0</b>	<b>3,028</b>	<b>100.0</b>	<b>3,473</b>	<b>100.0</b>

*Tabel 10(12): Invloed op de adjectivische buigingsalternantie van het onderscheid tussen primair en secundair beklemtoonde substantivische beginlettergrepen*

De frequentieverdeling in tabel 10(12) is significant ( $\chi^2 = 46.2264$ ,  $df = 1$ ,  $p = 1.053e-11$ ) en bevestigt de hypothese:  $RR_{\text{verbogen}} = 1.2150$  [1.1346,1.3012];  $OR = 0.4749$  [0.3826,0.5895]. Deze cijfers bewijzen dat de buigings-*e* significant vaker geselecteerd wordt vóór een substantief met een primair beklemtoonde beginlettergreep dan vóór een substantief met een secundair beklemtoonde beginlettergreep. Wanneer we het buigingsprofiel van de substantieven met een nevenklemtoon op de beginlettergreep in tabel 10(12) vergelijken met het buigingsprofiel van de substantieven met een onbeklemtoonde beginlettergreep in tabel 10(11) stellen we vast dat het aandeel onverbogen vormen significant groter is bij de substantieven met een secundair beklemtoonde beginlettergreep ( $\chi^2 = 45.9652$ ,  $df = 1$ ,  $p = 1.204e-11$ ;  $RR_{\text{onverbogen}} = 1.211$  [1.0384,1.4124];  $OR = 1.3183$  [1.0508,1.6541]). De verklaring voor het onverwacht grote aandeel onverbogen vormen bij de substantieven waarvan de beginlettergreep een nevenklemtoon draagt, is waarschijnlijk lexicaal of semantisch van aard: bij de attestaties met een onverbogen adjectief is er een aanzienlijk aantal observaties waarvan de adjectief-nomensequentie als een naam fungeert, zoals de voorbeelden in (20) illustreren.

- (20) Belgisch kampioenschap, Europees Parlement, gerechtelijk apparaat, hoger onderwijs, intern reglement, onderwijzend personeel, opvoedend personeel, uitvoerend comité, Vlaams Parlement, wetenschappelijk onderzoek

Deze intuïtieve verklaring op basis van de naamstatus van het adjectief-nomenpaar wordt in het volgende hoofdstuk geverifieerd door middel van een stratumanalyse. We vervangen dan de intuïtieve notie van adjectief-nomenpaar met naamstatus door een objectieverbaar alternatief, met name de lexicale collocabiliteit van het adjectief-nomenpaar (zie paragraaf 11.2.3).



*Bespreking*

In het eerste deel van de prosodische analyse hebben we de empirische validiteit aangetoond van de drie hypothesen die we op basis van het literatuuroverzicht geformuleerd hebben:

- De kans op een onverbogen adjectief daalt naarmate de basisvorm van het adjectief langer wordt. Hoewel de een- en tweelettergrepige adjectieven minder onverbogen vormen selecteren dan het algemeen gemiddelde, vertonen ze onderling significante verschillen ten aanzien van de verdeling van beide buigingsvarianten, zodat we een driedeling eenlettergrepig vs. tweelettergrepig vs. meerlettergrepig voorstellen.
- Zowel bij een onbeklemtoonde adjectivische eindlettergreep als bij een onbeklemtoonde substantivische beginlettergreep wordt de onverbogen vorm significant vaker geselecteerd dan bij de beklemtoonde tegenhanger. Bij de adjectieven is deze tendens dubbel zo groot als bij de substantieven.

Ondanks hun significante effect op de buigingsalternantie definiëren bovenstaande factoren geen enkele context met een absolute voorkeur voor de onverbogen variant. Dit is een empirisch argument voor de uitgebreide prosodische analyse in de volgende paragraaf. Daarnaast tonen de resultaten dat men de  $\chi^2$ -statistiek in grote materiaalverzamelingen met de nodige omzichtigheid moet gebruiken en zich niet mag blind staren op de zeer lage *p*-waarden (zie ook paragraaf 7.1).

**10.2.2 Prosodische conditionering van de buigingsalternantie**

Na het toetsen van de ‘ritmische’ of ‘eufonische’ hypothesen vatten we nu de eigenlijke prosodische analyse aan. Deze analyse breidt het prosodische domein uit van de lettergreep naar de metrische voet en de fonologische constituent om de algemene prosodische restricties op de buigingsalternantie te ontsluiten (zie paragraaf 10.1.1). Eerst identificeren we de prosodische restricties op het niveau van de metrische voet, c.q. het patroon aan de rechterwoordrand van het adjectief. Daarna kijken we naar de buigingseffecten op het niveau van de fonologische constituent, c.q. het patroon vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon.

*Structuur metrische voet*

We starten de prosodische analyse met de extensie van het prosodische domein van de lettergreep naar de metrische voet, meer bepaald het prosodische patroon gevormd door de lettergrepen vanaf de meest rechtse adjectivische klemtoon, ongeacht of het om een hoofd- of een nevenklemtoon gaat. We herhalen dat deze patronen de structuur van de (onverbogen) basisvorm van het adjectief voorstellen: omdat we op zoek zijn naar de prosodische restricties op de buigingsalternantie, bestuderen we hoe de structuur van de adjectivische basisvorm de selectie van het buigingsmorfeem bepaalt. Dus, om bijvoorbeeld de prosodische structuur te

verkrijgen van de verbogen vorm van een adjectivische basisvorm met ‘sw’ als rechterwoordrand, zoals *vorig*, moeten we een ‘w’ toevoegen voor de doffe lettergreep met de buigings-*e*, zoals in *vorig-e*: ‘sww<sub>★</sub>’ (‘w<sub>★</sub>’ staat voor de doffe lettergreep met de buigings-*e*).

De materiaalverzameling op basis van het CGN bevat de volgende prosodische patronen aan de adjectivische rechterwoordrand. De relevante lettergrepen zijn onderstreept.

- ‘s’, zoals in *gróót*, *héél*, *internationáál*, *rimpellòòs*, *Vlááms*, maar ook *ánd(e)r*, *dónk(e)r* (waar de haakjes de gedeleerde sjwa aanduiden; zie paragraaf 10.1.2)
- ‘sw’, zoals in *Ántwerps*, *Bélgisch*, *gelijknámig*, *voormálig*, *vórig*, maar ook *duíd(e)lijk*, *záng(e)rig* (waar de haakjes de gedeleerde sjwa aanduiden)
- ‘sww’, zoals in *ákelig*, *álgemeen*, *dágelijks*, *lichámelijk*, *stóffelijk*

In paragraaf 10.2.1 hebben we al vastgesteld dat een beklemtoonde eindlettergreep (‘s’ als prosodisch patroon) een inhiberende factor is voor het gebruik van de onverbogen vorm (zie tabel 10(8)). Voor de patronen ‘sw’ en ‘sww’ kunnen we vooropstellen dat de keuze van de onverbogen vorm waarschijnlijker wordt naarmate het aantal onbeklemtoonde lettergrepen aan de rechterwoordrand groter wordt, zodat we de volgende hiërarchie voor de selectie van de onverbogen vorm verkrijgen:<sup>18</sup>

$$(21) \quad 's' < 'sw' < 'sww'$$

Tabel 10(13) vat de buigingsprofielen voor de prosodische patronen aan de rechterrands van (de basisvorm van) het adjectief samen.

	's'		'sw'		'sww'		Σ	
<b>verbogen</b>	2,807	83.79	884	69.12	119	35.52	3,810	76.75
<b>onverbogen</b>	543	16.21	395	30.88	216	64.48	1,154	23.25
<b>Σ</b>	<b>3,350</b>	<b>100.0</b>	<b>1,279</b>	<b>100.0</b>	<b>335</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 10(13): Invloed op de adjectivische buigingsalternantie van het prosodische patroon aan de rechterwoordrand van het adjectief

De gradatie die we in (21) hebben voorgesteld, doorstaat de toetsing met de gegevens in tabel 10(13): het aandeel onverbogen vormen stijgt gestaag naarmate het aantal onbeklemtoonde lettergrepen aan de rechterwoordrand van het adjectief toeneemt. Wanneer het prosodische patroon aan de rechterwoordrand van het adjectief ‘sww’ is, stellen we daarenboven een absolute voorkeur voor de onverbogen vorm vast. De distributie in tabel 10(13) is significant ( $\chi^2 = 453.9676$ ,  $df = 2$ ,  $p < 2.2e-16$ ). Voor de berekening van het  $RR_{\text{onverbogen}}$  en de OR, nemen we als ijkpunt de adjectieven met ‘s’ als rechterwoordrand.

	<b>RR<sub>onv</sub></b>	<b>95% BI</b>	<b>OR</b>	<b>95% BI</b>
‘sw’	1.9053	[1.7027,2.1321]	2.3099	[1.9880,2.6838]
‘sww’	3.9779	[3.5612,4.4434]	9.3832	[7.3671,11.951]

Tabel 10(14): *RR<sub>onverbogen</sub> en OR voor de prosodische patronen aan de rechterwoordrand van het adjectief met het patroon ‘s’ als referentiepunt*

De tabellen 10(13) en 10(14) bewijzen de descriptieve meerwaarde van de metrische voet aan de rechterwoordrand van het adjectief: het  $RR_{onverbogen}$  en de OR voor het patroon ‘sww’ zijn respectievelijk tweemaal en viermaal groter dan hun tegenhangers voor het patroon ‘sw’. Bovendien bedraagt de OR voor ‘sww’ het drievoud van de OR die de beklemtoonde en onbeklemtoonde eindlettergrepen vergelijkt (zie paragraaf 10.2.1, tabel 10(8)).

Naast deze algemene opmerkingen dringen zich de volgende vaststellingen op. Een sequentie van twee onbeklemtoonde lettergrepen lijkt geen zware schending van de prosodische principes van het Nederlands te zijn, zoals kan worden afgeleid uit de duidelijke voorkeur voor de verbogen vorm bij de adjectieven met ‘sw’ als rechterrond: voor iedere onverbogen vorm telt dit patroon 2.23 verbogen vormen. Deze vaststelling strookt met Booij’s (1998, 2002b) hiërarchie van de prosodische restricties: de conditie die voorschrijft dat alle lettergrepen tot een metrische voet behoren (*ParseSyl*), is laag gerangschikt in (3). Verbogen vormen als *vorige* krijgen dan de volgende structuur:  $[[vo]_{\sigma}[ri]_{\sigma}]_v[ge]_{\sigma}$  (‘[sw]<sub>v</sub>w<sub>★</sub>’), met één voetloze lettergreep, namelijk de lettergreep met de buigings-*e*. De opeenvolging van twee voetloze lettergrepen, zoals in de verbogen vorm *duidelijke*  $[[dui]_{\sigma}[de]_{\sigma}]_v[l ij]_{\sigma}[ke]_{\sigma}$  (‘[sw]<sub>v</sub>ww<sub>★</sub>’), is daarentegen wel problematisch, zoals de absolute voorkeur voor de onverbogen buigingsvariant bij de adjectieven waarvan de basisvorm op het patroon ‘sww’ eindigt, aantoont. De prosodische structuur van de adjectieven waarvan de basisvorm op ‘sww’ eindigt, vertoont altijd één voetloze lettergreep (‘[sw]<sub>v</sub>w’). De toevoeging van de buigings-*e* veroorzaakt een opeenvolging van twee voetloze lettergrepen (‘[sw]<sub>v</sub>ww<sub>★</sub>’). Merk op dat het patroon ‘sww’ (met één voetloze lettergreep) wel in een basisvorm voorkomt (*duidelijk*, *vormelijk*), maar het patroon ‘swww’ (met twee voetloze lettergrepen) niet, tenminste niet in onze materiaalverzameling. Deze constatering sluit aan bij de drempel voorgesteld door onder andere Booij (1992), Cockx (2000) en Royen (1953): deze auteurs beschouwen de opeenvolging van twee onbeklemtoonde lettergrepen in de adjectivische basisvorm als voorwaarde voor de selectie van het onverbogen buigingsmorfeem (zie ook paragraaf 1.1.2.2). Het prosodische criterium dat Permentier & Van Den Eynden (1998) voorstellen voor het gebruik van de onverbogen vorm, namelijk een basisvorm die op een sequentie van drie onbeklemtoonde lettergrepen eindigt, lijkt daarentegen te streng volgens onze gegevens. Zoals we in de vorige paragraaf al vermeld hebben, wordt de sterke voorkeur voor de verbogen vorm bij de adjectieven met een beklemtoonde eindlettergreep mee bepaald door het realiseren van een optimale tweelettergrepige trochee aan de rechterwoordrand (‘[sw<sub>★</sub>]<sub>v</sub>’).

We sluiten de analyse van het prosodische patroon aan de rechterwoordrand van het adjectief af door na te gaan of de concrete invullingen van de beklemtoonde lettergrepen ('1': hoofdklemtoon tegenover '2': nevenklemtoon) en de onbeklemtoonde lettergrepen ('3': beklemtoonbaar/sjwaloos tegenover '0': onbeklemtoonbaar) voor een verdere verfijning zorgen (zie paragraaf 10.1.2 en tabel 10(4) voor de codering van de abstracte en concrete klemtoonpatronen). Eerst specificeren we de beklemtoonde lettergrepen. De resultaten worden samengevat in tabel 10(15). Merk op dat ditmaal de rijvariabele de responsvariabele is en de kolomvariabele de afhankelijke variabele. De relatieve cijfers worden bijgevolg niet per kolom maar per rij berekend. De OR vergelijkt binnen ieder abstract patroon de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  voor de nevenklemtoon met de hoofdklemtoon (de verschillende abstracte patronen worden door een stippellijn gescheiden).

Prosodisch patroon	Buigingsprofiel				OR	
	abstract	concreet	verbogen	onverbogen	OR	95% BI
's'	'1'	2,514	85.02	443	14.98	1.94 [1.51,2.48]
	'2'	293	74.55	100	25.45	
'sw'	'1w'	824	69.95	354	30.05	1.59 [1.04,2.41]
	'2w'	60	59.41	41	40.59	
'sww'	'1ww'	117	35.35	214	64.65	0.55 [0.08,3.93]
	'2ww'	2	50.00	2	50.00	

Tabel 10(15): Invloed op de adjectivische buigingsalternantie van het verschil tussen hoofd- en nevenklemtoon in de prosodische patronen aan de rechterwoordrand van het adjectief

Als we de OR's in deze tabel bekijken, nemen we een verminderend effect van het onderscheid tussen de hoofd- en de nevenklemtoon waar: voor het patroon 's' is er een duidelijk significant verschil tussen de hoofd- en de nevenklemtoon, voor het patroon 'sw' is dit onderscheid nog net significant, en voor het patroon 'sww' verdwijnt de significantie. Samengevat: hoe verder de klemtoon van de rechterwoordrand verwijderd is, des te minder het uitmaakt of het een hoofd- of een nevenklemtoon is.

Bij de concrete invulling van de onbeklemtoonde lettergrepen verwachten we het omgekeerde effect: hoe groter het aantal opeenvolgende onbeklemtoonde lettergrepen, des te zwaarder de doffe sjwa doorweegt, en des te vaker de onverbogen vorm gekozen wordt. De gegevens in tabel 10(16) bevestigen deze hypothese maar ten dele. De OR's vergelijken de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  voor de doffe lettergrepen ('0') met de andere, sjwaloze onbeklemtoonde lettergrepen ('3').

Prosodisch patroon		Buigingsprofiel				OR	
abstract	concreet	verbogen		onverbogen		OR	95% BI
'sw'	's3'	175	58.82	133	43.18	0.49	[0.37,0.63]
	's0'	709	73.02	262	26.98		
'sww'	's33'	19	52.78	17	47.22	2.22	[1.11,4.47]
	's00'	100	33.44	199	66.56		

Tabel 10(16): Invloed op de adjectivische buigingsalternantie van het verschil tussen onbeklemtoonde en doffe lettergrepen in de prosodische patronen aan de rechterwoordrand van het adjectief

De OR's tonen aan dat bij een opeenvolging van twee doffe lettergrepen ('s00') het effect van het prosodische patroon 'sww' verdubbelt in vergelijking met een opeenvolging van twee sjwaloze onbeklemtoonde lettergrepen ('s33').<sup>19</sup> Bij de adjectieven op 'sw' vertoont echter niet de doffe ('s0') maar de sjwaloze onbeklemtoonde ('s3') eindlettergreep een grotere voorkeur voor het onverbogen buigingsmorfeem. In de vorige paragraaf hebben we aangevoerd dat het grote aandeel onverbogen realisaties van de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep ('s0') wellicht lexicaal/semantisch gemotiveerd is (zie paragraaf 10.2.1, (19)).

#### Structuur fonologische constituent

Na de analyse van het prosodische patroon aan de rechterrands van het adjectief vestigen we nu de aandacht op de fonologische constituent, meer bepaald op het prosodische patroon gevormd door de lettergrepen vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon. Dit betekent dat we de prosodische patronen aan de rechterrands van het adjectief, 's', 'sw' en 'sww', combineren met de patronen aan de linkerrands van het substantief, 's', 'ws' en 'wsw'. We brengen in herinnering dat '[' de woordgrens identificeert: links van '[' staat het adjectivische patroon, rechts het substantivische patroon. De onderstreepte lettergrepen identificeren de relevante patronen binnen de fonologische constituent. In de materiaalverzameling komen deze acht patronen op het niveau van de fonologische constituent voor:

- 's|s', zoals in *Amerikááns vlégtuig*, *Bríts páár*, *enórm lándgoed*
- 's|ws', zoals in *polítiek gekíssebis*, *verbáál penséél*
- 's|wsw', zoals in *sociáál begeleidingsplan*
- 'sw|s', zoals in *préeg(e)lig hándschrift*, *slápend kínd*, *tróebel wáter* (waarbij de haakjes de gedeleerde sjwa aanduiden)
- 'sw|ws', zoals in *mímisch aspéct*, *theorétisch probléém*
- 'sw|wsw', zoals in *theorétisch begeleidingsgedoe*
- 'sww|s', zoals in *kóninklijk échtbaar*, *wérkelijk léven*
- 'sww|ws', zoals in *geréchtelijk systéém*, *zángerig accént*

Het belangrijkste aandachtspunt bij de conditionering van de buigingsalternantie op het niveau van de fonologische constituent is de impact van een dreigende klemtoonbotsing ('s|s') op de selectie van de buigings-*e*: aangezien een klemtoon-

botsing een inbreuk vormt op de optimale prosodische structuur (zie paragraaf 10.1.1) nemen we aan dat een dreigende klemtoonbotsing de selectie van de buigings-*e* motiveert. Daarnaast stellen we de vraag in welke mate de structuur van de fonologische constituent het gebruik van beide buigingsmorphemen adequater conditioneert dan de patronen aan de adjectivische rechterrاند. Hierbij zijn we in het bijzonder geïnteresseerd in het gewicht van het adjectivische en het substantivische deel van het prosodische patroon op het niveau van de fonologische constituent. Voor de accentuering hebben we immers vastgesteld dat het buigings-effect van de beklemtoning van de adjectivische eindlettergreep dubbel zo groot is als het overeenkomstige effect van de beklemtoning van de substantivische beginlettergreep (zie paragraaf 10.2.1).

A \ N									
		's'		'ws'		'wws'		$\Sigma_{\text{ADJ}}$	
's'	verb	2,017	85.54	789	80.18	1	12.50	2,807	83.79
	onverb	341	14.46	195	19.82	7	87.50	543	16.21
	$\Sigma$	<b>2,358</b>	<b>100.0</b>	<b>984</b>	<b>100.0</b>	<b>8</b>	<b>100.0</b>	<b>3,350</b>	<b>100.0</b>
'sw'	verb	631	70.74	253	65.54	0	0.00	884	69.12
	onverb	261	29.26	133	34.46	1	100.0	395	30.88
	$\Sigma$	<b>892</b>	<b>100.0</b>	<b>386</b>	<b>100.0</b>	<b>1</b>	<b>100.0</b>	<b>1,279</b>	<b>100.0</b>
'sww'	verb	86	38.57	33	29.46	0	0.00	119	35.52
	onverb	137	61.43	79	70.54	0	0.00	216	64.48
	$\Sigma$	<b>223</b>	<b>100.0</b>	<b>112</b>	<b>100.0</b>	<b>0</b>	<b>0.00</b>	<b>335</b>	<b>100.0</b>
$\Sigma_{\text{NOM}}$	verb	2734	78.72	1075	72.54	1	11.11	3,810	76.75
	onverb	739	21.28	407	27.46	8	88.89	1,154	23.25
	$\Sigma$	<b>3,473</b>	<b>100.0</b>	<b>1,482</b>	<b>100.0</b>	<b>9</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 10(17): Invloed op de adjectivische buigingsalternantie van het prosodische patroon tussen de meest rechtse adjectivische klemtoon en de meest linkse substantivische klemtoon

Tabel 10(17) toont de verdeling van beide buigingsvarianten in de verschillende prosodische patronen op het niveau van de fonologische constituent. Alvorens we de resultaten bespreken, moet de structuur van tabel 10(17) bondig worden toegelicht. Deze tabel bevat twee rijvariabelen: het prosodische patroon aan de rechterwoordrand van het adjectief ('s' vs. 'sw' vs. 'sww') en de adjectivische buiging (verbogen vs. onverbogen), wat nog steeds de responsvariabele is. De kolomvariabele bevat de prosodische patronen aan de linkerwoordrand van het substantief ('s' vs. 'ws' vs. 'wws'). De relatieve frequenties identificeren het aandeel van beide buigingsvarianten in iedere fonologische constituent. Voor iedere fonologische constituent vermelden we eveneens de marginaal, zodat we ons een duidelijk beeld kunnen vormen van de verdeling van de observaties over de verschillende prosodische patronen (deze rijen zijn vetgedrukt). We verduidelijken de opbouw van de tabel aan de hand van een voorbeeld: de eerste cel die we links bovenaan in de tabel tegenkomen ( $n = 2,017$ ), identificeert het aantal verbogen vormen voor het proso-

dische patroon ‘s|s’; in deze context selecteert het adjectief in 85.54% van de gevallen de buigings-*e*. Gezien het grote aantal vrijheidsgraden en de aanwezigheid van lege cellen is het niet zinvol een globale  $\chi^2$ -test uit te voeren.

Laten we eerst de marginalen bekijken. De vergelijking van de marginaal voor de adjectivische prosodische patronen ( $\Sigma_{\text{Adj}}$ ), die gelijk is aan tabel 10(13), met de marginaal voor de substantivische prosodische patronen ( $\Sigma_{\text{NOM}}$ ) bevestigt de eerdere vaststelling dat het prosodische effect van het adjectief sterker is dan het prosodische effect van het substantief (zie paragraaf 10.2.1). De verticale dimensie, met de verschillende adjectivische patronen, is de hoofddimensie. Langs deze dimensie vinden we de belangrijke kwantitatieve effecten terug, ‘s’ vs. ‘sw’ vs. ‘sww’, met een omkering van de relatieve verhouding tussen beide buigingsvarianten bij ‘sww’ (zie ook tabel 10(13)). De horizontale dimensie, met de verschillende substantivische patronen, is een secundaire dimensie die de adjectivische effecten vergroot, maar die niet voor breuklijnen in de tabel zorgt: als we het substantivische patroon ‘wws’ omwille van de dataschaarste buiten beschouwing laten, stellen we vast dat het onderscheid tussen ‘s’ en ‘ws’ aan de linkerrand van het substantief enkel significant is na een adjectief dat op ‘s’ eindigt ( $\chi^2 = 14.3933$ ,  $df = 1$ ,  $p = 0.0001483$ ). Hoewel het substantivische patroon ‘ws’ in vergelijking met ‘s’ telkens voor een stijging van het aandeel onverbogen vormen zorgt na een adjectief eindigend op ‘sw’ ( $\chi^2 = 3.1715$ ,  $df = 1$ ,  $p = 0.07494$ ) of ‘sww’ ( $\chi^2 = 2.3133$ ,  $df = 1$ ,  $p = 0.1283$ ), is deze tendens in geen van beide gevallen significant, wellicht mede door de lage absolute frequenties. De restricties op de structuur van het prosodische woord, c.q. de metrische voet aan de rechterraand van het adjectief, prevaleren op de restricties op de structuur van de fonologische constituent. Deze prioriteit van het prosodische woord op de fonologische constituent sluit aan bij het facultatieve karakter van de meeste condities die betrekking hebben op de fonologische constituent (Booij 1995: 161-164; Gussenhoven 1983a: 315-316; Trommelen & Zonneveld 1989: 245-254).

Zoals we hierboven hebben aangegeven, verdient één patroon in tabel 10(17) extra aandacht, met name het patroon ‘s|s’. Het buigingsprofiel van dit patroon bevat het kleinste aandeel onverbogen vormen (14.46%). De reden hiervoor is het vermijden van een klemtoonbotsing: door de selectie van de buigings-*e* worden de opeenvolgende klemtonen op de adjectivische eindlettergreep en de substantivische beginlettergreep gescheiden, met het afwisselende patroon ‘sw<sub>\*</sub>|s’ als resultaat. Dit verklaart wellicht waarom de distinctie tussen ‘s’ en ‘ws’ aan de linkerwoordrand van het substantief in tabel 10(17) een significant onderscheid definieert bij de adjectieven met ‘s’ als rechterwoordrand: doordat er bij het patroon ‘s|ws’ geen gevaar voor een klemtoonbotsing is, wordt het verbogen buigingsmorfeem significant minder vaak geselecteerd dan bij het patroon ‘s|s’ ( $RR_{\text{onverbogen}} = 1.3703$  [1.1683,1.6073];  $OR = 1.4619$  [1.2038,1.7753]).

Het vermijden van de klemtoonbotsing is het meest dwingend bij twee botsende hoofdklemtonen. Uit de voorbeelden aangehaald in Booij (1995: 161-164),

Gussenhoven (1983a) en Trommelen & Zonneveld (1989: 245-254) kan worden afgeleid dat dit het prototypische geval is voor de klemtoonbotsing op het niveau van de fonologische constituent. Het buigingsprofiel voor '1|1' telt slechts 10.81% onverbogen vormen. De buigingsprofielen van de andere contexten die een klemtoonbotsing definiëren, wijken daarentegen nauwelijks van het algemeen gemiddelde af: 21.62% onverbogen vormen bij '2|2', 23.17% bij '2|1', en zelfs 30.41% bij '1|2'. Doordat deze drie niet-prototypische contexten met een klemtoonbotsing significant ondervertegenwoordigd zijn ( $n_{2|2} = 37$ ;  $n_{2|1} = 245$ ;  $n_{1|2} = 263$ ) ten overstaan van de prototypische context ( $n_{1|1} = 1,813$ ), moet de impact van hun afwijkende buigingsprofiel gerelativeerd worden. Daarnaast hebben we tijdens de analyse van de beklemtoning van de substantivische beginlettergreep vastgesteld dat het adjectief significant vaker onverbogen is vóór een substantief met een secundair beklemtoonde beginlettergreep. Na de controle van de concrete instanties hebben we gesuggereerd dat de prosodische conditionering in deze gevallen ondergeschikt is aan de lexicale/semantische kenmerken van het adjectief-nomenpaar (zie paragraaf 10.2.1).

Na de discussie van de relevante buigingstendensen op het niveau van de fonologische constituent, kwantificeren we de impact van de verschillende prosodische patronen in tabel 10(17). We beschouwen de context met het grootste aandeel verbogen vormen als het referentiepunt. Dit is de context waar de klemtoonbotsing moet worden vermeden, met name 's|s'. Alle  $RR_{\text{onverbogen}}$ 's en OR's in tabel 10(18) berekenen het effect van een prosodisch patroon ten overstaan van deze referentiecontext. Tabel 10(18) heeft dezelfde structuur als tabel 10(17), met dit verschil dat de celwaardes niet de absolute en relatieve frequenties van beide buigingsvarianten uitdrukken, maar wel het  $RR_{\text{onverbogen}}$ , de OR en hun respectievelijke 95% BI'en (De schuine strepen duiden de contexten aan waarvoor geen RR of OR berekend kan worden. Dit zijn de lege cellen in tabel 10(17)).

A \ N		's'	95% BI	'ws'	95% BI	'wws'	95% BI
's'	$RR_{\text{onv}}$	referentiepunt		1.47	[1.25,1.73]	6.50	[4.90,8.61]
	OR	referentiepunt		1.59	[1.30,1.93]	44.96	[5.5,366.7]
'sw'	$RR_{\text{onv}}$	2.02	[1.76,2.33]	2.56	[2.15,3.04]	/	/
	OR	2.45	[2.03,2.94]	3.38	[2.65,4.30]	/	/
'sww'	$RR_{\text{onv}}$	4.56	[3.94,5.28]	4.88	[4.18,5.69]	/	/
	OR	10.23	[7.62,13.74]	14.16	[9.28,21.6]	/	/

Tabel 10(18):  $RR_{\text{onverbogen}}$  en OR voor de verschillende patronen op het niveau van de fonologische constituent met het patroon 's|s' als referentiepunt

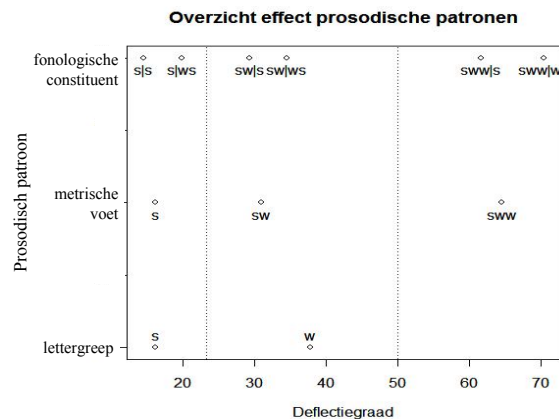
Als we van het ijkpunt in tabel 10(18) vertrekken, stellen we een graduele toename van het  $RR_{\text{onverbogen}}$  en de OR vast. In overeenstemming met bovenstaande analyse, nemen de waarden sterker toe op de verticale as dan op de horizontale as in tabel 10(18). Het substantivische patroon 'wws' vormt hierop een uitzondering, maar het



zeer ruime BI voor de OR van dit patroon wijst echter op dataschaarste. Als we het prosodische patroon met het kleinste aandeel onverbogen adjectieven en het prosodische patroon met het grootste aandeel onverbogen adjectieven vergelijken, stellen we net geen vervijfvoudiging van de kans op een onverbogen adjectief vast en een OR gelijk aan 14.16. De vaststelling dat alle RR's en OR's in tabel 10(18) significant zijn, onderstreept eens te meer de sterke neiging om de buigings-*e* te selecteren in contexten waar een klemtoonbotsing dreigt.

#### *Algemene bespreking*

Na de afzonderlijke analyses van de prosodische patronen op het niveau van de lettergreep, de metrische voet en de fonologische constituent, voegen we het effect van deze drie prosodische domeinen samen in één figuur. Figuur 10(1) is een schematische voorstelling van het buigingseffect van de patronen in deze domeinen.



*Figuur 10(1): Overzicht van het buigingseffect van de prosodische patronen op het niveau van de lettergreep, de metrische voet en de fonologische constituent*

Op de verticale as zetten we de verschillende prosodische domeinen uit: de lettergreep, de metrische voet en de fonologische constituent. De X-as kwantificeert de deflectiegraad van een prosodisch patroon op een schaal van 0 tot 100. De deflectiegraad van een prosodisch patroon is gelijk aan het percentage onverbogen vormen in het buigingsprofiel van dit patroon: een patroon met een deflectiegraad gelijk aan 0 selecteert altijd het verbogen adjectief; een patroon met een deflectiegraad gelijk aan 100 selecteert altijd het onverbogen adjectief. De gelabelde punten situeren de verschillende prosodische patronen in deze tweedimensionale ruimte.<sup>20</sup> De verticale stippellijnen verdelen de figuur in drie zones: de meest linkse zone bevat de patronen waar de verbogen vorm vaker dan het algemeen gemiddelde geselecteerd wordt; de centrale zone identificeert de patronen waar de onverbogen vorm vaker dan het algemeen gemiddelde voorkomt; de rechtse zone identificeert de contexten met een absolute voorkeur voor de onverbogen variant.

Wanneer we de verdeling van de prosodische patronen volgens de verticale dimensie bekijken, stellen we vast dat de twee extensies van het prosodische domein die we hebben voorgesteld in feite niets anders zijn dan een tweevoudige verfijning van de basisoppositie ‘s’ tegenover ‘w’ op het niveau van de lettergreep (c.q. de distinctie tussen een beklemtoonde en een onbeklemtoonde eindlettergreep):

- De metrische voet verfijnt de onbeklemtoonde eindlettergreep door het patroon vanaf de meest rechtse beklemtoonde adjectivische lettergreep te modelleren. Op dit niveau vertakt ‘w’ in ‘sw’ en ‘sww’, terwijl ‘s’ onveranderd blijft. De driedeling die op dit niveau ontstaat, vormt de ruggengraat van de prosodische conditionering van de buigingsalternantie: de drie patronen op dit niveau zijn niet alleen optimaal over de drie zones in figuur 10(1) verdeeld, maar vormen bovendien de basis van de drie clusters op het niveau van de fonologische constituent.
- De fonologische constituent verfijnt de patronen op het niveau van de metrische voet door naar de lettergrepen te kijken die op de adjectivische eindlettergreep volgen, meer bepaald de lettergrepen tot en met de meest linkse substantivische klemtoon. Deze vertakking leidt tot de vorming van drie paren op het hoogste niveau in figuur 10(1): in ieder paar staat het patroon met een beklemtoonde substantivische beginlettergreep links en het patroon met een onbeklemtoonde substantivische beginlettergreep rechts. Deze onderlinge posities bevestigen bovenstaande analyse dat de restricties op de structuur van het prosodische woord (i.e. de voet aan de rechterwoordrand van het adjectief) prevaleren op de restricties op de structuur van de fonologische constituent: het patroon aan de adjectivische rechterwoordrand definieert drie duidelijk onderscheiden clusters; binnen iedere cluster verfijnt het substantivische deel van de fonologische constituent vervolgens het basiseffect van het adjectivische deel.

Wanneer we de spreiding van de punten, c.q. de prosodische contexten, over de X-as beschouwen, zien we dat de afstand tussen de extreme punten groter wordt bij iedere uitbreiding van het prosodische domein. De afstand vergroot vooral doordat er in het rechtse gebied van figuur 10(1) nieuwe punten gecreëerd worden. Anders gezegd: de dubbele extensie van het prosodische domein komt in eerste instantie neer op de identificatie van condities die de selectie van het gemarkeerde onverbogen buigingsmorfem motiveren.

Op basis van bovenstaande analyse, zoals samengevat in figuur 10(1), formuleren we drie prosodische condities op de selectie van beide buigingsmorfemen. Eén, taalgebruikers streven naar het realiseren van een tweelettergrepige trochee aan de rechterwoordrand van het adjectief. Dit resulteert enerzijds in een groot aandeel verbogen vormen bij de adjectieven die op een beklemtoonde lettergreep eindigen, en anderzijds in een significante stijging van het aandeel onverbogen vormen bij de adjectieven die op het prosodische patroon ‘sw’ eindigen. Niettemin is bij deze laatste groep elementen de verbogen variant dominant. Als

verklaring hebben we gewezen op de lage rangschikking van de conditie *ParseSyl* in de hiërarchie voorgesteld door Booij (1998, 2002b) (zie (3), paragraaf 10.1.1): één voetloze onbeklemtoonde lettergreep wordt slechts als een lichte overtreding van de prosodische welgevormdheidscondities beschouwd. Twee, een opeenvolging van twee voetloze onbeklemtoonde lettergrepen veroorzaakt daarentegen wel een zware inbreuk op de prosodische structuur van een Nederlands woord. Dit kan worden afgeleid uit de absolute voorkeur voor de onverbogen buigingsvariant bij de adjectieven met het prosodische patroon ‘sww’ als rechterwoordrand. Drie, er is een sterke neiging om een klemtoonbotsing te vermijden bij een (primair) beklemtoonde adjectivische eindlettergreep en een (primair) beklemtoonde substantivische beginlettergreep (patroon: ‘1|1’). Als gevolg daarvan wordt de buigings-*e* vaker dan het algemeen gemiddelde geselecteerd om beide klemtonen te scheiden en een afwisselend klemtoonpatroon te realiseren.

Bovenstaande analyse doet vermoeden dat de prosodische structuur van het adjectief en van de fonologische constituent gevormd door de adjectief-nomensequentie een voldoende factor is voor de selectie van beide buigingsvarianten: bij een dreigende klemtoonbotsing is er een sterk uitgesproken voorkeur voor de verbogen buigingsvariant en bij een dreigende opeenvolging van drie onbeklemtoonde lettergrepen (i.e. twee voetloze lettergrepen) is de onverbogen vorm dominant. De OR’s die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in deze tegengestelde contexten vergelijken, variëren tussen 10.23 en 14.16. Bij deze vaststelling moeten we echter een kanttekening plaatsen: we hebben namelijk aangevoerd dat het grote aandeel onverbogen vormen bij de substantieven met een nevenklemtoon op de beginlettergreep en bij de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep wellicht lexicaal/semantisch gemotiveerd is (zie paragraaf 10.2.1). Dit voedt het vermoeden dat de prosodische condities in sommige contexten ondergeschikt zijn aan de lexicale/semantische kenmerken van de adjectief-nomensequentie.

### ***10.2.3 Partiële conditionering van prosodische factoren***

We sluiten het onderzoek naar de prosodische conditionering van de buigingsalternantie af met de analyse van de partiële conditionering van twee prosodische factoren. Eerst testen we de hypothese dat het deflecterende effect van een onbeklemtoonde adjectivische eindlettergreep versterkt wordt door een bezittelijk voornaamwoord in determinatorpositie (ANS 1997; Raidt 1968; zie ook paragraaf 1.1.1 en paragraaf 1.2.1). Daarna kijken we in welke mate het buigingseffect van de adjectivische lengtegraad een neveneffect is van de prosodische structuur van de metrische voet aan de rechterwoordrand van het adjectief.

We testen eerst of een bezittelijk voornaamwoord in determinatorpositie een versterkend effect heeft op de keuze van de onverbogen buigingsvariant bij adjectieven met een onbeklemtoonde eindlettergreep. We gebruiken de volgende

variabelen bij de stratumanalyse (zie paragraaf 7.2 voor een algemene uiteenzetting over de stratumanalyse):

- responsvariabele ( $y$ ): buigingsalternantie (verbogen tegenover onverbogen)
- verklarende variabele ( $x$ ): beklemtoning adjectivische eindlettergreep (beklemtoond tegenover onbeklemtoond)
- stratumvariabele ( $z$ ): woordsoort determinator (bezittelijk voornaamwoord tegenover andere)

Tijdens de eerste stap van de stratumanalyse berekenen we de associatie tussen de verklarende variabele en de responsvariabele, terwijl de stratumvariabele gecontroleerd wordt. De *Cochran-Mantel-Haenszel*-statistieken bevestigen de resultaten van de bivariate analyse op basis van tabel 10(8): ook wanneer het effect van de woordsoort van de determinator gecontroleerd wordt, komt het onverbogen buigingsmorfem significant vaker voor bij de adjectieven met een onbeklemtoonde eindlettergreep ( $MH-\chi^2 = 284.2144$ ,  $df = 1$ ,  $p < 2.2e-16$ ;  $CMH-OR = 3.1408$  [2.7410,3.5990]). De  $p$ -waarde voor de *Breslow-Day*-test ( $p = 0.04267$ ) wijst op een lichte, maar significante conditionering van het buigingseffect van de adjectivische eindlettergreep door de woordsoort van de determinator. De exacte aard van deze partiële conditionering specificeren we in de volgende stap van de stratumanalyse (zie tabel 10(19c)).

In de tweede stap van de stratumanalyse wordt de marginale tabel (i.e. de tabel die enkel de associatie tussen de verklarende variabele en de responsvariabele toont, zonder de stratumvariabele in rekening te nemen; zie tabel 10(18)) opgedeeld in twee stratumtabellen, een eerste voor de observaties waar de determinatorpositie door een bezittelijk voornaamwoord wordt bezet (tabel 10(19a)) en een tweede voor de observaties waar een ander element de determinatorpositie invult (tabel 10(19b)). De frequentieverdeling voor de contexten met een bezittelijk voornaamwoord als determinator ziet er als volgt uit ( $\chi^2 = 55.0891$ ,  $df = 1$ ,  $p = 1.152e-13$ ):

	beklemtoond		onbeklemtoond		$\Sigma$	
<b>verbogen</b>	340	86.96	98	58.33	438	78.35
<b>onverbogen</b>	51	13.04	70	41.67	121	21.65
<b><math>\Sigma</math></b>	<b>391</b>	<b>100.0</b>	<b>168</b>	<b>100.0</b>	<b>559</b>	<b>100.0</b>

Tabel 10(19a): Invloed van de beklemtoning van de adjectivische eindlettergreep op de buigingsalternantie in het stratum gedefinieerd door een bezittelijk voornaamwoord in determinatorpositie

De observaties waar het element in de determinatorpositie geen bezittelijk voornaamwoord is, leveren de volgende frequentieverdeling op ( $\chi^2 = 232.6337$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	beklemtoond		onbeklemtoond		$\Sigma$	
<b>verbogen</b>	2,467	83.37	905	62.59	2,467	83.37
<b>onverbogen</b>	492	16.63	541	37.41	492	16.63
<b><math>\Sigma</math></b>	<b>2,959</b>	<b>100.0</b>	<b>1,446</b>	<b>100.0</b>	<b>2,959</b>	<b>100.0</b>

*Tabel 10(19b): Invloed van de beklemtoning van de adjectivische eindlettergreep op de buigingsalternantie in het stratum gedefinieerd door een bepaald lidwoord, een aanwijzend voornaamwoord of een genitief in determinatorpositie*

In tabel 10(19c) vergelijken we de OR's voor beide stratumtabellen:

Stratum	OR	95% BI
bez.vnw.	4.76	[3.11,7.28]
–bez.vnw.	2.99	[2.60,3.46]
CMH-OR	3.14	[2.74,3.60]

*Tabel 10(19c): Samenvatting van de stratumanalyse voor de conditionering door de woordsoort van de determinator van het buigingseffect van de beklemtoning van de adjectivische eindlettergreep*

De OR is groter in het stratum gedefinieerd door de NP's met een bezittelijk voornaamwoord als determinator dan in het complementaire stratum. Op basis van de overlappende 95% BI'en moeten we echter concluderen dat dit effect – wellicht mede als een gevolg van dataschaarste in het stratum gedefinieerd door de NP's met een bezittelijk voornaamwoord in determinatorpositie – niet significant is en dat we bijgevolg de hypothese moeten verwerpen.<sup>21,22</sup>

We ronden de prosodische analyse af door terug te keren naar de eerste factor die we in dit hoofdstuk getest hebben, met name de lengte van het adjectief (zie paragraaf 10.2.1, tabel 10(6)). We verifiëren in welke mate de invloed van de lengte van het adjectief een neveneffect is van de aan- en afwezigheid van de verschillende prosodische patronen aan de adjectivische rechterwoordrand. Idealiter testen we zulke interacties door middel van een stratumanalyse. Voor deze factor is dit technisch uitgesloten omdat niet alle prosodische patronen bij alle lengtegraden voorkomen: het patroon 'sww' is per definitie uitgesloten bij een- en tweelettergrepige adjectieven, zodat het onmogelijk is het  $RR_{\text{onverbogen}}$  en de OR te berekenen omwille van de lege cellen in de tabel. Daarom beperken we deze analyse tot een tabel die de gecombineerde invloed van de adjectivische lengtegraad en de prosodische patronen aan de rechterwoordrand van het adjectief in kaart brengt. De kolomvariabele in tabel 10(20) stelt de verschillende adjectivische prosodische patronen voor ('s' vs. 'sw' vs. 'sww'), de rijvariabele combineert de lengte van het adjectief ( $n(\sigma) = 1$ ,  $n(\sigma) = 2$ ,  $n(\sigma) \geq 3$ ) en de responsvariabele (verbogen vs. onverbogen). De relatieve cijfers berekenen het aandeel van beide buigingsvarianten voor iedere combinatie van een adjectivische lengtegraad en een adjectivisch prosodisch patroon.

lengte	pros.patroon						
		's'	'sw'	'sww'			
<b>n(<math>\sigma</math>) = 1</b>	<b>verb</b>	1,698	91.14	0	0.00	0	0.00
	<b>onverb</b>	165	8.87	0	0.00	0	0.00
	<b><math>\Sigma</math></b>	<b>1,863</b>	<b>100.0</b>	<b>0</b>	<b>0.00</b>	<b>0</b>	<b>0.00</b>
<b>n(<math>\sigma</math>) = 2</b>	<b>verb</b>	528	88.44	548	74.56	0	0.00
	<b>onverb</b>	69	11.56	187	25.44	0	0.00
	<b><math>\Sigma</math></b>	<b>597</b>	<b>100.0</b>	<b>735</b>	<b>100.0</b>	<b>0</b>	<b>0.00</b>
<b>n(<math>\sigma</math>) <math>\geq</math> 3</b>	<b>verb</b>	581	65.28	336	61.76	119	35.52
	<b>onverb</b>	309	34.72	208	38.24	216	64.48
	<b><math>\Sigma</math></b>	<b>890</b>	<b>100.0</b>	<b>544</b>	<b>100.0</b>	<b>335</b>	<b>100.0</b>

Tabel 10(20): Gecombineerde invloed op de adjectivische buigingsalternantie van de lengte van het adjectief en het prosodische patroon aan de rechterwoordrand van het adjectief

Tabel 10(20) toont een genuanceerd beeld met betrekking tot de interactie tussen de lengte van het adjectief en de klemtoonpatronen aan de rechterwoordrand van het adjectief. Aan de ene kant is het patroon 's', dat het gebruik van de onverbogen variant faciliteert, – per definitie – oververtegenwoordigd bij de eenlettergrepige adjectieven en is het patroon 'sww' met een absolute voorkeur voor de onverbogen variant – per definitie – uitsluitend aanwezig bij de meerlettergrepige adjectieven. De zeer sterke voorkeur voor de verbogen vorm bij de eenlettergrepige adjectieven kan met andere woorden verklaard worden vanuit de algemene prosodische tendens tot de vorming van een tweelettergrepige trocheïsche voet: 'sw<sub>★</sub>'. Doordat de meerlettergrepige adjectieven de enige groep vormen die het patroon 'sww' vertonen, zal het aandeel onverbogen adjectieven in deze groep groter zijn omwille van de neiging om een opeenvolging van twee voetloze lettergrepen te vermijden.

Aan de andere kant tonen de cijfers aan dat het lengte-effect niet volledig tot de aan- of afwezigheid van prosodische patronen gereduceerd kan worden: niet alleen eindigt meer dan de helft van de meerlettergrepige adjectieven op een beklemtoonde eindlettergreep ('s'), maar de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  zijn bovendien significant groter bij de meerlettergrepige adjectieven met een beklemtoonde eindlettergreep dan bij de eenlettergrepige adjectieven, zoals de OR's in tabel 10(21) aantonen. De OR's in deze tabel vergelijken de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bij de twee- en meerlettergrepige adjectieven op een beklemtoonde eindlettergreep enerzijds met deze odds bij de eenlettergrepige adjectieven anderzijds (zie kolom 1 in tabel 10(20)).

	OR	95% BI
's' tweelettergrepig VS. 's' eenlettergrepig	1.34	[0.99,1.81]
's' meerlettergrepig VS. 's' eenlettergrepig	5.47	[4.43,6.76]

Tabel 10(21): OR's voor de twee- en meerlettergrepige adjectieven met een beklemtoonde eindlettergreep t.o.v. de eenlettergrepige adjectieven

We besluiten dan ook dat het buigingseffect van de adjectivische lengtegraad in grote mate – maar niet volledig – herleid kan worden tot het prosodische patroon aan de rechterwoordrand van het adjectief.

### 10.3 Conclusie

Alvorens de belangrijkste resultaten van het prosodische onderzoek op een rijtje te zetten, vatten we bondig het algemene analytische kader samen. Deze analyse is niet beperkt tot het falsificeren van de hypothesen die geformuleerd zijn op basis van de behandeling van de ‘eufonische’ of ‘ritmische’ factoren in de literatuur, maar stelt zich tot doel de prosodische restricties op het gebruik van het verbogen en onverbogen adjectief te identificeren. Hiervoor baseren we ons op de inzichten in de metrische fonologie. We maken meer bepaald gebruik van de prosodische condities die in *Optimality Theory* aangereikt worden, zonder evenwel een analyse in die trant na te streven (paragraaf 10.1.1). Voor de prosodische annotatie van de observaties in de materiaalverzameling hebben we gebruik gemaakt van de prosodische codes in de *Elektronische Grote Van Dale* die we vervolgens op een zestal punten hebben aangevuld (paragraaf 10.1.2).

De feitelijke empirische analyse bestaat uit drie delen: de falsificatie van de ‘ritmische’ of ‘eufonische’ hypothesen (paragraaf 10.2.1), de identificatie van de prosodische restricties op de buigingsalternantie (paragraaf 10.2.2) en een stratumanalyse (paragraaf 10.2.3).

In het eerste deel van de empirische analyse zijn de verschillende ‘eufonische’ en ‘ritmische’ hypothesen die we in het literatuuroverzicht naar voren hebben geschoven aan de corpusgegevens getoetst (paragraaf 10.2.1):

- Het aandeel onverbogen vormen neemt toe in functie van de lengte van het adjectief, waarbij drie verschillende lengtegraden voor significante distincties zorgen: eenlettergrepig vs. tweelettergrepig vs. meerlettergrepig.
- Vergeleken met een beklemtoonde adjectivische eindlettergreep is een onbeklemtoonde eindlettergreep een faciliterende factor voor het gebruik van de onverbogen buigingsvariant.
- Vergeleken met een beklemtoonde substantivische beginlettergreep is een onbeklemtoonde beginlettergreep een faciliterende factor voor de selectie van de onverbogen buigingsvariant.

In het tweede deel van de empirische analyse (paragraaf 10.2.2) hebben we het prosodische domein uitgebreid van het niveau van de lettergreep, zoals in bovenvermelde hypothesen, naar achtereenvolgens het niveau van de metrische voet (i.e. het patroon gevormd door de lettergrepen vanaf de meest rechtse adjectivische klemtoon) en het niveau van de fonologische constituent (i.e. het prosodische patroon vanaf de meest rechtse adjectivische klemtoon tot en met de meest linkse substantivische klemtoon). Deze dubbele extensie ten aanzien van het bestaande onderzoek heeft aangetoond dat bovenstaande hypothesen niet volstaan om de

prosodische conditionering van de adjectivische buigingsalternantie in kaart te brengen: de lettergreep beantwoordt in veel gevallen niet aan de minimale metrische structuur die vereist is om de buigingsalternantie optimaal te conditioneren. Op het niveau van de fonologische constituent prevaleert het prosodische patroon aan de rechterwoordrand van het adjectief op het prosodische patroon aan de linkerwoordrand van het substantief, dat hoofdzakelijk het buigingseffect van het adjectivische patroon versterkt.

De keuze van beide buigingsvarianten wordt in het CGN door de volgende drie algemene prosodische restricties gestuurd:

- het realiseren van een tweelettergrepige trocheïsche voet aan de rechterwoordrand van het adjectief: Deze prosodische conditie faciliteert zowel de selectie van het verbogen adjectief als de selectie van het onverbogen adjectief. Enerzijds vergemakkelijkt dit criterium het gebruik van de verbogen vorm bij adjectieven met een beklemtoonde eindlettergreep om een tweelettergrepige trochee aan de rechterwoordrand te realiseren. Anderzijds zorgt dit criterium voor een lichte – maar significante – stijging van het aandeel onverbogen vormen bij de adjectieven met het patroon ‘sw’ als rechterwoordrand. Hoewel de buigings-*e* bij adjectieven met een basisvorm op ‘sw’ voetloos is, vertoont deze groep adjectieven een absolute voorkeur voor de verbogen vorm. Met andere woorden: één voetloze onbeklemtoonde lettergreep wordt slechts als een lichte inbreuk op de prosodische structuur van het Nederlands beschouwd.
- het vermijden van twee opeenvolgende voetloze onbeklemtoonde lettergrepen aan de rechterwoordrand van het adjectief: In tegenstelling tot één voetloze onbeklemtoonde lettergreep vormt een opeenvolging van twee voetloze onbeklemtoonde lettergrepen wel een zware inbreuk op de prosodische structuur van het Nederlands, zoals blijkt uit de absolute voorkeur voor de onverbogen vorm bij de adjectieven met het patroon ‘sww’ aan de rechterwoordrand.
- het vermijden van een klemtoonbotsing: Bij de opeenvolging van een beklemtoonde adjectivische eindlettergreep en een beklemtoonde substantivische beginlettergreep motiveert het vermijden van een klemtoonbotsing de selectie van de buigings-*e*.

Naast de abstracte klemtoonpatronen, die de basis van de prosodische analyse vormen, hebben we ook de invloed van de concrete patronen beschouwd. Aan de ene kant zorgt de concrete invulling van de abstracte klemtoonpatronen voor een verdere verfijning van de vastgestelde tendensen:

- De tendens om twee opeenvolgende voetloze onbeklemtoonde lettergrepen te vermijden, wordt versterkt als de basisvorm van het adjectief op twee doffe lettergrepen eindigt.



- De invloed van het onderscheid tussen hoofd- en nevenklemtoon aan de adjectivische rechterwoordrand vermindert naarmate de beklemtoonde lettergreep verder van de woordrand verwijderd is.
- Het vermijden van een klemtoonbotsing is het meest dwingend bij twee hoofdklemtonen (i.e. bij het patroon '1|1').

Aan de andere kant vertonen de concrete patronen niet altijd het verwachte buigingsprofiel:

- Bij het patroon 'sw' aan de adjectivische rechterwoordrand vertonen de doffe eindlettergrepen ('s0') geen grotere voorkeur voor de onverbogen vorm dan de sjwaloze onbeklemtoonde eindlettergrepen ('s3').
- Bij een substantivische beginlettergreep die een nevenklemtoon draagt, is het aandeel onverbogen vormen significant groter dan bij een onbeklemtoonde substantivische beginlettergreep.

De inspectie van de adjectief-nomenparen die deze patronen realiseren, suggereert dat naast de prosodische structuur ook de lexicale/semantische eigenschappen van de adjectief-nomensequentie een belangrijke factor zijn bij de keuze van de onverbogen vorm. Deze intuïtieve lexicale/semantische verklaring wordt in het volgende hoofdstuk door middel van een stratumanalyse geverifieerd (zie paragraaf 11.2.3). Hierbij vervangen we de intuïtieve notie van adjectief-nomenpaar met naamstatus door een objectiveerbaar alternatief, met name de lexicale collocabiliteit van het adjectief-nomenpaar.

De stratumanalyse vormt het derde deel van de empirische analyse (paragraaf 10.2.3). Deze analyse heeft als doelstelling de conditionering van het effect van een eerste verklarende parameter door een tweede verklarende parameter te meten. Aan de hand van deze techniek hebben we de hypothese weerlegd dat de adjectieven met een onbeklemtoonde eindlettergreep vaker de onverbogen buigingsvariant kiezen in een NP met een bezittelijk voornaamwoord als determinator dan in een andere NP, al moet worden opgemerkt dat dit voor een deel het gevolg is van het geringe aantal NP's met een bezittelijk voornaamwoord in determinatorpositie. Tot slot hebben we getest in hoeverre het buigingseffect van de adjectivische lengtegraad gereduceerd kan worden tot de aan- en afwezigheid van prosodische patronen aan de rechterwoordrand van het adjectief die de keuze van één van beide buigingsvormen vergemakkelijken. Hoewel een stratumanalyse om technische redenen uitgesloten is, hebben we aangetoond dat het adjectivische lengte-effect voor een deel door de prosodische patronen aan de rechterwoordrand beïnvloed wordt, zonder dat dit lengte-effect echter volledig als een prosodisch neveneffect kan worden afgedaan. In de multivariate analyse zullen we bijgevolg een complexe verklarende factor creëren die de adjectivische woordlengte en het prosodische patroon aan de rechterwoordrand van het adjectief combineert (zie paragraaf 14.1, tabel 14(1)).

## Noten

1. Net als in het literatuuroverzicht (zie paragraaf 1.1.2.2) markeren we de lettergreep met de hoofdklemtoon of de primaire klemtoon door een *accent aigu* ‘´’ op de klinker te plaatsen. Op de klinker van de lettergrepen die de eventuele nevenklemtoon of secundaire klemtoon dragen, plaatsen we een *accent grave* ‘`’. Onbeklemtoonde lettergrepen zijn accentloos.
2. Een algemene inleiding tot de prosodische fonologie van het Nederlands is Trommelen & Zonneveld (1989).
3. De reden voor het aannemen van een onderliggend nul-buigingsmorfeem is dubbel. Ten eerste, de vorm *open-e* is de optimale vorm voor de nominalisatie. De adjectivische en de genominaliseerde vorm worden dan onderscheiden door [★] als basisvorm van het nominaliserende suffix te poneren en -∅ als basisvorm van het adjectivische buigings-suffix. De prosodisch welgevormde vorm *open-∅* schendt bij een nominalisatie bijgevolg de hoogst gerangschikte conditie *MorphI*, die vereist dat het onderliggende morfeem -*e* zichtbaar is in de uiteindelijke vorm. Bijgevolg is *open-e* de meest optimale vorm voor de nominalisatie. Ten tweede, het volstaat een blik op tabel 10(3) te werpen om vast te stellen dat voor andere adjectieven, zoals *groen*, [★] en niet -∅ als basisvorm van het adjectivische buigingsuffix moet worden geponeerd.
4. Variationele fenomenen vormen binnen OT een onderwerp van discussie. Hinskens, van Hout & Wetzels (1997: 8) vermelden twee mogelijke oplossingen voor de modellering van variationele fenomenen in OT, met name een partiële hiërarchie van de condities en zwevende condities. In het eerste geval wordt slechts een deel van de condities strikt geordend, zodat meerdere uitvoervormen mogelijk zijn. In het tweede geval kan een conditie een variabele plaats innemen binnen een bepaalde zone van de hiërarchie.
5. Voor een algemene beschrijving van de klemtoontoekenning in het Nederlands verwijzen we naar Booij (1995: 96-124), De Schutter (1993), Kooij & van Oostendorp (2003: 105-127) en Trommelen & Zonneveld (1989, 1999). Zonneveld (2003) geeft een kritisch overzicht van de verschillende theoretische benaderingen van dit complexe probleem.
6. Deze gegevens zijn ons via Dirk Geeraerts door *Van Dale Lexicografie* ter beschikking gesteld. Wij wensen hierbij *Van Dale Lexicografie* en Dirk Geeraerts te bedanken.
7. De code ‘04<sub>2</sub>’ voor *proces* duidt aan dat de primaire klemtoon (‘1’) uit het EGVD-klemtoonpatroon in een secundaire klemtoon (‘2’) verandert in de samenstelling.
8. Wanneer *superkatholiek* als substantief gebruikt wordt, draagt het eerste lid de hoofdklemtoon *súperkatholiek* (klemtoonpatroon: ‘10202’).
9. De reden voor het gelijkaardige gedrag van geprefigeerde adjectieven en samengestelde adjectieven is dat het prefix bij de syllabificatie van een woord, net zoals de leden van een samenstelling, een afzonderlijk prosodisch domein vormt (cf. *supra*). Dit is: het prefix gedraagt zich zoals het linkerlid van een samenstelling. De suffixen vormen – met

- uitzondering van *-achtig*, *-baar* en *-loos* – geen apart prosodisch domein (Booij 1995: 111).
10. Men spreekt in termen van “in het algemeen” (ANS 1997: 698), “seems to be obligatory” (Booij 1995: 162), “(redelijk) verplicht” (Trommelen & Zonneveld 1989: 247). Enkele prefixen vormen een uitzondering op deze algemene neiging tot twijfel en vaagheid: Booij (1995: 123) oordeelt dat het prefix *on-* meestal klemtoondragend is in attributieve positie (op voorwaarde dat de afleiding morfologisch transparant is); Schultink (1977, 1980) vermeldt dat intensifiërende prefixen, zoals *door-* in *dóornat* en *in-* in *ínslecht*, altijd de hoofdklemtoon dragen.
  11. Trommelen & Zonneveld (1989: 246) kenmerken de klemtoonverschuiving bij samenstellingen als “verplicht”. Booij (1995: 118) stelt dat de hoofdklemtoon altijd op het eerste lid valt wanneer de structuur van het adjectief niet transparant is, zoals in *zéziek*, *schietgraag* of *kléurrijk*.
  12. Booij (1995: 119-122) groepeerde de prefixen in twee klassen in functie van hun klemtoonbeïnvloedende gedrag: de uitheemse prefixen enerzijds en de ‘genaturaliseerde’ en inheemse prefixen anderzijds. Afleidingen met het eerste type prefix zijn niet transparant. Deze prefixen vormen bijgevolg geen apart prosodisch woord, zodat ze bij de klemtoon-toekenning samen met de stam één prosodisch woord vormen (bv. *a-* zoals in *apatisch*). De tweede groep bestaat uit de inheemse prefixen (bv. *on-*) en de ‘genaturaliseerde’ prefixen. De genaturaliseerde prefixen zijn van oorsprong vreemde prefixen wier gedrag identiek is aan dat van de inheemse prefixen (bv. *anti-*, *pro-*, *contra-*, *multi-*, *intra-*).
  13. Gussenhoven (1983a: 603) vormt een uitzondering op de algemene neiging tot voorzichtigheid: hij is van mening dat “to all intents and purposes, ST [Stress Retraction; JT] is obligatory”.
  14. Het onderscheid tussen ‘s’ voor beklemtoond en ‘w’ voor onbeklemtoond is afkomstig uit de Angelsaksische literatuur waar men het over ‘s(trong)’ en ‘w(eak)’ lettergrepen heeft. Deze oppositie is in de Nederlandse literatuur overgenomen (*inter al.* Trommelen & Zonneveld 1989 en Kooij & van Oostendorp 2003).
  15. We wijzen erop dat geen enkel van de annotatieproblemen die gepaard (kunnen) gaan met het gebruik van de EGVD-klemtoonpatronen zich voordoet bij de codering van *vegetarisch menu*.
  16. Omdat het enigszins vreemd lijkt de basisvorm van gereduceerde vormen zoals *andr-e* (gereduceerde vorm van *ander-e*; cf. sjwa-deletie) als eenlettergrepig te beschouwen, hebben we gecontroleerd of er significante verschillen zijn tussen het buigingseffect van de adjectivische lengtegraad op basis van de niet-gereduceerde vormen (i.e. *ander-e*;  $n(\sigma) = 2$ ) en de adjectivische lengtegraad op basis van de gereduceerde vormen (i.e. *andr-e*;  $n(\sigma) = 1$ ):
    - De observaties zijn niet op een significant verschillende manier over de adjectivische lengtegraden verdeeld voor de gereduceerde vormen en de niet-gereduceerde vormen, ook niet als we de zes- en zevenlettergrepige adjectieven

omwille van hun lage frequentie ( $n_{\text{zeslettergrepig}} = 18$ ;  $n_{\text{zevenlettergrepig}} = 2$ ) bij de vijflettergrepige adjectieven voegen ( $\chi^2 = 2.0282$ ,  $df = 4$ ,  $p = 0.7306$ ).

- Voor geen enkele adjectivische lengtegraad vertonen de buigingsvarianten een significant verschillende frequentieverdeling bij de niet-gereduceerde vormen en de gereduceerde vormen (de kleinste  $p$ -waarde, deze voor de tweelettergrepige adjectieven, is gelijk aan 0.23).

Aangezien er geen significant verschil is tussen de lengtegraden op basis van de gereduceerde vormen en de niet-gereduceerde vormen, rapporteren we alleen de lengtegraden voor de gereduceerde vormen (zie tabel 10(6)).

17. Deze driedeling wordt bevestigd door de resultaten van de partiële  $\chi^2$ -statistiek in tabel 10(i) (zie paragraaf 8.1.2) die de tweelettergrepige adjectieven als een overgangscategorie tussen de eenlettergrepige en de meerlettergrepige adjectieven bestempelt: bij de eenlettergrepige adjectieven komt de verbogen vorm significant vaker voor dan verwacht en treffen we de onverbogen vorm significant minder vaak aan dan verwacht; de meerlettergrepige adjectieven vertonen het tegengestelde patroon, namelijk een significante ondervertegenwoordiging van de verbogen vorm en een significante oververtegenwoordiging van de onverbogen vorm; de tweelettergrepige adjectieven selecteren alleen de onverbogen vorm significant minder vaak dan verwacht en hebben geen significant effect op de keuze van de verbogen vorm (de scharnierwaarde 1 behoort tot het BI, waardoor het vastgestelde effect voor deze cel niet significant is).

	$n(\sigma) = 1$	$n(\sigma) = 2$	$n(\sigma) \geq 3$
<b>verbogen</b>	[1.1417,1.2333]	[0.9968,1.1081]	[0.7217,0.8044]
<b>onverbogen</b>	[0.3238,0.4381]	[0.7281,0.9254]	[1.6633,1.9015]

Tabel 10(i): 95% BI'en partiële  $\chi^2$ -test voor de verdeling van de adjectivische buigingsvarianten over de adjectivische lengtegraden

18. We gebruiken intentioneel '<' in plaats van '>>' omdat we geen OT-model nastreven (zie paragraaf 10.1.1). In (21) worden trouwens prosodische patronen voorgesteld en geen prosodische condities zoals in OT.
19. Het prosodische patroon 's33' vraagt om enige toelichting. Dit patroon voegt de patronen 's03' ( $\text{verbogen}/\text{onverbogen} = 12/15$ ) en 's33' ( $\text{verbogen}/\text{onverbogen} = 7/2$ ) samen. Deze samenvoeging is gemotiveerd door de lage absolute frequenties van deze patronen en de afwezigheid van een significant verschil met betrekking tot de keuze van de buigingsvariant (*Fisher Exact*-test:  $p = 0.1279$ ). Het is opvallend dat de mogelijkheid om een tweelettergrepige trochee aan de rechterwoordrand te construeren door de adjectieven met het patroon 's03' te verbuigen niet ten volle geëxploiteerd wordt: *algemeen* zou door toevoeging van de buigings-*e* de prosodisch optimale vorm  $[[\text{al}]_{\sigma}[\text{ge}]_{\sigma}]_{\sigma}[[\text{me}]_{\sigma}[\text{ne}]_{\sigma}]_{\sigma}$  vertonen.
20. Omwille van de zeer lage absolute frequenties hebben we de patronen 's|wws' ( $n = 8$ ), 'sw|wws' ( $n = 1$ ) en 'sww|wws' ( $n = 0$ ) niet in figuur 10(1) opgenomen.
21. Om uit te sluiten dat het stratumeffect op zijn beurt geconditioneerd is door de ongelijke verdeling van de patronen 'sw' en 'sww' over de twee onderscheiden strata ('bez.vnw.' vs. '-bez.vnw'), hebben we de stratumanalyse gerepliceerd voor de prosodische patronen aan de rechterwoordrand van het adjectief ('s' vs. 'sw' vs. 'sww'). Merk op dat we niet

langer een  $2 \times 2 \times k$ -datamatrix gebruiken maar een  $2 \times l \times k$ -datamatrix, waardoor de CMH-OR en de *Breslow-Day*-statistiek niet berekend kunnen worden. Als gevolg berekenen we in iedere stratumtabel  $l-1$  OR's in plaats van één OR (i.e. 2-1), zoals in een standaard stratumanalyse (op basis van een  $2 \times 2 \times k$ -datamatrix). De adjectieven met 's' als rechterwoordrand fungeren als referentiepunt voor het berekenen van de OR's voor 'sw' en 'sww'. Tabel 10(ii) vat de resultaten van deze stratumanalyse samen.

Stratum	'sw' vs. 's'		'sww' vs. 's'	
	OR	95% BI	OR	95% BI
bez.vnw	3.72	[2.35,5.89]	12.2	[5.70,26.2]
-bez.vnw	2.18	[1.86,2.56]	9.09	[7.04,11.7]
marginale tabel	2.31	[1.99,2.68]	9.38	[7.37,11.9]

*Tabel 10(ii): Samenvatting van de stratumanalyse voor de partiële conditionering van het buigingseffect van het prosodische patroon aan de adjectivische rechterwoordrand door de woordsoort van de determinator*

In de NP's met een bezittelijk voornaamwoord als determinator is het deflectorende effect van 'sw' en van 'sww' groter dan in de NP's met een andere determinator, maar ook hier overlappen de 95% BI'en voor de OR's in beide strata, zodat we besluiten dat de verschillen tussen de strata niet significant zijn.

22. We signaleren dat de randsignificantie van de *Breslow-Day*-statistiek niet wordt bevestigd op het niveau van de stratumtabellen. Dit is een algemene vaststelling: telkens wanneer de  $p$ -waarde voor de *Breslow-Day*-test op een zwakke partiële conditionering wijst, wordt deze niet ondersteund door de (overlappende) 95% BI'en van de OR's.

# **Hoofdstuk 11:**

## **Lexicale/semantische conditionering**

De lexicale/semantische restricties op de keuze van het verbogen en het onverbogen adjectief hebben in de literatuur over de adjectivische buigingsalternantie een dubbelzinnig statuut. Aan de ene kant is de lexicale/semantische conditionering de factor die in de literatuur het vaakst aan bod komt (zie paragraaf 1.2.2) en die introspectief aan de meest diepgravende analyses onderworpen wordt (Blom 1994, 1995; Booij 2002a, 2002c; Honselaar 1980). Aan de andere kant wordt de lexicale/semantische conditionering van de buigingsalternantie expliciet verwijderd uit het bestaande kwantitatief-empirisch onderzoek (Lebrun & Schurmans-Swillen 1966; de Rooij 1980a).<sup>1</sup> De reden die hiervoor wordt opgegeven is de moeilijkheid – zo niet onmogelijkheid – om deze factor op objectieerbare wijze te operationaliseren. Het is dan ook enigszins verwonderlijk dat alle idiomatische adjectief-nomenparen in de bovengenoemde empirische analyses, ondanks de identificatieproblemen, schijnbaar probleemloos uit de materiaalverzameling verwijderd worden. Voor de bespreking van deze analyses verwijzen we naar het literatuuroverzicht (paragraaf 1.2.2). Voor de stelling dat de adjectivische positie in adjectief-nomenparen met naamstatus vaak door een relationeel adjectief bezet wordt (Booij 2002a; 2002c; Heynderickx 2001), ontbreekt voorlopig eveneens kwantitatief-empirische evidentie.

De eerste paragraaf van dit hoofdstuk behandelt de problematiek die ontstaat bij de identificatie van de adjectief-nomenparen met naamstatus (11.1.1). Om hieraan te verhelpen stellen we een kwantitatieve procedure voor om vaste verbindingen, dus ook adjectief-nomenparen met naamstatus en idiomatische adjectief-nomenparen, op een objectieerbare wijze te identificeren. Daarnaast beschrijven we de criteria die we hanteren om de relationele adjectieven te identificeren (11.1.2). In een tweede paragraaf becommentariëren we de resultaten van het empirisch onderzoek (11.2).

Dit hoofdstuk vormt de eerste fase van het onderzoek naar de lexicale en semantische conditionering van de buigingsalternantie. In deel IV komen we uitgebreid op dit onderwerp terug. In voorliggend hoofdstuk ligt de nadruk op de implementatie van een objectieerbare procedure voor de empirische analyse van de lexicale/semantische restricties op de buigingsalternantie, waarbij we uitdrukkelijk het gebruik van intuïtieve noties zoals idiomatisch adjectief-nomenpaar of adjectief-nomenpaar met naamstatus verwerpen. Als alternatief kijken we naar het buigings-effect van lexicale collocaties. De keuze voor dit alternatief roept onmiddellijk de vraag op of de lexicale collocaties een louter lexicaal fenomeen zijn, dan wel of er semantische tendenties aan de grondslag van deze patronen liggen. Aangezien we ons in dit hoofdstuk hoofdzakelijk op de verantwoording en de operationalisering

van de lexicale collocaties concentreren, komt deze vraag niet in dit hoofdstuk maar in het vierde deel van dit proefschrift aan bod. Bijgevolg hebben we het in dit hoofdstuk over de *lexicale/semantische* conditionering van de buigingsalternantie, waarbij we geen uitspraak doen over de exacte aard van deze conditionering. De reden hiervoor is dat het CGN zowel kwantitatief als kwalitatief minder geschikt is om deze vraag te beantwoorden: kwantitatief is het aantal lexicale collocaties relatief beperkt, zoals duidelijk wordt tijdens de empirische analyse (zie paragraaf 11.2.1); kwalitatief verzamelt het CGN uitingen in zeer diverse gesprekssituaties (zie paragraaf 6.1, tabel 6(1)) met lectische variatie (zowel geografisch als stilistisch) en weinig thematische consistentie als gevolg. In deel IV, waar we ons uitsluitend op de lexicale en semantische conditionering van de buigingsalternantie concentreren, maken we gebruik van een tweede corpus, met name (de edities van januari tot en met september van de jaargang 2003) van de Vlaamse kwaliteitskrant *De Standaard*. Hoewel de vraag naar de semantische motivatie van het buigingseffect van de lexicale collocaties in dit hoofdstuk niet expliciet gesteld wordt, formuleren we een aanzet tot antwoord op basis van de analyse van het buigingseffect van de relationele adjectieven, die onder andere door Booij (2002a) en Heynderickx (2001) geassocieerd worden met adjectief-nomenparen met naamstatus (zie paragrafen 11.1.2 en 11.2.2).

## 11.1 Hypotheses en operationalisering

We behandelen eerst de implementatie van de lexicale/semantische kenmerken van het adjectief-nomenpaar (11.1.1). Daarna bespreken we de operationalisering van de relationaliteit van het adjectief (11.1.2).

### 11.1.1 *Lexicale/semantische kenmerken van het adjectief-nomenpaar*

Voor de lexicale/semantische conditionering van de buigingsalternantie hebben we de volgende hypothese voorgesteld: de onverbogen vorm markeert het adjectief-nomenpaar als een semantische eenheid die als een naam fungeert (zie paragraaf 1.1.3.1). Introspectief onderzoek definieert deze eenheden in termen van niet-compositionele semantiek: de betekenis van de adjectief-nomencombinatie is niet afleidbaar uit de betekenis van de samenstellende delen (Broekhuis 1999; Odijk 1992). Ter illustratie worden uitdrukkingen zoals *stoffelijk overschot*, *bijvoeglijk naamwoord* en *lijdend voorwerp* aangehaald. De introspectieve analyses wekken de indruk dat idiomen, uitdrukkingen met een niet-compositionele semantiek of adjectief-nomencombinaties met naamstatus probleemloos en ondubbelzinnig geïdentificeerd kunnen worden. We zullen aanvoeren dat idiomen en vaste verbindingen niet eenduidig identificeerbaar zijn. Bovendien zijn de voorbeelden die in het introspectieve onderzoek geciteerd worden beperkt in aantal en spelen de auteurs bij de keuze van hun voorbeelden vaak leentjebuurt bij elkaar (zie paragraaf 1.2.2). Overigens hebben we ook zelf in wat voorafging gebruik gemaakt van een

puur intuïtieve afbakening van lexicale/semantische kenmerken: in de prosodische analyse hebben we adjectief-nomenparen met naamstatus geciteerd om bepaalde buigingstendenties te verklaren. We hebben hierbij wel iedere vorm van kwantificatie vermeden: de geciteerde voorbeelden dienen enkel om aan te geven dat er iets anders in het geding is dan de prosodische conditionering (zie paragraaf 10.2.1 en paragraaf 10.2.2.1). In dit hoofdstuk onderwerpen we deze intuïtieve verklaring aan een empirische analyse, waarbij we gebruik maken van de voorgestelde objectiveerbare lexicale/semantische criteria (zie paragraaf 11.2.3).

De bespreking van de operationalisering van de lexicale/semantische kenmerken van het adjectief-nomenpaar bestaat uit twee delen. Eerst argumenteren we dat de bestaande benaderingen een objectiveerbare en repliceerbare procedure – één van de methodologische doelstellingen van dit onderzoek – in de weg staan (11.1.1.1). Vervolgens stellen we een alternatieve, distributionele benadering voor (11.1.1.2).

#### 11.1.1.1 Problemen bij de identificatie van lexicale idiomen

Onderzoek naar idiomatische uitdrukkingen heeft een aantal problemen aan het licht gebracht die in traditionele, introspectieve analyses doorgaans over het hoofd worden gezien. Ten eerste, er bestaat geen algemene definitie van idiomaticiteit. Everaert (1993: 4) noteert dat iedere auteur “zijn/haar variant” van idiomaticiteit gebruikt. Het betreft met andere woorden een rekbaar begrip. Nunberg, Sag & Wasow (1994: 492) wijzen eveneens op de vaagheid van dit begrip:

“In actual linguistic discourse and lexicographical practice, ‘idiom’ is applied to a fuzzy category defined on the one hand by ostention of prototypical examples like English *kick the bucket*, *take care of NP*, or *keep tabs on NP*, and on the other by implicit opposition to related categories like formulae, fixed phrases, collocations [...]”

Laat dit net de werkwijze zijn die we hebben vastgesteld bij het merendeel van de analyses van de lexicale/semantische conditionering van de adjectivische buigingsalternantie: de evidentie voor de stelling dat het onverbogen adjectief lexicale idiomen of adjectief-nomenparen met naamstatus identificeert, is vaak beperkt tot enkele prototypische gevallen, type *stoffelijk overschot*, *zelfstandig naamwoord* of *lijdend voorwerp*. Deze werkwijze roept onmiddellijk de volgende vragen op met betrekking tot de conditionering van de buigingsalternantie:

- Hoe strikt moet de lexicale/semantische conditionering, los van de operationaliseringsproblematiek, geïnterpreteerd worden? In de literatuur worden uiteenlopende standpunten verdedigd die variëren van een zeer strikte conditionering, waarbij het gebruik van de onverbogen vorm als idiosyncratisch wordt beschouwd (de Haas & Trommelen 1993; Van Eynde 2003), tot een actief semantisch mechanisme dat op creatieve wijze kan worden aangewend (Blom 1994, 1995; Booij 2002a, 2002c; Honselaar 1980).



- Hoe systematisch is het gebruik van de onverbogen vorm in lexicale collocaties? Terwijl de meeste auteurs het gebruik van de onverbogen buigingsvariant bij adjectief-nomenparen met naamstatus als een tendentie beschouwen (*inter al.* Booij 2002a, 2002c; Schultink 1962), vermoedt Odijk (1992: 202) een discrete oppositie: “when the adjective is combined with the noun [een enkelvoudig *het*-woord in een definiete NP; JT] to form a N’ [een gelexicaliseerde NP; JT], it gets no *-e*”.

Ten tweede, het hoofdkenmerk van idiomaticiteit, met name de niet-compositionele semantiek, is geen discreet maar wel een gradueel gegeven: Gibbs (1995) wijst op metaforische en metonymische relaties tussen de letterlijke, compositionele betekenis en de figuurlijke, niet-compositionele betekenis van idiomen; Nunberg, Wasow & Sag (1994) introduceren de distinctie tussen “idiomatically combining expressions” en “idiomatic phrases” om aan te geven dat de meeste idiomen een minimum aan compositionaliteit en transparantie bezitten. Geeraerts (1995) combineert beide kritieken op de niet-compositionele betekenis van idiomen door middel van de noties isomorfie (i.e. de één-op-één-relatie tussen vorm en betekenis) en motivatie (i.e. de transparantie van de relatie tussen vorm en betekenis). De voorbeelden in (1) illustreren de vier logische combinaties van de waarden van dit begrippenpaar: isomorf én gemotiveerd (1a), isomorf én niet-gemotiveerd (1b), niet-isomorf én gemotiveerd (1c), en niet-isomorf én niet-gemotiveerd (1d). In (1a) is er een transparante metaforische relatie tussen de letterlijke betekenis ‘de koe bij de horens vatten’ en de figuurlijke betekenis ‘een probleem in de kern aanpakken’. Daarnaast is er een één-op-één-relatie tussen de samenstellende delen van de uitdrukking en de verschillende betekenisaspecten: ‘koe’:‘probleem’, ‘horens’:‘kern van het probleem’, ‘vatten’:‘aanpakken’. Voor (1d) is een analoge analyse zonder enige twijfel uitgesloten.

- (1) a. de koe bij de horens vatten  
 b. met spek schieten  
 c. met de handen in het haar zitten  
 d. de kat de bel aanbinden  
 (voorbeelden overgenomen uit Geeraerts 1995: 62)

Deze voorbeelden bewijzen dat er verschillende gradaties van idiomaticiteit en semantische (in)transparantie zijn. Toegepast op adjectief-nomenparen is het zeer moeilijk om in onze materiaalverzameling uitdrukkingen te vinden die in (1d) passen: de voorbeelden die we intuïtief als meest idiomatisch aanvoelen, *stoffelijk overschot*, *zelfstandig naamwoord*, *goede woord* en *rechte pad*, zijn naar onze mening isomorf én gemotiveerd. Dit is wellicht een gevolg van de basisfunctie van een NP: de identificatie van een referent/concept (N) waaraan een eigenschap wordt toegekend (A). Zelfs wanneer we “exosemantische woordgroepen” beschouwen (van Sterkenburg 1993), blijken de meeste voorbeelden een isomorfe relatie te vertonen tussen de betekenis en de samenstellende delen. Dit is het geval bij de

voorbeelden in (2a). Bovendien is er een transparante metaforische relatie tussen de letterlijke en de figuurlijke betekenis bij de meeste van deze voorbeelden: *vliegend fort* ('zware bommenwerper'), *dooie pier* ('loom iemand'), *zwarte kunst* ('toverij'), *linke soep* ('riskante boel'). De isomorfie en transparantie in de voorbeelden in (2b) is minder duidelijk. Deze voorbeelden zijn echter zeer beperkt in aantal, getuige het ontbreken van zulke tournures in onze materiaalverzameling.

- (2) a. kleine boodschap, vliegend fort, blauwe maandag, afgelikte boterham, dooie pier, flauwe kul, linke soep, zwarte kunst  
 b. gillende keukenmeid, blinde vink, hard gelag, halve zool, halve gare (voorbeelden overgenomen uit Van Sterkenburg 1993: 133)

Ten derde, de identificatie van idiomatieit gaat gepaard met idiolectische verschillen (Moon 1998: 8; Schenk 1995: 256-257). Everaert (1993: 5-7), die idiomatieit door de praktische bril van de lexicograaf bekijkt, toont aan dat woordenboeken en zelfs lexicografen onderling niet consistent zijn bij de identificatie van idiomaten.

Ten vierde, een alternatieve benadering is de identificatie van idiomatische uitdrukkingen op basis van hun syntactische defectiviteit (Schenk 1995). Nunberg, Sag & Wasow (1994) beschouwen syntactische defectiviteit als een vormelijk neveneffect van semantische defectiviteit: de gradatie in semantische transparantie en niet-compositionaliteit vertaalt zich in een verschillende graad van toepasbaarheid van syntactische operaties. Matthews (1991) toont aan dat de verschillende syntactische tests tot contrasterende resultaten leiden. Het is dan ook niet verwonderlijk dat Everaert (1993: 18) syntactische en morfologisch defectiviteit niet als een noodzakelijke voorwaarde beschouwt voor de definitie van vaste uitdrukkingen, een begrip dat trouwens ruimer is dan idioom (zie paragraaf 11.1.1.2).

Dit korte – en onvolledige – overzicht van de problemen die gepaard gaan met de definitie van idiomatieit toont aan dat we te maken hebben met een problematisch concept dat niet op eenduidige wijze gedefinieerd kan worden, laat staan dat de individuele attestaties probleemloos in het reële taalgebruik geïdentificeerd kunnen worden. De (niet-)compositionaliteit van de betekenisstructuur is bijgevolg uitgesloten als identificatiecriterium van de observaties waar het adjectief-nomenpaar een vaste uitdrukking is of een naamstatus heeft. In de volgende paragraaf stellen we dan ook een alternatieve benadering voor die gebaseerd is op de lexicale associatie tussen het adjectief en het substantief.

#### *11.1.1.2 Een alternatieve benadering: de lexicale collocabiliteit van het adjectief-nomenpaar*

We stellen de alternatieve benadering in drie stappen voor. Eerst definiëren we het concept lexicale collocatie en plaatsen we dit begrip in de ruimere context van de benoemingsmechanismen in het Nederlands. Daarna becommentariëren we de

operationalisering van deze notie. Tot slot lichten we de statistische techniek toe die we gebruiken om de lexicale collocabiliteit van adjectief-nomenparen te meten.

### *Definitie*

De centrale vraag luidt: hoe identificeren we op een objectieve en repliceerbare wijze de adjectief-nomenparen met naamstatus? Bij het beantwoorden van deze vraag volgen we het voorstel van Everaert (1993) om de notie van idiomaticiteit uit te breiden tot lexicale collocaties of vaste verbindingen. Dit is: “een combinatie van twee of meer woorden waarvan de verbondenheid gefixeerd is gegeven de betekenis van de woordgroep in een taalgemeenschap” (Everaert 1993: 18). Deze notie omvat ook tournures met een volkomen transparante betekenis die niettemin lexicaal gefixeerd zijn, zoals *in arren moede*, *iemand een loer draaien* of *te berde brengen* (Everaert 1993: 8-9). Een vergelijkbare definitie vinden we bij Geeraerts (1986: 134), die lexicale collocabiliteit definieert als “idiosyncratische restricties op de verbindbaarheid van lexicale elementen”.<sup>2</sup>

Voor de lexicale/semantische conditionering van de buigingsalternantie doen we een beroep op lexicale collocaties. We concentreren ons hierbij op de lexicale fixatie binnen adjectief-nomenparen. De algemeenheid van deze begrippen is geschikt voor een kwantitatieve implementatie en sluit een semantische interpretatie/analyse niet *a priori* uit, zoals we zullen aantonen in het lexicaal en semantisch onderzoek in deel IV. Semantische en syntactische restricties op de combineerbaarheid van elementen zijn in deze optiek niets anders dan generalisaties over verschillende lexicale preferentiepatronen. De lexicale items die het direct object van het werkwoord *eten* instantiëren, zullen bijvoorbeeld allemaal inanimaat zijn en voedsel aanduiden.

De operationalisering van de adjectief-nomenparen met naamstatus als lexicale collocaties sluit aan bij het onderzoek naar syntactische benoemingsmechanismen in het Nederlands. De Caluwe (1990) en Hüning (2004) wijzen op het gebruik van gelexicaliseerde woordgroepen van het type A+N (i.e. adjectief-nomenparen), een van oorsprong syntactisch mechanisme, wanneer de morfologische benoemingsprocédés te kort schieten: *dikke darm*, *zure regen*, *harde schijf*. Toegepast op de buigingsalternantie bij neutra betekent dit dat het onverbogen buigingsmorfeem signaleert dat de adjectief-nomensequentie als een benoemingsmechanisme, i.e. als een adjectief-nomenpaar met naamstatus, fungeert en niet als een typische syntactische naamwoordelijke groep. Voor een semantische categorisatie van gelexicaliseerde adjectief-nomenparen verwijzen we naar Van Sterkenburg (1993). De semantische interpretatie van de adjectief-nomencollocaties behandelen we in hoofdstuk 17.

Alvorens we de identificatie van de lexicale collocaties toelichten, vestigen we de aandacht op de functie van het collocatieonderzoek om eventuele misverstanden te vermijden. Ten eerste, we zijn niet geïnteresseerd in het ontwerpen van een algemeen theoretisch en descriptief kader voor adjectief-nomenparen met naamstatus. Voor aanzetten tot een algemeen kader verwijzen we naar Booij (2002a), De Caluwe (1990), Hüning (2004) en Van Sterkenburg (1993). Ten tweede, als gevolg

van de intuïtieve basis van de meeste criteria voor de identificatie van idiomen en als gevolg van het gebrek aan empirisch onderzoek naar adjectief-nomencollocaties in het Nederlands beschikken we niet over een objectief referentiekader om de kwaliteit van de geïdentificeerde collocaties te evalueren. We beperken ons bijgevolg tot het voorstellen van een kwantitatieve procedure die we evalueren op basis van de impact op de buigingsalternantie, wat tot nader order de basisdoelstelling van dit onderzoek is.

### *Identificatie*

De volgende vraag luidt dan: hoe operationaliseren we de notie van lexicale collocatie? We beschouwen de opeenvolging van een adjectief en een substantief als een lexicale collocatie enkel en alleen indien ze vaker samen voorkomen dan kan worden aangenomen op basis van hun afzonderlijke frequenties. Dit is een distributionele oplossing: op basis van de distributies van het adjectief en het substantief in het corpus meten we of de frequentie van het adjectief-nomenpaar significant groter is dan verwacht. Woordparen die lexicaal gefixeerd zijn, scoren hoger voor deze maat dan woordparen die tot stand komen als gevolg van de vrije combinatie van elementen. De centrale notie bij deze identificatieprocedure van lexicale collocaties is de habitualiteit van het adjectief-nomenpaar: lexicale collocaties zijn adjectief-nomenparen die verankerd zijn in het taalgebruik van een taalgemeenschap.

Everaert (1993: 20), die nochtans de notie vaste verbinding prefereert boven de notie idioom (cf. *supra*), is een kwantitatieve analyse niet genegen: “En let wel, ik heb het hier niet over een statistische interpretatie van de notie vaste verbinding”. Zijn reserve ten overstaan van statistisch onderzoek hangt wellicht samen met de kwantitatieve implementatie die hij voor ogen heeft:

“Aantonen dat een bepaalde woordcombinatie statistisch gezien veel voorkomt, is iets anders dan aantonen dat iets een vaste verbinding is [...]. De woordgroepen *de minister heeft voorgesteld* zal (in geschreven taal) misschien statistisch veel frequenter voorkomen dan *de vrouw heeft voorgesteld* of *de vogel is gevlogen* maar dat maakt de eerstgenoemde nog niet noodzakelijkerwijs een vaste verbinding en het betekent ook niet dat de laatstgenoemde géén vaste verbinding zou zijn.”

Bij de bespreking van de verschillende kwantitatieve associatiematen zullen we dadelijk betogen dat frequentietellingen geen betrouwbare indicator van lexicale collocabiliteit zijn. Daarnaast beperken we de materiaalverzameling tot een deelcorpus dat uitsluitend NP's bevat om de ruis tot een minimum te beperken. Tot slot wijzen we op een belangrijk verschil tussen Everaerts doelstelling en onze doelstelling: in tegenstelling tot Everaert, die in het geciteerde artikel een lexicografisch perspectief aanneemt, zijn wij niet geïnteresseerd in een (zo volledig mogelijke) lijst met vaste uitdrukkingen om deze in een woordenboek op te nemen, maar enkel in de

lexicale associatiesterkte tussen het adjectief en het substantief in de observaties in onze materiaalverzameling en de invloed van deze associatiesterkte op de selectie van het buigingsmorfeem. Niettemin zijn we van oordeel dat Everaerts (1993) voorstel om vaste verbindingen op een intuïtieve basis te identificeren de ontwikkeling van een objectiveerbare procedure bemoeilijkt: het probleem wordt enkel verlegd van de intuïtieve identificatie van idiomen naar de intuïtieve identificatie van vaste verbindingen. Het uiteindelijke criterium om te beslissen of een wending al dan niet een vaste uitdrukking is, blijft ten gronde impressionistisch en bijgevolg aleatoir.

### *Kwantificatie*

Dit brengt ons bij de laatste vraag: hoe meten we de lexicale associatie tussen twee elementen? Met andere woorden: hoe kwantificeren we de toevalligheid van de associatie en de associatiekracht tussen twee opeenvolgende lexicale elementen? De lezer die niet in de technische details geïnteresseerd is, kan naar de volgende paragraaf overgaan.

In de (Angelsaksische) corpuslinguïstiek bestaat er een lange traditie van collocatieonderzoek, in eerste instantie voor de identificatie van recurrente patronen die in woordenboeken (Sinclair 1995) en grammatica's (Biber et al. 1999) worden opgenomen als 'vaste verbinding' of als voorbeeld. Verschillende maten zijn in de loop der jaren voorgesteld: de absolute frequentie van de sequentie, de *pointwise mutual information* (MI; Church & Hanks 1990; Stubbs 1995), de *t*-test (Hunston 2001; Stubbs 1995), Pearsons  $\chi^2$ -test, de *Fisher Exact*-test (Pedersen 1996) en de *log likelihood*-test ( $G^2$ ; Dunning 1993). Het is zeer moeilijk – zo niet onmogelijk – om uit deze reeks tests de meest geschikte te kiezen, te meer omdat algemene besprekingen (Manning & Schütze 2002: 151-189) en empirische evaluaties (Evert & Krenn 2001) nog steeds een zeldzaamheid zijn.<sup>3</sup> We gebruiken de *log likelihood*-statistiek voor de identificatie van de adjectief-nomencollocaties. We verantwoorden deze keuze door enerzijds de tekortkomingen van de andere maten aan te duiden en door anderzijds op de geschiktheid van de *log likelihood*-statistiek te wijzen voor de identificatie van recurrente adjectief-nomenparen. Hierbij maken we hoofdzakelijk gebruik van Manning & Schütze (2002) en Evert & Krenn (2001).

Bij de bespreking van de verschillende technieken verwijzen we door middel van  $w_1$  en  $w_2$  naar de twee items waartussen de associatie wordt berekend. Wanneer we de verschillende maten beschouwen, vallen onmiddellijk de frequentietellingen en de MI-index (i.e. de *pointwise mutual information*) op: in tegenstelling tot de andere technieken zijn dit geen associatiematen. Ze meten bijgevolg niet de kracht van de associatie tussen twee elementen en ze bevatten geen indicatie over de toevalligheid van het samen voorkomen van de elementen in kwestie. Frequentietellingen kijken enkel naar de positieve evidentie, c.q. de contexten waar  $w_1$  en  $w_2$  samen voorkomen, en negeren de negatieve evidentie, c.q. de contexten waar deze items niet samen voorkomen. De hoge frequentie van een adjectief-nomenpaar kan met

andere woorden een neveneffect zijn van de hoge woordfrequenties van het adjectief en/of het substantief. Bijgevolg is het gebruik van de absolute frequentie snel opgegeven ten voordele van de MI-index en de associatiematen. De MI-index is evenmin een associatiemaat en berekent de informatie die  $w_1$  over  $w_2$  verstrekt. Met andere woorden: des te hoger de MI-index, des te voorspelbaarder de informatie in  $w_2$  wordt op basis van de informatie in  $w_1$ . Het belangrijkste nadeel van de MI-index is dat lage frequenties consistent overschat worden.

Laten we nu naar de associatietests kijken. De  $t$ -test valt af omdat deze test een normale distributie veronderstelt, wat doorgaans niet het geval is in natuurlijke taal. Daarnaast leent deze maat zich beter voor de identificatie van grammaticale collocaties dan voor de identificatie van lexicale collocaties (Stubbs 1995). De  $\chi^2$ -test is dan weer minder betrouwbaar bij lage frequenties. De  $G^2$ -statistiek is het meest geschikt voor onze doelstelling. In vergelijking met de MI-index en de  $\chi^2$ -statistiek doet de  $G^2$ -statistiek het goed bij lage frequenties (Dunning 1993; Manning & Schütze 2002). Bij een diagnostisch onderzoek dat de accuratesse van bovenvermelde maten (met uitzondering van de *Fisher Exact*-test) vergelijkt bij de identificatie van verschillende types lexicale collocaties, wordt het beste resultaat voor adjectief-nomencollocaties in het Duits bereikt door de  $G^2$ -statistiek, zowel bij hoge frequenties ( $f(\text{AN}) \geq 5$ ) als bij lage frequenties ( $2 \leq f(\text{AN}) < 5$ ) (Evert & Krenn 2001). Evert & Krenn (2001: 194-195) signaleren wel dat de  $G^2$ -statistiek niet geschikt is voor de identificatie van *hapax legomena*.<sup>4</sup> Rest nog de *Fisher Exact*-statistiek (zie ook paragraaf 7.1). Als exacte test stelt deze statistiek geen vereisten aan de frequentie van  $w_1$  en  $w_2$ , of aan de frequentieverdeling van het materiaal. Het belangrijkste nadeel van deze test is de computationele kostprijs: net doordat het een exacte test is, stoelt ze op een zware berekening. Het voordeel van de *Fisher Exact*-test ten overstaan van de  $G^2$ -statistiek ligt bij de identificatie van de *hapax legomena* (en de *dis legomena*). Voor de *hapax legomena* volgen we de vuistregel van Weeber, Vos & Baayen (2000) om deze elementen niet in collocatieonderzoek op te nemen. Door hun lage frequentie rijst bovendien de vraag wat de representativiteit van *hapax legomena* en *bis legomena* is. Tot slot tonen Weeber, Vos & Baayen (2000) aan dat de resultaten van de *Fisher Exact*-test en de  $G^2$ -statistiek vergelijkbaar zijn voor de identificatie van collocaties tussen items met een lage frequentie.

Om kort te gaan: de  $G^2$ -statistiek is niet alleen een betrouwbare associatiemaat bij lage frequenties (*hapax* en *dis legomena* buiten beschouwing gelaten), maar levert bovendien de beste resultaten op bij de identificatie van adjectief-nomencollocaties.

De  $G^2$ -statistiek, i.e. de *log likelihood*-ratio, berekent de ratio tussen  $H_0$  (dit is: er is geen significante associatie tussen  $w_1$  en  $w_2$ ) en de waarschijnlijkheid van de geobserveerde gegevens gegeven  $H_1$  (dit is: er is wel een significante associatie tussen  $w_1$  en  $w_2$ ).

	$w_1$	$\neg w_1$	$\Sigma$
$w_2$	$a_{11}$	$a_{12}$	$a_{1+}$
$\neg w_2$	$a_{21}$	$a_{22}$	$a_{2+}$
$\Sigma$	$a_{+1}$	$a_{+2}$	$a_{++}$

Tabel 11(1): Schematische voorstelling van een contingentietabel

Op basis van de schematische contingentietabel in tabel 11(1), wordt de  $G^2$ -statistiek door middel van de volgende formule berekend:<sup>5</sup>

$$(3) \quad G^2(w_1, w_2) = 2(a_{11} \ln(a_{11}) + a_{12} \ln(a_{12}) + a_{21} \ln(a_{21}) + a_{22} \ln(a_{22}) + a_{++} \ln(a_{++}) - a_{1+} \ln(a_{1+}) - a_{2+} \ln(a_{2+}) - a_{+1} \ln(a_{+1}) - a_{+2} \ln(a_{+2}))$$

Het resultaat van deze formule is een score die de associatiekracht tussen  $w_1$  en  $w_2$  kwantificeert: hoe groter de score voor  $G^2(w_1, w_2)$ , hoe sterker de lexicale associatie tussen beide elementen. Doordat de  $G^2$ -statistiek asymptotisch verdeeld is volgens de  $\chi^2$ -distributie kan de significantie van de associatie bepaald worden: voor  $G^2(w_1, w_2) \geq 3.84$  (i.e. de kritische waarde voor  $\alpha = 0.05$  in een  $2 \times 2$ -contingentietabel) is de associatie tussen  $w_1$  en  $w_2$  significant en spreken we van een significante lexicale collocatie.<sup>6</sup> In dit onderzoek reserveren we de term lexicale collocatie voor de adjectief-nomenparen met een significante  $G^2$ -score. Daarnaast kunnen we binnen de verzameling observaties met een significante  $G^2$ -score een rangorde aanbrenge in functie van de grootte van deze score. Dit beantwoordt aan de vaststelling dat idiomatieit en – bij uitbreiding – lexicale collocabiliteit graduele noties zijn (zie paragraaf 11.1.1.1).

Toegepast op de lexicale/semantische conditionering van de adjectivische buigingsalternantie hebben we de volgende informatie nodig voor ieder adjectief-nomenpaar  $A_i N_j$  in de materiaalverzameling:

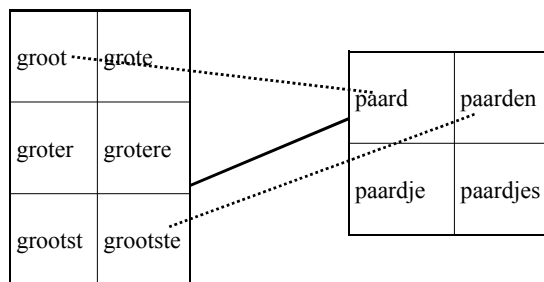
- het aantal voorkomens van  $A_i N_j$  ( $a_{11}$  in tabel 11(1))
- het aantal voorkomens van  $A_i$  zonder  $N_j$ , i.e.  $A_i \neg N_j$  ( $a_{21}$ )
- het aantal voorkomens van  $N_j$  zonder  $A_i$ , i.e.  $\neg A_i N_j$  ( $a_{12}$ )
- het aantal bigrammen<sup>7</sup> zonder  $A_i$  én zonder  $N_j$ , i.e.  $\neg A_i \neg N_j$  ( $a_{22}$ )

Doordat we enkel kijken naar de opeenvolging van  $A_i N_j$  is de co-text beperkt tot één positie links van  $N_j$  en één positie rechts van  $A_i$ . We berekenen de  $G^2$ -scores bijgevolg op basis van bigrammen met een *span* gelijk aan 1.<sup>8</sup>

Er zijn nog twee technische restricties op het gebruik van de  $G^2$ -statistiek. Ten eerste, de  $G^2$ -statistiek is een tweezijdige test. Een significante score voor  $G^2(w_1, w_2)$  betekent dat de associatie tussen  $w_1$  en  $w_2$  significant afwijkt van de  $H_0$  die vooropstelt dat er geen associatie is. Met andere woorden: een significante  $G^2$ -score kan betekenen dat  $w_1$  en  $w_2$  significant vaker dan verwacht samen voorkomen maar ook dat ze significant minder vaak dan verwacht samen voorkomen. Aangezien we enkel geïnteresseerd zijn in de adjectief-nomenparen die significant vaker dan verwacht voorkomen, moet ieder adjectief-nomenpaar (AN) met een significante  $G^2$ -score aan de volgende conditie voldoen:  $f_{\text{geobserveerd}}(\text{AN}) > f_{\text{verwacht}}(\text{AN})$ . Ten tweede, rekening

houdend met het onderzoek van Evert & Krenn (2001) bekijken we enkel de adjectief-nomensequenties waarvoor de frequentie in alle cellen in de contingentietabel gelijk aan of groter dan 3 is.

De lexicale collocabiliteit wordt berekend op het niveau van het lemma.<sup>9</sup> Dit betekent dat bij de berekening van de associatiekracht tussen het adjectief en het substantief alle inflectionele realisaties in beschouwing worden genomen. Daartegenover staat de berekening van de collocabiliteit op het woordvormniveau, waar de associatiekracht tussen de afzonderlijke morfosyntactische realisatievormen wordt berekend. Dit onderscheid wordt in figuur 11(1) geïllustreerd door de volle lijn en de stippellijn: de volle lijn duidt aan dat de associatiemaat berekend wordt tussen alle mogelijke inflectionele realisatievormen van het adjectief *groot* enerzijds en alle mogelijke inflectionele realisatievormen van het substantief *paard* anderzijds; de stippellijnen illustreren de maat voor de lexicale associatie tussen de woordvormen *groot*<sub>[POS,ONV]</sub> en *paard*<sub>[ENK,-DIM]</sub>, en tussen de woordvormen *grootste*<sub>[SUP,VERB]</sub> en *paarden*<sub>[MV,-DIM]</sub>.



Figuur 11(1): Lexicale collocaties op lemma-niveau en op woordvormniveau

We voeren drie argumenten aan om de keuze voor de berekening van de lexicale collocabiliteit op lemma-niveau te motiveren. Ten eerste, als abstracte basiseenheid van de verschillende geïnflecteerde woordvormen is het lemma de meest geschikte representatie van de conceptuele inhoud van een lexicaal item.<sup>10</sup> Op de semantische interpretatie van de lexicale collocaties komen we uitgebreid terug in deel IV van dit proefschrift, waar we de vraag stellen of de lexicale collocaties een louter lexicaal fenomeen zijn dan wel of er semantische regelmatigheden aan de basis van de collocatiepatronen geponeerd kunnen worden. Ten tweede, de berekening van collocabiliteit op het niveau van de woordvorm levert een vertekend beeld op als gevolg van de structurele dominantie van de verbogen vorm in attributieve positie (zie paragraaf 3.1, figuur 3(1)). Het verbogen adjectief wordt in vijf van de zes morfosyntactische configuraties van de NP gebruikt, terwijl de onverbogen pendant slechts in één configuratie de regel is (zie figuur 3(1)) en in drie configuraties de (laag-frequente) gemarkeerde variant is (zie figuur 3(2)). Ten derde, *het*-woorden, de nominale hoofden die het onverbogen adjectief het vaakst selecteren, vertegenwoordigen slechts 27% van de Nederlandse substantieven.<sup>11</sup> Het is



bijgevolg plausibel dat in het reële taalgebruik – waarvan een corpus een staalname is – het aantal NP's met een verbogen adjectief groter is dan het aantal NP's met een onverbogen adjectief. Vertaald naar het collocatieonderzoek komt dit neer op sterkere associatiescores voor de observaties met een onverbogen adjectief dan voor de observaties met een verbogen adjectief doordat de distributie van de onverbogen adjectieven – in attributieve positie – *a priori* selectiever is dan de distributie van de verbogen adjectieven. De corpusgegevens staven deze assumptie: de gemiddelde collocatiele associatiescore op woordvormniveau is hoger voor de adjectief-nomenparen met een onverbogen adjectief (mediaan  $G^2 = 25.05$ ; gemiddelde  $G^2 = 107.30$ ) dan voor de adjectief-nomenparen met een verbogen adjectief (mediaan  $G^2 = 12.120$ ; gemiddelde  $G^2 = 17.840$ ).

Voor ieder adjectief-nomenlemmapaar onderscheiden we het *typeniveau* en het *tokenniveau*. Het typeniveau is het niveau waarop de verschillende adjectief-nomenparen geïdentificeerd worden. Op dit niveau heeft ieder adjectief-nomenlemmatype, voortaan AN-lemmatype, hetzelfde gewicht, ongeacht wat de reële frequentie is van het lemmatype in kwestie. Als we bijvoorbeeld een woordenlijst met adjectief-nomenparen in het Nederlands willen opstellen, is dit het relevante niveau. Het tokenniveau houdt rekening met de verschillende realisaties van de AN-lemmatypes, dit is: de frequentie van een AN-lemmatype in het corpus.

Tot slot lichten we de materiaalverzameling toe die de empirische basis vormt voor de berekening van de  $G^2$ -scores. Op basis van de 5<sup>e</sup> pre-release van het CGN hebben we een deelcorpus gecompileerd dat alleen NP's bevat met de structuur [DET? A N]<sub>NP</sub>. Dit corpus bevat 45,040 NP's, die goed zijn voor 102,253 woorden. Deze inkringing van het CGN tot een CGN<sub>NP</sub> vermindert de ruis bij de berekening van de lexicale associatiekracht voor de adjectief-nomenparen door de contexten waarin de beoogde combinatie onmogelijk kan optreden – i.e. alle bigrammen die niet tot een NP behoren – *a priori* uit te sluiten. Het gevolg is dat cel  $a_{22}$  in tabel 11(1) aanzienlijk kleiner wordt. Bigrammen die vanuit het perspectief van onze doelstelling zinloos zijn, omdat ze noch een adjectief noch een substantief bevatten, worden op deze wijze niet in de berekening van de collocatiele sterkte opgenomen, zoals de lemmaparen *worden doen*, *gisteren gaan*, *zijn niet of ik zijn*, die allen tot cel  $a_{22}$  in tabel 11(1) zouden behoren.

Op basis van voorgaande discussie over het gebruik van de  $G^2$ -statistiek worden alleen de adjectief-nomenparen (AN) die aan de volgende criteria voldoen als een lexicale collocatie beschouwd:

- in alle cellen van de contingentietabel voor AN is de frequentie minimum 3
- $G^2(A,N) \geq 3.84$  (i.e. de kritische waarde voor het significantieniveau gelijk aan 0.05 in een 2x2-contingentietabel)
- $f_{\text{geobserveerd}}(A,N) > f_{\text{verwacht}}(A,N)$

De adjectief-nomenparen die niet aan deze noodzakelijke en voldoende voorwaarden beantwoorden, beschouwen we niet als (significante) lexicale

collocaties. Tot slot merken we op dat deze definitie van het concept lexicale collocatie ruimer is dan introspectieve definities (zie paragraaf 11.2.1).

### *11.1.1.3 Lexicale collocaties als operationalisering van de lexicale/semantische kenmerken van het adjectief-nomenpaar*

We hebben betoogd dat traditionele semantische en syntactische tests geen perspectief bieden op een objectieerbare procedure voor de identificatie van de lexicale/semantische kenmerken van het adjectief-nomenpaar (11.1.1.1). We stellen een distributioneel en kwantitatief alternatief voor op basis van het begrip lexicale collocatie: als lexicale collocatie beschouwen we alle adjectief-nomenparen die significant vaker voorkomen dan kan worden aangenomen op basis van de afzonderlijke woordfrequenties van het adjectief en het substantief (11.1.1.2). Deze werkwijze biedt de volgende methodologische voordelen:

- De lexicale collocabiliteit van het adjectief-nomenpaar is een objectieerbare en repliceerbare operationalisatie van een vaag en ongrijpbaar begrip (hetzij idioom, hetzij adjectief-nomenpaar met naamstatus). Lezers kunnen bezwaren hebben tegen bepaalde methodologische keuzes, maar weten hoe ze de gerapporteerde resultaten moeten interpreteren.
- Voor ieder adjectief-nomenpaar in de materiaalverzameling kan de significantie en de sterkte van de lexicale associatie berekend worden. Hiermee komen we tegemoet aan de bezwaren tegen de traditionele niet-compositionaliteitsvisie op idiomaticiteit. Deze tegenwerpingen wijzen onder andere op de gradeerbaarheid van de associatie tussen lexicale items.

De behandeling van semantische factoren is een algemeen probleem bij corpus-onderzoek bij ontstentenis van een algemene objectieve procedure om semantische effecten te testen in een empirische en kwantitatieve onderzoekscontext.<sup>12</sup> Associatiematen vormen vaak het enige alternatief dat een maximale objectieerbaarheid waarborgt. Bovendien is de adequaatheid van associatiematen bij de identificatie van semantisch relevante patronen al meermaals aangetoond, zowel op de syntagmatische als tussen verschillende lexicale items (Sinclair 1991; Stubbs 1995, 2001) als op de paradigmatische als tussen items en grammaticale constructies (Stefanowitsch & Gries 2003, te verschijnen).

Met deze informatie in het achterhoofd stellen we voor om de hypothese als volgt te herformuleren:

- De kans op het onverbogen buigingsmorfeem is groter bij de adjectief-nomenparen die een significante lexicale collocatie vormen dan bij de adjectief-nomenparen die een losse, c.q. niet-significante, sequentie vormen.
- De kans op het onverbogen buigingsmorfeem stijgt in functie van de sterkte van de collocabiliteit: hoe sterker de lexicale associatie tussen het adjectief en het substantief, des te groter de kans dat de onverbogen vorm van het adjectief geselecteerd wordt.

### 11.1.2 *Relationaliteit van het adjectief*

De tweede lexicale/semantische conditie op de buigingsalternantie die we op basis van het CGN bestuderen, betreft de semantische klasse van het adjectief, meer bepaald de vraag of het een relationeel of een kwalificerend adjectief is. Booij (2002a, 2002c) en Heynderickx (2001) signaleren namelijk dat in adjectief-nomencombinaties die een semantische eenheid vormen vaak een relationeel adjectief de adjectivische positie bezet.

Relationele adjectieven worden meestal gedefinieerd in contrast met kwalificerende adjectieven: terwijl een kwalificerend adjectief de entiteit aangeduid door het substantief modificeert, drukt een relationeel adjectief een relatie uit tussen twee (nominale) entiteiten, met name de (nominale) entiteit aangeduid door het substantief en de (nominale) entiteit aangeduid door het denominale relationele adjectief. Heynderickx (2001: 133) vat de semantische interpretatie van de sequentie van een relationeel adjectief en een substantief als volgt samen:

“Om een RA [relationeel adjectief; JT] te interpreteren moet men het altijd beschouwen samen met het nomen dat het modificeert. Er wordt immers een betrekking gecreëerd tussen de conceptuele entiteiten die door de basis [het substantief aan de basis van het relationeel adjectief; JT] en de kern [het nominale hoofd; JT] opgeroepen worden. De betekenis van een RA-N-combinatie bestaat bijgevolg uit drie elementen: de conceptuele entiteiten opgeroepen door de basis en de kern en de betrekking tussen beide.”

We verduidelijken het verschil tussen kwalificerende en relationele adjectieven aan de hand van een voorbeeld. De kwalificerende adjectieven in (4) modifieren de referent waaraan het substantief refereert. Deze adjectieven specificeren de referent door het toekennen van een bepaalde eigenschap aan de referent in kwestie, zoals aangetoond door de volgende parafrases: *het huis is groot*, *het jasje is groen* en *het besje is oud*.

- (4) het grote huis, het groene jasje, het oude besje
- (5) het koninklijk besluit, het Amerikaanse leger, het politieke debat

Relationele adjectieven hebben geen modifierende functie, maar verbinden twee substantivische entiteiten. De voorbeelden in (5) kunnen bijgevolg niet geparafraseerd worden met een predicatieve constructie: *\*het besluit is koninklijk*, *\*het leger is Amerikaans* en *\*het debat is politiek*.<sup>13</sup> In de parafrase van deze voorbeelden komt noodzakelijkerwijs een voorzetsel voor dat de relatie tussen beide concepten expliciteert: ‘het besluit van de koning’, ‘het leger van (de Verenigde Staten van) Amerika’ en ‘het debat over (de) politiek/door politici’. Deze parafrases tonen aan dat de invulling van de semantische relatie tussen beide concepten zeer divers is en bovendien contextueel bepaald is. Heynderickx (2001: 133) besluit dat

de taalgebruiker op basis van zijn/haar encyclopedische kennis uit een “niet-opsombaar aantal betrekkingen” moet putten.

Uit de vaststelling van Booij (2002a, 2002c) en Heynderickx (2001) dat adjectief-nomenparen met naamstatus vaak een relationeel adjectief bevatten, leiden we de hypothese af dat het buigingsprofiel van relationele adjectieven meer overbogen vormen bevat dan het buigingsprofiel van kwalificerende adjectieven. Deze afgeleide hypothese kunnen we echter positief invullen. Als we ervan uitgaan dat het prototypische adjectief de referent van het substantief semantisch modificeert, is een relationeel adjectief semantisch gemarkeerd.<sup>14</sup> Het gebruik van het onverbogen adjectief in zulke contexten is dan gemotiveerd als indicator van de gemarkeerde semantische situatie waar het adjectief geen extra informatie over een referent in het discours verschaft, maar een relatie tussen twee nominale concepten aanbrengt. We moeten uiteraard de eventuele partiële conditionering tussen het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar en de relationaliteit van het adjectief testen. De vaststelling dat in adjectief-nomenparen met naamstatus, c.q. lexicale collocaties, vaak relationele adjectieven gebruikt worden, zegt immers niets over het voorkomen van relationele adjectieven in de overige adjectief-nomenparen, laat staan over het effect van relationele adjectieven op de buigingsalternantie.

De identificatie van de relationele adjectieven in een corpus is geen sinecure doordat we herhaaldelijk een beroep moeten doen op intuïtieve oordelen. Daarom stellen we de volgende stapsgewijze procedure voor om een maximale intersubjectiviteit te garanderen bij de codering van de relationele adjectieven.<sup>15</sup> De annotatie van de relationele adjectieven in de materiaalverzameling gebeurt in twee fases. Tijdens de eerste fase identificeren we de potentiële relationele adjectieven op basis van hun morfologische structuur. Relationele adjectieven zijn afgeleid van een substantief door middel van één van de suffixen in tabel 11(2) (Heynderickx 2001: 135).<sup>16</sup>

De potentiële relationele adjectieven worden vervolgens getest met betrekking tot hun morfologische, syntactische en semantische eigenschappen (Heynderickx 2001: 17-88):

- de afwezigheid van trappen van vergelijking: *\*het zeer politiek(e)/politieker(e) lot, het meest politiek(e)/politiekste lot*
- de negatie door middel van *niet-* in plaats van *on-*: *\*het onpolitieke lot vs. het niet-politieke lot* (in contrast tot *het politieke lot*)
- de onmogelijkheid om een relationeel adjectief in predicatieve positie te plaatsen: *\*het lot is politiek*
- de onmogelijkheid om een relationeel adjectief en een kwalificerend adjectief in een nevenschikkend verband te plaatsen: *\*het goed(e) en politiek(e) lot vs. het economisch(e) en politiek(e) lot*

Anders gezegd: relationele adjectieven ontberen de morfologische en syntactische eigenschappen die typisch met adjectieven geassocieerd worden. We

kunnen deze criteria echter niet als noodzakelijke en voldoende voorwaarden hanteren. De enige constante eigenschap van relationele adjectieven is dat “ze een betrekking stichten, denominaal zijn en niet met een KA [kwalificerend adjectief; JT] kunnen worden gecoördineerd” (Heynderickx 2001: 88). We zijn bijgevolg genoodzaakt om de semantische kenmerken in beschouwing te nemen, te meer omdat denominale adjectieven polyseem kunnen zijn tussen een relationele en kwalificerende lezing (Heynderickx 2001: 116-125): in een adjectief-nomenpaar zoals *zomerse dag* kan het denominale adjectief zowel relationeel (‘een dag in de zomer’) als kwalificerend (‘een dag als in de zomer’) zijn. Omdat het onderscheid niet altijd duidelijk uit de context is af te leiden, kiezen we bij twijfelgevallen consistent voor de prototypische niet-relationele interpretatie. Merk tot slot op dat de nominale basisvorm van het relationele adjectief niet altijd ondubbelzinnig identificeerbaar is, zoals in *mobiel*, *politiek* of *katholiek*.

Suffix	Voorbeeld
-aal	<i>het regionaal structuurplan, het sociaal beleid, het neuraal netwerk</i>
-(i)aans	<i>het cartesiaans assenstelsel</i>
-air	<i>het universitair gebeuren,</i>
-eel	<i>het emotionele aspect, het seksuele taboe</i>
-ief	<i>het instinctieve systeem, het repressieve toezicht, het preventief toezicht</i>
-iek	<i>het politieke lot, het katholiek wetboek</i>
-iel	<i>ons mobiel netwerk</i>
-ig	<i>het tienjarig bestaan, het bolvormig licht</i>
-isch	<i>het technisch onderwijs, het limbische systeem, het medisch urgentieteam</i>
-lijk	<i>het bisschoppelijk purper, het ouderlijk gezag, het menselijk brein</i>
-ling	<i>het mondeling verslag, het mondeling examen</i>
-s	<i>het Zweedse parlement, het bijbelse scheppingsverhaal</i>

Tabel 11(2): Suffigering van de relationele adjectieven

Hoewel we de onderscheiding tussen de relationele en de kwalificerende adjectieven onder de noemer semantische eigenschappen van het adjectief behandelen, is deze distinctie bepalend voor de betekenis van de NP. De kernvraag luidt immers: wordt de betekenis/referentie van het nominale hoofd gemodificeerd of met een ander substantivisch concept in verband gebracht? Op deze wijze vormt de analyse van het buigingsprofiel van de relationele adjectieven de aanzet tot de semantische analyse van de buigingsalternantie in deel IV.

adjectief-nomenpaar	adjectief
lexicale collocabiliteit	relationeel vs. kwalificerend adjectief

Tabel 11(3): Overzicht van de operationalisering van de lexicale/semantische conditionering van de buigingsalternantie in het CGN

Tabel 11(3) vat de lexicale/semantische parameters samen.

## 11.2 Resultaten en bespreking

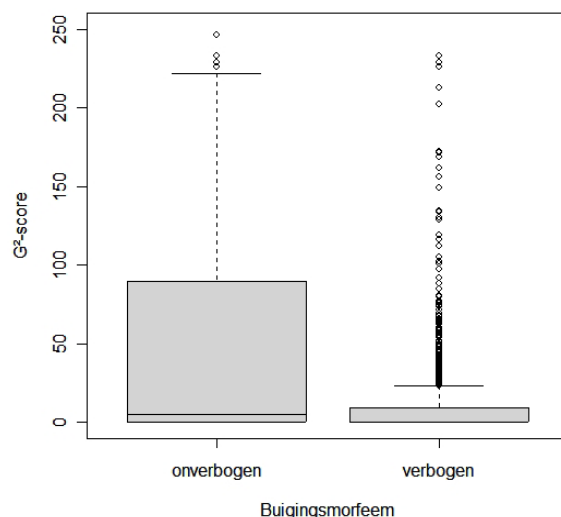
De bespreking van de resultaten van het empirisch onderzoek bestaat uit drie delen. Eerst analyseren we het buigingseffect van de lexicale collocaties (11.2.1). Vervolgens kijken we naar de keuze van het buigingsmorfeem bij de relationele adjectieven (11.2.2). We sluiten de empirische analyse af met een stratumanalyse om de lexicale/semantische conditionering van het – vanuit prosodisch oogpunt – onverklaarbare buigingsprofiel van enkele prosodische patronen te testen (11.2.3). In het vorige hoofdstuk hebben we immers op intuïtieve basis aangevoerd dat het onverwachte buigingseffect bij sommige prosodische patronen een gevolg is van de naamstatus van de adjectief-nomenparen die deze patronen realiseren.

### 11.2.1 *Lexicale/semantische conditionering van het adjectief-nomenpaar*

Alvorens de concrete invloed van de lexicale collocabiliteit op de buigingsalternantie te bespreken, kijken we naar enkele overzichtsstatistieken voor de  $G^2$ -statistiek. De  $G^2$ -scores bestrijken een gebied variërend van 0 tot en met 1721.0 (mediaan  $G^2 = 45.950$ ; gemiddelde  $G^2 = 109.800$ ).

Figuur 11(2) toont de verdeling van de verschillende waarden van de  $G^2$ -statistiek over de verbogen en de onverbogen adjectieven aan de hand van een boxplot. De X-as in figuur 11(2) toont twee boxplots, één voor de AN-lemmatokens met een onverbogen adjectief (linker boxplot) en één voor de AN-lemmatokens met een verbogen adjectief (rechter boxplot). Op de Y-as zetten we de  $G^2$ -scores uit.<sup>17</sup> Van een boxplot kan men de kwartielen, de mediaan en de grootste en de kleinste waarde aflezen. De mediaan markeert het midden van een geordende frequentieverdeling. Dit is: wanneer we alle meetpunten rangschikken van de laagste  $G^2$ -score naar de hoogste  $G^2$ -score ligt 50% van de meetpunten onder de mediaan en 50% erboven. Wanneer we de meetpunten niet in twee maar in vier deelverzamelingen groeperen, verkrijgen we kwartielen: het eerste kwartiel identificeert het eerste kwart van de meetpunten, het tweede kwartiel het tweede kwart, enz. Deze vier deelverzamelingen worden in een boxplot gevisualiseerd. De mediaan wordt aangeduid door de verticale lijn in de grijze box. De grijze rechthoek (box) toont het tweede kwartiel (beneden de mediaan) en het derde kwartiel (boven de mediaan). De snorharige uitlopers identificeren het eerste (uitloper beneden de box) en het vierde kwartiel (uitloper boven de box). Daarnaast toont de plot nog enkele *outliers*, punten die te ver van de mediaan afwijken en de boxplot nodeloos zouden uitrekken.

De AN-lemmatokens die niet voldoen aan de criteria die we in paragraaf 11.1.1.2 geformuleerd hebben om van een significante lexicale collocatie te spreken, krijgen een  $G^2$ -score gelijk aan 0. Dit wil zeggen dat we deze AN-lemmatokens niet als lexicale collocaties beschouwen.



Figuur11(2): Verdeling van de scores van de  $G^2$ -statistiek over beide buigingsvarianten

De boxplots in figuur 11(2) beantwoorden niet volledig aan bovenstaande beschrijving van de boxplot: bij de boxplot voor de AN-lemmatokens met een onverbogen adjectief ontbreekt de snorharige uitloper voor het eerste kwartiel beneden de box; bij de boxplot voor de AN-lemmatokens met een verbogen adjectief ontbreken het eerste en het tweede kwartiel. Dit wil zeggen dat meer dan een kwart van de AN-lemmatokens met een onverbogen adjectief en meer dan de helft van de AN-lemmatokens met een verbogen adjectief een  $G^2$ -score gelijk aan 0 heeft en bijgevolg geen significante lexicale collocatie vormt volgens de criteria die we in paragraaf 11.1.1.2 geformuleerd hebben. Er zijn slechts 482 AN-lemmatypes (13.87%) die aan deze criteria voldoen. Deze 482 types vertegenwoordigen 1,166 tokens (33.60%). Deze kleine aantallen zijn een gevolg van de zeer lage token/type-ratio: gemiddeld wordt ieder AN-lemmatype door 1.42 AN-lemmatokens gerealiseerd; ruim 2,924 (84.14%) AN-lemmatypes zijn *hapax legomena*.<sup>18</sup> De reden hiervoor is dat het CGN het taalgebruik van verschillende sprekers in uiteenlopende gesprekssituaties verzamelt (zie paragraaf 6.1). Het gevolg is een grote heterogeniteit van gespreksonderwerpen en lage token/type-ratio's. Het relatief kleine aantal AN-lemmatypes en AN-lemmatokens met een significante  $G^2$ -score motiveert het vierde deel van dit onderzoek, met name de analyse van de lexicale en semantische conditionering van de buigingsalternantie op basis van *De Standaard*.<sup>19</sup>

De boxplot voor de AN-lemmatokens met een onverbogen adjectief bestrijkt een grotere zone op het  $G^2$ -continuüm dan de tegenhanger voor de AN-lemmatokens met een verbogen adjectief: behalve het minimum, dat voor beide buigingsvarianten gelijk aan 0 is, zijn de waarden van de  $G^2$ -statistiek groter bij de onverbogen adjectieven dan bij de verbogen adjectieven. Deze verdeling is conform met de

hypothese: de AN-lemmatokens met een onverbogen adjectief hebben een hogere  $G^2$ -score dan de AN-lemmatokens met een verbogen adjectief, dit wil zeggen ze hebben een hogere collocabiliteit.

Na deze verkennende analyse van de relatie tussen de  $G^2$ -score en de keuze van het adjectivische buigingsmorfeem, toetsen we nu de hypothese dat adjectief-nomenparen die een lexicale collocatie vormen significant vaker het onverbogen buigingsmorfeem selecteren dan adjectief-nomenparen die geen lexicale collocatie vormen en de hypothese dat de kans op een onverbogen adjectief groeit naarmate de lexicale associatiesterkte tussen het adjectief en het substantief groter wordt.

#### *Lexicale collocabiliteit*

Het onderzoek naar het buigingseffect van de lexicale collocabiliteit gebeurt in twee stappen. In eerste instantie vergelijken we de buigingsprofielen van de adjectief-nomenparen, i.e. van de AN-lemmatokens, die een lexicale collocatie vormen en van de adjectief-nomenparen die geen lexicale collocatie vormen volgens de criteria die we in paragraaf 11.1.1.2 hebben voorgesteld. Daarna spitsen we ons toe op de AN-lemmatokens die een lexicale collocatie vormen om na te gaan of het aandeel onverbogen vormen stijgt naarmate de lexicale associatie tussen het adjectief en het substantief sterker wordt.

Tabel 11(4) bevat de buigingsprofielen voor de observaties waar het adjectief-nomenpaar een (significante) lexicale collocatie vormt ('lex.col') en de observaties waar het adjectief-nomenpaar geen (significante) lexicale collocatie vormt ('-lex.col'). De frequentieverdeling in tabel 11(4) is significant ( $\chi^2 = 186.0279$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	-lex.col		lex.col		$\Sigma$	
<b>verbogen</b>	2,722	82.58	1,088	65.23	3,810	76.75
<b>onverbogen</b>	574	17.42	580	34.77	1,154	23.15
<b><math>\Sigma</math></b>	<b>3,296</b>	<b>100.0</b>	<b>1,668</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

*Tabel 11(4): Invloed van de lexicale collocabiliteit van het adjectief-nomenpaar op de adjectivische buigingsalternantie*

De kans op een onverbogen vorm verdubbelt wanneer het adjectief en het substantief significant colloqueren:  $RR_{\text{onverbogen}} = 1.99$  [1.81,2.20]. De vergelijking van de AN-lemmatokens die een significante collocatie vormen met de AN-lemmatokens die geen significante collocatie vormen, resulteert in een OR die gelijk is aan 2.53 [2.21,2.89].

Deze scores stijgen wanneer we ons concentreren op de deelverzameling met een significante  $G^2$ -score. Omdat de significante  $G^2$ -waarden een groot gebied bestrijken (de  $G^2$ -score van de observaties met een adjectief-nomencollocatie behoort tot het interval [3.84,1721.0]), verdelen we de AN-lemmatokens in kwartielen op basis van hun  $G^2$ -score.<sup>20</sup> Het aandeel onverbogen vormen in tabel 11(5) klimt progressief naarmate de lexicale associatiesterkte (i.e. de  $G^2$ -score) tussen het adjectief en het



substantief stijgt, met een absolute voorkeur voor de onverbogen vorm in het vierde kwartiel ('lex.col<sub>kw.4</sub>') ( $\chi^2 = 263.7014$ ,  $df = 3$ ,  $p < 2.2e-16$ ).

	lex.col <sub>kw.1</sub>	lex.col <sub>kw.2</sub>	lex.col <sub>kw.3</sub>	lex.col <sub>kw.4</sub>	$\Sigma_{LEX.COL}$					
<b>verb</b>	370	88.52	303	73.19	260	62.80	155	36.73	1,088	65.23
<b>onverb</b>	48	11.48	111	26.81	154	37.20	267	63.27	580	34.77
<b><math>\Sigma</math></b>	<b>418</b>	<b>100.0</b>	<b>414</b>	<b>100.0</b>	<b>414</b>	<b>100.0</b>	<b>422</b>	<b>100.0</b>	<b>1,668</b>	<b>100.0</b>

Tabel 11(5): Invloed van de sterkte van de lexicale collocabiliteit van het adjectief-nomenpaar op de adjectivische buigingsalternantie

De OR's in tabel 11(6) zijn berekend ten overstaan van de observaties zonder significante collocabiliteit ('-lex.col' in tabel 11(4)):

	RR <sub>onv</sub>	95% BI	OR	95% BI
lex.col <sub>kw.1</sub> vs. -lex.col	0.66	[0.50,0.87]	0.62	[0.45, 0.84]
lex.col <sub>kw.2</sub> vs. -lex.col	1.54	[1.29,1.84]	1.74	[1.37,2.20]
lex.col <sub>kw.3</sub> vs. -lex.col	2.14	[1.85,2.47]	2.81	[2.26,3.50]
lex.col <sub>kw.4</sub> vs. -lex.col	3.63	[3.27,4.03]	8.17	[6.57,10.15]

Tabel 11(6): RR's en OR's die het buigingseffect meten van de collocabiliteitskwartielen t.o.v. de afwezigheid van lexicale collocabiliteit ('-lex.col')

De RR<sub>onverbogen</sub>'s en de OR's in tabel 11(6) klimmen gestaag naarmate de lexicale associatiesterkte toeneemt, met een forse stijging in het hoogste kwartiel ('lex.col<sub>kw.4</sub>'). De enige uitzondering hierop vormen de AN-lemmatokens in het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>'): deze elementen worden significant minder vaak verbogen dan de AN-lemmatokens zonder significante G<sup>2</sup>-score. Hierop komen we dadelijk terug.

De resultaten van het collocatieonderzoek zijn conform met de hypothesen:

- De gemarkeerde onverbogen vorm identificeert op iconische wijze de situatie waar het adjectief en het substantief een lexicale collocatie vormen.
- De keuze van het onverbogen buigingsmorfeem wordt waarschijnlijker naarmate de associatiesterkte tussen het adjectief en het substantief toeneemt: de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  zijn ruim 8 maal groter bij de sterkst colloquerende adjectief-nomenparen dan bij de adjectief-nomenparen zonder significante associatie tussen het adjectief en het substantief (en de adjectief-nomenparen waarvoor de collocabiliteit niet berekend kan worden op basis van de criteria opgesteld voor het gebruik van de G<sup>2</sup>-statistiek).

Het buigingseffect van de lexicale collocabiliteit is niet absoluut of discreet, zoals sommige auteurs al dan niet expliciet beweren, maar gradueel: de selectie van de onverbogen vorm wordt waarschijnlijker naarmate de lexicale associatie tussen het adjectief en het substantief hechter wordt. Een blik op de concrete adjectief-nomenparen suggereert bovendien dat de reductie van de lexicale/semantische conditionering van de buigingsalternantie tot de adjectief-nomenparen met

naamstatus onhoudbaar is: adjectief-nomenparen zoals *half jaar, half uur, lang haar, warm eten, vorig hoofdstuk* kiezen onveranderd de onverbogen vorm, zonder dat er – naar onze mening – sprake is van een naamstatus. Daartegenover staan dan weer duidelijke voorbeelden van adjectief-nomenparen met – naar onze intuïtie – naamstatus die zonder uitzondering verbogen worden: *groene poldermodel, goede woord, Witte Huis, Rode Kruis* en *Oude Testament*.

#### *Lexicale specificiteit*

Het zal de lezer niet ontgaan zijn dat het aandeel onverbogen vormen in het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>': 11%) in tabel 11(5) significant kleiner is dan het aandeel onverbogen vormen voor de observaties waar het adjectief-nomenpaar geen lexicale collocatie vormt ('-lex.col': 17%) ( $\chi^2 = 8.9409$ ,  $df = 1$ ,  $p = 0.002788$ ). Dit is een neveneffect van het gehanteerde operationaliseringsprocédé: doordat we op basis van de distributionele kenmerken van de samenstellende elementen in een adjectief-nomenpaar berekenen of het al dan niet een lexicale collocatie is, hanteren we een definitie die ruimer is dan de traditionele notie van idiomaticiteit of adjectief-nomenpaar met naamstatus. Bijgevolg krijgen recurrente AN-lemmatypes zonder aperte semantische specialisatie, zoals de voorbeelden in (10), eveneens een significante  $G^2$ -score.

- (10) half jaar, heel jaar, ander land, groot huis, groot probleem, volgend jaar,  
heel leven, vorig jaar, goed boek

Deze groep bevat verscheidene attestaties met een verbogen adjectief. Opvallend is dat deze voorbeelden uit zeer algemene adjectieven en substantieven zijn opgebouwd, zodat we ons de vraag stellen of de lexicale specificiteit – of algemeenheid – van het adjectief en/of het substantief hier geen rol zou kunnen spelen. We stellen de volgende distributionele definitie van lexicale specificiteit voor: de lexicale specificiteit van een adjectief/substantief is omgekeerd evenredig met het aantal substantieven/adjectieven waarmee het voorkomt.

De operationalisatie van de lexicale specificiteit is gebaseerd op de distributionele kenmerken van het element in kwestie. We berekenen de lexicale specificiteit in twee fases. We beschouwen  $w_1$  in tabel 11(1) als het adjectief en  $w_2$  als het substantief. Tijdens de eerste fase wordt de lexicale specificiteit berekend van het adjectief en het substantief in iedere observatie afzonderlijk:

- lexicale specificiteit adjectief:  $P(w_2|w_1) = a_{11}/a_{1+}$
- lexicale specificiteit substantief:  $P(w_1|w_2) = a_{11}/a_{+1}$

Deze score varieert tussen 0 en 1. Een score van 1 duidt een maximale specificiteit aan, dit is: het adjectief/substantief komt in het corpus slechts met één substantief/adjectief voor. Een score die 0 benadert (0 is uitgesloten), wijst op een zeer lage lexicale specificiteit. In de tweede fase berekenen we het (ongewogen) gemiddelde voor de specificiteitsscores van alle observaties.

Tabel 11(7) vat de (ongewogen) gemiddelde specificiteitsscores voor het adjectief (A) en het substantief (N) samen voor ieder kwartiel op de lexicale collocabiliteitsschaal (zie tabel 11(5)):

	$\text{lex.col}_{\text{kw.1}}$	$\text{lex.col}_{\text{kw.2}}$	$\text{lex.col}_{\text{kw.3}}$	$\text{lex.col}_{\text{kw.4}}$	$\Sigma_{\text{LEX.COL}}$
<b>A</b>	0.0172	0.0927	0.1461	0.2730	0.1327
<b>N</b>	0.0287	0.0670	0.1562	0.2593	0.1282

*Tabel 11(7): Lexicale specificiteit van het adjectief en het substantief in de lexicale collocabiliteitskwartielen*

De gegevens in tabel 11(7) staven ons vermoeden dat het lage aandeel onverbogen vormen in het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>') in tabel 11(5) een neveneffect is van de algemene betekenis van zowel het adjectief als het substantief: gemiddeld komt een adjectief in het laagste kwartiel 'lex.col<sub>kw.1</sub>' met  $58.1$  ( $1/0.0172 = 58.1395$ ) verschillende substantieven voor tegen  $3.6$  in het hoogste kwartiel 'lex.col<sub>kw.4</sub>'; gemiddeld komt een substantief in het kwartiel 'lex.col<sub>kw.1</sub>' met  $34.8$  verschillende adjectieven voor tegen  $3.8$  in het kwartiel 'lex.col<sub>kw.4</sub>'. Deze scores voor de lexicale specificiteit in het laagste collocabiliteitskwartiel zijn bovendien lager dan bij de adjectief-nomenparen die geen lexicale collocatie vormen. In deze laatste verzameling bedraagt de gemiddelde specificiteitsscore respectievelijk  $0.0876$  voor het adjectief en  $0.092$  voor het substantief.

Een kritisch lezer kan opwerpen dat de verdeling van de buigingsmorphemen niets van doen heeft met de significantie en de hechtheid van de lexicale associatie tussen het adjectief en het substantief, maar door de lexicale specificiteit van het adjectief en/of het substantief bepaald wordt. We ontkennen niet dat er een relatie tussen beide maten bestaat, maar de lexicale collocabiliteit is gesofisticeerder dan de conditionele probabiteit die de specificiteitsberekening in feite is. Om deze stelling empirisch te onderbouwen, vergelijken we de onzekerheidscoëfficiënt voor de associatie tussen het buigingsmorfeem enerzijds en de in kwartielen opgedeelde conditionele probabiteit in de volledige materiaalverzameling ( $\Sigma_{\text{DB}}$ ) en in de deelverzameling bestaande uit de observaties waar het adjectief-nomenpaar een lexicale collocatie vormt ( $\Sigma_{\text{LEX.COL}}$ ) anderzijds.<sup>21</sup> De onzekerheidscoëfficiënt berekent de grootte van de associatie tussen twee nominale variabelen, meer bepaald de onzekerheid over de waarde van de responsvariabele (c.q. de keuze van het buigingsmorfeem) gegeven de kennis van de verklarende variabele (c.q. het kwartiel voor de lexicale specificiteit van het adjectief en het substantief). Het resultaat is een cijfer tussen 0 en 1: een score gelijk aan 0 betekent dat de verklarende variabele geen zekerheid geeft over de waarde van de responsvariabele, een score gelijk aan 1 betekent dat de verklarende variabele volledige zekerheid geeft over de waarde van de responsvariabele. Het exacte cijfer is moeilijk interpreteerbaar. Wat relevant is voor onze doelstellingen is de significantie en de onderlinge positionering van de verschillende onzekerheidscoëfficiënten.

	<b>Specificiteit A</b>	<b>Specificiteit N</b>
$\Sigma_{DB}$	0.045 [0.039,0.050]	0.035 [0.027,0.043]
$\Sigma_{LEX.COL}$	0.201 [0.185,0.217]	0.064 [0.056,0.071]

*Tabel 11(8): Onzekerheidscoëfficiënten en 95% BI'en voor de lexicale specificiteit van het adjectief en het substantief in de volledige materiaalverzameling ( $\Sigma_{DB}$ ) en in de deelverzameling met de lexicale collocaties ( $\Sigma_{LEX.COL}$ )*

Deze cijfers bevestigen dat de lexicale collocabiliteit niet gereduceerd kan worden tot de lexicale specificiteit van het adjectief en/of het substantief: zowel voor het adjectief als voor het substantief is de reductie van de onzekerheid met betrekking tot de keuze van het buigingsmorfeem significant groter in de deelverzameling met de lexicale collocaties dan in de volledige materiaalverzameling. Daarnaast blijkt de lexicale specificiteit van het adjectief een betere indicator te zijn van de buigingsalternantie dan de lexicale specificiteit van het substantief. We concluderen dat de lexicale specificiteit van het adjectief en het substantief ondergeschikt is aan de lexicale collocabiliteit tussen het adjectief en het substantief.

Tot slot kijken we naar de interne variatie binnen de AN-lemmatypes die een lexicale collocatie vormen. Hieronder verstaan we de consistentie waarmee de tokens die een AN-lemmatype realiseren een bepaalde buigingsvariant kiezen. Het aandeel AN-lemmatypes die consistent één buigingsmorfeem selecteren, bedraagt gemiddeld 83.61%. Uitgedrukt in absolute cijfers vertonen slechts 79 AN-lemmatypes die een lexicale collocatie vormen interne variatie.<sup>22</sup>

De resultaten van het collocatieonderzoek suggereren dat de lexicale collocaties niet geconstrueerd worden tijdens het derivatieve proces – door middel van regels van het type  $N' \rightarrow AP N$  – maar dat deze sequenties een eenheid in het lexicon vormen en als geheel in de zin worden ingepast. Dit sluit aan bij recente ontwikkelingen in het kader van de Constructiegrammatica, waar de strikte opdeling tussen een grammaticale en een lexicale component wordt verworpen (Croft & Cruse 2004; Tomasello 2000), en – minder recente – ontwikkelingen in de Angelsaksische corpuslinguïstische traditie, waar al geruime tijd op het sterk colloquerende karakter van het taalgebruik wordt gewezen (Sinclair 1991; Stubbs 2001). Deze visie stelt uiteraard het syntactisch voorspelbare en redundante karakter van congruentie in vraag (zie paragraaf 4.2): het buigingsmorfeem spelt dan niet langer de eigenschappen van de syntactische context uit, maar vormt een integraal onderdeel van de adjectief-nomencollocatie. Dit standpunt wordt bevestigd door de uiterst geringe interne variatie bij lexicale collocaties. Op deze problematiek komen we uitgebreid terug in hoofdstuk 17.

### 11.2.2 Relationaliteit van het adjectief

We beschouwen achtereenvolgens het effect van de relationele adjectieven op de buigingsalternantie en de interactie tussen de relationaliteit van het adjectief en de lexicale collocabiliteit van het adjectief-nomenpaar.

#### Relationele adjectieven

In deze paragraaf toetsen we de hypothese dat relationele adjectieven vaker onverbogen blijven dan kwalificerende adjectieven aan de empirische realiteit van het CGN. Tabel 11(10) stelt de buigingsprofielen voor die ontstaan op basis van de distinctie tussen relationele ('rel.adj') en kwalificerende ('¬rel.adj') adjectieven ( $\chi^2 = 429.5352$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	¬rel.adj		rel.adj		$\Sigma$	
<b>verbogen</b>	2,921	85.06	889	58.10	3,810	76.25
<b>onverbogen</b>	513	14.94	641	41.90	1,154	23.25
<b><math>\Sigma</math></b>	<b>3,434</b>	<b>100.0</b>	<b>1,530</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 11(10): Invloed op de adjectivische buigingsalternantie van de relationaliteit van het adjectief

De kans dat het onverbogen buigingsmorfeem gekozen wordt, is bijna driemaal groter bij een relationeel adjectief dan bij een kwalificerend adjectief:  $RR_{\text{onverbogen}} = 2.8045$  [2.5395,3.0971]. De OR bevestigt dit resultaat:  $OR = 4.1055$  [3.5754,4.7143].

We zetten vervolgens de tegenstelling tussen relationele en kwalificerende adjectieven uit tegen de verschillende kwartielen voor de collocabiliteitsscore. We vullen deze kwartielen aan met de observaties die geen lexicale collocatie vormen ('¬lex.col'). In tabel 11(11) worden de 95% BI'en voor de partiële  $\chi^2$ -statistiek gerapporteerd (tussen ronde haken staan de basisgegevens, i.e. de absolute frequenties van de verschillende cellen).

	¬lex.col	lex.col <sub>kw.1</sub>	lex.col <sub>kw.2</sub>	lex.col <sub>kw.3</sub>	lex.col <sub>kw.4</sub>
<b>rel.adj</b>	[0.881,0.988] (949)	[0.241,0.442] (44)	[1.186,1.588] (177)	[1.063,1.445] (160)	[1.329,1.746] (200)
<b>¬rel.adj</b>	[0.999,1.060] (2,347)	[1.167,1.419] (374)	[0.725,0.930] (237)	[0.781,0.993] (254)	[0.663,0.858] (222)

Tabel 11(11): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de relationele en kwalificerende adjectieven over de lexicale collocabiliteitskwartielen

De 95% BI'en voor de partiële  $\chi^2$ -statistiek tonen duidelijk aan dat de relationele adjectieven significant ondervertegenwoordigd zijn bij adjectief-nomenparen die geen lexicale collocatie vormen en bij de adjectief-nomenparen in het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>'). Vanaf het tweede collocabiliteitskwartiel

(‘lex.col<sub>kw.2</sub>’), daarentegen, treffen we significant meer relationele adjectieven aan dan verwacht. Dit komt niet over als een verrassing: doordat de relationele adjectieven een relatie tussen twee substantivische concepten oproepen, met name tussen het concept van het nominale hoofd en het concept van het basissubstantief van het denominale relationele adjectief, is de semantische relatie tussen het adjectief en het substantief hechter dan bij een louter kwalificerende relatie.

De combinatie van een relationeel adjectief en een substantief vormt vaak een nieuw concept, dat een taxonomische specialisatie is ten opzichte van het concept aangeduid door het nominale hoofd. Halliday (1994: 184-185) spreekt niet toevallig van “classifying adjectives”.<sup>23</sup> Dit is niet – vaak – het geval bij kwalificerende adjectieven die – in de typische gevallen – enkel nadere informatie over de substantivische referent of het substantivische concept verschaffen. De voorbeelden in (11) illustreren de taxonomische specialisatie van enkele adjectief-nomenparen met een relationeel adjectief.

- (11) mondeling examen, neuraal netwerk, technisch onderwijs

Het begrip in de adjectief-nomenparen in (11) identificeert een hyponiem van het begrip opgeroepen door het nominale hoofd: een mondeling examen is een specifiek type examen, in contrast tot een schriftelijk examen; het technisch onderwijs is een van de verschillende onderwijstypes naast het beroepsonderwijs, het algemeen secundair onderwijs en het kunstonderwijs; een neuraal netwerk is een specifiek soort algoritme dat de werking van het menselijk brein simuleert. Merk op dat twee van de vier onderwijstypes door een samenstelling worden aangeduid. Dit sluit aan bij de terugkerende opmerking in de literatuur dat adjectief-nomenparen met naamstatus vergelijkbaar zijn met samenstellingen. Net als bij NN-samenstellingen wordt de specifieke relatie opgeroepen tussen de twee nominale concepten in het relationele adjectief en het nominale hoofd ingevuld op basis van contextuele en encyclopedische informatie (zie paragraaf 11.1.2).

Deze gegevens zijn conform met de vooropgestelde hypotheses: ten eerste, relationele adjectieven worden significant minder frequent verbogen dan andere adjectieven; ten tweede, we treffen significant meer relationele adjectieven aan bij de lexicale collocaties dan bij de overige adjectief-nomenparen.

#### *Relationele adjectieven en lexicale collocabiliteit*

De relationele adjectieven zijn het sterkst vertegenwoordigd in de hoogste collocabiliteitskwartielen, waar de voorkeur voor de onverbogen vorm het grootst is. Het is dus niet uitgesloten dat de selectie van de onverbogen vorm bij de lexicale collocaties niet zozeer het gevolg is van de sterke lexicale associatie, maar dat het een neveneffect is van het voorkomen van relationele adjectieven in deze contexten. Indien deze veronderstelling correct is – en de lexicale collocabiliteit in ruime mate mede beïnvloed wordt door de aanwezigheid van relationele adjectieven – verwachten we een significante differentiëring tussen de OR’s voor de lexicale

collocabiliteit over de strata gedefinieerd door de aan- en afwezigheid van relationele adjectieven. Dit vermoeden wordt niet door de corpusgegevens bevestigd: op basis van de  $p$ -waarde van de *Breslow-Day*-test besluiten we dat het buigingseffect van de binaire lexicale collocabiliteit ('lex.col' vs. '¬lex.col') homogeen verdeeld is over de strata gedefinieerd door de aan- en afwezigheid van de relationele adjectieven (*Breslow-Day*-test:  $p = 0.3945$ ). De CMH-statistieken bevestigen de resultaten van de bivariate analyse op basis van de frequentieverdeling in tabel 11(4) (MH- $\chi^2 = 166.5421$ ,  $df = 1$ ,  $p < 2.2e-19$ ; CMH-OR = 2.5040 [2.1728,2.8858]).

We verfijnen deze binaire oppositie nu door de observaties met een significante collocabiliteitscore in kwartielen op te delen. Tabel 11(12a) toont de frequentieverdeling van de buigingsprofielen van de verschillende collocabiliteitsgraden in het stratum gedefinieerd door de relationele adjectieven:

	¬lex.col		lex.col <sub>kw.1</sub>		lex.col <sub>kw.2</sub>		lex.col <sub>kw.3</sub>		lex.col <sub>kw.4</sub>	
<b>verb</b>	626	65.96	38	86.36	115	64.97	64	40.00	46	23.00
<b>onverb</b>	323	34.04	6	23.64	62	35.03	96	60.00	154	77.00
<b>Σ</b>	<b>949</b>	<b>100.0</b>	<b>44</b>	<b>100.0</b>	<b>177</b>	<b>100.0</b>	<b>160</b>	<b>100.0</b>	<b>200</b>	<b>100.0</b>

Tabel 11(12a): Impact van de lexicale collocabiliteit van het adjectief-nomenpaar op de buigingsalternantie in het stratum gedefinieerd door de aanwezigheid van relationele adjectieven

Tabel 11(12b) toont de frequentieverdeling voor het stratum gedefinieerd door de kwalificerende adjectieven:

	¬lex.col		lex.col <sub>kw.1</sub>		lex.col <sub>kw.2</sub>		lex.col <sub>kw.3</sub>		lex.col <sub>kw.4</sub>	
<b>verb</b>	2,096	89.31	332	88.77	188	79.32	196	77.17	109	49.10
<b>onverb</b>	251	10.69	42	11.23	49	20.77	58	22.83	113	50.90
<b>Σ</b>	<b>2,347</b>	<b>100.0</b>	<b>374</b>	<b>100.0</b>	<b>237</b>	<b>100.0</b>	<b>254</b>	<b>100.0</b>	<b>222</b>	<b>100.0</b>

Tabel 11(12b): Impact van de lexicale collocabiliteit van het adjectief-nomenpaar op de buigingsalternantie in het stratum gedefinieerd door de afwezigheid van relationele adjectieven ('¬rel.adj')

Tabel 11(12c) vat het buigingseffect samen van de collocabiliteitskwartielen in beide strata aan de hand van de OR's en de 95% BI'en. Het referentiepunt bij de berekening van de OR's in tabel 11(12c) is de afwezigheid van lexicale collocabiliteit, i.e. het buigingsprofiel van de observaties zonder significante  $G^2$ -score of waarvoor geen  $G^2$ -score kan worden berekend op basis van de voorgestelde criteria ('¬lex.col').

Stratum	lex.col <sub>kw.1</sub>		lex.col <sub>kw.2</sub>		lex.col <sub>kw.3</sub>		lex.col <sub>kw.4</sub>	
	OR	95% BI	OR	95% BI	OR	95% BI	OR	95% BI
‘rel.adj’	0.30	[0.11,0.76]	1.04	[0.74,1.46]	2.91	[2.06,4.10]	5.33	[3.82,7.44]
‘-rel.adj’	1.06	[0.73,1.49]	2.17	[1.55,3.06]	2.47	[1.79,3.41]	8.65	[6.45,11.61]

Tabel 11(12c): Samenvatting van de stratumanalyse van de partiële conditionering van de lexicale collocabiliteit van het adjectief-nomenpaar door de aan- en afwezigheid van relationele adjectieven

De OR's in tabel 11(12c) zijn (niet-significant) groter in het stratum gedefinieerd door de afwezigheid van een relationeel adjectief dan in het stratum gedefinieerd door de aanwezigheid van een relationeel adjectief. Slechts in één context in tabel 11(12c) overlappen de 95% BI'en niet, met name in de context gedefinieerd door het tweede kwartiel ('lex.col<sub>kw.2</sub>'), waar de OR trouwens significant groter is in het stratum gedefinieerd door de afwezigheid van relationele adjectieven. Conclusie: het deflecterende effect van lexicale collocaties wordt niet significant geconditioneerd door de aan- of afwezigheid van een relationeel adjectief.

Nu wisselen we van perspectief en kijken we naar de mogelijke conditionering van het deflecterende effect van de relationele adjectieven door de lexicale collocabiliteit van het adjectief-nomenpaar. Het buigingseffect van de relationele adjectieven is homogeen in de twee deelverzamelingen gedefinieerd door de aan- en afwezigheid van lexicale collocaties ('-lex.col' vs. 'lex.col') (*Breslow-Day*-test:  $p = 0.3945$ ).<sup>24</sup> Het bivariate buigingseffect van de relationele adjectieven, zoals vastgesteld op basis van de frequentieverdeling in tabel 11(10), blijft behouden wanneer we de binaire lexicale collocabiliteit controleren ( $MH-\chi^2 = 409.0401$ ,  $df = 1$ ,  $p < 2.2e-16$ ;  $CMH-OR = 4.0759$  [3.5388,4.6947]).

Wanneer we het effect van de relationele adjectieven binnen de deelverzamelingen gedefinieerd door de verschillende collocabiliteitskwartielen voor de  $G^2$ -score beschouwen, duikt er een significante differentiëring op (*Breslow-Day*-test:  $p = 0.0066$ ). Wanneer we de OR's beschouwen die de odds <sup>onverbogen</sup>/<sub>verbogen</sub> bij de relationele adjectieven vergelijken met de kwalificerende adjectieven (tabel 11(13)), stellen we vast dat het significante resultaat voor de *Breslow-Day*-statistiek veroorzaakt wordt door het significant grotere effect van de relationele adjectieven in het derde collocabiliteitskwartiel ('lex.col<sub>kw.3</sub>').<sup>25</sup> Het effect van de relationele adjectieven is niet significant bij het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>'). Gezien het ruime BI is dit wellicht een gevolg van dataschaarste (slechts 44 relationele adjectieven tegen 374 kwalificerende adjectieven). Merk ten slotte op dat de OR stijgt in functie van de collocabiliteitsgraad, met uitzondering van de lichte daling in het hoogste kwartiel ('lex.col<sub>kw.4</sub>'). Dit is hoogstwaarschijnlijk een gevolg van de zeer hoge OR-waarde in het derde kwartiel ('lex.col<sub>kw.3</sub>'), zoals blijkt uit de vergelijking van de score voor de OR's in de kwartielen 'lex.col<sub>kw.4</sub>' en 'lex.col<sub>kw.2</sub>'.



<b>Stratum</b>	<b>OR</b>	<b>95% BI</b>
lex.col <sub>kw.1</sub>	1.25	[0.50,3.13]
lex.col <sub>kw.2</sub>	2.07	[1.33,3.21]
lex.col <sub>kw.3</sub>	5.07	[3.29,7.80]
lex.col <sub>kw.4</sub>	3.23	[2.12,4.92]

Tabel 11(13): *Buigingseffect van de relationele adjectieven in de verschillende collocabiliteitskwartielen*

De stratumanalyses hebben aangetoond dat beide lexicale/semantische factoren een significante impact behouden wanneer de andere gecontroleerd wordt. Het effect van de lexicale collocabiliteit van het adjectief-nomenpaar is bijgevolg niet herleidbaar tot de aan- of afwezigheid van een relationeel adjectief en *vice versa*. Aan de andere kant kunnen we niet negeren dat beide effecten significant vaker dan verwacht samen voorkomen en dat het deflecterende effect van de relationele adjectieven vergroot naarmate de lexicale collocabiliteit toeneemt.

Het deflecterende effect van de relationele adjectieven en de lexicale collocaties toont aan dat het onverbogen buigingsmorfeem een hechte lexicale/semantische relatie tussen het adjectief en het substantief identificeert. De lexicale collocaties selecteren de adjectief-nomenparen waar het adjectief en het substantief statistisch significant vaker dan verwacht samen voorkomen. Dit wijst op een sterke lexicale associatie tussen beide elementen. In verscheidene observaties wordt deze lexicale associatie semantisch geconcretiseerd door de relationele adjectieven: de buigings-*e* wordt vooral weggelaten bij de adjectieven die geen modifierende maar een relationele functie vervullen, i.e. de adjectieven die een relatie aanbrengen tussen twee substantivische concepten. De gemarkeerde onverbogen vorm signaleert aan de toehoorder dat het adjectief en het substantief een hechte semantische eenheid vormen, waarbij het element in adjectivische positie niet zozeer het concept of de referent aangeduid door het nominale hoofd modificeert, maar een relatie tussen twee substantivische concepten oproept, met name het concept aangeduid door het substantief aan de basis van het relationeel adjectief en het concept verwoord door het nominale hoofd.

De resultaten die in dit hoofdstuk gerapporteerd zijn, reiken tot slot enkele elementen aan die de semantische modellen voor de buigingsalternantie onderbouwen (Blom 1994, 1995; Booij 2002a, 2002c; Honselaar 1980). Het gebruik van de onverbogen vorm bij niet-modifierende relationele adjectieven is in sommige gevallen verbonden met een categoriserende functie van deze buigingspendant: verscheidene relationele adjectieven in de materiaalverzameling vervullen een categoriserende of subclassificerende functie van het nominale hoofd door het aan een ander concept te relateren (cf. *classifying adjective*). Een adjectief-nomencombinatie als *presidentieel vliegtuig* duidt een subtype vliegtuig aan door het te relateren aan een bepaalde persoon, met name de eigenaar of de unieke gebruiker (c.q. de president). De semantische modellen komen aan bod in hoofdstuk 17, waar

de vraag naar de semantische basis van (het buigingseffect van) de adjectief-nomencollocaties gesteld wordt.

### ***11.2.3 Stratumanalyse voor de lexicale/semantische conditionering van prosodische buigingseffecten***

In het derde en laatste deel van het empirisch onderzoek naar de lexicale/semantische conditionering van de buigingsalternantie analyseren we de eventuele partiële conditionering van het buigingseffect bij prosodische patronen door lexicale en/of semantische kenmerken. De prosodische analyse in het vorige hoofdstuk heeft bij de concrete klemtoonpatronen een aantal onverwachte tendenties aan het licht gebracht die we – voorlopig – geassocieerd hebben met adjectief-nomenparen met naamstatus. Aangezien we de intuïtieve notie adjectief-nomenpaar met naamstatus in dit hoofdstuk vervangen hebben door een objectieverbaar alternatief, met name de lexicale collocabiliteit van het adjectief-nomenpaar, verifiëren we de empirische adequaatheid van de intuïtieve verklaringen die we in het vorige hoofdstuk hebben aangereikt.

Dit zijn de twee prosodische contexten waar er een – vanuit prosodisch perspectief – onverwacht buigingseffect optreedt:

- de adjectieven met een doffe eindlettergreep (prosodisch patroon ‘s0’) selecteren significant vaker het onverbogen buigingsmorfeem dan de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep (prosodisch patroon ‘s3’) (zie paragraaf 10.2.2.1, tabel 10(15))<sup>26</sup>
- vóór een substantief met een secundair beklemtoonde beginlettergreep komt de onverbogen buigingsvariant significant minder vaak voor dan vóór een substantief met een onbeklemtoonde beginlettergreep (zie paragraaf 10.2.1, tabel 10(12)).

We kijken eerst naar het significant kleinere aandeel onverbogen vormen bij de substantieven met een secundair beklemtoonde beginlettergreep ten aanzien van de substantieven met een onbeklemtoonde beginlettergreep. Om de intuïtieve lexicaal-semantische verklaring te toetsen die we in het vorige hoofdstuk voor deze onverwachte distributie hebben aangereikt, voeren we een stratumanalyse uit met de volgende variabelen:

- responsvariabele: adjectivische buigingsalternantie (verbogen vs. onverbogen)
- verklarende variabele: secundair beklemtoonde vs. onbeklemtoonde substantivische beginlettergreep
- stratumvariabele: lexicale collocabiliteit van het adjectief-nomenpaar (‘-lex.col’ vs. ‘lex.col’)

Op basis van de CMH-statistieken stellen we vast dat vóór een substantief met een secundair beklemtoonde beginlettergreep significant vaker een onverbogen adjectief voorkomt dan vóór een substantief met een onbeklemtoonde beginletter-

greep, ook wanneer het effect van de lexicale collocabiliteit gecontroleerd wordt (MH- $\chi^2 = 6.2960$ ,  $df = 1$ ,  $p < 0.0121$ ; CMH-OR = 1.349 [1.0688,1.7027]). De significante  $p$ -waarde voor de *Breslow-Day*-statistiek wijst echter op een partiële conditionering van het buigingseffect van de accentuering van de substantivische beginlettergreep door de lexicale collocabiliteit van het adjectief-nomenpaar ( $p = 0.0180$ ). Om de zin en de omvang van deze partiële conditionering te kennen, analyseren we de stratumtabellen voor de strata gedefinieerd door de afwezigheid (tabel 11(14a)) en de aanwezigheid (tabel 11(14b)) van adjectief-nomenparen die een lexicale collocatie vormen.

In het stratum gedefinieerd door de afwezigheid van lexicale collocaties heeft het onderscheid tussen een secundair beklemtoonde ('nevenklemtoon') en een onbeklemtoonde ('onbeklemtoond') substantivische beginlettergreep geen significante impact op de buigingsalternantie ( $\chi^2 = 0.1776$ ,  $df = 1$ ,  $p = 0.6735$ ):

	onbeklemtoond		nevenklemtoon		$\Sigma$	
<b>verbogen</b>	785	77.26	233	75.90	1,018	76.95
<b>onverbogen</b>	231	22.74	74	24.10	305	23.05
<b><math>\Sigma</math></b>	<b>1,016</b>	<b>100.0</b>	<b>307</b>	<b>100.0</b>	<b>1,323</b>	<b>100.0</b>

Tabel 11(14a): Invloed op de buigingsalternantie van het onderscheid tussen een onbeklemtoonde en een secundair beklemtoonde substantivische beginlettergreep in het stratum gedefinieerd door de afwezigheid van lexicale collocaties

In het stratum gedefinieerd door de aanwezigheid van adjectief-nomencollocaties is deze distinctie wel significant ( $\chi^2 = 11.0238$ ,  $df = 1$ ,  $p = 0.0008995$ ):

	onbeklemtoond		nevenklemtoon		$\Sigma$	
<b>verbogen</b>	291	61.26	62	44.93	353	57.59
<b>onverbogen</b>	184	38.74	76	55.07	260	42.41
<b><math>\Sigma</math></b>	<b>475</b>	<b>100.0</b>	<b>138</b>	<b>100.0</b>	<b>613</b>	<b>100.0</b>

Tabel 11(14b): Invloed op de buigingsalternantie van het onderscheid tussen een onbeklemtoonde en een secundair beklemtoonde substantivische beginlettergreep in het stratum gedefinieerd door de aanwezigheid van lexicale collocaties

Tabel 11(14c) vat de resultaten van deze stratumanalyse samen:

Stratum	OR	95% BI
¬lex.col	1.08	[0.80,1.46]
lex.col	1.93	[1.32,2.84]
CMH-OR	1.34	[1.07,1.70]

Tabel 11(14c): Samenvatting van de stratumanalyse naar de partiële conditionering door de lexicale collocabiliteit van het onderscheid tussen een onbeklemtoonde en een secundair beklemtoonde substantivische beginlettergreep

De significante  $p$ -waarde van de *Breslow-Day*-test wordt niet door de overlappende BI'en van de OR's bevestigd. We stellen voor om het resultaat van de *Breslow-Day*-test te laten primeren op de overlappende 95% BI'en in tabel 11(14c). Ten eerste, de frequentieverdeling in het stratum gedefinieerd door de afwezigheid van lexicale collocabiliteit is niet significant; ten tweede, de overlapping tussen de 95% BI'en in tabel 11(14c) is wellicht een gevolg van dataschaarste in het stratum gedefinieerd door de aanwezigheid van lexicale collocabiliteit, zoals het relatief brede BI voor dit stratum aantoont. We verwijzen eveneens naar de absolute frequenties in tabel 11(14b). We besluiten dat het significant grotere aandeel onverbogen vormen bij de substantieven met een secundair beklemtoonde beginlettergreep ten overstaan van de substantieven met een onbeklemtoonde beginlettergreep ten dele samenhangt met de lexicale collocabiliteit van het adjectief-nomenpaar, waarvan we in dit hoofdstuk het deflecterende effect hebben aangetoond (zie paragraaf 11.2.1).

De tweede stratumanalyse test de partiële conditionering door de lexicale collocabiliteit van het buigingseffect van het onderscheid tussen de adjectieven met een doffe eindlettergreep (prosodisch patroon 's0') en de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep (prosodisch patroon 's3'). Ook hier gaan we uit van een lexicaal/semantisch effect aan de basis van het significant grotere aandeel onverbogen vormen bij de adjectieven op 's3' vergeleken met de adjectieven op 's0'.

De CMH-statistieken bevestigen het resultaat van de bivariate analyse: ook wanneer het effect van de lexicale collocabiliteit gecontroleerd wordt, selecteren de adjectieven met het patroon 's0' als rechterraand significant minder vaak de onverbogen vorm dan de adjectieven met het patroon 's3' als rechterwoordrand ( $MH-\chi^2 = 30.1444$ ,  $df = 1$ ,  $p < 2.2e-16$ ;  $CMH-OR = 0.4622$  [0.3502,0.61]). Daarnaast weerlegt de grote  $p$ -waarde voor de *Breslow-Day*-test de hypothese dat er sprake is van een significant stratumeffect ( $p = 0.6544$ ).

We herinneren de lezer eraan dat het merendeel van de adjectieven met het patroon 's3' als meest rechtse voet met *-isch* gesuffigeerd is (zie paragraaf 10.2.1, (19)). Dit is één van de suffixen die in het Nederlands gebruikt worden bij de vorming van relationele adjectieven (zie tabel 11(2)). Derhalve doen we de stratum-analyse over, maar ditmaal met de relationaliteit van het adjectief als stratum-variabele.

De CMH-statistieken weerleggen het resultaat van de bivariate analyse (zie tabel 10(15)): wanneer we de relationaliteit van het adjectief controleren, is het onderscheid tussen de adjectieven met 's0' en 's3' als rechterwoordrand niet langer significant ( $MH-\chi^2 = 0.4457$ ,  $df = 1$ ,  $p = 0.5044$ ;  $CMH-OR = 0.8607$  [0.5892,1.2575]). De grote  $p$ -waarde voor de *Breslow-Day*-test wijst bovendien de mogelijkheid van een partiële conditionering af ( $p = 0.2086$ ). De OR's voor beide strata hebben een tegengestelde zin, maar geen van beide is significant: wanneer we de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  voor de adjectieven op 's0' met de adjectieven op 's3'

vergelijken, stellen we een niet-significante stijging vast bij de niet-relatieve adjectieven (OR = 2.9348 [0.3729,23.0971]) en een niet-significante daling bij de relationele adjectieven (OR = 0.8246 [0.5607,1.2127]). Het zeer ruime BI in het stratum gedefinieerd door de niet-relatieve adjectieven wijst op dataschaarste (en een bijna-perfecte scheiding). Uit een nadere inspectie van de verdeling van de relationele adjectieven over de adjectieven met een doffe eindlettergreep ('s0') en de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep ('s3') blijkt dat de laatste groep nauwelijks 14 niet-relatieve adjectieven omvat. Op basis van deze constatering rijst het vermoeden dat de relationele adjectieven significant oververtegenwoordigd zijn in de groep adjectieven die op het patroon 's3' eindigen

Deze hypothese controleren we door een partiële  $\chi^2$ -test uit te voeren op de contingentietabel die ontstaat wanneer we de relationaliteit van het adjectief uitzetten tegen het onderscheid tussen de adjectieven met een doffe eindlettergreep ('s0') en de adjectieven met een sjwaloze onbeklemtoonde ('s3') eindlettergreep. Tabel 11(15) bevat de 95% BI'en voor de partiële  $\chi^2$ -statistiek (tussen ronde haken staan de basisfrequenties voor iedere cel).

	's3'	's0'
<b>¬rel.adj</b>	[0.0345,0.1098] (14)	[1.2385,1.35] (791)
<b>rel.adj</b>	[2.3172,2.834] (294)	[0.4324,0.5679] (180)

Tabel 11(15): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de relationele adjectieven over de adjectieven met 's0' en 's3' als meest rechtse voet

De 95% BI'en in tabel 11(15) tonen aan dat de relationele adjectieven significant vaker dan verwacht het patroon 's3' en significant minder vaak dan verwacht het patroon 's0' als meest rechtse metrische voet hebben. De niet-relatieve adjectieven vertonen het omgekeerde beeld. We concluderen dat het significant grotere aandeel onverbogen vormen bij de adjectieven op een sjwaloze onbeklemtoonde eindlettergreep in ruime mate samenhangt met de significante oververtegenwoordiging van de – deflecterende – relationele adjectieven bij de adjectieven met 's3' als meest rechtse voet.

De resultaten van de stratumanalyses bevestigen de intuïtieve verklaring die we in het vorige hoofdstuk hebben voorgesteld voor twee – vanuit een prosodisch perspectief – onverwachte buigingseffecten bij concrete prosodische patronen:

- het significant grotere aandeel onverbogen adjectieven vóór de substantieven met een secundair beklemtoonde beginlettergreep ten aanzien van de substantieven met een onbeklemtoonde beginlettergreep is voor een deel het gevolg van een significant groter deflecterend effect van deze elementen bij de lexicale collocaties

- het significant grotere aandeel onverbogen vormen bij de adjectieven met het patroon ‘s3’ als rechterwoordrand ten overstaan van de adjectieven met het patroon ‘s0’ als rechterwoordrand hangt samen met de significante oververtegenwoordiging van de relationele adjectieven in de eerste groep adjectieven

### 11.3 Conclusie

In dit hoofdstuk hebben we de lexicale/semantische restricties op het gebruik van het verbogen en het onverbogen adjectief bestudeerd. We hebben twee types condities behandeld: de lexicale/semantische conditionering van het adjectief-nomenpaar en de semantische conditionering van het adjectief.

De lexicale/semantische conditionering van het adjectief-nomenpaar wordt in de literatuur meestal behandeld in termen van idiomatische adjectief-nomenparen of in termen van adjectief-nomenparen die een semantische eenheid vormen (11.1.1). De semantische eenheid of de idiomatische van het adjectief-nomenpaar wordt in de literatuur introspectief vastgesteld op basis van de niet-compositionele betekenisstructuur. We hebben aangevoerd dat dit semantische criterium in tegenspraak is met de analytische transparantie en replicerbaarheid die we in dit onderzoek nastreven. Als alternatief hebben we een distributioneel criterium voorgesteld: de lexicale collocabiliteit van het adjectief-nomenpaar. We hebben een kwantitatieve operationalisering voor deze notie gepresenteerd: alle adjectief-nomenparen waarvan de samenstellende delen significant vaker dan verwacht samen voorkomen, beschouwen we als een lexicale collocatie. Voor het kwantificeren van de significantie en de sterkte van de lexicale associatie tussen het adjectief en het substantief gebruiken we de  $G^2$ -statistiek (i.e. de  $-2 \log \text{likelihood}$ -ratio).

De hypothesen die we in verband met de lexicale collocabiliteit van het adjectief-nomenpaar naar voor hebben geschoven, doorstaan de toetsing aan de corpusgegevens (11.2.1): het onverbogen adjectief wordt significant vaker geselecteerd indien het adjectief-nomenpaar een lexicale collocatie vormt; het aandeel onverbogen buigingsvormen stijgt progressief naarmate de lexicale associatie tussen het adjectief en het substantief sterker is. Met andere woorden: bij neutra markeert het onverbogen buigingsmorfeem gelexicaliseerde adjectief-nomenparen, i.e. adjectief-nomensequenties die een lexicale eenheid vormen. Voorbeelden zoals *half uur*, *vorig jaar* of *volgend jaar*, die consistent de onverbogen buigingsvariant selecteren, tonen aan dat de lexicale/semantische conditionering niet gereduceerd kan worden tot idiomatische uitdrukkingen met een niet-compositionele semantiek. Tot slot hebben we aangetoond dat de lexicale collocaties intern nauwelijks variatie vertonen: bijna 85% van de AN-lemmatypes vertoont geen interne buigingsvariatie. Dit betekent dat deze vormen consistent met één van beide buigingsmorfemen gebruikt worden.

Deze resultaten tonen duidelijk het belang aan van het lexicale/semantische niveau voor de conditionering van de buigingsalternantie: de leden van een taalgemeenschap lijken een stabiel beeld te hebben van de adjectivische buigingsvorm die in een AN-lemmatype geselecteerd wordt. Op basis van deze resultaten rijst de vraag naar het statuut van de adjectivische buiging en – bij uitbreiding – van congruentie in het algemeen. Er is een duidelijke lexicale conditionering van de buigingsalternantie, die bovendien niet gereduceerd kan worden tot een opsomming van idiomatische uitzonderingen: ten eerste, het aandeel lexicale collocaties is proportioneel te groot (ongeveer één op drie observaties is een collocatie); ten tweede, de kans op een onverbogen buigingsmorfeem stijgt progressief naarmate de lexicale associatiesterkte toeneemt; ten derde, het deflecterende gedrag van de relationele adjectieven, die significant oververtegenwoordigd zijn bij de lexicale collocaties, waar hun deflecterende effect bovendien toeneemt in de hogere collocabiliteitskwartielen, suggereert een semantische regelmaat aan de basis van de lexicale patronen.

De vaststelling van Booij (2002a) en Heynderickx (2001) dat relationele adjectieven frequent de adjectivische positie bezetten in adjectief-nomenparen met naamstatus ligt aan de basis van de hypothese dat relationele adjectieven een faciliterende factor zijn voor het gebruik van het onverbogen buigingsmorfeem (11.2.1). De empirische gegevens ondersteunen deze hypothese: met de selectie van het onverbogen buigingsmorfeem signaleert de taalgebruiker dat het adjectief geen modifierende functie ten overstaan van de substantivische entiteit vervult, maar dat het adjectief een relatie tussen twee substantivische concepten oproept (11.2.2).

We besluiten dat het vormelijk gemarkeerde onverbogen adjectief op iconische wijze een niet-prototypische relatie tussen adjectief en substantief aanduidt:

- Het onverbogen buigingsmorfeem signaleert dat het adjectief en het substantief een lexicale eenheid vormen. Dit is: de adjectief-nomensequentie is niet het resultaat van een syntagmatische operatie maar wordt als één entiteit, inclusief buigingsmorfeem, geactiveerd en in de zin ingevoegd.
- Het onverbogen buigingsmorfeem signaleert dat het adjectief een relatie tussen twee substantivische concepten identificeert.

Beide functies zijn met elkaar verbonden, zoals blijkt uit de significante oververtegenwoordiging van relationele adjectieven in de adjectief-nomenparen in de hoogste collocabiliteitskwartielen (c.q. 'lex.col<sub>kw.2</sub>', 'lex.col<sub>kw.3</sub>' en 'lex.col<sub>kw.4</sub>'). Dit is mogelijkwijs een gevolg van de 'betrekking stichtende' betekenis van relationele adjectieven. Het gebruik van een relationeel adjectief in een lexicale collocatie gaat daarenboven vaak gepaard met een taxonomische betekenis-specialisatie, waarbij het concept aangeduid door de adjectief-nomensequentie een specialisatie is van het concept aangeduid door het nominale hoofd (cf. *classifying adjective*).

Tot slot hebben we de intuïtieve verklaring die we in het vorige hoofdstuk hebben aangereikt om enkele – vanuit prosodisch perspectief – onverwachte

buigingstendenties te verklaren, empirisch getest in het licht van de voorgestelde lexicale/semantische analyse (11.2.3). Het onverwacht grote aandeel onverbogen adjectieven in het buigingsprofiel van de substantieven met een secundair beklemtoonde beginlettergreep en in het buigingsprofiel van de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep is voor een deel lexicaal/semantisch gemotiveerd: voor de eerste groep hebben we door middel van een stratumanalyse aangetoond dat het deflecterende effect van de lexicale collocabiliteit significant groter is bij de substantieven met een secundair beklemtoonde lettergreep dan bij de substantieven met een onbeklemtoonde beginlettergreep; voor de tweede groep hebben we bewezen dat de adjectieven met een sjwaloze onbeklemtoonde eindlettergreep (patroon 's3') significant vaker een (deflecterend) relationeel adjectief zijn dan de adjectieven met een doffe eindlettergreep (patroon 's0').

Zoals we in de inleidende paragraaf hebben aangekondigd, vormt dit hoofdstuk de eerste fase van het onderzoek naar de lexicale en semantische conditionering van de buigingsalternantie. De klemtoon ligt in deze fase op de operationalisering van de lexicale/semantische conditionering van de buigingsalternantie; de semantische interpretatie van (het buigingseffect van) de lexicale collocaties vormt het onderwerp van deel IV van dit proefschrift. We laten bijgevolg voorlopig in het midden of de lexicale collocaties louter lexicale effecten identificeren dan wel of er semantische patronen aan de basis van het buigingseffect van de adjectief-nomenclollocaties geponeerd moeten worden. De empirische analyse bevat niettemin een aantal aanwijzingen dat het buigingseffect van de lexicale collocaties niet tot een louter lexicaal fenomeen gereduceerd kan worden. Ten eerste, de hogere collocabiliteitskwartielen, waar het deflecterende effect het grootst is, zijn significant met het relationeel adjectief geassocieerd. Ten tweede, de lexicale collocaties in het algemeen en de hogere collocabiliteitskwartielen in het bijzonder worden gerealiseerd door adjectieven en substantieven met een hoge score voor de lexicale specificiteit. Deze vaststellingen lijken erop te wijzen dat het deflecterende effect van de lexicale collocaties op zijn minst voor een deel semantisch gemotiveerd is. Dit is een bijkomende empirische motivatie voor de uitgebreide semantische analyse in deel IV.

Rekening houdend met de dataschaarste – voor slechts 482 van de 3,475 AN-lemmatypes kan een significante  $G^2$ -score berekend worden – en de lectische en thematische diversiteit van de materiaalverzameling breiden we het onderzoek naar de lexicale/semantische conditionering van de buigingsalternantie in het vierde deel van deze dissertatie uit door gebruik te maken van een 20M woorden tellend corpus op basis van de Vlaamse kwaliteitskrant *De Standaard* (zie hoofdstuk 15). We verwachten dat het taalgebruik in de edities van één jaargang van een krant niet alleen lectisch maar ook thematisch homogener is dan in een corpus dat uiteenlopende gesprekssituaties verzamelt, zoals dat in het CGN het geval is. Door deze kwantitatieve en kwalitatieve nadelen is de materiaalverzameling op basis van



het CGN minder geschikt als basis voor de semantische interpretatie van (het buigingseffect van) de lexicale collocaties.

De lexicale/semantische analyse op basis van het CGN is echter allesbehalve overbodig. Ten eerste, door het louter lexicaal-semantische karakter van de *De Standaard*-analyse, is de CGN-analyse onontbeerlijk om het buigingseffect van de lexicale/semantische restricties ten overstaan van de andere verklarende factoren te positioneren. Ten tweede, een sterke verklarende factor (cf. OR-scores) zoals de lexicale collocabiliteit zorgt hoogstwaarschijnlijk voor een significante en sterke reductie van de variatie in het logistische-regressiemodel.

Voor het uitgebreide onderzoek naar de semantische conditionering van de buigingsalternantie en van de lexicale collocabiliteit van het adjectief-nomenpaar in deel IV formuleren we de volgende hypothesen:

- De lexicale collocabiliteit van het adjectief-nomenpaar heeft een deflecterend effect. Niet alleen selecteren adjectief-nomenparen die een lexicale collocatie vormen significant vaker het onverbogen adjectief dan adjectief-nomenparen die geen lexicale collocatie vormen, maar de kans op een onverbogen adjectief stijgt progressief naarmate de sterkte van de lexicale associatie tussen het adjectief en het substantief groter wordt (zie hoofdstuk 16).
- Het buigingseffect van de lexicale collocaties kan niet gereduceerd worden tot een louter lexicaal verschijnsel. Aan de basis van het deflecterende effect van de adjectief-nomencollocaties bevinden zich semantische patronen, waaronder de aanwezigheid van relationele en categoriserende adjectieven in de adjectivische positie (zie hoofdstuk 17).

## Noten

1. Degryse (1959) rept met geen woord van de semantische conditionering van de buigingsalternantie.
2. Net als voor idiomen bestaan er ook voor lexicale collocaties meerdere definities waarvan de specificiteit varieert in functie van het concrete onderzoek. Voor een overzicht verwijzen we naar Van der Wouden (1992, 2001).
3. De laatste jaren is er een toename van vergelijkende studies naar de functionaliteit van deze maten in specifieke onderzoekscontexten. Om enkele voorbeelden te citeren: Weeber, Vos & Baayen (2000) vergelijken de  $G^2$ -test, de *Fisher Exact*-test en de MI-index bij de identificatie van collocaties tussen items met een zeer lage frequentie; Stubbs (1995) vergelijkt de *t*-test en de MI-index voor de extractie van lexicale en grammaticale collocaties.
4. *Hapax legomena* zijn patronen die slechts éénmaal in een corpus voorkomen; *dis legomena* zijn elementen met twee attestaties.
5. Voor de wiskundige afleiding van deze formule en het rationale achter *maximum likelihood*-statistieken, waartoe ook de *log likelihood*-statistiek hoort, verwijzen we naar Dunning (1993).
6. Let wel: de  $G^2$ -statistiek is in feite de  $-2 \log$  *likelihood*-ratio (zie ook formule in (3)). Het is deze  $-2 \log$  *likelihood*-ratio en niet de gewone *log likelihood*-ratio die een  $\chi^2$ -verdeling vertoont. Voor de mathematische details verwijzen we opnieuw naar Dunning (1993).
7. Een bigram is een koppel gevormd door twee elementen, i.c.  $w_1$  en  $w_2$  of  $A_i$  en  $N_j$ .
8. De *span* is de afstand (in woorden) tussen de verschillende elementen in een collocatie-relatie. Doordat we enkel geïnteresseerd zijn in adjectief-nomensequenties is de *span* in dit onderzoek gelijk aan 1.
9. In de morfologische traditie spreekt men doorgaans van het lexeem (zie ook paragraaf 4.1).
10. Deze stellingname vinden we ook terug in de psycholinguïstiek (Levelt 1989: 188-192).
11. Dit percentage is op basis van de CELEX-materiaalverzameling berekend (Baayen, Piepenbrock & van Rijn 1993). De 91,503 substantivische lemmata opgenomen in CELEX zijn als volgt over beide grammaticale genera verdeeld: 24,819 (27.12%) substantivische lemmata zijn *het*-woorden en 66,622 (72.81%) substantivische lemmata zijn *de*-woorden. Tot slot is er nog een restgroep van 62 bigenerische substantieven.
12. Hampe (2000) en Stefanowitsch (2003) kampen duidelijk met de methodologische problemen die voortvloeien uit het gebrek aan een objectieve semantische testprocedure. Hoewel in beide papers expliciet een systematische operationalisering en kwantificatie van verklarende parameters wordt nagestreefd – een doelstelling die gerealiseerd wordt voor respectievelijk pragmatische en discours-functionele factoren – wordt het

semantische effect geanalyseerd aan de hand van een traditionele intuïtieve en impressionistische argumentatie.

13. In contrastieve contexten wordt het predicatieve gebruik aannemelijker: *Het leger waarvan de resten vorige week gevonden zijn, is Amerikaans.*
14. De ANS (1997: 381) definieert een adjectief als “een woord dat een nadere bijzonderheid van zelfstandigheden aangeeft [...] De nadere bijzonderheid die aan een zelfstandigheid toegeschreven wordt, kan een eigenschap (een kenmerk, een kwaliteit en dergelijke) zijn, of een toestand.” De Schutter (1968) vertrekt van een analoge modifierende adjectivische functie bij het opstellen van zijn semantische categorisatie van het Nederlandse adjectief.
15. Geeraerts (1989: 57-59) bepleit de toepassing van strikte criteria om semantisch onderzoek van het niveau van de subjectiviteit naar het niveau van de intersubjectiviteit te tillen. Rekening houdend met de relatief grote inbreng van de persoonlijke appreciatie en de encyclopedische kennis van de onderzoeker bij de annotatie van deze semantische parameter verkiezen we om van intersubjectiviteit te spreken veeleer dan van objectiviteit.
16. Aan de opsomming in tabel 11(2) moeten we in feite nog twee suffixen toevoegen: *-er* en *-en*. Rekening houdend met de onverbuigbaarheid van de adjectieven die door middel van deze suffixen zijn afgeleid, zijn deze suffixen irrelevant voor onze doelstellingen: *het norbertijner/\*norbertijnere klooster, het zilveren/\*zilverene lepeltje* (zie paragraaf 3.2).
17. We brengen in herinnering dat de maximale  $G^2$ -score 1721.0 en niet 250 bedraagt, zoals gesuggereerd wordt in figuur 11(2). Omdat het slechts enkele *outliers* zijn die een  $G^2$ -score groter dan 250 hebben, nemen we ze niet in de figuur op.
18. Traditioneel berekent men de type/token-ratio. Aangezien wij niet geïnteresseerd zijn in de creativiteit van het taalgebruik, maar net in de repetitiviteit ervan, keren we deze ratio om.
19. Deze lage cijfers sluiten een aparte berekening van de lexicale collocabiliteit in de verschillende lectisch gedefinieerde deelcorpora uit, zoals dit bijvoorbeeld het geval is in Stefanowitsch & Gries (te verschijnen).
20. Dit zijn de  $G^2$ -scores voor de verschillende kwartielgrenzen:
  - minimum:  $G^2 = 3.841$
  - 1<sup>e</sup> kwartiel:  $G^2 = 21.700$
  - mediaan (= 2<sup>e</sup> kwartiel):  $G^2 = 45.950$
  - 3<sup>e</sup> kwartiel:  $G^2 = 102.200$
  - maximum:  $G^2 = 1721.000$
21. De onzekerheidscoëfficiënt is één van de statistische maten om de grootte van de associatie tussen twee nominale variabelen te berekenen. Doordat de  $\chi^2$ -score sterk beïnvloed wordt door de grootte van het staal, is deze statistiek minder aangewezen om de sterkte van de associatie te vergelijken tussen twee stalen met een uiteenlopende omvang. Bij de bivariate analyses maken we doelbewust geen gebruik van de onzekerheidscoëfficiënt: enerzijds berekenen we het  $RR_{\text{onverbogen}}$  en de OR voor de verschillende

waarden van de verklarende variabelen; anderzijds is de score van de onzekerheidscoëfficiënt moeilijk interpreteerbaar.

De onzekerheidscoëfficiënt is gebaseerd op de berekening van de entropie. De onzekerheidscoëfficiënt (OC) wordt als volgt berekend:

$$(i) \quad OC = \frac{U(Y) - U(Y|X)}{U(Y)}$$

Deze formule meet de afname van de onzekerheid bij de responsvariabele ( $U(Y)$ ) onder invloed van de informatie aanwezig in de verklarende variabele ( $U(Y|X)$ ). De onzekerheid wordt berekend op basis van de entropiemaat ( $H$ ), waarbij  $i$  gelijk is aan het aantal waarden van  $Y$ :

$$(ii) \quad U(Y) = H(Y) = -\sum_i (Y_i) \log_2(P(Y_i))$$

22. Evert, Heid & Spanger (2004) hebben vastgesteld dat de instanties van Duitse adjectief-nomenclollocaties duidelijke morfosyntactische voorkeuren vertonen voor het getal en de grammaticale casus. Dit is: de meeste tokens van een gegeven AN-lemmatype realiseren hetzelfde getal en drukken dezelfde grammaticale casus uit.
23. Let wel: de termen *classifying adjective* en relationeel adjectief zijn geen synoniemen. Het begrip *classifying adjective* omvat de meeste relationele adjectieven, maar ook adjectieven als *rood* en *houten* in *rode wijn* en *houten trein*. Voor een uitgebreide discussie verwijzen we naar paragraaf 17.1.
24. Aangezien we ten overstaan van de vorige stratumanalyse de stratumvariabele en de verklarende variabele onderling verwisselen, ligt het voor de hand dat het resultaat van de *Breslow-Day*-test identiek is. De OR is immers een symmetrische maat (zie paragraaf 7.1).
25. De vijf stratumtabellen voor de vier lexicale collocabiliteitskwartielen en de verzameling zonder lexicale collocaties, die als referentiepunt dient bij de berekening van de OR's in tabel 11(13), kunnen worden afgeleid uit de stratumtabellen 11(12a) en 11(12b) door voor iedere collocabiliteitsgraad de buigingsprofielen van de relationele adjectieven (tabel 11(12a)) en de niet-relationele adjectieven (tabel 11(12b)) naast elkaar te plaatsen.
26. Merk op dat we alleen de adjectieven in beschouwing nemen waarvan de basisvorm op één onbeklemtoonde lettergreep eindigt, dit is: de oppositie tussen 's0' en 's3' als concrete realisaties van het abstracte patroon 'sw'. De adjectieven die op twee onbeklemtoonde lettergrepen eindigen – met de oppositie tussen 's00' en 's33' als concrete instanties van het abstracte patroon 'sww' – hoeven niet in de stratumanalyse opgenomen te worden omdat ze de frequentieverdeling vertonen die we op prosodische gronden verwachten, met name een significant groter aandeel onverbogen vormen bij het patroon 's00' dan bij het patroon 's33' (zie paragraaf 10.2.2.1, tabel 10(15)).

## **Hoofdstuk 12:**

### **Lectische conditionering**

Na de behandeling van de taalinterne factoren concentreren we ons nu op de taalexterne of lectische factoren. De literatuur vermeldt de nationale variëteit van het Nederlands en het register als factoren die de keuze tussen het verbogen adjectief en het onverbogen adjectief bepalen. We hebben bovendien vastgesteld dat deze factoren nauw met elkaar verweven zijn (zie paragrafen 1.1.4 en 1.1.5). Bijgevolg zullen we beide lectische variabelen gezamenlijk behandelen. Alvorens we de lectische conditionering van de buigingsalternantie aanvatten, bespreken we de status van lectische factoren in het algemeen en de relatie tussen geografische en stilistische variabelen in het bijzonder (12.1). Daarna becommentariëren we de hypothesen en de wijze waarop ze gefalsificeerd worden (12.2). In de derde paragraaf toetsen we de hypothesen over de lectische functie van de adjectivische buigingsmorfemen bij neutra aan de empirische realiteit van het CGN (12.3).

#### **12.1 Lectische betekenisaspecten**

Wij beschouwen lectische factoren als een integraal onderdeel van de betekenisstructuur van een lexicaal item of een grammaticaal patroon, dus ook van de buigingsmorfemen. We hebben in het methodologische hoofdstuk betoogd dat de ‘ongecontroleerde’ aard van corpusgegevens de analyse van eventuele lectische condities impliceert, te meer omdat de meeste corpora, zo ook het CGN, het taalgebruik van verschillende sprekers en verschillende situationele contexten cumuleren (zie paragraaf 5.2). Deze stellingname is veeleer uitzonderlijk in de taalkunde. De studie van de stilistische en geografische aspecten van taal is nog steeds grotendeels in handen van de stilistiek, de sociolinguïstiek en de dialectologie, drie taalkundige disciplines die sinds de opkomst en bloei van de autonome grammaticamodellen in de jaren ‘60 naar de periferie van de theoretische taalkunde zijn verwezen, zodat het gros van het theoretisch taalkundig onderzoek zich op de introspectieve beschrijving van de kerngrammatica kon concentreren (Geeraerts 2003b). Het laatste decennium stellen we een recontextualisatie van deze perifere componenten vast, onder andere in het kader van de Cognitieve Taalkunde (Geeraerts 2003b, 2005).

Geeraerts (2005) haalt twee bijkomende argumenten aan waarom de lectische dimensie een integraal onderdeel vormt van het onderzoek naar het reële taalgebruik. Ten eerste, zelfs wanneer de corpuslinguïst niet in de lectische aspecten geïnteresseerd is, dan nog moet h/zij deze factoren bestuderen om hun effect te

neutraliseren. Van *ease of articulation*-factoren, zoals de fonologische conditionering van de buigingsalternantie, kan bijvoorbeeld worden aangenomen dat hun effect groter is in informele contexten dan in formele contexten waar de productiedruk minder groot is en waar meer zorg aan de vormelijke aspecten van de talige productie kan worden besteed. Ten tweede, Geeraerts beperkt de betekenis van een talig teken niet tot de denotationele betekenis die de betekeniskenmerken groepeerst die betrekking hebben op de referentie of de taalkundige betekenis van een teken. Naast de denotationele betekenis onderscheidt hij verschillende niet-denotationele betekenisaspecten, waaronder de stilistische betekenis van een talig teken. De stilistische betekenis verschaft informatie over de gesprekscontext en de relatie tussen de gesprekspartners. De stilistische betekenisaspecten worden geïllustreerd door de volgende paren van denotationele synoniemen: *rijwiel/fiets, aanvang/begin, vader/ouweheer*. In deze paren is het eerste item formeler dan het tweede.<sup>1</sup> De keuze tussen de elementen in deze paren wordt bepaald door de gesprekscontext of door het beeld dat de spreker van zichzelf wil ophangen. Aan de stilistische betekenisdimensie voegen we de geografische betekenisdimensie toe. We verduidelijken de geografische betekenisaspecten aan de hand van de volgende denotationele synoniemenparen: *fruitsap/vruchtensap, dokkeren/ratelen, hesp/ham, luik/onderdeel, pollepel/soeplepel*. Het verschil tussen de elementen in deze paren is geografisch: het eerste woord is de Belgisch Nederlandse variant, het tweede woord de Nederlands Nederlandse variant. Deze alternanties bieden de spreker de gelegenheid om de eigen nationale identiteit uit te drukken.<sup>2</sup>

De geografische variëteiten worden gedefinieerd op basis van de taalgemeenschap, zeg maar de groep van sprekers die deze variëteit als communicatiemiddel gebruiken. De registers of stilistische variëteiten zijn contextueel gedefinieerd: in functie van een onderwerp, een gesprekspartner, een gespreksituatie, enz. kiezen taalgebruikers een bepaald register. Hudson (1996: 45) parafraseert deze visie als “varieties according to user” voor de regionale variëteiten en “varieties according to use” voor de stilistische variëteiten.

Aan beide types lectische variëteiten wordt vaak een verschillende status toegekend: een spreker wordt op basis van zijn/haar geboorte-, school- of woonplaats bij één bepaalde regionaal gedefinieerde taalgemeenschap ingedeeld en selecteert in functie van de specifieke gesprekscontext een register binnen deze regionale variëteit. Met andere woorden: de regionale taalvariëteiten vormen afzonderlijke systemen met een eigen stilistische stratificatie. Deze visie impliceert dat de meeste taalgebruikers slechts één nationale of geografische variëteit beheersen. Het strikte onderscheid tussen beide types lectische variabelen wordt onder andere door Hudson (1996) en Geeraerts (2005) in vraag gesteld. Zij betogen dat de verschillende types lectische variëteiten niet strikt kunnen worden afgebakend:

“We cannot rely on the view that stylistic variation could be consciously employed by most language users whereas dialect

variation could not because most speakers master only one dialect. In actual practice, many language users have an active command of more than one sociolect and/or dialect, and actively switch between the various elements of their lectal repertoire. At the same time, the repertoire of lects of the individual speakers in a linguistic community is not the same.” (Geeraerts 2005)

Geeraerts (2005) besluit dat de homogene taalgemeenschap waar alle sprekers één repertoire talige uitdrukkingsmiddelen delen niet verdedigbaar is. Als alternatief stelt hij het talige systeem van een taalgemeenschap voor als een diasysteem, dit wil zeggen als “a conglomerate of overlapping repertoires” (Geeraerts 2005).

Voor de operationalisering van de geografische en stilistische hypothesen hebben we echter nood aan discrete en – bijgevolg onvermijdelijk – geïdealiseerde geografische en stilistische distincties. De doelstelling van dit onderzoek is immers de identificatie van de lectische conditionering van de buigingsalternantie en niet de modellering van het diasysteem voor het Nederlands. Daartegenover staat dat we de stilistische variatie niet los van de regionale variatie (kunnen) beschouwen, zoals duidelijk zal worden bij de bespreking van de lectische onderzoekshypothesen in de volgende paragraaf.

## 12.2 Hypothesen en operationalisering

Eerst becommentariëren we de hypothesen die we op basis van de literatuur geformuleerd hebben. Hierbij gaan we uit van de historische evolutie van de adjectivische buiging in het algemeen en van de standaardisering(sachterstand) in het Belgisch Nederlands in het bijzonder (12.2.1). Daarna bespreken we hoe we bij de operationalisering van de lectische hypothesen gebruik maken van de macrostructuur van het CGN (12.2.2).

### 12.2.1 Hypothesen

Op basis van het literatuuroverzicht formuleren we één geografische en twee geografisch gedifferentieerde stilistische hypothesen:

- De onverbogen buigingsvorm is de Belgisch Nederlandse of Vlaamse variant, terwijl de verbogen buigingsvorm de Nederlandse variant is (zie paragraaf 1.1.4).
- In het Nederlands Nederlands wordt de onverbogen vorm geassocieerd met zeer formeel taalgebruik, in het bijzonder met literair en ambtelijk taalgebruik. De Rooij (1980a) beschouwt het gebruik van het onverbogen adjectief als een uiting van een “bijzondere [zeer formele; JT] stijl” (zie paragraaf 1.1.5). Gegeven de zeer formele en ambtelijke waarde die in de literatuur met de onverbogen buigingsvariant geassocieerd wordt, stellen we voor om deze hypothese uit te breiden met de stelling dat de buigings-

alternantie in het Nederlandse deelcorpus van het CGN niet stilistisch geconditioneerd is. Anders gezegd: in het Nederlands Nederlands, zoals weergegeven in het CGN, is de onverbogen buigingsvariant ofwel zeer formeel ofwel niet stilistisch gemotiveerd.

- Het Belgisch Nederlands vertoont de tegenovergestelde stilistische situatie. Net als in het Nederlands Nederlands is het verbogen adjectief de norm in de standaardtaal. De informele contexten worden daarentegen gekenmerkt door een frequent(er) gebruik van de onverbogen pendant (zie paragraaf 1.1.5).

Deze hypothesen illustreren de verstrengeling van de lectische betekenisaspecten: de stilistische waarde van het gemarkeerde onverbogen buigingsmorfem is tegengesteld in beide nationale variëteiten van het Nederlands. Deze complexe toestand is het gevolg van divergerende historische evoluties in de noordelijke en zuidelijke contreien van het Nederlandse taalgebied. Vooraleer we de hypothesen toelichten, schetsen we een algemeen beeld van deze divergerende historische evoluties.

De divergerende ontwikkelingen waarvan sprake is, beginnen aan het einde van het Middel-Nederlands en raken in de zeventiende eeuw in een stroomversnelling als gevolg van de politieke toestand. Terwijl in het Zuiden de sterke/onverbogen buigingsvariant het onzijdige genus markeert in een drie-generasysteem, verspreidt de (Hollandse) zwakke/verbogen pendant zich in de noordelijke dialecten (zie paragraaf 2.3.1). Na de val van Antwerpen in 1585 wordt dit verschil geïnstitutionaliseerd: vanaf de tweede helft van de zeventiende eeuw wordt het gebruik van het verbogen adjectief als de basisvorm bij definiëte enkelvoudige onzijdige substantieven veralgemeend in de ontluikende standaardtaal, waarvan het Zuiden gedurende drie eeuwen Spaanse, Oostenrijkse en Franse bezetting verstoken blijft. Toen het Nederlands als gevolg van de Vlaamse ontvoogdingsstrijd drie eeuwen later een officiële taal werd in België, hebben de integrationisten, de pleitbezorgers van de overname van de Nederlandse norm, het gehaald op de particularisten, de aanhangers van een aparte Vlaamse norm. Geeraerts (2001: 337-338) parafraseert de onderliggende redenering van deze keuze als volgt:

“Vlaanderen *heeft geen* Nederlandse standaardtaal, Vlaanderen *wil* een Nederlandse standaardtaal, er *is* een Nederlandse standaardtaal in Nederland, en dus *adopteert* Vlaanderen die bestaande standaardtaal.”  
[cursivering in de originele tekst]

Verschillende culturele en politieke initiatieven versterken de talige eenheid met Nederland: de organisatie van Nederlandse Congressen (vanaf 1849), waar gekozen wordt voor een gemeenschappelijke spelling en waar in 1860 het plan wordt opgevat voor het *Woordenboek der Nederlandsche Taal*. In 1980 ondertekenen België en Nederland het Taalunieverdrag waarin de culturele verbondenheid en de gemeenschappelijke taal benadrukt worden. De keuze voor de Nederlandse norm wordt ondersteund door een expliciete taalpolitiek die een hoogtepunt kent in de jaren ‘60



van de vorige eeuw met inspanningen in het onderwijs en taalpropaganda in de media. Deze initiatieven zorgen ervoor dat het Algemeen Nederlands een dagelijkse realiteit wordt in Vlaanderen.

De op convergentie gerichte taalpolitieke initiatieven in de jaren '60 en '70 van de twintigste eeuw hebben ertoe geleid dat de standaardisering in de formele registers van het Belgisch Nederlands is doorgedrongen. Het VRT-Nederlands wordt algemeen als richtnorm voor het Algemeen Belgisch Nederlands beschouwd: deze Belgisch Nederlandse norm sluit nauw aan bij de Noord-Nederlandse norm, maar vertoont voldoende – lexicale en fonetische (Geerts 1992: 19-80; Haeseryn 1996: 110-112; Taeldeman 1992: 47-48) – verschillen om de eigen, Vlaamse identiteit talig uit te drukken.<sup>3</sup> Naast deze convergerende tendens in de formele registers van het Belgisch Nederlands stellen we een divergerende, particularistische onderstroom vast: in de informele registers heeft de tussentaal de laatste twintig jaar vaste voet aan de grond gekregen. De tussentaal is het product van een Vlaamse substandaardisering die zich niet stoort aan de op convergentie gerichte taalpolitiek. Deze taalvariëteit beantwoordt aan een concrete communicatieve behoefte bij de modale Vlaming die zich (in alledaagse situaties) niet comfortabel voelt bij het formele VRT-Nederlands dat h/zij voor formele aangelegenheden reserveert. Geeraerts (2001: 343) spreekt van een “zondagse-pakmentaliteit”: “zoals een zondags pak is de hoogste taalnorm iets waarvan de noodzaak buiten kijf staat, maar je er echt goed in voelen doe je niet”. De Caluwe (2003) besluit dat de totstandkoming van de tussentaal de vrucht van de “natuurlijke zelfredzaamheid van een taalgemeenschap” is.

De reductie van de gestage opmars van de tussentaal tot een kwestie van taalcomfort is echter een simplificatie van de werkelijkheid. De verspreiding van de tussentaal is het gevolg van verschillende, met elkaar verweven factoren (De Caluwe 2003; Goossens 2000; Taeldeman 1992; Van de Velde 1996). Ten eerste, door de toenemende mobiliteit zijn de sociale contacten niet langer tot de eigen regio (c.q. streektaal) beperkt, zodat er nood is aan een supraregionale taal waarbij de taalgebruiker zich goed voelt. Daarnaast vult de tussentaal de leemte op die ontstaan is door de teloorgang van de dialecten, waardoor deze variëteit de moedertaal is van een groeiende groep jongeren. Ten derde, beleidsmatig wordt taalzorg in het onderwijs en in de media teruggeschroefd, zodat een groeiende groep Vlamingen de norm niet of onvoldoende beheerst. Ten vierde, de politieke emancipatie van Vlaanderen gaat gepaard met de drang tot zelfbevestiging, waarbij de tussentaal de eigenheid van de regio onderstreept, zelfs in het taalgebruik van politici en bedrijfsleiders. Tot slot, de commercialisering van het medialandschap sinds het einde van de jaren '80 populariseert het media-aanbod. In Vlaamse soaps en fictiereksen – ook op het openbare net – wordt naar de tussentaal gegrepen omdat deze variëteit natuurlijker is dan het VRT-Nederlands bij de evocatie van het alledaagse leven. Bovendien verdwijnen de Nederlandse netten – die steeds een contactpunt met de Nederlandse norm vormden – haast volledig uit de Vlaamse huiskamers. De

combinatie van de exogene herkomst van het VRT-Nederlands en de endogene origine van de tussentaal leidt tot een taalkloof: in het Belgisch Nederlands is de afstand tussen de formele en de informele registers groter dan in het Nederlands Nederlands (Geeraerts 2001, 2003a; Geeraerts, Grondelaers & Speelman 1999).

Tegen deze (beknopte en onvolledige) historische achtergrond kunnen we nu de lectische hypothesen toelichten. De situatie is het duidelijkst in het Nederlands Nederlands waar de verbogen vorm de basisvariant is in zowel de (noordelijke) streektalen als in de standaardtaal. De huidige (zeer) formele waarde van het onverbogen adjectief gaat terug op het gebruik van deze variant in het ambtelijke en literaire taalgebruik vanaf de zeventiende eeuw (zie paragraaf 2.3.2). Als we terugblikken op de bespreking van het dialectologische onderzoek naar de adjectivische buigingsalternantie (zie paragraaf 2.3.1) moeten we het beeld bijschaven dat de keuze tussen de verbogen en de onverbogen vormen van het adjectief samenvalt met de staatsgrens tussen Nederland en België. Ten eerste, het gebied waar de onverbogen variant de dominante streektalige vorm is, moet in noordelijke richting worden uitgebreid met Nederlands Limburg en Noord-Brabant. Ten tweede, naast de homogene onverbogen dialectzone ten zuiden van de rivieren en de homogene verbogen dialectzone ten noorden van de rivieren zijn er enkele menggebieden waar de onverbogen vorm niet na een bepaald lidwoord geselecteerd wordt, maar wel na een bezittelijk voornaamwoord kan voorkomen. Een eerste mengzone ligt in het noordoosten van Nederland: Friesland, Groningen en Oost-Drenthe. Zeeuws-Vlaanderen vormt de tweede mengzone in het zuidwesten. Het onderzoek van de geografische conditionering van de buigingsalternantie is bijgevolg niet beperkt tot de nationale variëteiten van het Nederlands, maar gaat na in welke mate de grote dialectregio's of streektalen voor een verdere differentiëring zorgen binnen de nationale variëteiten, in eerste instantie binnen het Nederlands Nederlands.<sup>4</sup>

In het Belgisch Nederlands is de situatie complexer als gevolg van de daarnet genoemde taalkloof. We beginnen met het formele taalgebruik. Toegepast op de buigingsalternantie betekent de overname van de noordelijke norm dat de endogene onverbogen variant van de Vlaamse dialecten verdrongen wordt door de exogene verbogen variant. Ondanks bovenvermelde inspanningen ter verspreiding van de noordelijke norm, vermelden analyses van het Algemeen Belgisch Nederlands steevast het gebruik van de onverbogen buigingsvariant als een algemeen kenmerk van het Belgisch Nederlands (Haeseryn 1996; de Rooij 1972) en van de tussentaal (Goossens 2000). Het is bijgevolg plausibel dat de onverbogen vorm vaker in de Belgisch Nederlandse variëteit van de gesproken standaardtaal wordt gebruikt dan in de Nederlands Nederlandse variëteit, waar de onverbogen vorm nauwelijks een regiolectische voet aan de grond heeft.

De stilistische hypothese voor het Belgisch Nederlands lijkt de taalkloof te weerspiegelen: in de formele registers, die het sterkst met het Nederlands Nederlands convergeren, is de exogene verbogen vorm dominant; in de informele registers kiest de Vlaming vaker het endogene onverbogen adjectief. Op basis van een analyse van

de adnominale buiging in de tussentaal is Goossens (2000) zelfs beducht voor een contextuele tweetaligheid in Vlaanderen tussen de informele Vlaamse tussentaal en de formele Nederlandse norm.<sup>5</sup> De vraag of we een apart model moeten poneren voor de beschrijving van de buigingsalternantie in beide nationale variëteiten van het Nederlands kunnen we pas na de multivariate analyse beantwoorden (zie paragraaf 14.2.4).

Naast de twee bivariate hypothesen over de globale stilistische restricties op het gebruik van beide buigingsvarianten, onderzoeken we ook in welke mate de formaliteit van de gesprekscontext het effect van andere verklarende factoren conditioneert. In eerste instantie kijken we naar de segmenteel-fonologische conditionering van de adjectivische buiging (zie hoofdstuk 9). Deze factoren vergemakkelijken of bemoeilijken het gebruik van de onverbogen vorm in functie van een articulatorisch flexibele foneemconfiguratie (*ease of articulation*). Het is aannemelijk dat het effect van deze factoren groter is in het informele taalgebruik dan in het formele taalgebruik, dat gekenmerkt wordt door een grotere controle van de vormelijke aspecten (van Marle 1997). Daarnaast testen we de stilistische differentiëring van de lexicale/semantische conditionering van de buigingsalternantie (hoofdstuk 11). Voor de lexicale collocaties schuiven we de volgende hypothese naar voor: het is plausibel dat het deflecterende effect van adjectief-nomencollocaties groter is in formele contexten omwille van de aanwezigheid van gespreksonderwerpen die gekenmerkt worden door het gebruik van ambtelijke en formulaïsche tournures. Daarnaast lijkt het aannemelijk dat de lexicale collocaties in de informele gespreksituaties vooral uit (zeer) algemene adjectieven en substantieven zijn opgebouwd en bijgevolg minder vaak de buigings-*e* achterwege laten (zie paragraaf 11.1.2, tabel 11(7)). Tot slot testen we ook de hypothesen dat het effect van de woordsoort van de determinator beperkt is tot de informele gespreksituaties (Raidt 1968) of tot de formele gespreksituaties (de Rooij 1980a) (zie paragrafen 1.1.1 en 1.2.1).

### 12.2.2 Operationalisering

We bespreken achtereenvolgens de operationalisering van de geografische (12.2.2.1) en de stilistische hypothesen (12.2.2.2). We brengen in herinnering dat we bij de operationalisering van deze hypothesen gebruik maken van geïdealiseerde, discrete geografische en stilistische distincties. Dit is onvermijdelijk bij de beschrijving van de lectische conditionering van de buigingsalternantie in een multivariate analyse.

#### 12.2.2.1 Operationalisering van de geografische differentiëring

Het onderscheid tussen beide nationale variëteiten van het Nederlands maakt deel uit van de architectuur van het CGN, dat Belgisch Nederlands en Nederlands Nederlands materiaal verzamelt (zie paragraaf 6.1). Rekening houdend met de noordelijke herkomst van de standaard en de numerieke dominantie van het Nederlands Nederlands beschouwen we de Nederlandse variëteit als referentiepunt

voor het berekenen van de impact van de nationale variëteit op de adjectivische buigingsalternantie.

Voor de verfijning van de geografische onderscheiding van de nationale variëteiten naar de regionale variëteiten maken we gebruik van de codering van de sprekersinformatie in het CGN.<sup>6</sup> Voor iedere spreker is – idealiter – de volgende informatie gecodeerd:

- geboorteplaats en -regio: de gemeente/stad en de regio waar de spreker geboren is
- schoolplaats en -regio: de gemeente/stad en de regio waar de spreker van 4 tot 16 gewoond heeft en naar school is gegaan
- woonplaats en -regio: de gemeente/stad en de regio waar de spreker woont op het ogenblik van zijn/haar bijdrage tot het CGN

We maken gebruik van de codering op het regionale niveau. De codering op het niveau van de geboorte-, school- en woonplaats is te fijnmazig, met dataschaarste als gevolg: niet alleen ontbreekt de informatie voor een zeer groot aantal meetpunten, maar de beschikbare informatie per ‘gemeentelijk/stedelijk’ meetpunt is te beperkt voor een kwantitatieve analyse van het taalgebruik.<sup>7</sup> De regionale distincties in het CGN vallen *grosso modo* samen met de Nederlandse en Vlaamse dialectgebieden, zoals ingedeeld in Hoppenbrouwers & Hoppenbrouwers (2001).<sup>8</sup> Tabel 12(1) geeft een overzicht van de regio’s die in het CGN onderscheiden worden en de provincies die ze omvatten.

Nat.var.	CGN-regio	Provincies
Ned. Ned.	Kerngebied	Noord-Holland, Zuid-Holland, West-Utrecht, Flevoland
	Periferie	Zeeland, Zeeuws-Vlaanderen, Oost-Utrecht, Gelders rivierenland, Veluwe, West-Friesland, Polders
	N.O. randgebied	Drenthe, Groningen, Friesland, Achterhoek en Gelderland, Overijssel
	Zuidelijk randgebied	Noord-Brabant en Limburg
Bel. Ned.	Kerngebied	Brabant en Antwerpen
	Oost-Vlaanderen	Oost-Vlaanderen
	West-Vlaanderen	West-Vlaanderen
	Limburg	Limburg

Tabel 12(1): Codering van de regionale informatie in het CGN

De Vlaamse regio’s komen in grote lijnen overeen met de dialectgebieden die traditioneel onderscheiden worden. De Nederlandse regio’s zijn een minder getrouwe weergave van het dialectlandschap: de vier onderscheiden regio’s bundelen verschillende streektalen. Vooral de regio’s ‘periferie’ en ‘noordoostelijk randgebied’ zijn zeer heterogeen en groeperen streektalen die in Hoppenbrouwers &

Hoppenbrouwers (2001) tot verschillende clusters behoren. Door de getrapte codering van de regionale informatie in het CGN kunnen we wel de streektalen isoleren die relevant zijn voor onderhavig onderzoek, met name de observaties die door een inwoner van Zeeuws-Vlaanderen ('periferie')<sup>9</sup>, of door een inwoner van Drenthe, Groningen of Friesland ('NO randgebied') geuit zijn. Deze informatie gebruiken we voor de verificatie van de dialectologische bevindingen van De Rooij (1980b) dat in deze streektalen na een bezittelijk voornaamwoord de onverbogen vorm vaker geselecteerd wordt dan in de andere Nederlandse streektalen (zie paragraaf 2.3.1). Tot slot zijn beide Limburgen, die in de dialectologie als één streektaal beschouwd worden, gescheiden in functie van de rijksgrens.

#### 12.2.2.2 Operationalisering van de stilistische differentiëring

Voor de falsificatie van de stilistische hypothesen maken we eveneens gebruik van de macrostructuur van het CGN. Het CGN is in essentie niets anders dan een verzameling van opnames in verschillende gesprekssituaties, zoals: een gesprek aan tafel, een interview met een leerkracht Nederlands, zakelijke onderhandelingen, een item in een radiojournaal, de rechtstreekse verslaggeving van een sportwedstrijd, een debat in het parlement, een toespraak bij de opening van het academiejaar, enz. Deze situaties zijn gegroepeerd in 14 situationeel gedefinieerde componenten, waarvan er 12 in de geraadpleegde 5<sup>e</sup> pre-release aanwezig zijn (zie paragraaf 6.1).

Component nr	omschrijving	Situatiele parameters			Inf.index
		interactie	voorbereiding	publiek	
1	conversaties	dia/multi	spont	priv	3
2	interviews	dia/multi	spont	priv	3
4	zakelijke onderhandelingen	dia/multi	spont	priv	3
5	interviews en discussies	dia/multi	vbr	pub	1
6	discussies, debatten, vergaderingen	dia/multi	spont	pub	2
7	lessen	dia/multi	spont	pub	2
9	spontaan commentaar	mono	spont	pub	1
10	reportages	mono	vbr	pub	0
11	niewsbulletins	mono	vbr	pub	0
12	beschouwingen	mono	vbr	pub	0
13	lezingen, toespraken	mono	vbr	pub	0
14	voorgelezen tekst	mono	vbr	pub	0

Tabel 12(2): Stilistische stratificatie in het CGN (5<sup>e</sup> pre-release)

Tabel 12(2) stelt de stilistische stratificatie van het CGN voor. De indeling in 14 situationele componenten sluit aan bij de definitie van talige registers als “varieties according to use” in paragraaf 12.1 (Biber 1988; Hudson 1996).

Naast het componentnummer en de componentomschrijving bevat tabel 12(2) drie situationele parameters die de CGN-componenten op basis van drie verschillende stilistische perspectieven in een (meer) formele en een (meer) informele deelverzameling groeperen. Deze drie situationele parameters zijn:

- *interactietype*: Naar de aard van de talige interactie worden monologen (‘mono’) en dialogen/multilogen (‘dia/multi’) onderscheiden. Het onderscheid tussen beide interactietypes beïnvloedt de productiedruk bij de spreker.<sup>10</sup> In een monoloog, waar de spreker de volledige controle over het gesprekskanaal heeft, is de productiedruk aanzienlijk lager dan in een dialoog of een multiloog waar de spreker rekening moet houden met de mogelijkheid van (onverwachte) interventies en interacties van zijn/haar gesprekspartners. Het onderscheid tussen monologische en dialogische/multilogische gespreksituaties vormt de stilistische basisdistinctie in het CGN. Uit tabel 12(2) kunnen we afleiden dat de twee andere situationele parameters in hoofdzaak een verdere verfijning van dit basisonderscheid bewerkstelligen.
- *voorbereiding tekst*: De bijdrage van een spreker in een gesprek kan zowel voorbereid (‘vbr’) als spontaan (‘spont’) zijn. Net als bij het interactietype beïnvloedt deze distinctie de controle van de spreker op de tekst en de inhoud. Wanneer een spreker een voorbereide tekst voorbrengt, controleert h/zij de inhoud, zodat h/zij de volle aandacht op de vorm kan richten. Doordat de voorbereiding meestal schriftelijk is – denk aan voordrachten en toespraken – komt dit meestal neer op het voorlezen van een geschreven en dus redactioneel gecontroleerde tekst. Deze distinctie valt in grote lijnen samen met het interactietype: in prototypische monologen maakt men gebruik van voorbereide en geschreven teksten, terwijl dialogen en multilogen typisch spontane genres zijn waar de gesprekspartners zonder (schriftelijke) voorbereiding in gesprek treden. Op het parallellisme tussen de voorbereiding van de tekst en het gesprekstype zijn er twee uitzonderingen, met name de componenten 5 en 9: component 5 bevat discussies en debatten in tv- en radioprogramma’s die in ruime mate voorbereid zijn; component 9 verzamelt verschillende types rechtstreekse verslaggeving, zoals sportverslaggeving en het commentaar bij een prinselijk huwelijk.
- *publiek*: De laatste situationele parameter definieert de relatie tussen de gesprekspartners. Op basis van de (niet-fysieke) afstand tussen de gesprekspartners worden publieke gespreksituaties (‘pub’) en privé-gespreksituaties (‘priv’) onderscheiden. Deze distinctie brengt enkel een verdere verfijning aan binnen de dialogen/multilogen (zie tabel 12(2)).

Deze drie situationele parameters<sup>11</sup> relateren we aan de sprekergebaseerde laboviaanse visie op formaliteit. Labov (1972a, 1994) definieert de formaliteit van een register op basis van de aandacht die de spreker aan de vormelijke aspecten van de uiting besteedt. Met andere woorden: formaliteit weerspiegelt het groeiende besef van de spreker hoe h/zij iets verwoordt, bovenop het besef wat h/zij verwoordt. Dit vormelijke bewustzijn van de spreker wordt in ruime mate bepaald door de controle die h/zij op de productie uitoefent: des te groter de controle op de interactie, de tekst, en de inhoud, des te meer aandacht aan de vormelijke aspecten kan worden besteed. Deze controle duiden we aan met de term productiedruk (zie ook noot 10). Toegepast op de drie situationele parameters in het CGN verkrijgen we de volgende verdeling van formele en informele gesprekscontexten:

	<b>Formeel</b>	↔	<b>Informeel</b>
<b>interactie</b>	monoloog	↔	dialogue/multiloog
<b>voorbereiding</b>	voorbereid	↔	spontaan
<b>publiek</b>	publiek	↔	privé

Tabel 12(3): Formele en informele waarden van de situationele parameters in het CGN

De formele en informele waarden voor het interactietype en de voorbereiding van de tekst zijn duidelijk: de formaliteit wordt rechtstreeks bepaald door de controle van de spreker op de interactie of de tekstuele inhoud. De registerdistinctie vanuit het standpunt van het publiek vraagt om enige toelichting: Bell (1984) argumenteert dat de stilistische keuze van de spreker in ruime mate bepaald wordt door het (geïntendeerde) publiek. Als we dit criterium op de distinctie publiek/privé toepassen, bepaalt de afstand tussen de gesprekspartners de formaliteit: voor een groter en niet-vertrouwd of onbekend publiek zal de spreker meer aandacht aan de vormelijke aspecten besteden.

Tot slot is er in tabel 12(2) naast de drie kolommen met de situationele parameters nog een laatste kolom met de informaliteitsindex ('inf.index'). De informaliteitsindex van een CGN-component is gelijk aan het aantal informele waarden bij de situationele parameters (zie tabel 12(3)). Dit is een score variërend tussen 0, wat de afwezigheid van informaliteit – en dus maximale formaliteit – aanduidt, en 3, wat op een maximale informaliteit wijst. Doordat de informaliteitsindex het midden houdt tussen de drie parallelle situationele parameters en de (te) fijnmazige twaalfdelige componentstructuur, vat hij de algemene invloed van de registervariatie in één stilistisch continuum samen.<sup>12</sup> Deze informaliteitschaal varieert van zeer formele voorbereide publieke monologen (informaliteitsindex = 0) tot spontane alledaagse privé-dialogen (informaliteitsindex = 3). Tabel 12(4) schematiseert het stilistische continuum gevormd door de informaliteitsindex. Er zijn twee varianten van de informaliteitsindex met de waarde 1: ofwel levert de tekstuele voorbereiding de informele waarde ('1<sup>a</sup>', component 9), ofwel levert het interactietype de informele waarde ('1<sup>b</sup>', component 5). Bij ontstentenis van een

significant buigingseffect (zie paragraaf 12.3.2.1), beschouwen we beide varianten als één homogene informaliteitsindex, c.q. informaliteitsindex 1.

Inf.index	Situatiele parameters		
	interactie	voorbereiding	publiek
0	mono	vbr	pub
1 <sup>a</sup>	mono	spont	pub
1 <sup>b</sup>	dia/multi	vbr	pub
2	dia/multi	spont	pub
3	dia/multi	spont	priv

Tabel 12(4): *Stilistisch continuum gerealiseerd door de informaliteitsindex*

Bij de interpretatie van de informaliteitsindex moeten we voor ogen houden dat deze schaal niet in de structuur van het CGN voorzien is: wij stellen de informaliteitsindex voor als een globale maat voor de informaliteit van een gesprekssituatie. De meest formele situaties (informaliteitsindex = 0) kunnen we het best omschrijven als ‘gesproken geschreven taal’ waar de vorm en inhoud op voorhand volledig voorbereid en gecontroleerd zijn. De concrete gesprekssituaties staven deze omschrijving: component 14 omvat uitsluitend voorgelezen werken uit de blindenbibliotheek en component 13 beslaat voordrachten en lezingen.<sup>13</sup> De meest informele situaties (informaliteitsindex = 3) zijn alledaagse, ongedwongen conversaties tussen kennissen, i.e. prototypische gesprekssituaties, zoals de contexten in component 1, die onder andere familiale tafelgesprekken bevat. Wanneer we de stilistische gradatie volgens de informaliteitsindex in tabel 12(4) beschouwen, nemen we een progressieve toename van de productiedruk waar: van de afwezigheid van productiedruk (informaliteitsindex 0 is tegelijk monologisch én voorbereid) over een intermediaire productiedruk (informaliteitsindex 1 betekent ofwel dialogisch ofwel spontaan, maar niet allebei tegelijk) tot een maximale productiedruk (de informaliteitsindexen 2 en 3 zijn gelijktijdig dialogisch én spontaan).

Tabel 12(5) vat de lectische factoren samen. Bij de stilistische factoren moet men voor ogen houden dat deze afzonderlijk in beide nationale variëteiten getest worden.

Regio	Stijl
nationale variëteit	interactietype
streektaalen	voorbereiding
	publiek
	informaliteitsindex

Tabel 12(5): *Overzicht van de lectische factoren op basis van het CGN*



### 12.3 Resultaten en bespreking

We bespreken eerst de geografische verspreiding van de buigingsmorfemen (12.3.1). Vervolgens bekijken we de stilistische aspecten van het verbogen en het onverbogen adjectief (12.3.2). Tot slot gaan we na in welke mate het buigingseffect van taalinterne factoren stilistisch gedifferentieerd is (12.3.3).

#### 12.3.1 Geografische differentiëring

We beginnen met het hoogste geografische niveau, met name de nationale variëteit van het Nederlands. Tabel 12(6) toont de verspreiding van beide buigingsmorfemen in het Belgisch Nederlands en het Nederlands Nederlands.

	Nederland		België		Σ	
<b>verbogen</b>	1,841	81.53	1,969	72.76	3,810	76.75
<b>onverbogen</b>	417	18.47	737	27.24	1,154	23.25
<b>Σ</b>	<b>2,258</b>	<b>100.0</b>	<b>2,706</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>

Tabel 12(6): Invloed van de nationale variëteit van het Nederlands op de adjectivische buigingsalternantie

Deze gegevens ondersteunen de hypothese: de onverbogen vorm komt significant frequenter voor in het Belgisch Nederlands dan in het Nederlands Nederlands ( $\chi^2 = 52.5449$ ,  $df = 1$ ,  $p = 4.205e-13$ ). Ondanks de zeer kleine  $p$ -waarde voor de  $\chi^2$ -statistiek is de impact van de nationale variëteit relatief beperkt:  $RR_{\text{onverbogen}} = 1.4748$  [1.326,1.6402];  $OR = 1.6525$  [1.4425,1.893]. Rekening houdend met de gesproken natuur van het materiaal en de endogene status van het onverbogen adjectief in Vlaanderen is dit een enigszins opmerkelijk resultaat.

De structuur van het CGN biedt een aanzet tot verklaring voor het relatief beperkte aandeel onverbogen vormen in het Belgisch Nederlands. Bij de voorstelling van het CGN hebben we reeds opgemerkt dat het materiaal voor de verschillende stilistische strata onevenredig over het Belgisch Nederlandse en het Nederlands Nederlandse deelcorpus verdeeld is (zie paragraaf 6.1, tabel 6(1)): de informele contexten zijn significant oververtegenwoordigd in de Nederlandse materiaalverzameling en significant ondervertegenwoordigd in de Belgische materiaalverzameling.<sup>14</sup> Deze verklaring sluit aan bij de resultaten van een pilotstudie: in het ConDiv-corpus geschreven Nederlands<sup>15</sup> komen de buigingsverschillen tussen het Belgisch Nederlands en het Nederlands Nederlands vooral in de informele, minder gecontroleerde registers tot uiting (Tummers, Speelman & Geeraerts 2005). De validiteit van deze verklaring wordt in het stilistische deel van het lectische onderzoek geverifieerd (zie paragraaf 12.3.2).

Na de analyse van de nationale variëteiten, vestigen we nu de aandacht op de streektalen in het Belgisch Nederlands en vooral in het Nederlands Nederlands. Bij de interpretatie van de regionale sprekersinformatie stellen zich twee praktische

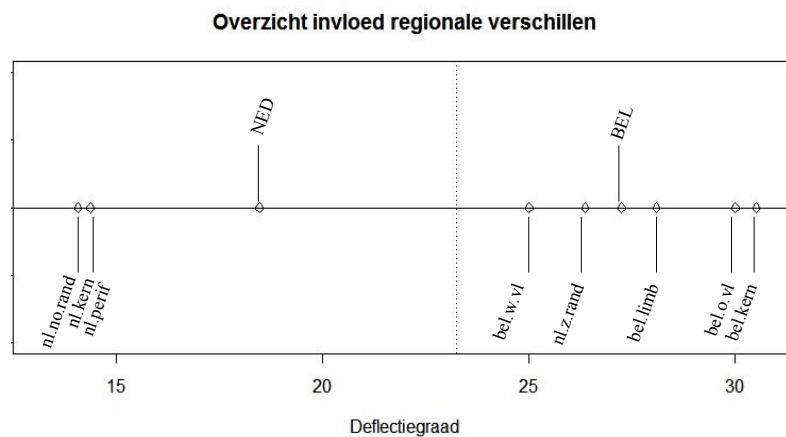
problemen. Ten eerste, de dialectkennis (c.q. regiolectkennis) van de spreker is niet gecodeerd in het CGN, zodat het onmogelijk is de effectieve invloed van de dialecten en streektalen te meten. Ten tweede, het is moeilijk om *a priori* tussen de verschillende regionale coderingen te kiezen: geboorteregio, schoolregio en woonregio.<sup>16</sup> Om beide praktische problemen maximaal te ondervangen, beperken we de analyse van de regiolectische invloeden tot de *honkvaste sprekers*. Dit zijn de sprekers die vanaf hun geboorte tot en met hun participatie aan het CGN in dezelfde regio gewoond hebben. De onderliggende redenering is de volgende: als er in het CGN regiolectische effecten aanwezig zijn, hebben we het meeste kans om deze te ontdekken bij de taalgebruikers die het sterkst op de eigen streek gericht zijn. Dit zijn in eerste instantie de sprekers die hun hele leven in dezelfde regio gewoond hebben en wier taalgebruik – idealiter – door slechts één streektaal beïnvloed is. Deze kwalitatief strenge keuze heeft echter een kwantitatieve keerzijde. We verliezen ongeveer de helft van onze observaties, vooral in Nederland: terwijl ruim 71.08% (n = 1,925) van de observaties in het Belgische deelcorpus gegenereerd zijn door een spreker die altijd in dezelfde regio gewoond heeft, bedraagt het aandeel observaties in het Nederlandse deelcorpus die geuit zijn door een honkvaste spreker slechts 33.48% (n = 756). Vooral in de noordoostelijke en zuidwestelijke gebieden, i.e. de menggebieden voor de buigingsalternantie, gaan veel observaties verloren, met als gevolg dat we te weinig gegevens hebben om de invloed van het bezittelijk voornaamwoord op basis van het taalgebruik van de honkvaste sprekers in deze inflectionele menggebieden te testen. Hierop komen we later terug.

We kijken eerst naar de Vlaamse regiolecten. Op basis van het bestaande dialectologische onderzoek (zie paragraaf 2.3.1) nemen we aan dat er geen interne regionale verschillen in Vlaanderen optreden. De Rooijs (1980b) kaarten bevatten slechts vijf uitzonderingen op de algemene verspreiding van de onverbogen vorm in Vlaanderen: voor drie meetpunten in West-Vlaanderen en twee meetpunten in Oost-Vlaanderen hebben de respondenten de verbogen vorm als basisvariant opgegeven. De afwezigheid van significante buigingsverschillen tussen de Vlaamse streektalen in het corpusmateriaal bevestigt dit vermoeden ( $\chi^2 = 2.4503$ ,  $df = 3$ ,  $p = 0.4843$ ).

De regionale vierdeling die in het CGN voor het Nederlands Nederlands is voorgesteld, vertoont wel significante verschillen ( $\chi^2 = 20.8327$ ,  $df = 3$ ,  $p = 0.0001140$ ). Dit effect komt volledig op het conto van Noord-Brabant en Limburg, de twee Nederlandse dialectzones waar de onverbogen vorm endogeen is: de onverbogen vorm wordt in het zuidelijke randgebied (i.e. Nederlands Limburg en Noord-Brabant; cf. tabel 12(1)) – statistisch gesproken – significant vaker dan verwacht geselecteerd (95% BI partiële  $\chi^2 = [1.2761, 2.5242]$ ). De buigingsprofielen voor de gebieden die in het dialectologische onderzoek als mengzones geboekstaafd staan, Zeeuws-Vlaanderen in het zuidwesten, en Drenthe, Groningen en Friesland in het noordoosten, vertonen geen significante stijging van de onverbogen buigingsvariant. Dit kan een gevolg zijn van dataschaarste, die bovendien de verificatie van de hypothese met betrekking tot het gebruik van de onverbogen vorm na een

bezittelijk voornaamwoord onmogelijk maakt. Daarom kijken we nu naar de gegevens op basis van de drie regionale coderingen, ongeacht of een spreker al dan niet naar een andere regio verhuisd is. Wanneer we alle uitingen samenvoegen van de sprekers die in Drenthe, Groningen en Friesland geboren zijn, vertonen ze voor de geboorteregio significant meer onverbogen vormen na een bezittelijk voornaamwoord dan de uitingen van sprekers die in een andere Nederlandse regio geboren zijn (*Fisher Exact*-test:  $p = 0.004509$ ) en significant meer onverbogen vormen na een bezittelijk voornaamwoord dan na een andere determinator (*Fisher Exact*-test:  $p = 0.002655$ ).<sup>17</sup> Ter controle hebben we dezelfde tests voor de andere geboorteregio's in Nederland uitgevoerd en het resultaat was telkens negatief.<sup>18</sup> Voor de schoolregio en de woonregio kan de  $H_0$  (dit is: het bezittelijk voornaamwoord definieert geen significant buigingsverschil in de menggebieden) niet verworpen worden.

Figuur 12(1) vat de resultaten van de regionale analyse samen: de X-as identificeert de deflectiegraad. De verticale stippellijn markeert de gemiddelde deflectiegraad voor de volledige materiaalverzameling. De punten identificeren de verschillende regio's die in het CGN onderscheiden worden (zie tabel 12(1)). De labels boven de horizontale lijn identificeren de gemiddelde deflectiegraad voor het Nederlands Nederlands ('NED') en het Belgisch Nederlands ('BEL') op basis van de gegevens in tabel 12(6). De labels beneden de horizontale lijn identificeren de deflectiegraad in de verschillende Nederlandse en Vlaamse streektalen (zie tabel 12(1)) op basis van de buigingsprofielen bij de honkvaste sprekers. We herinneren eraan dat de honkvaste sprekers goed zijn voor 71% van de Belgisch Nederlandse observaties en 33% van de Nederlands Nederlandse observaties. Het label bij ieder regionaal meetpunt bestaat uit twee delen: het deel voor de (eerste) punt identificeert de nationale variëteit ('nl': Nederland; 'bel': België); het deel na de (eerste) punt identificeert de CGN-regio zoals beschreven in tabel 12(1).



*Figuur 12(1): Overzicht van de streektalige verschillen bij de adjectivische buigingsalternantie*

Als we de verspreiding van de punten ten overstaan van de verticale stippellijn beschouwen, zien we dat de punten in het linker vlak de Nederlandse regio's voorstellen en de punten in het rechter vlak de Vlaamse regio's, met uitzondering van het meetpunt voor Noord-Brabant en Nederlands Limburg ('nl.z.rand') dat zich tussen de Belgisch Nederlandse meetpunten bevindt. De globale deflectiegraad voor beide nationale variëteiten ligt niet mooi in het midden tussen de verschillende regionale variëteiten. Dit is een gevolg van het verschillende gewicht van de regio's en van het feit dat de honkvaste sprekers (niet-significant) meer verbuigen in Nederland (Noord-Brabant en Limburg buiten beschouwing gelaten) en (niet-significant) minder verbuigen in Vlaanderen dan de niet-honkvaste sprekers. Met andere woorden: de honkvaste sprekers lijken de streektalige buigingsvariant te cultiveren.

De Nederlandse regio's in figuur 12(1) kunnen in twee groepen gesplitst worden: Noord-Brabant en Limburg ('nl.z.rand') aan de ene kant, en de andere regio's aan de andere kant. Dit sluit aan bij de dialectologische bevindingen: net als in Vlaanderen is de verbogen variant exogeen in de Zuid-Nederlandse provincies. We hebben in paragraaf 2.3.1 gezien dat de scheiding tussen de noordelijke en de zuidelijke streektalen niet samenvalt met de rijksgrens. De deflectiegraad voor het noordoosten van Nederland ('nl.no.rand') is onverwacht laag. Dit contrasteert met de significante toename van het onverbogen adjectief na een bezittelijk voornaamwoord in het taalgebruik van de sprekers die in deze regio geboren zijn (cf. *supra*). Het is niet uitgesloten dat deze uiterst linkse positie op de deflectieschaal in figuur 12(1) streektalig is: door het ontbreken van de sjwa-apocope in deze regio (Weijnen 1991: 275 kaart 45) is het plausibel dat het algemene buigingsprofiel voor deze regio meer vormelijk verbogen adjectieven (buigings-*e* én stam-*e*; zie ook paragraaf 2.2) bevat dan de buigingsprofielen voor de andere regio's in Nederland. Indien deze verklaring correct is, betekent dit dat bij de keuze van de buigingsvariant in het noordoosten van Nederland twee tegengestelde regiolectische factoren actief zijn: de afwezigheid van de sjwa-apocope faciliteert de keuze van de verbogen vorm, terwijl het aandeel van de onverbogen vorm stijgt na een bezittelijk voornaamwoord. In het taalgebruik van de honkvaste sprekers in de kern van Nederland ('nl.kern') en de ruime periferie ('nl.perif') komt de onverbogen variant minder vaak voor dan het Nederlandse algemeen gemiddelde.<sup>19</sup> Dit zijn de streektalen die nooit het onverbogen adjectief selecteren bij een enkelvoudig *het*-woord in een definiëte NP (zie paragraaf 2.3.1).

De Vlaamse regio's drommen samen in een relatief smalle zone op de X-as als gevolg van het ontbreken van significante onderlinge verschillen. Voor het verklaren van de positie van het West-Vlaams ('bel.w.vl') als minst deflecterende Vlaamse streektaal roepen we, net als voor het noordoosten van Nederland, het ontbreken van de sjwa-apocope in (Weijnen 1991: 275 kaart 45).<sup>20</sup>

In de volgende paragraaf wordt de geografische oppositie tussen het Belgisch Nederlands en het Nederlands Nederlands verfijnd door het toevoegen van de

stilistische dimensie. Empirisch onderzoek heeft immers aangetoond dat de lexicale (Geeraerts, Grondelaers & Speelman 1999) en inflectionele (Tummers, Speelman & Geeraerts 2005) afstand tussen beide nationale variëteiten vergroot naarmate het register informeler wordt.

### 12.3.2 *Stilistische differentiëring*

De rapportering van de resultaten van het stilistische onderzoek valt uiteen in twee delen. In deze paragraaf bespreken we de feitelijke stilistische conditionering van de buigingsalternantie. In paragraaf 12.3.3 kijken we naar de mogelijke differentiërende invloed van de formaliteit van het register op het buigingseffect van taalinterne factoren.

Rekening houdend met de vermoedelijk tegengestelde stilistische betekenis van de onverbogen variant in beide nationale variëteiten analyseren we de stilistische conditionering afzonderlijk voor het Belgisch Nederlands en het Nederlands Nederlands. We bespreken achtereenvolgens de verdeling van beide buigingsvarianten over de drie situationeel gedefinieerde registersverschillen: het interactietype van het gesprek, de voorbereiding van de tekst en de relatie tussen de spreker en het publiek. Daarna stellen we op basis van de informaliteitsindex een globaal beeld voor. De formele taalvariëteit doet telkens dienst als referentiepunt voor de berekening van het RR en de OR.

#### *Interactietype*

We bekijken eerst de invloed van het gesprekstype. In het Belgisch Nederlands komt de onverbogen vorm significant vaker voor in dialogen/multilogen dan in monologen ( $\chi^2 = 180.212$ ,  $df = 1$ ,  $p < 2.2e-16$ ), zoals voorzien in de hypothese.

	monoloog		dialog/multiloog		$\Sigma$	
<b>verbogen</b>	1,353	81.90	616	58.21	1,969	72.76
<b>onverbogen</b>	299	18.10	438	41.79	737	27.24
<b><math>\Sigma</math></b>	<b>1,652</b>	<b>100.0</b>	<b>1,054</b>	<b>100.0</b>	<b>2,706</b>	<b>100.0</b>

Tabel 12(7a): *Invloed van het interactietype op de adjectivische buigingsalternantie in het Belgisch Nederlands*

De odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in een dialoog bedragen het drievoud van hun tegenhanger in een monoloog: OR = 3.2492 [2.7265,3.8720]. Het  $RR_{\text{onverbogen}}$  is gelijk aan 2.3091 [2.0378,2.6166].

Laten we nu de situatie in het Nederlands Nederlands bekijken. Hoewel de hypothese het onverbogen adjectief niet als informeel bestempelt in het Nederlands Nederlands, is er een significante stijging van de onverbogen vorm in de dialogische/multilogische gesprekken ten opzichte van de monologen ( $\chi^2 = 40.0909$ ,  $df = 1$ ,  $p = 2.424e-10$ ;  $RR_{\text{onverbogen}} = 1.7794$  [1.4853,2.1318]; OR = 2.0231 [1.6256,2.5179]).

	monoloog		dialoog/multiloog		$\Sigma$	
<b>verbogen</b>	1,003	86.61	838	76.18	1,841	81.53
<b>onverbogen</b>	155	13.39	262	23.82	417	18.47
<b><math>\Sigma</math></b>	<b>1,158</b>	<b>100.0</b>	<b>1,100</b>	<b>100.0</b>	<b>2,258</b>	<b>100.0</b>

Tabel 12(7b): Invloed van het interactietype op de adjectivische buigingsalternantie in het Nederlands Nederlands

### Vorbereiding

De tweede situationele parameter, het onderscheid tussen een voorbereid en een spontaan gesprek, toont een vergelijkbaar beeld: zowel in het Belgisch Nederlands ( $\chi^2 = 122.5376$ ,  $df = 1$ ,  $p < 2.2e-16$ ) als in het Nederlands Nederlands ( $\chi^2 = 39.3282$ ,  $df = 1$ ,  $p = 3.582e-10$ ) komt de onverbogen vorm significant vaker voor in spontane gesprekken dan in voorbereide gesprekken.

	voorbereid		spontaan		$\Sigma$	
<b>verbogen</b>	1,516	78.84	453	57.85	1,969	72.76
<b>onverbogen</b>	407	21.16	330	42.15	737	27.24
<b><math>\Sigma</math></b>	<b>1,923</b>	<b>100.0</b>	<b>783</b>	<b>100.0</b>	<b>2,706</b>	<b>100.0</b>

Tabel 12(8a): Invloed van de tekstuele voorbereiding op de adjectivische buigingsalternantie in het Belgisch Nederlands

De scores voor de Belgisch Nederlandse deelverzameling in het CGN zijn conform met de hypothese:  $OR = 2.7134$  [2.2684,3.2458];  $RR_{onverbogen} = 1.9913$  [1.7678,2.2431]. De hypothese voor het Nederlandse materiaal wordt andermaal door de corpusgegevens tegengesproken: de zin van de associatie is tegengesteld aan het beeld dat in de literatuur naar voren geschoven wordt ( $OR = 2.0099$  [1.6150,2.5013];  $RR_{onverbogen} = 1.77$  [1.4775,2.1205]).

	voorbereid		spontaan		$\Sigma$	
<b>verbogen</b>	1,000	86.58	841	76.25	1,841	81.53
<b>onverbogen</b>	155	13.42	262	23.75	417	18.47
<b><math>\Sigma</math></b>	<b>1,155</b>	<b>100.0</b>	<b>1,103</b>	<b>100.0</b>	<b>2,258</b>	<b>100.0</b>

Tabel 12(8b): Invloed van de tekstuele voorbereiding op de adjectivische buigingsalternantie in het Nederlands Nederlands

### Publiek

Alleen voor het publieksperspectief doorstaan beide hypothesen de toetsing aan de corpusgegevens. In het Belgische deelcorpus komt de onverbogen vorm significant vaker voor in privé-gesprekken dan in publieke gesprekken ( $\chi^2 = 59.499$ ,  $df = 1$ ,  $p = 1.224e-14$ ;  $OR = 2.6030$  [2.0325,3.3337];  $RR_{onverbogen} = 1.859$  [1.6143,2.1407]).

	publiek		privé		Σ	
<b>verbogen</b>	1,812	75.03	157	53.58	1,969	72.71
<b>onverbogen</b>	603	24.97	136	46.41	739	27.29
<b>Σ</b>	<b>2,415</b>	<b>100.0</b>	<b>293</b>	<b>100.0</b>	<b>2,708</b>	<b>100.0</b>

Tabel 12(9a): Invloed van het publiekstype op de adjectivische buigingsalternantie in het Belgisch Nederlands

In het Nederlandse deelcorpus is er nauwelijks een verschil tussen het buigingsprofiel in publieke gesprekken en het buigingsprofiel in privé-gesprekken ( $\chi^2 = 2.256$ ,  $df = 1$ ,  $p = 0.1331$ ). Dit is de enige situationele parameter in het CGN waarvoor de – uitgebreide variant van de – stilistische hypothese voor het Nederlands Nederlands de toetsing aan het taalgebruik doorstaat.

	publiek		privé		Σ	
<b>verbogen</b>	1,234	80.65	607	83.83	1,841	81.53
<b>onverbogen</b>	296	19.35	121	16.62	417	18.46
<b>Σ</b>	<b>1,530</b>	<b>100.0</b>	<b>728</b>	<b>100.0</b>	<b>2,258</b>	<b>100.0</b>

Tabel 12(9b): Invloed van het publiekstype op de adjectivische buigingsalternantie in het Nederlands Nederlands

#### Informaliteitsindex

Tot slot bekijken we de impact van de informaliteitsindex (zie tabel 12(4)) op de buigingsalternantie. De gegevens voor het Belgisch Nederlands zijn conform met de hypothese: het aandeel onverbogen vormen groeit progressief naarmate de informaliteitsindex stijgt ( $\chi^2 = 184.3803$ ,  $df = 3$ ,  $p < 2.2e-16$ ), met net geen absolute voorkeur voor het onverbogen buigingsmorfeem in de meest informele contexten ('inf.ind.3').

	inf.ind.0		inf.ind.1		inf.ind.2		inf.ind.3	
<b>verbogen</b>	1,297	82.24	275	65.32	240	57.83	157	53.58
<b>onverbogen</b>	280	17.76	146	34.68	175	42.17	136	46.42
<b>Σ</b>	<b>1,577</b>	<b>100.0</b>	<b>421</b>	<b>100.0</b>	<b>415</b>	<b>100.0</b>	<b>293</b>	<b>100.0</b>

Tabel 12(10a): Invloed van de informaliteitsindex op de adjectivische buigingsalternantie in het Belgisch Nederlands

Als we naar het Nederlands Nederlandse materiaal kijken, stellen we vast dat de buigingsprofielen niet constant zijn over de verschillende graden van de informaliteitsindex (*Fisher Exact-test*:  $p < 2.2e-16$ ). De relatieve frequenties van de verbogen vorm in tabel 12(10b) vertonen een U-vormige curve: het aandeel verbogen vormen daalt geleidelijk vanaf informaliteitsindex 0 tot en met informaliteitsindex 2 om dan bij informaliteitsindex 3 weer te stijgen tot een niveau dat vergelijkbaar is met dat van de informaliteitsindex gelijk aan 0.

	<b>inf.ind.0</b>		<b>inf.ind.1</b>		<b>inf.ind.2</b>		<b>inf.ind.3</b>	
<b>verbogen</b>	996	86.68	11	73.33	227	62.02	607	83.38
<b>onverbogen</b>	153	13.32	4	26.67	139	37.98	121	16.62
<b>Σ</b>	<b>1,149</b>	<b>100.0</b>	<b>15</b>	<b>100.0</b>	<b>366</b>	<b>100.0</b>	<b>728</b>	<b>100.0</b>

Tabel 12(10b): Invloed van de informaliteitsindex op de adjectivische buigingsalternantie in het Nederlands Nederlands

We herinneren eraan dat informaliteitsindex '1' twee verschillende varianten heeft (cf. het onderscheid tussen '1<sup>a</sup>' en '1<sup>b</sup>' in tabel 12(4)): 'dialoog/multiloog & voorbereid & publiek' (component 5) en 'monoloog & spontaan & publiek' (component 9). Omdat het verschil tussen beide varianten van deze informaliteitsindex niet significant is, noch in het Belgisch Nederlands ( $\chi^2 = 3.0347$ ,  $df = 1$ ,  $p = 0.0815$ ) noch in het Nederlands Nederlands (*Fisher Exact*-test:  $p = 1$ ), beschouwen we deze twee varianten als één homogene categorie die gedefinieerd wordt door de aanwezigheid van één informeel kenmerk, hetzij voor het gesprekstype (c.q. dialoog/multiloog) hetzij voor de voorbereiding van de tekst (c.q. spontaan).

#### Algemene bespreking

Na de afzonderlijke behandeling van de verschillende stilistische factoren, kijken we nu naar het globale beeld voor de stilistische conditionering van de buigingsalternantie. In eerste instantie kijken we naar de verschillende stilistisch gedefinieerde OR's en hun 95% BI'en voor het Belgisch Nederlands en het Nederlands Nederlands. Deze gegevens worden samengevat in tabel 12(11) (het register na 'vs.' is telkens het referentiepunt voor de berekening van de OR).

	<b>Belgisch Nederlands</b>		<b>Nederlands Nederlands</b>	
	<b>OR</b>	<b>95% BI</b>	<b>OR</b>	<b>95% BI</b>
<b>dia/multi vs. mono</b>	3.22	[2.71,3.83]	2.02	[1.63,2.52]
<b>spont vs. vbr</b>	2.71	[2.27,3.25]	2.01	[1.61,2.50]
<b>privé vs. publiek</b>	2.60	[2.03,3.33]	0.83	[0.66,1.05]
<b>inf.ind.1 vs. inf.ind.0</b>	2.46	[1.94,3.12]	2.37	[0.74,7.53]
<b>inf.ind.2 vs. inf.ind.0</b>	3.38	[2.67,4.27]	3.98	[3.04,5.23]
<b>inf.ind.3 vs. inf.ind.0</b>	4.01	[3.08,5.22]	1.30	[0.99,1.69]

Tabel 12(11): Overzicht van de stilistische conditionering van de adjectivische buigingsalternantie in het Belgisch Nederlands en het Nederlands Nederlands

De OR's voor het Belgisch Nederlands zijn conform met de hypothese: op alle stilistische dimensies bevat het buigingsprofiel voor de informele(re) contexten significant meer onverbogen adjectieven dan het buigingsprofiel voor de formele(re) contexten. De verschillen op het informaliteitscontinuum nemen daarenboven gradueel toe: in het Belgisch Nederlands verviervoudigen de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  als we de minst formele informaliteitsindex ('inf.ind.3') met de meest formele informaliteitsindex ('inf.ind.0') vergelijken.



De hypothese voor de stilistische buigingsvariatie in het Nederlands Nederlands wordt niet door de corpusgegevens bevestigd. Het tegendeel is zelfs waar: de onverbogen vorm komt significant vaker voor in de informele contexten gedefinieerd door twee van de drie situationele parameters in het CGN, met name het interactietype en het publiek. Als we naar de informaliteitsindex kijken, stellen we alleen in het register met een informaliteitsindex gelijk aan 2 een significante toename van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vast. Voor het Nederlands Nederlands wordt deze informaliteitsindex alleen door CGN-component 6 gerealiseerd. Deze component groepeerde discussies, debatten en vergaderingen. De nadere inspectie van de concrete gespreksituaties leert ons dat 355 van de 366 observaties in deze component geuit zijn in een debat in de Tweede Kamer. Hierbij moeten twee kanttekeningen worden gemaakt.

Ten eerste, parlementaire debatten worden gekenmerkt door ambtelijk en technisch taalgebruik dat gepaard gaat met het gebruik van adjectief-nomenparen met naamstatus, c.q. lexicale collocaties, zoals de voorbeelden in (1) aantonen:

- (1) burgerlijk huwelijk, geregistreerd partnerschap, juridisch ouderschap, grondwettelijk recht, landelijk pakket

Om te verifiëren of component 6 daadwerkelijk meer lexicale collocaties bevat dan de andere componenten, zetten we de collocabiliteit ('lex.col' vs. '¬lex.col') van de observaties uit tegen de componenten in het Nederlandse deelcorpus van het CGN ('component.6<sub>NED</sub>' vs. '¬component.6<sub>NED</sub>'). De 95% BI'en voor de partiële  $\chi^2$ -statistiek in tabel 12(12) wijzen op een significante oververtegenwoordiging van lexicale collocaties in component 6 en een significante onderverteenwoordiging van adjectief-nomenparen die geen significante lexicale collocatie vormen in deze CGN-component.

	lex.col	¬lex.col
<b>component.6<sub>NED</sub></b>	[1.1316,1.5290]	[0.7292,0.9475]
<b>¬component.6<sub>NED</sub></b>	[0.8706,1.0016]	[0.9951,1.0675]

Tabel 12(12): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de lexicale collocaties over de componenten in het Nederlandse deelcorpus

In de overige componenten van het Nederlands Nederlandse deelcorpus rijst het tegengestelde beeld op, met dit verschil dat de BI'en niet significant zijn.<sup>21</sup>

Ten tweede, de stilistische classificatie van component 6 als 'spontaan & publiek & dialogisch' is betwistbaar. De debatten in de Tweede Kamer zijn vaak voorbereide interventies die bovendien door de strakke leiding van de kamer-voorzit(s)ter een sterk monologisch tintje hebben: de verschillende gesprekspartners leveren beurtelings hun bijdrage tot het debat, (meestal) zonder interferentie van andere kamerleden. Voeg hieraan de technische en ambtelijke stijl toe, en het betreft wellicht het meest formele taalgebruik in het Nederlandse deelcorpus.

Het vermoeden dat het onverwachte stilistische buigingseffect in het Nederlands Nederlands het resultaat is van een te informele codering van parlementaire debatten toetsen we door de buigingsprofielen te bekijken in de registers gedefinieerd op basis van de drie situationele parameters in het CGN met exclusie van component 6: indien het ambtelijk taalgebruik in component 6 aan de basis ligt van het stilistische effect dat we tot nu toe in het Nederlandse deelcorpus geobserveerd hebben, bevatten de stilistisch gedefinieerde deelverzamelingen na exclusie van component 6 homogene buigingsprofielen. Tabel 12(13) vat de resultaten samen:

	OR	95% BI
<b>interactie</b>	1.30	[1.01,1.68]
<b>voorbereiding</b>	1.29	[0.99,1.67]
<b>publiek</b>	1.28	[0.98,1.65]

*Tabel 12(13): Stilistische conditionering van de adjectivische buigingsalternantie in het Nederlands Nederlands na exclusie van component 6*

De gegevens in tabel 12(13) bevestigen onze veronderstelling. Primo, de stilistische variatie op basis van de voorbereiding van de tekst is niet langer significant; het verschil tussen de buigingsprofielen voor de registers gedefinieerd door het interactietype is randsignificant.<sup>22</sup> Secundo, de OR's in tabel 12(13) zijn kleiner dan hun tegenhangers in tabel 12(11). Samengevat: voor zover een effect gevonden wordt in overeenstemming met de stilistische hypothese ten aanzien van het Nederlands Nederlands, blijkt dit effect terug te voeren te zijn tot de significante oververtegenwoordiging van lexicale collocaties in deze contexten.

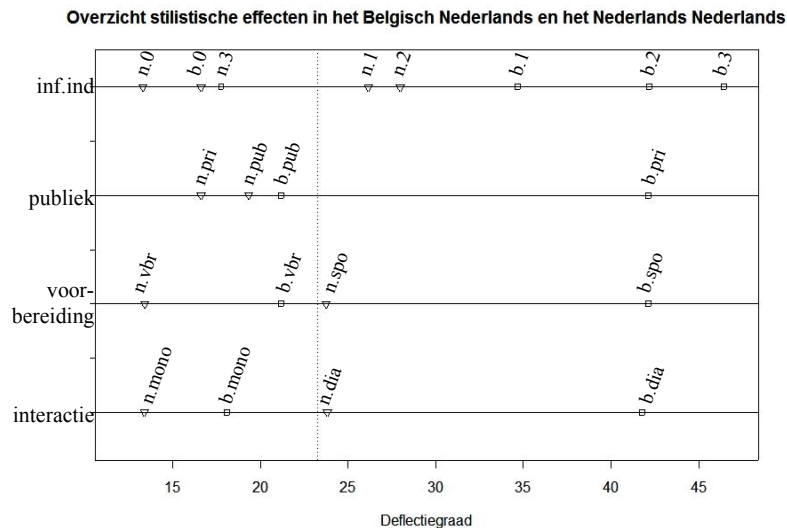
Tot slot kijken we naar het niet-significante deflecterende effect in de gesprekssituaties met informaliteitsindex 1 (zie tabel 12(11)). Dit effect hangt enerzijds samen met dataschaarste (n = 15) en anderzijds met het grote aandeel lexicale collocaties in dit register, dat items uit radiojournaals bevat: de vier observaties met een onverbogen adjectief zijn allemaal lexicale collocaties.

Naast de resumerende tabel met de OR's (tabel 12(11)) vatten we de stilistische verschillen in de twee nationale variëteiten van het Nederlands visueel samen in figuur 12(2). De X-as identificeert de deflectiegraad. De vier verschillende stilistische dimensies zijn verticaal uitgezet:

- interactietype: 'mono' (monoloog) tegenover 'dia' (dialog/multiloog)
- voorbereiding van de tekst: 'vbr' (voorbereid) tegenover 'spo' (spontaan)
- relatie van de spreker tot het publiek: 'pub' (publiek) tegenover 'pri' (privé)
- informaliteitsindex: het cijfer duidt de informaliteitsindex aan

Het label bij ieder meetpunt bestaat uit twee delen, die door een punt gescheiden zijn: vóór de punt wordt de nationale variëteit geïdentificeerd ('b': Belgisch Nederlands; 'n': Nederlands Nederlands), na de punt het register. De vierkantjes identificeren de Belgische meetpunten en de omgekeerde driehoekjes duiden de

Nederlandse meetpunten aan. De verticale stippellijn markeert de gemiddelde deflectiegraad van de gehele materiaalverzameling.



*Figuur12(2): Stilistische conditionering van de adjectivische buigingsalternantie in het Belgisch Nederlands en het Nederlands Nederlands*

Deze figuur schematiseert de globale lectische situatie in het Nederlandse taalgebied. We bespreken eerst het algemene geografische en stilistische beeld:

- De meetpunten voor het Belgisch Nederlands bevinden zich rechts ten overstaan van hun pendanten voor het Nederlands Nederlands. Deze globale positie bevestigt het resultaat van de geografische analyse: de onverbogen vorm is kenmerkend voor het Belgisch Nederlands (zie paragraaf 12.3.1).
- Het Belgisch Nederlands realiseert een hogere deflectiegraad in de informele(re) registers dan in de formele(re) registers, zowel voor de informaliteitsindex als voor de binaire stilistische opposities op basis van de drie situationele parameters in het CGN.
- Voor het Nederlands Nederlands moeten we voor ogen houden dat een eenduidige interpretatie van de CGN-componenten op een stilistisch continuum variërend van formeel naar informeel misleidend kan zijn. We hebben namelijk geargumenteerde dat het register met informaliteitsindex 'n.2' geen informeel(er) register is, zoals het cijfer voor de informaliteitsindex suggereert, maar net een zeer formeel register dat gekenmerkt wordt door een groot aandeel (deflecterende) lexicaal collocaties: in 'n.2' hangt het grote aandeel onverbogen vormen samen met het formulair taalgebruik van de Tweede-Kamerdebatten in component 6.

We vatten de bespreking van de verschillende stilistische dimensies aan op het meest globale niveau, met name op het niveau van de informaliteitsindex. Op dit

niveau onderscheiden we drie clusters. De eerste cluster bevindt zich in het gebied links van de verticale stippellijn en groepeert het formele Belgisch Nederlands ('b.0') samen met het formele Nederlands Nederlands ('n.0') en het zeer informele Nederlands Nederlands ('n.3'). Bij deze cluster moet worden opgemerkt dat het onderscheid tussen het formele Belgisch Nederlandse register en het formele Nederlands Nederlandse register in de vier stilistische dimensies significant is.<sup>23</sup> Daarnaast vertonen de buigingsprofielen van het formele Nederlandse register met informaliteitsindex 0 ('n.0') en van het informele Nederlandse register met informaliteitsindex 3 ('n.3') geen significant verschil (zie OR en BI in tabel 12(11)), zoals vooropgesteld door de hypothese.<sup>24</sup> De tweede cluster bestaat uit de drie meetpunten voor de informele Belgisch Nederlandse registers aan het rechteruiteinde van de grafiek ('b.1', 'b.2', 'b.3'). De derde, centrale cluster ('n.1', 'n.2') verzamelt enerzijds het Nederlandse register met zeer formele gespreksituaties die gekenmerkt worden door een ambtelijk en formulaïsch taalgebruik ('n.2') en waartoe component 6 behoort, en anderzijds het Nederlandse register waar de combinatie van dataschaarste en een groot aandeel lexicaal collocaties in een hoge deflectiegraad resulteert ('n.1').

Deze driedeling vormt de ruggengraat voor de buigingsverschillen op de andere stilistische dimensies, die het effect van de drie situationele parameters in het CGN visualiseren, i.e. interactietype, voorbereiding en publiek.

- Het informele Belgisch Nederlandse register ('b.dia', 'b.spo', 'b.pri') neemt op alle stilistische dimensies de meest rechtse positie in, op grote afstand van de andere Belgische en de Nederlandse registers.
- Het formele Belgisch Nederlandse register ('b.mono', 'b.vbr', 'b.pub') vormt een cluster samen met het Nederlandse materiaal. Alleen in het publieksperspectief zijn beide Belgische meetpunten ('b.pub', 'b.pri') rechts van de Nederlandse meetpunten ('n.pub', 'n.pri') gesitueerd. Enerzijds is dit het gevolg van de gelijke verdeling van de lexicaal collocaties over de publieke gesprekssituaties en de privé-gesprekssituaties in het Nederlandse materiaal, waardoor de lexicaal/semantisch bepaalde registerdistinctie verdwijnt. Het informele meetpunt ('n.pri') bevindt zich zelfs links van het formele meetpunt ('n.pub'). Anderzijds brengen we in herinnering dat deze situationele parameter uitsluitend een verdere onderscheiding binnen de dialogen/multilogen aanbrengt (zie paragraaf 12.2.2.2, tabel 12(2)), zodat een deel van de publieke gesprekken (c.q. formele waarde 'pub' voor de situationele parameter 'publiek') nog steeds dialogische/multilogische gesprekken (informele waarde 'dial' voor de situationele parameter 'interactietype') zijn.
- De Nederlandse registers 'n.dia' en 'n.spo' bezetten een centrale positie tussen de minst deflecterende cluster links en de meest deflecterende cluster rechts. In tegenstelling tot wat de stilistische interpretatie van de situationele parameters in tabel 12(3) suggereert, is dit geen informeel maar een formeel

buigingseffect als gevolg van het ambtelijk taalgebruik in CGN-component 6. Deze component is in de structuur van het CGN als dialogisch en spontaan gecodeerd, hoewel een nadere analyse van de gesprekken in deze component onthult dat het veeleer monologische en voorbereide interventies in de Tweede Kamer betreft.

Zoals aangetoond in het geografische onderzoek (zie paragraaf 12.2.1) kennen beide buigingsvarianten een verschillende verspreiding in beide nationale variëteiten. Daarnaast is de stilistische waarde van het gemarkeerde onverbogen buigingsmorfeem tegengesteld in beide nationale variëteiten. Het deflecterende stilistische effect dat we in het Nederlandse materiaal hebben waargenomen, valt grotendeels samen met de componenten die gekenmerkt worden door een groot aantal lexicale collocaties. Deze vaststelling sluit aan bij de hypothese die aan de onverbogen vorm in het Nederlands Nederlands een zeer formele en ambtelijke betekenis toekent: de componenten met een groot aandeel adjectief-nomenparen met naamstatus definiëren situaties en gespreksonderwerpen waar ambtelijk taalgebruik niet ongewoon is, zoals debatten in de Tweede Kamer.

In het Belgisch Nederlands is de verbogen vorm van het adjectief de standaardkeuze. Deze tendens is het sterkst in monologische gespreksituaties, die het vaakst de laagste, c.q. meest formele, informaliteitsindex realiseren (zie tabel 12(4)). In de informele(re) gesprekscontexten wordt de endogene variant significant vaker geselecteerd dan in de formele(re) contexten. De OR's in tabel 12(11) tonen aan dat het interactietype de belangrijkste stilistische conditionering is: wanneer een Vlaming dialogueert, verdrievoudigen de odds <sup>onverbogen</sup>/<sub>verbogen</sub>. Daarna volgt de voorbereiding van de tekst (OR = 2.71). Deze twee situationele parameters definiëren de productiedruk: in een dialoog/multiloog en in een spontaan gesprek verliest de spreker de controle op de interactie en de tekstuele inhoud, met een stijging van de productiedruk als gevolg. Bij een stijgende productiedruk vermindert automatisch de zorg voor de vormelijke aspecten. In het Belgisch Nederlands is de zorg voor het taalgebruik verstrengeld met het onderscheid tussen de standaardtaal en de tussentaal: het endogene substraat, inclusief de onverbogen vorm, wordt prominenter bij een hoge productiedruk. De externe en interne afbakening van de rechtse, informele cluster voor de informaliteitsindex in figuur 12(2), i.e. 'b.1', 'b.2' en 'b.3', bevestigt deze analyse. Extern wordt deze cluster als geheel gedefinieerd door de aanwezigheid (tegenover de afwezigheid) van productiedruk: 'b.1' betekent ofwel dialogisch én voorbereid ofwel spontaan én monologisch, terwijl 'b.0' voorbereid én monologisch betekent, wat neerkomt op een minimale (c.q. afwezige) productiedruk. Intern staat 'b.1' los van 'b.2' en 'b.3': terwijl in register 'b.1' slechts één van de twee waarden die de productiedruk verhogen, gerealiseerd wordt, zijn deze waarden simultaan aanwezig in de registers 'b.2' en 'b.3'. Dit zijn de informaliteitsindexen die tegelijk dialogisch én spontaan zijn, wat neerkomt op een maximale productiedruk of een minimale controle. De significantie van dit interne verschil tussen 'b.1' enerzijds en 'b.2' en 'b.3' anderzijds ( $\chi^2 = 4.6442$ ,  $df = 1$ ,  $p =$

0.03116) onderstreept het belang van de productiedruk bij de keuze tussen beide buigingsvarianten in het Belgisch Nederlands.<sup>25</sup>

We stellen vast dat de corpusgegevens aansluiten bij de taalkloof, niet-tegenstaande het CGN een staalname van de gesproken *standaardtaal* is. De grote waarden van de OR's in het Belgisch Nederlands in tabel 12(11) worden in figuur 12(2) gevisualiseerd als een grote(re) afstand tussen het formele Belgisch Nederlands en het Nederlands Nederlands enerzijds en het informele Belgisch Nederlands anderzijds. Bovendien zijn de OR's die het buigingsprofiel in de informele en de formele registers vergelijken, voor een aantal stilistische dimensies significant groter in de Belgische deelverzameling dan in de Nederlandse deelverzameling: de OR voor het gesprekstype, de OR voor het publiek en de OR die het register met informaliteitsindex 3 vergelijkt met het register met informaliteitsindex 0 (we verwijzen naar de 95% BI'en voor het Belgisch Nederlands en het Nederlands Nederlands in tabel 12(11)). Deze gegevens sluiten aan bij vroegere bevindingen op basis van lexicale (Geeraerts, Grondelaers & Speelman 1999) en inflectionele variabelen (Tummers, Speelman & Geeraerts 2005). De significantie van deze verschillen tussen beide nationale variëteiten kan echter niet zomaar aangegrepen worden als een bevestiging van Goossens' (2000) beduchtheid voor het ontstaan van een nieuwe tweetaligheid in Vlaanderen tussen een Nederlandse en een Vlaamse variëteit. Ten eerste, in geen enkel register van het Belgisch Nederlands is er een absolute voorkeur voor de onverbogen buigingsvariant. Dit sluit aan bij de algemene kenmerken van pluricentrische talen waar de grammatica de belangrijkste eenheids-scheppende factor is (zie paragraaf 12.2.1). Ten tweede, dit onderzoek heeft tot nu toe slechts één significant verschil tussen het Belgisch Nederlands en het Nederlands Nederlands blootgelegd: om historische redenen heeft de buigingsalternantie in het Belgisch Nederlands een stilistische functie ontwikkeld die in het Nederlands Nederlands ontbreekt. We moeten bijgevolg het multivariate onderzoek afwachten om de vraag te beantwoorden of er nood is aan een afzonderlijk model voor het Belgisch Nederlands (zie paragraaf 14.3.4). Tot slot veronderstelt de falsificatie van Goossens' vermoeden een uitgebreider onderzoek naar de interne stilistische stratificatie van het Belgisch Nederlands op basis van een gediversifieerde verzameling talige variabelen.

### ***12.3.3 Stilistische conditionering van taalinterne factoren***

Na de falsificatie van de globale stilistische hypotheses vestigen we nu de aandacht op de mogelijke partiële stilistische conditionering van het buigingseffect van taalinterne factoren. Drie soorten factoren passeren de revue. Eerst vergelijken we het effect van de woordsoort van de determinator in formele en informele gesprekscontexten (12.3.3.1). Daarna testen we de stilistische conditionering van de segmenteel-fonologische factoren, meer bepaald de algemene veronderstelling dat het effect van deze *ease of articulation*-factoren versterkt wordt in informele gesprekssituaties (12.3.3.2). Tot slot gaan we na in welke mate het buigingseffect

van de lexicale/semantische factoren, c.q. de lexicale collocabiliteit van het adjectief-nomenpaar en de relationaliteit van het adjectief, stilistisch gedifferentieerd is (12.3.3.3).

Om de stratumanalyse overzichtelijk te houden, gebruiken we één globale en binaire stilistische variabele: we reduceren de vierdelige informaliteitsindex, die de drie situationele parameters in het CGN combineert (zie tabel 12(4)), tot een binaire oppositie: formeel (informaliteitsindex = 0) tegenover niet-formeel (informaliteitsindex  $\geq 1$ ). Deze oppositie is significant, zowel voor het volledige corpus als voor de twee nationaal gedefinieerde deelcorpora.<sup>26</sup> De stratumanalyse wordt uitgevoerd op basis van het volledige corpus, c.q. zonder onderscheiding van het Belgisch Nederlands en het Nederlands Nederlands: we gaan ervan uit dat de eventuele stilistische conditionering van taalinterne factoren vergelijkbaar is in beide nationale variëteiten, rekening houdend met de analoge stilistische structuur van het CGN voor beide nationale deelverzamelingen.

### 12.3.3.1 Woordsoort determinant

Met betrekking tot de stilistische conditionering van het buigingseffect van de woordsoort van de determinant bestaan twee tegengestelde hypothesen: Raidt (1968) stelt dat de stijging van het aandeel onverbogen vormen na een bezittelijk voornaamwoord beperkt is tot de informele registers; De Rooij (1980a) van zijn kant is de mening toegedaan dat dit effect vooral in de formele registers tot uiting komt (zie paragraaf 1.1.1 en paragraaf 1.2.1 tabel 1(2)). Doordat het aantal waarden van de verklarende variabele groter is dan twee, kunnen de algemene stratumstatistieken (c.q. de *Cochran-Mantel-Haenszel*-statistieken en de *Breslow-Day*-test) niet gebruikt worden. De analyse is bijgevolg beperkt tot de strumtabellen.

In het formele register zijn beide buigingsmorfemen homogeen over de verschillende waarden van de verklarende variabele verspreid (*Fisher Exact*-test:  $p = 0.082$ ).<sup>27</sup>

	<i>het</i>		<i>aanw.vnw</i>		<i>bez.vnw</i>		<i>genitief</i>		$\Sigma$	
<b>verb</b>	1,743	83.80	198	89.59	338	83.05	14	83.78	2,293	84.12
<b>onverb</b>	337	16.20	23	10.41	69	16.95	4	22.22	433	15.88
<b><math>\Sigma</math></b>	<b>2,080</b>	<b>100.0</b>	<b>221</b>	<b>100.0</b>	<b>407</b>	<b>100.0</b>	<b>18</b>	<b>100.0</b>	<b>2,726</b>	<b>100.0</b>

Tabel 12(14a): Invloed van de woordsoort van de determinant op de adjectivische buigingsalternantie in het formele stratum

In het informele stratum is de frequentieverdeling wel significant (*Fisher Exact*-test:  $p = 0.0003494$ ). Net als bij de analyse van de marginale tabel in paragraaf 8.1.2 (zie tabel 8(1)) is de significantie het gevolg van de significante oververtegenwoordiging van het aantal verbogen vormen (95% BI partiële  $\chi^2 = [1.0077, 1.2664]$ ) en de significante ondervertegenwoordiging van het aantal onverbogen vormen (95% BI partiële  $\chi^2 = [0.5566, 0.867]$ ) na een aanwijzend voornaamwoord (de genitieven zijn

omwille van hun lage frequentie niet in de berekening van de partiële  $\chi^2$ -statistiek opgenomen).

	<i>het</i>		aanw.vnw		bez.vnw		genitief		$\Sigma$
<b>verb</b>	1,154	66.13	262	77.06	100	65.79	1	100.0	1,517 67.78
<b>onverb</b>	591	33.87	78	22.94	52	34.21	0	0.00	721 32.22
<b><math>\Sigma</math></b>	<b>1,745</b>	<b>100.0</b>	<b>340</b>	<b>100.0</b>	<b>152</b>	<b>100.0</b>	<b>1</b>	<b>100.0</b>	<b>2,238 100.0</b>

Tabel 12(14b): Invloed van de woordsoort van de determinator op de adjectivische buigingsalternantie in het informele stratum

Tabel 12(14c) vat de resultaten van de stratumanalyse samen. (De OR's zijn berekend met het bepaald lidwoord *het* als referentiepunt.) We herinneren de lezer eraan dat de OR's de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vergelijken.

Stratum	aanw.vnw. vs. <i>het</i>		bez.vnw. vs. <i>het</i>	
	OR	95% BI	OR	95% BI
formeel	0.60	[0.38,0.94]	1.06	[0.79,1.40]
informeel	0.58	[0.44,0.76]	1.02	[0.72,1.44]
marginale tabel <sup>28</sup>	0.68	[0.55,0.86]	0.86	[0.70,1.07]

Tabel 12(14c): Overzicht van de stilistische stratumanalyse voor het buigingseffect van de woordsoort van de determinator

De gegevens in tabel 12(14c) bevestigen noch de hypothese van Raid (1968) noch de hypothese van De Rooij (1980a) met betrekking tot de stilistische conditionering van het buigingseffect van het bezittelijk voornaamwoord: in geen van beide stilistische strata selecteert het bezittelijk voornaamwoord significant vaker het onverbogen buigingsmorfeem dan het ijkpunt, i.e. het bepaald lidwoord.<sup>29</sup>

Tot slot komen we terug op het relatief grote aandeel onverbogen vormen na *uw* (5 onverbogen adjectieven tegen 7 verbogen adjectieven). De veronderstelling dat dit een tussentaligheidseffect betreft (zie paragraaf 8.1.2), is conform met de corpusgegevens: de 5 onverbogen vormen komen allemaal in het Belgisch Nederlands voor, waarvan 4 in informele gesprekscontexten.

Om kort te gaan: het effect op de buigingsalternantie van de woordsoort van de determinator wordt nauwelijks door de formaliteit van de gesprekscontext beïnvloed, met uitzondering van het deflecterende effect van *uw*, dat sterk informeel gekleurd is.

### 12.3.3.2 Fonologische factoren

We bekijken achtereenvolgens de mogelijke stilistische conditionering van de segmenteel-fonologische factoren (zie hoofdstuk 9).<sup>30</sup> De eerste waarde is telkens de ijkwaarde:



- beginklank substantief: medeklinker tegenover klinker
- sonoriteit adjectivische eindmedeklinker: niet-sonoor (i.e. stemloze eind-obstruent) tegenover sonoor

De hypothese voorspelt dat het buigingseffect van deze fonologische factoren groter is in het informele taalgebruik dan in het formele taalgebruik (zie ook paragraaf 1.2.1, tabel 1(2)).

#### *Beginfoneem substantief*

De eerste fonologische factor waarvan we de stilistische differentiëring testen, is de beginklank van het substantief. We bekijken eerst de algemene stratumstatistieken. De significantie voor de MH- $\chi^2$  (MH- $\chi^2 = 39.7511$ ,  $df = 1$ ,  $p = 2.885e-10$ ) bevestigt het resultaat van Pearsons  $\chi^2$  voor de marginale tabel (zie paragraaf 9.2.1, tabel 9(2)): ook bij controle van het register wordt de onverbogen vorm significant vaker geselecteerd vóór een substantief dat met een klinker begint dan vóór een substantief dat met een medeklinker begint (CMH-OR = 1.84079 [1.5251,2.2218]). De kleine  $p$ -waarde voor de *Breslow-Day*-test ( $p = 0.0028$ ) duidt op een partiële associatie tussen het register en de beginklank van het substantief. Om de zin van deze associatie te identificeren, kijken we naar de stratumtabellen.

Tabel 12(15a) toont de frequentieverdeling in het formele stratum.

	medeklinker		klinker		$\Sigma$	
<b>verbogen</b>	2,074	84.41	219	81.41	2,293	84.12
<b>onverbogen</b>	383	15.59	50	18.59	433	15.88
<b><math>\Sigma</math></b>	<b>2,457</b>	<b>100.0</b>	<b>269</b>	<b>100.0</b>	<b>2,726</b>	<b>100.0</b>

*Tabel 12(15a): Invloed van de beginklank van het substantief op de adjectivische buigingsalternantie in het formele stratum*

Deze frequentieverdeling is niet significant ( $\chi^2 = 1.4156$ ,  $df = 1$ ,  $p = 0.2341$ ). De significantie van de MH- $\chi^2$  wordt bijgevolg in het informele stratum gerealiseerd, voorgesteld in tabel 12(15b) ( $\chi^2 = 47.1065$ ,  $df = 1$ ,  $p = 6.723e-12$ ).

	medeklinker		klinker		$\Sigma$	
<b>verbogen</b>	1,347	70.63	170	51.36	1,517	67.78
<b>onverbogen</b>	560	29.37	161	48.64	721	32.22
<b><math>\Sigma</math></b>	<b>1,907</b>	<b>100.0</b>	<b>331</b>	<b>100.0</b>	<b>2,238</b>	<b>100.0</b>

*Tabel 12(15b): Invloed van de beginklank van het substantief op de adjectivische buigingsalternantie in het informele stratum*

Tabel 12(15c) vat de stratumanalyse samen. We besluiten op basis van de samenvatting in tabel 12(15c) dat de hiaatvermijding vóór een substantief met een klinker als beginklank uitsluitend in de informele registers de selectie van het buigingsmorfem beïnvloedt. Deze conclusie sluit aan bij de hypothese dat het effect van fonologische factoren groter is in informele, minder gecontroleerde registers.<sup>31</sup>

Stratum	OR	95% BI
formeel	1.24	[0.89,1.71]
informeel	2.28	[1.80,2.89]
CMH-OR	1.84	[1.53,2.22]

Tabel 12(15c): Overzicht van de stilistische stratumanalyse voor het buigingseffect van de substantivische beginklank

#### *Sonoriteit adjectivische eindmedeklinker*

De algemene stratumstatistieken voor de sonoriteit van de adjectivische eindmedeklinker tonen aan dat het buigingseffect van deze parameter homogeen is in de twee stilistische strata (*Breslow-Day*-test:  $p = 0.0686$ ). Zowel in de formele registers als in de informele registers wordt de buigings-*e* significant vaker geselecteerd na een sonore eindmedeklinker dan na een stemloze eindobstruent (CMH-OR = 0.4640 [0.4045,0.5322]).<sup>32</sup>

#### *12.3.3.3 Lexicale/semantische kenmerken van het adjectief en het adjectief-nomenpaar*

De laatste taalinterne factoren waarvan we de stilistische conditionering testen, zijn de lexicale/semantische kenmerken van het adjectief en het adjectief-nomenpaar. Rekening houdend met het ambtelijke en formele karakter van de meeste voorbeelden die in de literatuur worden aangehaald ter illustratie van adjectief-nomenparen met naamstatus, is het plausibel dat deze elementen vooral in het formele taalgebruik morfologisch gemarkeerd worden. We bekijken achtereenvolgens de stilistische differentiëring van het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar en van de relationaliteit van het element in de adjectivische positie.

#### *Lexicale collocabiliteit van het adjectief-nomenpaar*

Wat de stilistische conditionering van het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar aangaat (zie paragraaf 11.2.1), hebben we de volgende hypothese voorgesteld: het lijkt plausibel dat het buigingseffect van de lexicale collocabiliteit groter is in het formele stratum omdat de lexicale collocaties formulaïsche sequenties identificeren die kenmerkend zijn voor een ambtelijk en formeel taalgebruik en omdat het daarenboven niet is uitgesloten dat de lexicale collocaties in het informele stratum gerealiseerd kunnen worden door algemene adjectieven en substantieven die minder vaak het onverbogen buigingsmorfem selecteren (zie paragraaf 11.2.1, tabel 11(7)).<sup>33</sup>

De grote  $p$ -waarde voor de *Breslow-Day*-test ontkracht de hypothese ( $p = 0.9464$ ): de OR's voor het buigingseffect van de lexicale collocaties zijn homogeen over beide stilistische registers. De CMH-statistieken bevestigen het effect van de bivariate analyse in hoofdstuk 11 (tabel 11(4)): het deflecterende effect van de

lexicale collocabiliteit is significant over de verschillende stilistische registers heen ( $MH-\chi^2 = 135.7392$ ,  $df = 1$ ,  $p < 2.2e-16$ ).<sup>34</sup>

#### *Relationeel adjectief*

De hypothese voorziet dat het deflecterende effect van de relationele adjectieven, die sterk geassocieerd zijn met de lexicale collocaties, groter is in het formele stratum. De oppositie relationeel adjectief tegenover kwalificerend adjectief is significant over de verschillende stilistische strata heen ( $MH-\chi^2 = 451.1059$ ,  $df = 1$ ,  $p < 2.2e-16$ ), waarbij de relationele adjectieven significant vaker het onverbogen buigingsmorfem selecteren dan de kwalificerende adjectieven ( $CMH-OR = 4.429$  [3.8365,5.1129]). De OR's zijn echter niet homogeen over de verschillende stilistische strata verdeeld (*Breslow-Day*-test:  $p = 9.095777e-04$ ). Om de zin en de omvang van het stratumeffect te weten, kijken we naar de stratumtabellen.

Tabel 12(16a) toont het effect van het onderscheid tussen relationele en kwalificerende adjectieven in het formele stratum ( $\chi^2 = 283.8903$ ,  $df = 1$ ,  $p < 2.2e-16$ ):

	$\neg$ rel.adj		rel.adj		$\Sigma$	
<b>verbogen</b>	1,729	92.12	564	66.43	2,293	67.78
<b>onverbogen</b>	148	7.88	285	33.57	433	32.22
<b><math>\Sigma</math></b>	<b>1,877</b>	<b>100.0</b>	<b>849</b>	<b>100.0</b>	<b>2,726</b>	<b>100.0</b>

Tabel 12(16a): Invloed van de relationele adjectieven op de adjectivische buigingsalternantie in het formele stratum

Tabel 12(16b) toont dezelfde frequentieverdeling, maar dan voor het informele stratum ( $\chi^2 = 179.0549$ ,  $df = 1$ ,  $p < 2.2e-16$ ):

	$\neg$ rel.adj		rel.adj		$\Sigma$	
<b>verbogen</b>	1,192	76.56	325	47.72	1,517	67.78
<b>onverbogen</b>	365	23.44	356	52.28	721	32.22
<b><math>\Sigma</math></b>	<b>1,557</b>	<b>100.0</b>	<b>681</b>	<b>100.0</b>	<b>2,238</b>	<b>100.0</b>

Tabel 12(16b): Invloed van de relationele adjectieven op de adjectivische buigingsalternantie in het informele stratum

Tabel 12(16c) vat de stilistische stratumanalyse voor de distinctie tussen relationele en kwalificerende adjectieven samen:

Stratum	OR	95% BI
formeel	5.85	[4.70,7.29]
informeel	3.58	[2.96,4.33]
CMH-OR	4.43	[3.84,5.11]

Tabel 12(16c): Overzicht van de stilistische stratumanalyse voor het buigingseffect van de relationaliteit van het adjectief

Het deflectorende effect van de onderscheiding tussen relationele en kwalificerende adjectieven is significant groter in het formele stratum dan in het informele stratum.<sup>35</sup>

Kortom: terwijl het effect van de lexicale collocabiliteit constant is over de gebinariseerde informaliteitsindex, zijn de resultaten van de stratumanalyse voor de relationele adjectieven conform met de hypothese dat het lexicale/semantische effect op de buigingsalternantie sterker is in formele gesprekssituaties.

## 12.4 Conclusie

Na de verantwoording en situering van de lectische variabelen in het geheel van het onderzoek (paragraaf 12.1), hebben we de historische achtergrond geschetst bij de tegengestelde stilistische waarde van het onverbogen adjectief in beide nationale variëteiten van het Nederlands (paragraaf 12.2.1). Als gevolg van divergerende evoluties heeft de onverbogen vorm een zeer formele waarde in het Nederlands Nederlands en een informele waarde in het Belgisch Nederlands. We hebben aangevoerd dat de stilistische waarde van de buigingsalternantie in het Belgisch Nederlands de Vlaamse taalkloof weerspiegelt: de verbogen vorm is de exogene en formele variant, terwijl de onverbogen vorm de endogene en informele variant is. Voor de lectische annotatie van de observaties in de materiaalverzameling maken we gebruik van de macrostructuur van het CGN, die zowel geografische als stilistische distincties codeert (paragraaf 12.2.2).

De geografische verdeling van de adjectivische buigingsmorfemen bij neutra is conform met de hypothese dat sprekers van het Belgisch Nederlands de onverbogen vorm vaker selecteren dan sprekers van het Nederlands Nederlands (paragraaf 12.3.1). Dit is een gevolg van het verschillende statuut van beide buigingsvarianten in Vlaanderen: terwijl de verbogen buigingsvorm bij een *het*-woord het product is van de integrationistische vernederlandsingspolitiek die de noordelijke norm adopteerde, is de onverbogen variant de endogene vorm van de Vlaamse dialecten. Niettemin is het globale verschil tussen beide nationale variëteiten relatief beperkt wanneer we geen rekening houden met de stilistische stratificatie.

De Vlaamse streektaalen vertonen onderling geen significante verschillen. In Nederland is het geografische beeld complexer. In de Zuid-Nederlandse provincies, Noord-Brabant en Limburg, selecteren de honkvaste taalgebruikers significant vaker de onverbogen vorm dan in de overige regio's van Nederland. Dit sluit aan bij de dialectologische realiteit: net als in Vlaanderen is de onverbogen vorm de endogene vorm in deze dialectgebieden. De noordoostelijk dialectzone (Friesland, Drenthe en Groningen), die De Rooij (1980b) met betrekking tot de buigingsalternantie als een menggebied identificeert, vertoont een diffuus beeld: aan de ene kant is dit de enige regio in Nederland waar, wellicht onder invloed van de streektaal, de onverbogen vorm significant vaker geselecteerd wordt na een bezittelijk voornaamwoord; aan de

andere kant bezetten deze gebieden de uiterst linkse, c.q. flecterende, positie op het geografische deflectiecontinuum in figuur 12(1), wellicht als gevolg van het ontbreken van de sjwa-apocope. Bij de sprekers uit Zeeuws-Vlaanderen, een regio die door De Rooij (1980b) eveneens als een menggebied wordt beschouwd, is er geen enkel spoor van regiolectische invloeden in het algemene taalgebruik. In de andere Nederlandse regio's wordt, in overeenstemming met de dialectologische gegevens, de verbogen vorm vaker dan het Nederlandse algemeen gemiddelde gebruikt. We concluderen dat de regiolectische invloed in het CGN relatief beperkt is, zowel in Vlaanderen als in Nederland. Dit is wellicht – ten dele – een gevolg van het opzet van de samenstellers van het CGN om een staalname te verzamelen van de algemene Nederlandse standaardtaal (zie paragraaf 6.1).

Het stilistische deel van het lectische onderzoek heeft enerzijds de globale stilistische conditionering van de buigingsalternantie in het Belgisch Nederlands en het Nederlands Nederlands in kaart gebracht (paragraaf 12.3.2) en anderzijds de mogelijke stilistische conditionering van de woordsoort van de determinator en van de fonologische en de lexicale/semantische factoren getest (paragraaf 12.3.3).

De corpusgegevens ondersteunen de stelling dat de stilistische functie van de buigingsvarianten verschillend is in beide nationale variëteiten van het Nederlands. Als gevolg van de divergerende historische evoluties in de noordelijke en zuidelijke dialecten van het Nederlandse taalgebied is de stilistische waarde van beide buigingsvarianten tegengesteld in het Belgisch Nederlands en het Nederlands Nederlands: de noordelijke, Hollandse dialecten kennen enkel de verbogen vorm, zodat de onverbogen vorm er als zeer formeel en ambtelijk wordt aangevoeld; door de overname van de noordelijke norm is de 'regelmatige' verbogen vorm voor de modale Vlaming een exogeen element, zodat h/zij in informele contexten – al dan niet bewust – sneller de onverbogen vorm selecteert.

De hypothese dat de onverbogen variant in het Nederlands Nederlands een zeer formele waarde heeft, wordt in eerste instantie weerlegd door de hoge frequentie van de onverbogen vorm in contexten die we op basis van de algemene stilistische kenmerken als informeel getypeerd hebben. De nadere inspectie van de concrete gespreksituaties en van de concrete adjectief-nomenparen werpt een ander licht op de stilistische waarde van de onverbogen buigingsvariant in het Nederlands Nederlands. Ten eerste, de gespreksituaties met het grootste aandeel onverbogen adjectieven, c.q. debatten in de Tweede Kamer, zijn – naar onze mening – in het CGN te sterk met informele kenmerken geassocieerd: we zijn van oordeel dat de Tweede-Kamerdebatten veeleer monologisch dan dialogisch zijn, en veeleer voorbereid dan spontaan. Ten tweede, in parlementaire debatten wordt een zeer formeel en ambtelijk taalgebruik gehanteerd, dat onder andere gekenmerkt wordt door het veelvuldige gebruik van lexicale collocaties, waarvan we het deflecterende effect in het vorige hoofdstuk hebben aangetoond. We concluderen dat de

corpusgegevens conform zijn met de stilistische hypothese die de onverbogen vorm in het Nederlands Nederlands als ambtelijk en formeel karakteriseert.

In het Belgisch Nederlands is het verbogen adjectief de basisvariant in de formele(re) standaardtaal. Naarmate de gesprekscontext informeler wordt, stellen we een progressieve toename van het onverbogen buigingsmorfem vast, zonder dat deze endogene, Vlaamse variant echter dominant wordt. We hebben dit stijgende gebruik van de onverbogen variant in de informele(re) registers verbonden met de productiedruk: de Vlaming kiest sneller de endogene onverbogen buigingsvariant wanneer h/zij onder een grote(re) productiedruk staat, en bijgevolg minder aandacht aan de vorm kan besteden, dan wanneer h/zij de interactie en/of de tekstuele inhoud controleert, en er zo meer ruimte ontstaat om de vormelijke aspecten van het taalgebruik te verzorgen.

We hebben het lectische onderzoek afgesloten met de analyse van de potentiële stilistische conditionering van taalinterne factoren (12.3.3). De invloed van de woordsoort van de determinator is niet stilistisch gedifferentieerd (12.3.3.1). We hebben aangetoond dat het effect van de lexicale collocabiliteit van het adjectief-nomenpaar homogeen is over de verschillende stilistische strata, terwijl het deflecterende effect van de relationele adjectieven significant groter is in de formele gesprekssituaties (12.3.3.3). Wat de stilistische conditionering van de segmenteel-fonologische factoren aangaat (12.3.3.2), wordt de verwachte stijging van het effect van deze factoren niet eenduidig door de corpusgegevens ondersteund: de tendens om een hiaat te vermijden vóór een substantief met een beginklinker door de selectie van de onverbogen vorm is significant groter in het informele taalgebruik; het flecterende effect van de sonoriteit van de adjectivische eindmedeklinker blijft daarentegen constant over de verschillende stilistische strata.

Met de analyse van de lectische conditionering van de buigingsalternantie sluiten we deel II af. In dit deel hebben we ons toegespitst op de analyse van de condities die het gebruik van de adjectivische buigingsvarianten bij neutra beïnvloeden. Hierbij lag de klemtoon op de operationalisering, de kwantificering en de interpretatie van het buigingseffect van de afzonderlijke verklarende factoren. Zoals aangegeven in de inleiding bij dit tweede deel, vormen de bivariate analyses niet het einddoel van de empirische analyse, maar het eerste deel in een methodologisch tweeluik: nu we de verschillende potentiële restricties op de adjectivische buigingsalternantie bij neutra geïdentificeerd en geoperationaliseerd hebben, kijken we in het volgende deel van dit proefschrift naar de cumulatieve impact van deze factoren op de keuze van het buigingsmorfem (zie hoofdstuk 14).

## Noten

1. Er is een verschil tussen de eerste twee paren en het laatste paar. Het eerste element in de eerste twee paren, *rijwiel* en *aanvang*, is (zeer) formeel en de stilistische waarde van het tweede element is neutraal; in het derde paar heeft het eerste element (*vader*) een neutrale stilistische waarde en heeft het tweede element (*ouweheer*) een (zeer) informele waarde. Deze drie paren illustreren de verscheidenheid aan stilistische schakeringen die mogelijk zijn in het taalgebruik.
2. De stellingname dat het gebruik van een taalkundig teken mee bepaald wordt door lectische factoren vinden we terug in de pragmatiek, een andere taalkundige deeldiscipline die door de autonome grammaticamodellen naar de periferie werd verdrongen: “[P]ragmatics [...] [can] be conceived as the study of the mechanisms and motivations behind any of the choices made when using language (at the level of phonology, morphology, syntax, semantics, whether they are variety-internal options or whether they involve regionally, socially, or functionally distributed types of variation)” (Verschuere 1987: 36). Dit citaat bevestigt de relevantie van lectische betekenisaspecten bij de conditionering van het taalgebruik. Kristiansen (2003) analyseert in het theoretische kader van de Cognitieve Taalkunde de lectische betekenisaspecten als een onderdeel van de conceptuele structuur van het talige teken.
3. Verscheidene bijdragen in Clyne (1992) tonen aan dat de uitspraak en het (alledaagse) lexicon de taalkundige domeinen zijn die het meeste variatie vertonen en bijgevolg worden aangewend bij de expressie van de nationale eigenheid. De grammaticale domeinen, in de eerste plaats de syntaxis en de inflectionele morfologie, vertonen daarentegen een grote transnationale stabiliteit.
4. We spreken in dit hoofdstuk liever van regionale invloeden dan van dialectische invloeden omdat een dialectologische analyse de doelstellingen van dit proefschrift ruimschoots overschrijdt. Bovendien is enkel de regionale informatie in het CGN bruikbaar voor onze doeleinden (zie ook paragraaf 12.2.2.1).
5. Zo Goossens’ (2000) vrees bewaarheid wordt, zou de taalsituatie in Vlaanderen een uitzondering vormen op de algemene tendentie in pluricentrische talen, waar de grammaticale domeinen, inclusief de inflectionele morfologie, voor transnationale stabiliteit en eenheid zorgen (Clyne 1992).
6. We houden eraan Ronny Keulen te bedanken voor zijn waardevolle suggesties bij de analyse van de regionale verspreiding van de adjectivische buigingsvarianten.
7. Dit vormt een belangrijk methodologisch verschilpunt met het traditionele dialectologische onderzoek waar de verschillende (geïdealiseerde) regionale taalsystemen in kaart worden gebracht. Bijgevolg volstaat in de traditionele dialectologie de talige intuïtie van één dialectspreker per geografisch meetpunt. Ook in dialectometrisch onderzoek, zoals Hoppenbrouwers & Hoppenbrouwers (2001), worden afstanden doorgaans tussen regionale taalsystemen gemeten.

8. De classificatie van de Belgische en Nederlandse streektalen in Hoppenbrouwers & Hoppenbrouwers (2001) is gebaseerd op fonetische en fonologische criteria.
9. Merk op dat Zeeuws-Vlaanderen en Zeeland samen als één regio gecodeerd zijn in het CGN.
10. Onder productiedruk verstaan we de druk op de spreker om rekening houdend met mogelijke interacties van de gesprekspartners een inhoud over te brengen. Deze druk is omgekeerd evenredig met de controle van de spreker op de gespreksituatie en de gespreksinhoud: de druk is minimaal (en de controle maximaal) bij een monologisch gesprek op basis van een voorbereide tekst; de druk is maximaal (en de controle minimaal) bij een dialogisch/multilogisch gesprek over een niet-voorbereid onderwerp. Terwijl de spreker zich in de eerste situatie volledig op de talige vorm kan concentreren, is dit in de tweede situatie niet het geval doordat h/zij een inhoud moet construeren en rekening moet houden met mogelijke interventies van gesprekspartners.
11. In de definitieve versie van het CGN is er een vierde situationele parameter voorzien voor een verdere opsplitsing van de spontane privé-dialogen/multilogen, met name het onderscheid tussen 'direct' (componenten 1 en 2) en 'distanced' (componenten 3 en 4). Het volstaat een blik te werpen op de frequenties in tabel 6(1) om vast te stellen dat de waarde 'distanced' ondervertegenwoordigd is in de 5<sup>e</sup> pre-release die wij raadplegen.
12. We zijn ons ervan bewust dat het samenvoegen van de drie situationele parameters tot één informaliteitsindex haaks staat op de multidimensionaliteit van het stilistische model van Biber (1988). Ter verantwoording vestigen we de aandacht op de tegengestelde onderzoeksperspectieven in het onderzoek van Biber en onderhavig onderzoek: terwijl de multidimensionale stilistische ruimte bij Biber berekend wordt op basis van het gedrag van een reeks items en constructies, zijn wij geïnteresseerd in de stilistische conditionering van één talige alternantie. We zijn op zoek naar een optimale kalibrering van het stilistische continuüm voor het verklaren van de keuze tussen beide buigingsvarianten. Daarbovenop komt de vaststelling dat de drie situationele parameters in het CGN veeleer inclusief dan complementair zijn: de situationele parameters vormen geen complementaire dimensies (zoals bij Biber), maar een stapsgewijze verfijning van één basisdimensie, met name de distictie monoloog tegenover dialoog/multiloog.
13. Bij het beluisteren van de CGN-geluidsbestanden tijdens de prosodische annotatie van de observaties in de materiaalverzameling (zie paragraaf 10.1.1) is het opgevallen dat de optionele sjwadeletie nagenoeg niet voorkomt in deze componenten, met als gevolg een zeer verzorgde uitspraak waarbij elke lettergreep duidelijk gearticuleerd wordt (*het drein ★r ★g ★jong*).
14. Als we de structuur van het CGN, zoals voorgesteld in tabel 6(1), tot een 2x2-contingentietabel herleiden door de componenten 1 tot en met 9 (informeel: informaliteitsindex  $\geq 1$ ) en de componenten 10 tot en met 14 (formeel: informaliteitsindex = 0) samen te voegen, verkrijgen we de volgende 95% BI'en voor de partiële  $\chi^2$ -statistiek (tussen ronde haakjes staan de basisgegevens).



	<b>Belgisch Nederlands</b>	<b>Nederlands Nederlands</b>
<b>Informeel</b>	[0.8917,0.8945] (1,092,580)	[1.0888,1.0914] (1,580,627)
<b>Formeel</b>	[1.2741,1.2799] (602,794)	[0.7641,0.7684] (428,694)

Tabel 12(i): 95% BI en partiële  $\chi^2$ -statistiek voor de verdeling van de formele en informele CGN-componenten over het Belgische en het Nederlandse deelcorpus

In het Belgisch Nederlandse materiaal zijn de formele componenten significant oververtegenwoordigd en de informele contexten significant ondervertegenwoordigd. Het Nederlandse deelcorpus vertoont het tegengestelde beeld: de omvang van de formele contexten is significant kleiner dan verwacht, terwijl de omvang van de informele contexten significant groter is dan verwacht.

15. Het ConDiv-corpus is speciaal samengesteld voor het onderzoek naar de convergente en divergente tendenties tussen het Belgisch Nederlands en het Nederlands Nederlands. Deze materiaalverzameling van om en bij de 40M woorden bestrijkt verschillende registers, die variëren van het zeer formele en gecontroleerde taalgebruik in kwaliteitskranten tot het zeer spontane en informele taalgebruik van *Inter Relay Chat*-discussiefora (Grondelaers et al. 2000).
16. Zowel in het Belgisch Nederlandse materiaal als in het Nederlands Nederlandse materiaal zorgt de geografische classificatie op basis van de geboortestreek voor de beste differentiëring tussen beide buigingsvarianten. De regionale differentiëring op basis van de geboortestreek bestuderen, reduceert het contact met de streektaal tot de leeftijd van 4 jaar: van 4 tot en met 16 jaar spreekt het CGN namelijk van schoolregio (zie paragraaf 12.2.2.1).
17. Hoppenbrouwers & Hoppenbrouwers (2001: 58) sorteren deze streektalen op basis van de fonetische en fonologische kenmerken in twee groepen: het Fries vormt een aparte cluster en het Gronings en het Drenths behoren tot de Nedersaksische cluster. Wij achten het verdedigbaar om het materiaal voor deze drie streektalen samen te voegen: ten eerste, deze streektalen vertonen een gelijkaardig buigingsgedrag (de Rooij 1980b); ten tweede, indien we deze drie streektalen niet samenvoegen, is een significant effect *a priori* uitgesloten als gevolg van dataschaarste.
18. Tabel 12(ii) vat de OR's en de 95% BI'en samen voor de verschillende regio's in Nederland. Het externe perspectief vergelijkt het buigingsprofiel na een bezittelijk voornaamwoord voor de regio in kwestie met de andere Nederlandse regio's; het interne perspectief vergelijkt het buigingsprofiel binnen de regio in kwestie na een bezittelijk voornaamwoord en na een andere determinator. Enkel de sprekers die geboren zijn in de dialectgebieden Drenthe, Groningen en Friesland selecteren in beide perspectieven de onverbogen vorm significant vaker na een bezittelijk voornaamwoord. De zeer ruime 95% BI'en zijn een gevolg van het kleine aantal bezittelijke voornaamwoorden in de verschillende regio's. (Een OR gelijk aan nul betekent dat de regio in kwestie geen enkele attestatie bevat van een onverbogen vorm na een bezittelijk voornaamwoord.)

Voor de schoolregio en de woonregio is geen enkele van de OR's in tabel 12(ii) significant, ook niet voor Drenthe, Groningen en Friesland.

Regio in Nederland	Extern perspectief		Intern perspectief	
	OR	95% BI	OR	95% BI
Kerngebied	1.51	[0.76,3.01]	1.31	[0.73,2.35]
Periferie zonder Zeeuws-Vlaanderen	0.78	[0.31,1.97]	0.88	[0.35,2.21]
Zuidelijk randgebied	0.02	[0.01,0.07]	0.57	[0.25,1.33]
NO randgeb. zonder Drenthe, Groningen en Friesland	1.10	[0.36,3.36]	1.19	[0.36,3.87]
Zeeuws-Vlaanderen	0.00	[0,+∞[	0.00	[0,+∞[
Drenthe, Groningen en Friesland	4.31	[1.66,11.15]	5.29	[1.89,14.83]

Tabel 12(ii): Invloed van het bezittelijk voornaamwoord op de adjectivische buiging in de Nederlandse regio's

19. Het kerngebied en – in mindere mate het perifere gebied – zijn de gebieden waarheen de uitwijkingen uit het zuidwesten en het noordoosten in hoofdzaak lijken te trekken. Dit leidt in het kerngebied en het perifere gebied tot een niet-significante stijging van de onverbogen buigingsvariant in de buigingsprofielen voor de woonregio vergeleken met de geboorteregio (kerngebied:  $\chi^2 = 0.216$ ,  $df = 1$ ,  $p = 0.6421$ ; periferie:  $\chi^2 = 0.6846$ ,  $df = 1$ ,  $p = 0.408$ ).
20. Het relatief grote aandeel onverbogen vormen in het buigingsprofiel van de honkvaste sprekers in Oost-Vlaanderen ('bel.o.vl') is problematisch voor deze verklaring: aangezien de verschillende isoglossen voor de af- en aanwezigheid van de sjwa-apocope in *laat*, *vinger*, *kaars* en *zoon*, tussen Gent en Aalst lopen (Taeldeman 1978: 50-52, kaart 19) verwachten we een vergelijkbaar buigingsprofiel in West- en Oost-Vlaanderen, wat in het CGN echter niet het geval is.
21. We hebben deze test overgedaan voor het Belgisch Nederlands. In het Belgische deelcorpus zijn de lexicale collocaties homogeen over component 6 en de andere componenten verspreid ( $\chi^2 = 1.4978$ ,  $df = 1$ ,  $p = 0.221$ ). In het Belgisch Nederlandse deelcorpus bevat het buigingsprofiel van de meest informele CGN-component, c.q. component 1, het grootste aandeel onverbogen vormen. Deze component omvat uitsluitend spontane conversaties.
22. Deze randsignificantie wordt bevestigd door de niet-significante  $p$ -waarde van de  $\chi^2$ -statistiek:  $\chi^2 = 3.8115$ ,  $df = 1$ ,  $p = 0.0509$ .
23. Dit zijn de resultaten van de  $\chi^2$ -tests die de buigingsprofielen voor het Belgisch Nederlands en het Nederlands Nederlands vergelijken in de formele registers op basis van de vier stilistische variabelen:
- informaliteitsindex = 0:  $\chi^2 = 9.4748$ ,  $df = 1$ ,  $p = 0.002083$
  - interactietype = 'monoloog':  $\chi^2 = 10.8232$ ,  $df = 1$ ,  $p = 0.001002$
  - voorbereiding = 'voorbereid':  $\chi^2 = 28.4847$ ,  $df = 1$ ,  $p = 9.444e-08$
  - publiek = 'publiek':  $\chi^2 = 16.5099$ ,  $df = 1$ ,  $p = 4.84e-05$
- Merk op dat deze  $p$ -waarden kleiner zijn dan de  $p$ -waarden voor de andere berekeningen in dit onderzoek.

24. Hierbij moet de lezer voor ogen houden dat het Nederlands Nederlandse register met informaliteitsindex 0 minder formeel is dan het register met informaliteitsindex 2, dat gekenmerkt wordt door ambtelijk taalgebruik.
25. Deze interpretatie houdt geen rekening met de actieve, bewuste keuzes die een spreker in een gesprek maakt: h/zij ondergaat niet alleen de configuratie van de gesprekscontext, maar neemt actief deel aan het gesprek. Bijgevolg wendt de spreker de stilistische waarde van talige tekens actief aan om zich binnen de gesprekscontext ten overstaan van zijn/haar gesprekspartners te situeren (Bell 1984) en om een beeld van zijn sociale en cognitieve persoonlijkheid te verspreiden (Coupland 1980; Eckert 2000; Kristiansen 2003). De stilistische betekenis van elementen reflecteert dus niet alleen de formaliteit – en de daarmee samenhangende productiedruk – van de gesprekscontext, maar definieert op actieve wijze de persoonlijkheid van de spreker en de relatie tussen de gesprekspartners.
- In deze optiek zou de Vlaming de verbogen vorm als een strategie kunnen aanwenden om afstand te creëren ten overstaan van zijn gesprekspartners. Omgekeerd zou het gebruik van de onverbogen vorm niet alleen een uiting van de eigen identiteit, maar ook een solidariteitsmechanisme kunnen zijn: de keuze voor de onverbogen vorm is dan een (bewuste) strategie om een gemoedelijke en vertrouwelijke sfeer tussen de gesprekspartners te scheppen. Voorlopige evidentie ter ondersteuning van deze interpretatie is de vaststelling dat de spontane, dialogische privé-gesprekken de hoogste deflectiescores vertonen.
- De gevoerde empirische analyse is echter ontoereikend om het eventuele gebruik van het onverbogen adjectief als solidariteitsmechanisme te beschrijven: enerzijds vraagt de verificatie van deze veronderstelling een analyse van de dynamiek van gesprekken; anderzijds moeten de intenties van de sprekers door middel van experimenteel onderzoek en/of enquêtes worden vastgesteld (zie onder andere Geeraerts, Grondelaers & Speelman 1999: 133-145). Deze vraagstelling overschrijdt echter ruimschoots de doelstellingen van dit onderzoek.
26. Tabel 12(iii) toont de OR's die de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in het formele en het informele stratum vergelijken voor het gehele CGN en beide nationale deelcorpora:

	OR	95% BI
Belgisch Nederlands	3.15	[2.64,3.75]
Nederlands Nederlands	2.03	[1.49,2.14]
CGN	2.51	[2.19,2.88]

Tabel 12(iii): Invloed van de gebinariseerde informaliteitsindex op de adjectivische buigingsalternantie

Merk op dat het effect van de niet-formele gespreksituaties significant verschillend is in het Belgisch Nederlands en het Nederlands Nederlands. Dit sluit aan bij de resultaten van het stilistische onderzoek (zie tabel 12(11)). We kunnen de binaire oppositie dus als een adequate samenvatting van de stilistische distincties in het CGN beschouwen.

27. Wanneer we de observaties met een geanteponeerde genitief als determinator omwille van de dataschaarste buiten beschouwing laten, is de distributie evenmin significant ( $\chi^2 = 5.4752$ ,  $df = 2$ ,  $p = 0.06473$ ).

28. De OR's voor de marginale tabel zijn berekend op basis van de frequentieverdeling in tabel 8(1).
29. We wijzen erop dat de zin van de OR voor de marginale tabel tegengesteld is aan de zin van de OR's voor de stratumtabellen. Deze schijnbare paradox is het gevolg van de scheve verdeling van de observaties met een bezittelijk voornaamwoord als determinator over de verschillende registers: de analyse van de verdeling van de verschillende determinatoren over de stilistische strata leert ons dat de attestaties met een bezittelijk voornaamwoord in determinatorpositie significant oververtegenwoordigd zijn in de formele registers (95% BI partiële  $\chi^2 = [1.2058, 1.4533]$ ) en significant onderverteenwoordigd zijn in de informele registers (95% BI partiële  $\chi^2 = [0.507, 0.6952]$ ).
30. We hebben ook de mogelijke stilistische differentiëring van het prosodische patroon aan de rechterwoordrand van het adjectief en van het prosodische patroon op het niveau van de fonologische constituent getest (zie paragraaf 10.2.2). In geen van beide gevallen hebben we een significant verschillend effect tussen het formele en het informele stratum geïdentificeerd.
31. Wanneer we de strumanalyse voor beide nationale variëteiten afzonderlijk uitvoeren, zien we dat het effect in beide nationale variëteiten groter is in de informele contexten dan in de formele contexten, maar dat dit onderscheid enkel significant is in het Belgisch Nederlands (*Breslow-Day-test*:  $p = 0.0031$ ) en niet in het Nederlands Nederlands (*Breslow-Day-test*:  $p = 0.1610$ ).
32. Wanneer we beide nationale variëteiten afzonderlijk beschouwen, is het effect van de adjectivische eindmedeklinker telkens homogeen over beide stilistische strata verdeeld in het Belgisch Nederlands (*Breslow-Day-test*:  $p = 0.0837$ ) en in het Nederlands Nederlands (*Breslow-Day-test*:  $p = 0.3845$ ).
33. Volgens deze gedachtegang zouden de lexicale collocaties in de informele registers vooral vaste uitdrukkingen zijn die samenhangen met een beperkte(re) woordenschat.
34. We hebben deze test overgedaan voor de twee nationale variëteiten afzonderlijk. In de twee gevallen benadert de  $p$ -waarde van de *Breslow-Day-test* de waarde 1, wat op een homogene distributie van de OR's wijst (*Breslow-Day-test*<sub>BelgischNederlands</sub>:  $p = 0.99$ ; *Breslow-Day-test*<sub>NederlandsNederlands</sub>:  $p = 0.9514$ ). Het effect van de lexicale collocabiliteit is constant over de verschillende registers.
35. Als we deze analyse overdoen in de twee nationaal gedefinieerde deelcorpora verkrijgen we een analogo resultaat: zowel in het Belgisch Nederlands als in het Nederlands Nederlands is het deflecterende effect van de relationele adjectieven significant groter in de formele contexten.

## **DEEL III**

### **MODELLERING VAN DE BUIGINGSALTERNANTIE: INTERACTIE EN COMPETITIE**

In het vorige deel hebben we de condities op het gebruik van de adjectivische buigingsvarianten bij neutra in afzonderlijke bivariate analyses blootgelegd. Het zwaartepunt van deze analyses lag enerzijds bij de operationalisering en de kwantificering en anderzijds bij de interpretatie van de verschillende verklarende factoren. Aan de hand van stratumanalyses hebben we eventuele partiële conditioneringen tussen het buigingseffect van verklarende factoren bestudeerd.

De bivariate analyses zijn echter een idealisering en kunnen onmogelijk het eindpunt zijn van een corpuslinguïstische studie van de adjectivische buigingsalternantie. De bivariate analyses bestuderen immers het afzonderlijke effect van één verklarende factor zonder rekening te houden met het effect van de andere factoren, hoewel iedere observatie in het corpus gekenmerkt wordt door de combinatie van de verschillende verklarende factoren: aangezien iedere observatie één waarde heeft voor iedere verklarende factor, wordt de uiteindelijke keuze van het buigingsmorfem bepaald door een complex netwerk waar de verschillende verklarende factoren interageren en concurreren.

In dit derde deel presenteren we een multivariate analyse om de verklarende factoren in een globaal model te integreren dat rekening houdt met de onderlinge interacties en de competitie tussen deze factoren. Concreet betekent dit dat we een logistische-regressiemodel construeren. Deze multivariate analyse vormt samen met de bivariate analyses een methodologisch tweeluik: na de operationalisering, de kalibrering en de interpretatie van de afzonderlijke factoren in de bivariate analyses in het vorige deel, modelleren we nu het netwerk dat de selectie van het buigingsmorfem bij neutra bepaalt. Dit model meet het effect op de buigingsalternantie van iedere factor, terwijl het effect van de andere factoren gecontroleerd wordt. Daarnaast kunnen we de descriptieve adequaatheid van het voorgestelde model – en bij uitbreiding van het gevoerde onderzoek – evalueren.

Op basis van dit geïntegreerde, multivariate model kunnen we verschillende vragen beantwoorden die niet gesteld kunnen worden in bivariaat kwantitatief onderzoek, laat staan in introspectief onderzoek:

- Welke factoren hebben een significante impact op de buigingsalternantie? Dit wil zeggen: welke factoren zijn nodig om een optimaal model te construeren?
- Wat is de impact van de verschillende significante factoren in het geïntegreerde model?
- Hoeveel variatie kunnen we op basis van het geïntegreerde model verklaren? Met andere woorden: hoe aleatoir is de buigingsalternantie? En: hoe adequaat is het model dat we construeren?
- Wordt de buigingsalternantie in Nederland en Vlaanderen door verschillende onderliggende systemen bepaald? Met andere woorden: zijn er factoren met een significant verschillende impact in het Belgisch Nederlands en het Nederlands Nederlands?

Het multivariate onderzoek dat in dit derde deel gepresenteerd wordt, omvat twee hoofdstukken. Het eerste hoofdstuk (hoofdstuk 13) is technisch van aard en stelt de gebruikte multivariate techniek voor, met name de logistische regressie, waarbij we een bijzondere aandacht aan de interpretatie van het resulterende regressiemodel besteden. In het tweede hoofdstuk (hoofdstuk 14) rapporteren we het logistische-regressiemodel voor de buigingsalternantie in het gesproken standaard Nederlands. In eerste instantie kijken we naar het effect van de factoren in het model. Daarnaast confronteren we dit model met de bestaande attitudes ten aanzien van de buigingsalternantie (zie paragraaf 1.2.2). Tot slot belichten we de descriptieve en methodologische meerwaarde van multivariaat corpusonderzoek evenals mogelijke interpretatieve valstrikken bij bivariate analyses.

## Hoofdstuk 13:

# Multivariate statistiek: logistische regressie

In het eerste hoofdstuk van het derde deel van dit proefschrift stellen we de statistische techniek voor die we gebruiken bij de multivariate analyse van de adjectivische buigingsalternantie, met name de logistische regressie. Deze techniek biedt de mogelijkheid om de cumulatieve impact van verschillende verklarende variabelen op de responsvariabele te modelleren.

Regressietechnieken worden gebruikt om de impact van één (enkelvoudige regressie) of meerdere (meervoudige regressie) verklarende variabelen op een responsvariabele te modelleren. Het resultaat van een regressieanalyse is een statistisch model waarin de significantie en het effect van de verschillende verklarende variabelen wordt berekend. Daarnaast biedt de constructie van een statistisch model het voordeel dat de door het model verklaarde variatie kan worden gemeten. Deze score kan worden aangewend om de adequaatheid van het model – en bij uitbreiding van het onderliggende empirisch onderzoek – te evalueren.

We introduceren de regressietechniek aan de hand van de intuïtief eenvoudigere lineaire regressie (13.1). Daarna bespreken we de technische aspecten en de interpretatie van de logistische regressie (13.2).

### 13.1 Lineaire regressie

Een enkelvoudige lineaire regressie wordt gebruikt om de relatie tussen twee continue numerische variabelen te modelleren. De regressieanalyse berekent de best passende rechte door de meetpunten in een tweedimensionale ruimte (Y-as voor de responsvariabele en X-as voor de verklarende variabele). De best passende rechte is de rechte die de som van de afstanden tot de meetpunten minimaliseert.<sup>1</sup> De positie van de regressierechte kan als een wiskundige vergelijking worden uitgedrukt:  $y = a + bx + fout$ , waarbij  $a$  het intercept is en  $b$  de regressiecoëfficiënt of de schatter van de verklarende variabele  $x$ . Het intercept identificeert het punt waar de regressierechte de Y-as snijdt; de regressiecoëfficiënt identificeert de hellingsgraad of gradiënt van de regressierechte (bij  $b = 0$  heeft  $x$  geen invloed op de waarde van  $y$ ); *fout* drukt de onverklaarde variatie uit (het zogenaamde residu). Als we dit model uitbreiden met een tweede verklarende variabele, verkrijgen we een meervoudige lineaire regressie:  $y = a + b_1x_1 + b_2x_2 + fout$ . Deze vergelijking identificeert niet langer de best passende rechte in een tweedimensionale ruimte maar het best passende vlak (gedefinieerd door  $a$ ,  $b_1$  en  $b_2$ ) in een driedimensionale ruimte.

Technisch gesproken kunnen  $n$  variabelen in een regressiemodel worden opgenomen, alleen kan de resulterende regressievergelijking bij meer dan drie dimensies niet langer gevisualiseerd worden.

We verduidelijken bovenstaande uiteenzetting over de lineaire regressie aan de hand van een concreet voorbeeld uit de economische wetenschappen (de metingen die de basis van dit voorbeeld vormen, zijn overgenomen uit Chandan, White & Whyte 1998: 456). Op basis van metingen in 64 landen zijn we geïnteresseerd in de invloed van het gemiddelde bruto nationaal inkomen per hoofd (BNI, uitgedrukt in dollar) op de kindersterfte (uitgedrukt in het aantal kinderen op duizend die vóór hun vijfde jaar sterven). Dit is de resulterende regressievergelijking:

$$(1) \quad \text{kindersterfte} = 157.3617 - 0.011362 * \text{BNI}$$

We zijn hier in eerste instantie geïnteresseerd in de regressiecoëfficiënt van het BNI. Dit cijfer drukt het effect van het BNI op de kindersterfte uit:  $-0.011362$ . In een lineaire-regressiemodel kunnen de regressiecoëfficiënten rechtstreeks geïnterpreteerd worden. De negatieve oriëntatie wijst op een negatieve correlatie tussen de kindersterfte en het BNI. Dit bevestigt de intuïtie dat de kindersterfte afneemt naarmate het gemiddelde BNI stijgt. De waarde van deze regressiecoëfficiënt,  $0.011362$ , betekent dat voor iedere dollar dat het BNI stijgt de kindersterfte met  $0.011$  per duizend kinderen daalt. Deze associatie is significant ( $p < 2.2e-16$ ).<sup>2</sup> Aangezien de vergelijking in (1) een statistisch model voorstelt, kan ook de significantie en de verklarende waarde ( $r^2$ ) van het regressiemodel berekend worden. De  $p$ -waarde van het lineaire-regressiemodel is gelijk aan  $0.000824$ , wat op een zeer significant model wijst.<sup>3</sup> De  $r^2$ -score is gelijk aan  $0.1528$ . Dit betekent dat  $15.28\%$  van de variatie in de materiaalverzameling verklaard wordt door de regressierechte in (1).<sup>4</sup> Dus: ondanks de hoge significantie van de regressiecoëfficiënt en van het regressiemodel is de verklarende kracht relatief beperkt.

Rekening houdend met de geringe verklarende kracht van het enkelvoudige regressiemodel voegen we een tweede verklarende variabele aan het model toe, met name de gemiddelde alfabetiseringsgraad van de vrouwen (AGV, uitgedrukt als het percentage van het totale aantal vrouwen). Dit levert de volgende meervoudige regressievergelijking op:

$$(2) \quad \text{kindersterfte} = 263.6571 - 0.00568 * \text{BNI} - 2.231508 * \text{AGV}$$

We bespreken eerst het effect van de verklarende variabelen. Daarna evalueren we het lineaire-regressiemodel.

Zowel het bruto nationaal inkomen (BNI) als de vrouwelijke alfabetiseringsgraad (AGV) hebben een significante impact op de kindersterfte:  $p_{\text{BNI}} = 0.0061$ ;  $p_{\text{AGV}} = 1.55e-15$ . Wanneer we de regressiecoëfficiënten beschouwen, vallen drie punten op. Ten eerste, beide verklarende variabelen hebben – zoals kan worden aangenomen – een negatieve regressiecoëfficiënt. Ten tweede, het effect van de vrouwelijke alfabetiseringsgraad is significanter dan het effect van het BNI: de kindersterfte



neemt met 2.23 per duizend kinderen af voor ieder extra percent aan geletterde vrouwen, terwijl de kindersterfte slechts met 0.0056 per duizend kinderen afneemt voor iedere dollar dat het BNI toeneemt. Ten derde, wanneer we de regressiecoëfficiënt van het BNI in de regressievergelijkingen in (1) en (2) naast elkaar plaatsen, stellen we vast dat de coëfficiënt in (2) beïnvloed wordt door de vrouwelijke alfabetiseringsgraad. Het feit dat er twee verklarende variabelen zijn, impliceert dat deze elkaar wederzijds beïnvloeden: in een meervoudige lineaire regressie wordt namelijk het cumulatieve effect van de verschillende verklarende variabelen berekend. Deze wederzijdse beïnvloeding tussen de verschillende verklarende variabelen in het meervoudige regressiemodel onderstreept eens te meer de noodzaak van multivariaat onderzoek (zie ook paragraaf 5.2.3).

Er rest ons nog één vraag: welke impact heeft het toevoegen van de tweede verklarende variabele op het regressiemodel? De  $p$ -waarde ( $p < 2.2e-16$ ) toont aan dat de regressievergelijking in (2) op significante wijze de kindersterftegraad verklaart. Vergeleken met het enkelvoudige model in (1) neemt de (gecorrigeerde)  $r^2$ -score aanzienlijk toe:  $r^2 = 0.6986$ , wat betekent dat 69.86% van de kindersterfte in de bestudeerde materiaalverzameling door het lineaire-regressiemodel in (2) verklaard wordt.<sup>5</sup>

## 13.2 Logistische regressie

Aangezien het onderzoek naar de buigingsvariatie niet aan de statistische condities op het gebruik van een lineaire regressie voldoet – we maken bijvoorbeeld enkel gebruik van nominale variabelen, die per definitie niet normaal verdeeld zijn –, gebruiken we de binomiale logistische regressie. Dit is de regressietechniek die gebruikt wordt bij een binomiale responsvariabele en één of meerdere categorische en/of numerische verklarende variabelen (Agresti 1996: 103-144; Paolillo 2001: 153-173; Rietveld & van Hout 1993: 327-360).

We lichten achtereenvolgens de constructie van het logistische-regressiemodel (13.2.1) en de interpretatie van de resultaten van de logistische-regressieanalyse toe (13.2.2).

### 13.2.1 Constructie logistische-regressiemodel

De logistische regressie is een lineaire regressie op de logit van de odds in plaats van op de oorspronkelijke waarde van de responsvariabele, c.q. een kans. Gezien onze interesse in de impact van de verklarende variabelen op de onverbogen buigingsvariant definiëren wij de responsvariabele van de logistische regressie als volgt:

$$(3) \quad \text{logit}(P(\text{onverbogen.adjectief})) = \ln\left(\frac{P(\text{onverbogen.adjectief})}{1 - P(\text{onverbogen.adjectief})}\right)$$

De motivatie voor de logittransformatie is dubbel. Ten eerste, in een additief model, waar de waarde van de  $y$ -variabele wordt bepaald door het sommeren van het effect van de verklarende variabelen, zoals in het lineaire-regressiemodel in (2), kan de waarde gegenereerd door de regressievergelijking buiten het interval  $[0,1]$  vallen, hetgeen per definitie uitgesloten is bij kansen. Het tweede argument voor de logittransformatie is technisch van aard: een lineaire regressie zou onstabiele regressiecoëfficiënten genereren doordat het residu niet constant is.

De regressievergelijking ziet er na de logittransformatie als volgt uit:

$$(4) \quad \ln\left(\frac{P(\text{onverbogen.adjectief})}{1-P(\text{onverbogen.adjectief})}\right) = a + b_1 * \ln\left(\frac{P(x_1)}{1-P(x_1)}\right) + \dots + b_n * \ln\left(\frac{P(x_n)}{1-P(x_n)}\right)$$

Laten we nu de verklarende variabelen bekijken. De verklarende variabelen in ons onderzoek zijn nominale variabelen. Dit zijn variabelen die voor geen enkele rekenkundige operatie in aanmerking komen (in tegenstelling tot de numerische variabelen in (1) en (2)). Het gebruik van dit type variabelen vereist een aparte codering en rapportering. In dit onderzoek zijn we geïnteresseerd in het gedrag van de responsvariabele in de verschillende groepen gedefinieerd door een verklarende variabele. Hiervoor worden de verklarende variabelen in dummy-variabelen omgezet. We verduidelijken dit eerst aan de hand van een binomiale variabele: één van de waarden van de verklarende variabele is de ijkwaarde, die als vergelijkingspunt fungeert bij de beoordeling van de regressiecoëfficiënt van de andere waarde van deze variabele. Het ijkpunt krijgt de dummy-waarde 0 en de gemarkeerde waarde de dummy-waarde 1. Deze techniek hebben we tot nu toe impliciet toegepast in het bivariaat onderzoek: het RR en de OR berekenen in welke mate het aandeel van het onverbogen adjectief of van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  stijgen/dalen bij de gemarkeerde waarde van een verklarende variabele ten overstaan van de referentiewaarde. Bij een multinomiale variabele is het gebruik van de dummy-variabelen complexer. Stel een multinomiale variabele  $X$ , met drie waarden:  $X_1$ ,  $X_2$ , en  $X_3$ . Als we  $X_1$  als ijkwaarde nemen, hebben we twee dummy-variabelen nodig ( $3 - 1 = 2$ ),  $D_1$  en  $D_2$ , om deze multinomiale variabele in binomiale variabelen om te zetten (Rietveld & van Hout 1993: 101-105). In de regressievergelijking zijn het de dummy-variabelen en niet de oorspronkelijke variabelen die de  $x$ -waarden invullen.

Waarde variabele	Dummy-variabelen	
	$D_1$	$D_2$
$X_1$ (ijkpunt)	0	0
$X_2$	1	0
$X_3$	0	1

Tabel 13(1): Codering van nominale variabelen door middel van dummy-variabelen

De waarde die als referentiepunt fungeert,  $X_1$ , heeft voor alle dummy-variabelen de waarde 0. In de uitvoer van de logistische-regressieanalyse worden de regressie-

coëfficiënten van de waarden  $X_2$  en  $X_3$  geïnterpreteerd ten overstaan van  $X_1$ . Door het gebruik van dummy-variabelen worden multinomiale nominale variabelen in de regressievergelijking in verschillende binomiale variabelen opgesplitst.

We illustreren de dummy-codering aan de hand van twee voorbeelden, een eerste voor een binomiale variabele (relationaliteit van het adjectief; zie paragraaf 11.1.2) en een tweede voor een multinomiale variabele (prosodisch patroon adjectivische rechterwoordrand; zie paragraaf 10.2.2). Bij een binomiale variabele hebben we één dummy-variabele nodig ( $2 - 1 = 1$ ): één van de twee waarden van de verklarende variabele fungeert als referentiepunt (i.e. ‘ $\neg$ adj.rel’). Deze waarde krijgt 0 als dummy-waarde toegewezen. De andere waarde (i.e. ‘adj.rel’) heeft 1 als dummy-waarde.

Waarde variabele	Dummy-variabele
$\neg$ adj.rel (ijkpunt)	0
adj.rel	1

Tabel 13(2): Dummy-codering voor de relationaliteit van het adjectief (binomiale variabele)

Zoals we hierboven hebben aangegeven, hebben we deze dummy-codering impliciet toegepast bij de bivariate analyse van het buigingseffect van de relationaliteit van het adjectief (en andere binomiale verklarende variabelen): we hebben het buigingsprofiel van de relationele adjectieven (dummy = 1 in tabel 13(2)) vergeleken met dat van de niet-relationele of kwalificerende adjectieven (dummy = 0 in tabel 13(2)).

Voor de dummy-codering van de multinomiale variabele voor het prosodische patroon aan de rechterwoordrand van het adjectief, met de waarden ‘s’, ‘sw’ en ‘sww’, hebben we twee dummy-variabelen nodig,  $D_1$  en  $D_2$  ( $3 - 1 = 2$ ). Tabel 13(3) schematiseert de dummy-codering voor de waarden van het prosodische patroon aan de rechterwoordrand van het adjectief:

Waarde variabele	Dummy-variabelen	
	$D_1$	$D_2$
‘s’ (ijkpunt)	0	0
‘sw’	1	0
‘sww’	0	1

Tabel 13(3): Dummy-codering voor het prosodische patroon aan de rechterwoordrand van het adjectief (multinomiale variabele)

Ook deze dummy-codering hebben we al impliciet toegepast in de bivariate analyse: bij de prosodische analyse hebben we het buigingseffect van de patronen ‘sw’ (dummy-codering = (1,0) in tabel 13(3)) en ‘sww’ (dummy-codering = (0,1)) gekwantificeerd ten overstaan van het referentiepunt ‘s’ (dummy-codering = (0,0)).

Tot slot bespreken we de constructie van het logistische-regressiemodel. De drie belangrijkste manieren om de regressievergelijking te construeren, zijn de volgende:

- manuele invoer: De onderzoeker bepaalt zelf welke variabelen in het regressiemodel worden opgenomen. Dit is een belangrijk verschilpunt met de *stepwise*-methoden, waar het regressiemodel zelf berekent welke variabelen worden opgenomen.
- *forward stepwise*: Het systeem vertrekt van een leeg model, i.e. een model dat geen enkele verklarende variabele bevat, en voegt één voor één de verklarende variabelen toe in functie van hun verklarende waarde, te beginnen met de variabele die de meeste variatie verklaart.<sup>6</sup> Als alle variabelen volgens dit criterium zijn opgenomen of als geen van de resterende variabelen de variatie op significante wijze reduceert, stopt het algoritme.
- *backward stepwise*: Het systeem vertrekt van een model dat alle potentiële verklarende variabelen bevat en test vervolgens welke variabelen verwijderd kunnen worden zonder dat de verklarende kracht van het model op significante wijze afneemt.

De *stepwise*-methoden worden meestal in de exploratieve fase van het onderzoek gebruikt om na te gaan welke factoren de grootste reductie van de variatie bewerkstelligen. Voor de constructie van het uiteindelijke regressiemodel wordt meestal de voorkeur gegeven aan de manuele invoer van gegevens. Ten eerste, een *stepwise*-model bepaalt op zuiver mathematische gronden welke factoren in het model worden opgenomen en welke niet. Ten tweede, indien men de empirische validiteit van een theoretisch model wil testen, is het nodig dat de onderzoeker de relevante factoren zelf in het model invoert.<sup>7</sup>

We combineren de *forward stepwise* en de manuele constructiemethoden op de volgende wijze: eerst voeren we een *forward stepwise*-regressie uit om te meten hoeveel variatie de verschillende factoren verklaren (zie paragraaf 14.1.2); daarna construeren we een manueel model dat alle verklarende factoren omvat om de (significantie van de) impact van de verschillende factorwaarden te kwantificeren (zie paragraaf 14.1.3). Op deze wijze construeren we het regressiemodel dat de meeste buigingsvariatie verklaart. Dit sluit aan bij de basisdoelstelling van dit onderzoek: het construeren van een optimaal model om de conditionering van de adjectivische buigingsalternantie bij neutra in kaart te brengen. Naast dit globale regressiemodel zullen we ook verschillende alternatieve modellen construeren met verklarende variabelen die geselecteerd worden in functie van een specifieke visie op de buigingsalternantie (zie paragraaf 1.2.2) of in functie van een algemene congruentietheorie, zoals de discursieve visie op congruentie voorgesteld door Barlow (1991, 1999) (zie paragraaf 4.2.2). Door deze alternatieve modellen te vergelijken met het globale, theorieonafhankelijke basismodel kunnen we de empirische validiteit meten van de verschillende visies op de buigingsalternantie en op congruentie in het algemeen.<sup>8</sup>

### 13.2.2 Interpretatie logistische-regressiemodel

Net als een lineaire-regressiemodel verschaft een logistische-regressiemodel informatie over het model als geheel en over de impact van de afzonderlijke variabelen. Door het gebruik van nominale variabelen en de logittransformatie zijn de rapportering en de interpretatie lichtjes afwijkend van de lineaire regressie.

Voor de evaluatie van het gehele regressiemodel zijn er twee maten: de verklarende kracht van het model en de voorspellende kracht van het model. De verklarende kracht van het model berekent de hoeveelheid variatie die het regressiemodel verklaart door de niet-verklaarde variatie (residu) af te trekken van de totale variatie in de materiaalverzameling. Voor dat verschil wordt vervolgens de significantie berekend. De variatie wordt net als bij de parameterselectie in het *stepwise*-algoritme berekend aan de hand van de *maximum likelihood estimation*, meer bepaald aan de hand van de  $-2 \log \text{likelihood}$  (zie ook paragraaf 11.1.1.2). Dit komt neer op het testen van de  $H_0$  dat het logistische-regressiemodel geen effect heeft op de responsvariabele:  $H_0$  stelt dan dat  $b_1 = b_2 = \dots = b_n = 0$ .<sup>9</sup> De tweede maat die gebruikt wordt bij de evaluatie van het regressiemodel is de voorspellende kracht. Hiervoor gebruiken we de *c*-maat: deze maat drukt het percentage observaties in de materiaalverzameling uit waarvoor het model de correcte buigingsvariant voorspelt.<sup>10</sup> In het regressiemodel wordt voor iedere observatie de logit berekend op basis van zijn parameterwaarden: bij een positieve logit voorspelt het model succes (c.q. een onverbogen adjectief), bij een negatieve waarde mislukking (c.q. een verbogen adjectief). Dit resultaat wordt vervolgens geëvalueerd tegen de reële buigingsvariant.

Bij de evaluatie van het effect van de verklarende variabelen moet er rekening mee worden gehouden dat de regressiecoëfficiënten in een logitmodel berekend zijn. Daarom worden ze getransformeerd tot een OR, meer bepaald tot de OR die het effect op de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  kwantificeert van de parameterwaarde in kwestie vergeleken met de ijkwaarde. De OR wordt als volgt berekend:

$$(5) \quad \text{OR} = e^b; \text{ waarbij } b \text{ de regressiecoëfficiënt is}$$

Op basis van deze formule krijgt een negatieve regressiecoëfficiënt (c.q. een daling van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ ) een OR tussen 0 en 1 en een positieve regressiecoëfficiënt (c.q. een stijging van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ ) een OR groter dan 1 (voor de interpretatie van de OR verwijzen we naar paragraaf 7.1.1). Bij de interpretatie van dit effect moet men rekening houden met de significantie, uitgedrukt als de *p*-waarde of het 95% BI voor de OR.<sup>11</sup>

Tot slot vestigen we de aandacht op de mogelijkheid dat een variabele in z'n geheel een significante reductie van de variatie veroorzaakt – en dus in het *stepwise*-regressiemodel wordt opgenomen – maar dat sommige parameterwaarden niet significant van de ijkwaarde verschillen. De impact van een verklarende variabele in z'n geheel leiden we af uit de *forward stepwise*-analyse die de variabelen selecteert in functie van hun bijdrage tot de reductie van de totale variatie; de significantie van

de verschillende parameterwaarden wordt daarentegen berekend op basis van het effect op de odds <sup>onverboden</sup>/<sub>verboden</sub>, of de logit van deze odds om precies te zijn, van de parameterwaarde in kwestie vergeleken met de ijkwaarde van deze parameter.

We gebruiken het *Design*-pakket in R voor de constructie van de logistische-regressiemodellen (Harrell 2001). Dit pakket omvat daarenboven ook diagnostische maten voor de logistische regressie, zoals de *Variance Inflation Factor* (zie paragraaf 14.3.4 noot 5) en de *Influential Observations* (zie paragraaf 14.3.4 noot 4). Bij de bespreking van de logistische-regressiemodellen in het volgende hoofdstuk concentreren we ons op de modelstatistieken en de regressiecoëfficiënten. De resultaten van de diagnostische maten, die niet essentieel zijn voor een goed begrip van de logistische-regressiemodellen, worden in voetnoten behandeld.

## Noten

1. De positie van de regressierechte wordt berekend aan de hand van de methode van de *ordinary least squares*.
2. In een lineaire regressie wordt de  $p$ -waarde van de regressiecoëfficiënten door middel van de  $t$ -test berekend.
3. De significantie van de reductie van de variatie gerealiseerd door een lineaire-regressiemodel wordt berekend aan de hand van de  $F$ -statistiek. Dit is het resultaat voor de regressierechte in (1):  $F$ -statistiek = 12.37 voor 1 en 62 df,  $p = 0.000824$ .
4. De  $r^2$ -score is, zoals de naam aangeeft, het kwadraat van de lineaire correlatie-index  $r$ . De absolute waarde van  $r$  en  $r^2$  varieert tussen 0 en 1: 0 wijst op een totale afwezigheid van associatie/verklaring van de variatie, 1 wijst op een volledige associatie/verklaring van de variatie. Vanaf een  $r^2$ -score gelijk aan 0.60 à 0.65 spreekt men doorgaans van een degelijk model. Tot slot signaleren we dat we de gecorrigeerde  $r^2$ -waarde gerapporteerd hebben. De  $r^2$ -waarde die niet gecorrigeerd is voor het aantal vrijheidsgraden is gelijk aan 0.1663.
5. Dit zijn de modelstatistieken voor de regressievergelijking in (2):
  - residu standaardfout: 41.71 voor 61 df
  - $r^2 = 0.7082$ ; gecorrigeerde  $r^2 = 0.6986$
  - $F$ -statistiek = 74.03 voor 2 en 61 df,  $p < 2.2e-16$ .
6. De regressiecoëfficiënten worden berekend aan de hand van de *maximum likelihood estimation*. De *likelihood*-functie berekent de waarschijnlijkheid dat het regressiemodel de distributie van de materiaalverzameling vertoont. *Maximum likelihood estimation* past iteratief de regressiecoëfficiënten in het model aan en selecteert de waarden die het model met de grootste waarschijnlijkheid genereren.
 

We corrigeren de verklaarde variatie aan de hand van het AIC (*Akaike Information Criterion*): dit is een correctie op de *log likelihood ratio* met penalisatie voor het aantal vrijheidsgraden. Het AIC wordt vaak gebruikt als een *goodness of fit*-maat voor de vergelijking van statistische modellen.
7. Voor een voorbeeld verwijzen we naar Gries (2003). Gries selecteert de verklarende variabelen die hij in het multivariaat model opneemt op basis van de globale psycholinguïstische hypothese die hij wil testen. Een consequentie van Gries' onderzoeksdoelstelling is dat verschillende verklarende factoren die in de bivariate analyses een significant effect hebben *a priori* van het multivariaat model worden uitgesloten, omdat ze niet in het gefalsificeerde psycholinguïstische model passen. Dit is een belangrijk verschilpunt met onze onderzoeksdoelstelling, met name het construeren van een optimaal model voor de conditionering van de buigingsalternantie. Voor een uitgebreide discussie verwijzen we naar Heylen, Tummers & Geeraerts (te verschijnen).
8. Voor de statistische vergelijking van (twee) logistische-regressiemodellen gebruiken we de ANOVA-techniek (*ANalysis Of VAriance*). Dit is een variant op de lineaire regressie

waar de verklarende variabelen categorische in plaats van numerische variabelen zijn (Woods, Fletcher & Hughes 1986; zie ook paragraaf 17.2).

Wij gebruiken de ANOVA om de verklarende kracht van logistische-regressiemodellen te vergelijken: de responsvariabele is de (door de regressiemodellen) verklaarde variatie, de verklarende variabele is het onderscheid tussen twee regressiemodellen, c.q. het globale theorieonafhankelijke model aan de ene kant en de alternatieve modellen die een bepaalde buigingstheorie of congruentietheorie implementeren aan de andere kant. Op deze wijze meten we niet alleen het verschil in verklaarde variatie, maar evalueren we eveneens of dit verschil significant is.

R voorziet een specifieke functie voor het vergelijken en evalueren van de verklarende kracht van twee logistische-regressiemodellen, met name `anova.glm` (*glm* staat voor *generalized linear model*, de familie van statistische modellen waartoe de logistische regressie behoort).

9. Voor de score van de verklarende kracht van een logistische-regressiemodel is er geen intuïtieve interpretatie, zoals dat het geval is voor de  $r^2$ -score bij een lineaire regressie.
10. De  $c$ -maat berekent in feite de oppervlakte onder de ROC-curve (*Receiver Operating Characteristic*). Deze curve zet de valse positieven uit tegen de valse negatieven. Positief geformuleerd: de ROC-curve zet de gevoeligheid (correct voorspelde successen) van een model uit tegen 1 - de specificiteit (correct voorspelde mislukkingen). Het resultaat van de  $c$ -maat bevindt zich altijd tussen 0.5 en 1.0. Bij 0.5 trekt de ROC-curve een diagonaal, wat op een snelle toename van de valse positieven én de valse negatieven wijst. Een score die 1.0 benadert, duidt op een geringe toename van de valse positieven en de valse negatieven, en *vice versa* op een grote gevoeligheid en een grote specificiteit (c.q. een kleine 1 - specificiteit). In tegenstelling tot de score voor de verklarende kracht is de score voor de voorspellende kracht van een logistische-regressiemodel wel een intuïtief interpreteerbaar cijfer.

Bij de interpretatie van de ROC-curve, c.q. van de  $c$ -score, volgen we deze algemene vuistregel:

- $c \in [0.50, 0.75[$ : matig model
- $c \in [0.75, 0.92[$ : goed model
- $c \in [0.92, 0.97[$ : zeer goed model
- $c \in [0.97, 1.00]$ : uitstekend model

11. Dit komt neer op het testen van de  $H_0$  dat  $b_i = 0$  (waarbij  $b_i$  de regressiecoëfficiënt van de dummy-variabele is). De significantie van het effect van een parameterwaarde wordt berekend aan de hand van de *Wald*-statistiek. Computationeel is de *Wald*-statistiek een  $z$ -test op basis van het kwadraat van het quotiënt van de logistische-regressiecoëfficiënt van een parameterwaarde en zijn standaardfout (Agresti 1996: 88).



# Hoofdstuk 14:

## Multivariate analyse van de adjectivische buigingsalternantie bij neutra

De bespreking van de resultaten van de logistische-regressieanalyse bestaat uit vier delen. Eerst lichten we de parametrisering van de verklarende factoren toe (14.1). Daarna stellen we het resultaat van de *forward stepwise*-regressie voor (14.2). De volgende stap bestaat uit de presentatie en de bespreking van de resultaten van de feitelijke, manuele regressieanalyses (14.3). We sluiten dit hoofdstuk af met een algemene bespreking van de modellering van de adjectivische buigingsalternantie bij neutra in het standaard gesproken Nederlands, waarvan het CGN een staalname is (14.4).

### 14.1 Parametrisering van de verklarende factoren

Tabel 14(1) geeft een overzicht van de factoren die we in deel II bestudeerd hebben. De eerste kolom somt de factoren op met tussen haakjes een korte beschrijving. De tweede kolom bevat de referentiewaarde. Dit is de waarde die als referentiepunt fungeert voor het bepalen van de invloed van de (andere) waarden van een factor (i.e. bij de interpretatie van de regressiecoëfficiënten en de OR's; zie ook paragrafen 7.1 en 13.2.2). Als referentiepunt ('referentie') gebruiken we dezelfde waarden als bij de bivariate analyses. De laatste kolom somt de andere waarden van de parameter in kwestie op.

Enkele parameters in tabel 14(1) vragen om een woordje toelichting. De sonoriteit van de adjectivische eindmedeklinker ('a.sonor') hebben we binair geoperationaliseerd als het onderscheid tussen de aan- en afwezigheid van sonoriteit. In de bivariate analyse is dit namelijk het belangrijkste scheidingspunt gebleken (zie paragraaf 9.2.2). Voor de berekening van het effect van de lexicale collocabiliteit ('lex.col') gebruiken we de afwezigheid van een significante lexicale associatie tussen het adjectief en het substantief als referentiepunt ('lex.col= $\neg$ lex.col'). De andere waarden van deze parameter identificeren de vier onderscheiden kwartielen voor de collocatiesterkte tussen het adjectief en het substantief. Het prosodische patroon aan de linkerwoordrand van het substantief ('n.pros') hebben we als een binomiale factor ('n.pros=s' vs. 'n.pros=w') geoperationaliseerd omdat het patroon 'wws' te infrequent is. Het patroon 'w' is bijgevolg een generalisatie over 'ws' en 'wws'.

Factor	Referentie	Waarden
det.cat (woordsoort determinator)	<i>het</i>	bez.vnw, aanw.vnw, genitief
a.comp (comparatief in positie A)	neen	ja
n.inf (genominaliseerde infinitief in positie N)	neen	ja
n.dim (diminuering N)	neen	ja
n.genus (genus/bigenericiteit N)	<i>het</i> -woord	bigenerisch
a.sonor (sonoriteit eindmedeklinker A)	neen	ja
n.beginklank (beginklank N)	medeklinker	klinker
a.len.pros (lengte en prosodisch patroon A)	een.s	twee.s, twee.sw, meer.s, meer.sw, meer.sww
n.pros (prosodisch patroon linkerrand N)	s	w
reg.inf.index (regio & informaliteitsindex)	nl.0	nl.1, nl.2, nl.3, bel.0, bel.1, bel.2, bel.3
lex.col (lexicale collocabiliteit AN)	-lex.col	kw.1, kw.2, kw.3, kw.4
a.rel (relationaliteit A)	neen	ja
a.geleed (morfologische geleedheid A)	neen	ja

*Tabel 14(1): Parametrisering van de verklarende factoren voor de logistische-regressieanalyse*

Tabel 14(1) bevat twee complexe factoren, met name ‘a.len.pros’ en ‘reg.inf.index’. Deze complexe factoren combineren twee factoren uit het bivariaat onderzoek. We kijken eerst naar ‘a.len.pros’. Deze factor combineert de lengtegraad van het adjectief (‘een’:  $n_{\sigma} = 1$ , ‘twee’:  $n_{\sigma} = 2$ , ‘meer’:  $n_{\sigma} \geq 3$ ) en het prosodische patroon aan de rechterwoordrand van het adjectief (‘s’, ‘sw’, ‘sww’). De code ‘twee.sw’ betekent dat het adjectief in kwestie tweelettergrepig is en op het patroon ‘sw’ eindigt. In de bivariate analyse hebben we aangetoond dat deze factoren een sterke interactie vertonen: eenlettergrepige adjectieven realiseren enkel het patroon ‘s’, terwijl we het patroon ‘sww’ uitsluitend bij meerlettergrepige adjectieven aantreffen (zie paragraaf 10.2.3).<sup>1</sup> We combineren de adjectivische lengte met het prosodische patroon van het adjectief en niet met dat van de fonologische constituent, omdat we dan een explosie van factorwaarden zouden hebben. Bovendien realiseren verschillende waarden van het prosodische patroon van de fonologische constituent een perfecte scheiding (‘sw|wsw’ en ‘sww|wsw’), met onstabiele regressiecoëfficiënten als gevolg.<sup>2</sup> Om het prosodische patroon aan de linkerwoordrand van het substantief niet te verliezen, coderen we dit als een aparte parameter (‘n.pros’).

De tweede complexe factor balt de twee lectische factoren samen tot één variabele ‘reg.inf.index’ (regio/nationale variëteit én informaliteitsindex). Deze keuze is ingegeven door de tegengestelde stilistische waarde van de onverbogen buigingsvariant in beide nationale variëteiten (zie paragraaf 12.3.2). De gespreks-situaties met een informaliteitsindex gelijk aan 0 in het Nederlands Nederlands

vormen het referentiepunt. Dit register is gecodeerd als ‘nl.0’, wat staat voor het Nederlands Nederlandse (‘nl’) register met informaliteitsindex 0 (‘0’). Wat de geografische verschillen aangaat, kijken we enkel naar de nationale variëteiten van het Nederlands. De streektaalige invloed nemen we om twee redenen niet in de regressieanalyse op: primo, de strategie om enkel de uitingen van honkvaste sprekers te beschouwen leidt – vooral in de Nederlands Nederlandse deelverzameling – tot een aanzienlijke reductie van de beschikbare gegevens (zie paragraaf 12.2.1); secundo, voor 14.48% van de observaties in de Nederlandse deelverzameling en voor 6.32% van de observaties in de Belgische deelverzameling ontbreekt een streektaalige code. Wat het stilistische deel van deze complexe factor betreft, hebben we de meest globale stilistische factor geselecteerd, met name de informaliteitsindex die de drie situationele parameters uit het CGN combineert (zie paragraaf 12.2.2.2, tabel 12(4)).

De laatste factor in tabel 14(1) die om nadere toelichting vraagt, is ‘a.geleed’. Dit is een binomiale parameter voor de morfologische geleedheid van het adjectief met als waarden geleed (‘a.geleed=ja’) en niet-geleed (‘a.geleed=neen’). Hoewel we het buigingseffect van de morfologische structuur van het adjectief bestudeerd hebben, leek het ons verdedigbaar om hieraan geen afzonderlijk hoofdstuk te besteden, omdat de tendenties die we vaststellen bij de morfologische structuur van het adjectief grotendeels herleidbaar zijn tot het prosodische patroon aan de rechterwoordrand van het adjectief (bv. beklemtoonde tegenover onbeklemtoonde suffixen, samenstellingsklemtoon) en/of de relationaliteit van het adjectief (bv. een onverbogen adjectief met een – secundair – beklemtoond suffix is meestal een relationeel adjectief).<sup>3</sup> We nemen deze factor in de regressieanalyse op om de correctheid na te gaan van onze stelling dat het effect van deze factor epifenomenaal is ten overstaan van de prosodische (c.q. de adjectivische lengte en het prosodische patroon aan de rechterwoordrand van het adjectief) en semantische (c.q. de relationaliteit van het adjectief) factoren.

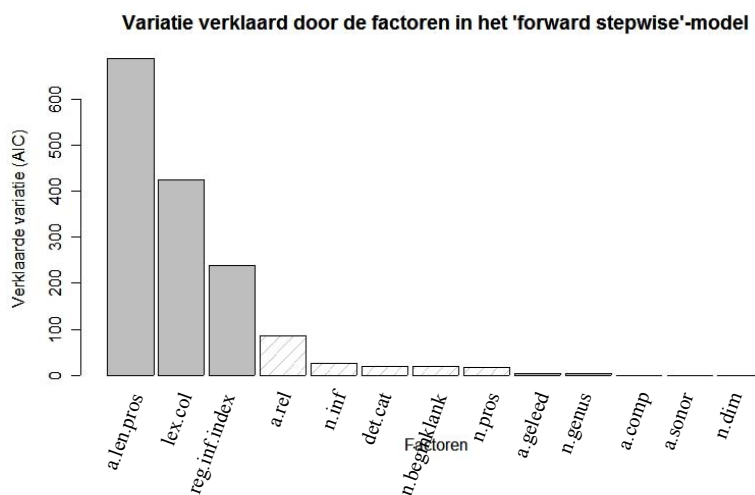
We hebben alle variabelen als nominale variabelen gecodeerd. Hoewel de lexicale collocabiliteit, de lengte van het adjectief en de informaliteitsindex een codering als ordinale of numerische variabele niet uitsluiten, hebben we voor een maximale consistentie en vergelijkbaarheid met de bivariate analyses gekozen.

De responsvariabele in de logistische-regressieanalyse is net zoals bij de bivariate analyses de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ , of om preciezer te zijn de logit van deze odds (zie paragraaf 13.2.1).

## 14.2 Impact van de verklarende factoren: *forward stepwise*-regressie

Zoals beschreven in de technische uiteenzetting in het vorige hoofdstuk (zie paragraaf 13.2.1) gebruiken we de *forward stepwise*-regressietechniek om te meten welke factoren de onverklaarde variatie in de materiaalverzameling het meest reduceren. Figuur 14(1) schematiseert de bijdrage die de verschillende factoren in

tabel 14(1) tot de reductie van de onverklaarde (buigings)variatie in de materiaalverzameling realiseren. De totale variatie of de onverklaarde variatie in het lege model, i.e. het model zonder enige verklarende variabele, is gelijk aan 5,383.4. De Y-as stelt de variatie voor (gecorrigeerd aan de hand van het AIC; zie paragraaf 13.2.1 noot 6) die iedere factor verklaart. Op de X-as worden de variabelen gerangschikt in functie van hun verklarende kracht: de meest linkse factor is de factor die het eerst door het *forward stepwise*-algoritme in het model wordt opgenomen, de op één na meest linkse factor is de tweede factor die door het *forward stepwise*-algoritme in het model wordt opgenomen, ... De drie factoren rechts in de figuur worden niet in het *forward stepwise*-model opgenomen omdat ze geen significante reductie van de onverklaarde variatie realiseren.



*Figuur 14(1): Variatie verklaard door de verschillende factoren in het forward stepwise-regressiemodel*

Op basis van figuur 14(1) bundelen we de verklarende factoren in vier klassen:

- *primaire factoren*: het prosodische patroon van het adjectief gecombineerd met de adjectivische lengtegraad, de lexicale collocabiliteit van het adjectief-nomenpaar, de lectische factor (nationale variëteit & informaliteitsindex)
- *secundaire factoren*: de relationaliteit van het adjectief, het voorkomen van een genominaliseerde infinitief in de nominale kernpositie, de woordsoort van de determinator, de beginklank van het substantief, het prosodische patroon aan de linkerwoordrand van het substantief
- *tertiaire factoren*: de morfologische geleedheid van het adjectief, het genus van het substantief

- *niet-significante factoren*: de trappen van vergelijking van het adjectief (c.q. stellende trap vs. vergrotende trap), de sonoriteit van de adjectivische eind-medeklinker, de diminuering van het substantief

Wanneer we de rangorde van de factoren in figuur 14(1) beschouwen, dringen zich enkele vaststellingen op. Ten eerste, de combinatie van de adjectivische lengte en het prosodische patroon aan de rechterwoordrand van het adjectief realiseert de grootste reductie van de variatie en wordt op ruime afstand gevolgd door de lexicale collocabiliteit van het adjectief-nomenpaar. De prominente positie van de prosodische factor is niet verwonderlijk rekening houdend met de gesproken natuur van het gebruikte materiaal. Daarnaast bevestigt dit resultaat de conclusie van de bivariate analyse dat de prosodische conditionering wellicht een voldoende factor is voor de selectie van beide buigingsvarianten (zie paragraaf 10.2.2). Merk op dat dit resultaat echter in tegenspraak is met de conclusies van de stratumanalyse in het lexicale/semantische hoofdstuk (paragraaf 11.2.3), waar we geopperd hebben dat bepaalde (concrete) prosodische buigingstendenties geconditioneerd worden door de lexicale/semantische kenmerken van het adjectief en het adjectief-nomenpaar. Ten tweede, de twee lexicale/semantische factoren zijn hoog gerangschikt in figuur 14(1). Dit onderstreept het belang van de lexicale/semantische conditionering van de buigingsalternantie en vormt, naast de ruime aandacht die de semantische conditionering in de literatuur geniet (zie paragrafen 1.1.3 en 1.2.2), een bijkomende motivatie voor de lexicale en semantische analyse in het volgende deel van dit proefschrift (deel IV). Ten derde, de lectische conditionering – een parameter die in de literatuur slechts zelden aan bod komt – bezet de derde positie in de rangorde en is bijgevolg een prominente verklarende factor voor de buigingsalternantie in het gesproken standaard Nederlands. Deze drie primaire factoren realiseren het grootste deel van de verklaarde variatie.

Tot slot, de tertiaire factoren identificeren zwakke maar niettemin significante effecten ( $0.05 \geq p > 0.001$ ). De classificatie van de morfologische geleedheid van het adjectief als tertiaire factor toont aan dat het buigingseffect van deze factor epifomonaal is ten overstaan van de semantiek (c.q. de relationaliteit) en de prosodische structuur van het adjectief. De *forward stepwise*-analyse bevat bijkomende evidentie voor deze stelling: de verklarende kracht van de morfologische geleedheid van het adjectief neemt telkens aanzienlijk af wanneer de prosodische structuur van het adjectief (i.e. de selectie van ‘a.len.pros’ in stap 1 van het *forward stepwise*-algoritme) en de relationaliteit van het adjectief (i.e. de selectie van ‘a.rel’ in stap 4 van het *forward stepwise*-algoritme) in het regressiemodel worden opgenomen. Dit betekent dat de variatie verklaard door de morfologische geleedheid grotendeels vervat zit in de prosodische structuur en de relationaliteit van het adjectief.

### 14.3 Logistische-regressiemodel voor de adjectivische buigingsalternantie bij neutra

Voor de eigenlijke logistische regressie voeren we alle factoren in tabel 14(1) manueel in het regressiemodel in. Het algoritme berekent het effect van iedere factorwaarde, rekening houdend met het effect van de andere factor(waard)en in het model. We analyseren achtereenvolgens de verklarende en voorspellende kracht van het model (14.3.1), het buigingseffect van de verschillende factorwaarden (14.3.2) en de voorspellende kracht van de verschillende combinaties van factorwaarden (14.3.3). We sluiten het multivariaat onderzoek af met een afzonderlijke regressie-analyse voor het Nederlands Nederlands en het Belgisch Nederlands (14.3.4). Op deze wijze willen we achterhalen of beide nationale variëteiten een afzonderlijk model voor de buigingsalternantie bij neutra vereisen.

De gerapporteerde regressiemodellen bevatten alleen de factoren met een significante impact op de buigingsalternantie. Na een eerste analyse op basis van alle variabelen in tabel 14(1) hebben we het model herberekend na manuele verwijdering van de niet-significante verklarende variabelen.

#### 14.3.1 Modelstatistieken

We bekijken eerst de algemene modelstatistieken (zie ook paragraaf 13.2.2). De verklarende en de voorspellende kracht geven een algemeen beeld van de adequaatheid van het logistische-regressiemodel.

De gegevens voor de verklarende kracht wijzen op een zeer significante reductie van de onverklaarde variatie (gecorrigeerd aan de hand van het AIC):

$$\begin{aligned}
 (1) \quad \text{variatie}_{\text{VERKLAARD}} &= \text{variatie}_{\text{TOTAAL}} - \text{variatie}_{\text{RESIDU}} \\
 &= 5,383.4 - 3,779.7 \\
 &= 1,603.7 \\
 \text{df} &= 25, p < 0.00001
 \end{aligned}$$

Dit resultaat wordt bevestigd door de voorspellende kracht van het model:  $c = 0.855$ . Dit cijfer overschrijdt ruimschoots de drempelwaarde van 0.72 om van een ‘goed model’ te kunnen spreken (zie paragraaf 13.2.2 noot 10). Het kan op de volgende wijze geïnterpreteerd worden: voor 85.5% van de observaties in de materiaalverzameling voorspelt het logistische-regressiemodel de correcte buigingsvariant.

We concluderen dat voorliggende analyse een adequate modellering is van de adjectivische buigingsalternantie bij neutra in het gesproken standaard Nederlands.<sup>4</sup> Het gebruik van beide buigingsmorfemen is bijgevolg niet aleatoir, zoals gesuggereerd wordt door sommige bronnen in de neerlandistiek (zie paragraaf 14.3.3 voor een uitgebreide discussie): gesteld dat men de gepaste methode (c.q. kwantitatief-empirisch onderzoek) en de gepaste statistische technieken (c.q. multivariate technieken) gebruikt, is de adjectivische buigingsalternantie bij neutra een verschijnsel dat op adequate wijze in kaart kan worden gebracht. In de volgende

paragraaf kijken we naar de regressiecoëfficiënten om de impact van de factor(waard)en op de keuze van het adjectivische buigingsmorfeem te meten.

### 14.3.2 *Regressiecoëfficiënten: het effect van de factorwaarden op de buigingsalternantie*

Na de bespreking van de algemene modelstatistieken in de vorige paragraaf richten we onze aandacht nu op de verklarende variabelen. De impact van de verklarende variabelen wordt samengevat in tabel 14(2). Op basis van deze informatie kunnen we de tendenties blootleggen die aan de basis liggen van het gebruik van beide buigingsvarianten in het gesproken standaard Nederlands.<sup>5</sup>

Factor	Coëf.	S.F.	Wald Z	OR	95% BI OR	p
(Intercept)	-4.285	0.1662	-25.78			0.0000
<i>det.cat=aanw.vnw</i>	-0.144	0.1422	-1.02	0.8652	[0.6547,1.1435]	0.3090
<i>det.cat=bez.vnw</i>	0.7064	0.1364	5.18	2.0266	[1.5512,2.6478]	0.0000
<i>det.cat=genitief</i>	0.4499	0.6773	0.66	1.5681	[0.4157,5.9145]	0.5065
<i>n.inf=ja</i>	1.4758	0.2642	5.59	4.3745	[2.6062,7.3425]	0.0000
<i>n.genus=bigen</i>	-0.496	0.2392	-2.08	0.6085	[0.3807,0.9725]	0.0379
<i>n.beginklank=klink</i>	0.5373	0.1203	4.46	1.7113	[1.3516,2.1668]	0.0000
<i>a.len.pros=twee.s</i>	0.3130	0.1946	1.61	1.3675	[0.9337,2.0027]	0.1078
<i>a.len.pros=twee.sw</i>	1.1954	0.1668	7.16	3.3048	[2.3829,4.5835]	0.0000
<i>a.len.pros=meer.s</i>	1.0489	0.1602	6.55	2.8545	[2.0852,3.9075]	0.0000
<i>a.len.pros=meer.sw</i>	1.9085	0.1788	10.67	6.7429	[4.7494,9.5732]	0.0000
<i>a.len.pros=meer.sww</i>	2.9127	0.1915	15.21	18.406	[12.645,26.792]	0.0000
<i>n.pros=w</i>	0.4121	0.0886	4.65	1.5099	[1.2691,1.7965]	0.0000
<i>reg.inf.index=nl.1</i>	1.1741	0.6157	1.91	3.2352	[0.9678,10.814]	0.0565
<i>reg.inf.index=nl.2</i>	1.0585	0.1693	6.25	2.882	[2.068,4.0164]	0.0000
<i>reg.inf.index=nl.3</i>	0.8642	0.1593	5.42	2.3731	[1.7366,3.2427]	0.0000
<i>reg.inf.index=bel.0</i>	0.2649	0.128	2.07	1.3033	[1.014,1.6751]	0.0385
<i>reg.inf.index=bel.1</i>	0.9846	0.1654	5.95	2.6767	[1.9352,3.7023]	0.0000
<i>reg.inf.index=bel.2</i>	1.6458	0.1598	10.29	5.1851	[3.7902,7.0934]	0.0000
<i>reg.inf.index=bel.3</i>	2.2318	0.1824	12.23	9.3166	[6.5159,13.321]	0.0000
<i>lex.col=kw.1</i>	0.2043	0.1862	1.10	1.2266	[0.8514,1.7671]	0.2728
<i>lex.col=kw.2</i>	0.6219	0.1438	4.32	1.8624	[1.4048,2.4692]	0.0000
<i>lex.col=kw.3</i>	1.3140	0.1405	9.35	3.7210	[2.8252,4.9008]	0.0000
<i>lex.col=kw.4</i>	2.1982	0.1402	15.68	9.0087	[6.8439,11.858]	0.0000
<i>a.rel=ja</i>	0.9044	0.098	9.22	2.4704	[2.0384,2.994]	0.0000
<i>a.geleed=ja</i>	0.3479	0.1607	2.16	1.416	[1.0334,1.9404]	0.0304

Tabel 14(2): *Regressiecoëfficiënten van de factorwaarden in het globale logistische-regressiemodel op basis van het CGN*

Alvorens we de interpretatie van de kwantitatieve gegevens aanvatten, lichten we de structuur van tabel 14(2) toe. De eerste kolom ('Factor') somt de verschillende factorwaarden in het regressiemodel op. Het effect van de factorwaarden op de responsvariabele, de (logit van de) odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ , moet men steeds tegen de achtergrond van de referentiewaarde van de factor in kwestie interpreteren (zie tabel 14(1)). De tweede kolom ('Coëf.') stelt de regressiecoëfficiënt voor: dit is de coëfficiënt in de logistische-regressievergelijking (i.e. de  $b_i$  in de abstracte regressievergelijkingen in paragraaf 13.1). De derde kolom ('S.F.>') bevat de standaardfout op de regressiecoëfficiënt. Deze maat is nodig voor het berekenen van het 95% BI van de OR. In de vierde kolom wordt de *Wald Z*-statistiek gerapporteerd die gebruikt wordt bij de berekening van de  $p$ -waarde. De laatste drie kolommen zijn het belangrijkste bij de interpretatie van de regressieanalyse: de OR (berekend op basis van de regressiecoëfficiënt; cf. (5) in paragraaf 13.2.2), het 95% BI voor de OR, en de  $p$ -waarde. De bespreking is op deze grootheden gebaseerd. De parameterwaarden die geen significante impact op de buigingsalternantie hebben, zijn gecursiveerd: dit zijn de parameterwaarden met een  $p$ -waarde groter dan 0.05 en een 95% BI voor de OR dat de scharnierwaarde 1 bevat (we verwijzen naar paragraaf 7.1 voor de interpretatie van de  $p$ -waarde en de 95% BI'en van de OR).<sup>6</sup>

We nemen de verschillende factoren afzonderlijk in beschouwing. Aan de hand van deze analyse kunnen we de belangrijkste onderzoeksvraag beantwoorden: welke tendenties liggen aan de basis van het gebruik van het verbogen en het onverbogen adjectief bij neutra?

#### *Woordsoort determinant*

De eerste factor in tabel 14(2) toont onmiddellijk de noodzaak van multivariaat onderzoek aan. De bivariate analyse (zie paragraaf 8.1.2) heeft – in tegenspraak met de hypothese – geen deflecterend effect van het bezittelijk voornaamwoord aangetoond. Telkens bleek er ook een significant verschil te zijn tussen de buigingsprofielen van het bepaald lidwoord en het aanwijzend voornaamwoord. De logistische-regressieanalyse toont het tegengestelde beeld: ten eerste, de OR voor het bezittelijk voornaamwoord identificeert een significante stijging van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  in vergelijking met het bepaald lidwoord (OR = 2.0266 [1.5512,2.6478]); ten tweede, het regressiemodel wijst geen significant verschil aan tussen het bepaald lidwoord en het aanwijzend voornaamwoord (OR = 0.8652 [0.6547,1.1435]). Dit betekent dat de corpusgegevens conform zijn met beide hypothesen die we op basis van de literatuurstudie geformuleerd hebben (zie paragrafen 1.1.1 en 8.1.1).

Het resultaat van de logistische-regressieanalyse sluit aan bij de historische ontwikkelingen (zie hoofdstuk 2). Na het bezittelijk voornaamwoord stellen we sinds het Middel-Nederlands een tendentie vast tot het gebruik van de sterke/onverbogen adjectivische vorm. Tot halverwege de vorige eeuw schreven veel grammatica's het gebruik van de onverbogen vorm voor na een bezittelijk voornaamwoord. Het gelijkaardige buigingsgedrag na een bepaald lidwoord en een aanwijzend



voornaamwoord sluit dan weer aan bij de diachrone evolutie van het bepaald lidwoord uit het aanwijzend voornaamwoord in het Oud-Nederlands.

Het buigingsprofiel van de genitief vertoont tot slot geen significant verschil vergeleken met dat van het bepaald lidwoord (OR = 1.5681 [0.4157,5.9145]). Hierbij moeten we, gezien het ruime 95% BI voor de OR, rekening houden met dataschaarste.

Het contrasterende effect van een verklarende parameter in de bivariate analyse en de multivariate analyse staat in de statistiek bekend als *Simpsons paradox* (Agresti 1996: 54-55). Bij de algemene bespreking van de multivariate analyse (cf. *infra*) reiken we een verklaring aan voor deze tegenstrijdigheid.

#### *Genominaliseerde infinitief*

De OR voor de genominaliseerde infinitief wijst op een duidelijk deflecterend effect: we stellen ruim een verviervoudiging van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vast (OR = 4.3745 [2.6062,7.3425]). In het bivariaat onderzoek hebben we aangevoerd dat de selectie van de onverbogen buigingsvariant gefaciliteerd wordt door het (semi-) adverbiale statuut van het adjectief bij een genominaliseerde infinitief (zie paragraaf 8.3.2).

#### *Genus/bigenericiteit substantief*

Het genus van het element dat de nominale kernpositie bezet, definieert een significant effect: bij een bigenerisch substantief komt de verbogen vorm significant vaker voor dan bij een *het*-woord (OR = 0.6085 [0.3807,0.9725]). Het 95% BI voor de OR en de *p*-waarde (*p* = 0.0379) duiden echter aan dat het om een zwak effect gaat. Rekening houdend met de grootte van de materiaalverzameling en de zeer kleine *p*-waarden in de bivariate analyses en bij de andere factorwaarden in het regressiemodel in tabel 14(2) concluderen we dat het genuseffect zeer zwak is. Gegeven de omvang van de materiaalverzameling (n = 4,964) moet er sprake zijn van significantie op het  $\alpha$ -niveau 0.001 voordat we met stelligheid een sterk effect (kunnen) claimen.

#### *Beginklank substantief*

Vóór een substantief dat met een klinker begint, wordt significant vaker de onverbogen buigingsvariant geselecteerd dan vóór een substantief dat met een medeklinker begint (OR = 1.7113 [1.3516,2.1668]). Dit resultaat bevestigt het effect dat we in het bivariaat onderzoek hebben vastgesteld (zie paragraaf 9.2.1). De selectie van het onverbogen buigingsmorfeem vóór een substantief met een klinker als beginklank hebben we verklaard als een instantie van hiaatvermijding. We vestigen nogmaals de aandacht op de stilistische conditionering van dit effect: de hiaatvermijding gaat enkel in de informele registers (informaliteitsindex > 0) gepaard met een significante daling van het aandeel verbogen vormen (zie paragraaf 12.3.3.2).

*Adjectivische lengtegraad en prosodisch patroon adjectief*

De OR's staven de resultaten van de bivariate analyse, zowel voor het prosodische patroon aan de rechterwoordrand van het adjectief (zie paragraaf 10.2.2) als voor de lengtegraad van het adjectief (paragraaf 10.2.1). De adjectieven met 's' als meest rechtse voet tonen bij elk van de onderscheiden adjectivische lengtegraden de grootste voorkeur voor de verbogen variant. De adjectieven die op één onbeklemtoonde eindlettergreep eindigen, tonen een significant sterkere voorkeur voor de onverbogen buigingsvariant ( $OR_{\text{twee.sw}} = 3.3048$  [2.3829,4.5835];  $OR_{\text{meer.sw}} = 6.7429$  [4.7494,9.5732]). Bij een opeenvolging van twee onbeklemtoonde lettergrepen aan de rechterrandaan van de adjectivische basisvorm zien we een forse stijging van de OR ( $OR_{\text{meer.sw.w}} = 18.406$  [12.645,26.792]).

Het lengte-effect kan niet tot het prosodische patroon aan de rechterwoordrand van het adjectief gereduceerd worden: we stellen een toename van de OR vast wanneer we de patronen 's' en 'sw' bij de twee- en de meerlettergrepige adjectieven vergelijken ( $OR_{\text{twee.s}} = 1.3675 < OR_{\text{meer.s}} = 2.8545$ ;  $OR_{\text{twee.sw}} = 3.3048 < OR_{\text{meer.sw}} = 6.7429$ ). Een uitzondering op dit lengte-effect is het niet-significante verschil tussen een- en tweelettergrepige adjectieven met een beklemtoonde eindlettergreep ( $OR_{\text{twee.s}} = 1.3675$  [0.9337,2.0027]).

*Prosodisch patroon substantief*

De resultaten van de bivariate analyse voor het buigingseffect van het prosodische patroon aan de linkerwoordrand van het substantief (zie paragrafen 10.2.1 en 10.2.2) worden bevestigd door de multivariate analyse. Aan de ene kant heeft het prosodische patroon aan de linkerwoordrand van het substantief een significant effect op de buigingsalternantie: vóór een onbeklemtoonde substantivische beginlettergreep wordt de onverbogen buigingsvariant significant vaker geselecteerd dan vóór een beklemtoonde substantivische beginlettergreep ( $OR = 1.5099$  [1.2691,1.7965]). Aan de andere kant is de waarde van de OR voor het substantivische patroon aanzienlijk kleiner dan de OR's voor het adjectivische patroon. Kortom: het prosodische patroon aan de rechterwoordrand van het adjectief definieert het basiseffect; het prosodische patroon van het substantief zorgt hooguit voor een versterking van het adjectivische basiseffect (zie ook paragraaf 10.2.2).

Wanneer we de gegevens voor de adjectivische en de substantivische prosodische patronen samenvoegen, stellen we vast dat de regressiecoëfficiënten voor deze patronen conform zijn met de prosodische restricties die we op basis van de bivariate analyse geformuleerd hebben:

- het vermijden van een klemtoonbotsing: Deze prosodische restrictie kan enkel indirect uit de regressieanalyse worden afgelezen. Bij iedere adjectivische lengtegraad vertoont het patroon 's' de laagste OR. Daarnaast stijgt de OR op significante wijze wanneer het substantief met een 'w' begint. Het cumulatieve effect van het adjectivische en het substantivische patroon onderbouwt de stelling dat de buigings-*e* geselecteerd kan worden om een klemtoonbotsing te vermijden: alle adjectivische en substantivische

prosodische patronen waar geen klemtoonbotsing kan optreden, selecteren significant vaker het onverbogen adjectief dan de patronen die we wel bij een klemtoonbotsing kunnen aantreffen.

- het vermijden van voetloze lettergrepen: Bij de patronen waar de buigings-*e* de enige voetloze lettergreep zou zijn (c.q. ‘sw’ als meest rechtse voet van de adjectivische basisvorm), stellen we een significante toename van de onverbogen vorm vast om een optimaal prosodisch patroon aan de rechterwoordrand van het adjectief te realiseren. We nemen minstens een verdrievoudiging van deze deflecterende tendentie waar wanneer de eventuele buigings-*e* de klinker van de tweede opeenvolgende voetloze lettergreep zou zijn (c.q. ‘sww’ als meest rechtse voet van de adjectivische basisvorm). Dit verschil tussen de OR’s weerspiegelt de verschillende impact van beide inbreuken tegen de prosodische welgevormdeheidsprincipes: terwijl één voetloze lettergreep (i.e. ‘[sw]<sub>v</sub>w<sub>★</sub>’ als verbogen vorm) als een lichte inbreuk op de prosodische structuur van een Nederlands woord wordt gevoeld, beschouwen de leden van de Nederlandse taalgemeenschap twee opeenvolgende voetloze lettergrepen (i.e. ‘[sw]<sub>v</sub>ww<sub>★</sub>’ als verbogen vorm) als een zware inbreuk die moet worden vermeden indien mogelijk, zelfs wanneer hiervoor de door de grammatica voorziene buigings-*e* wordt weggelaten.

*Lectische variatie: nationale variëteit en register*

We kijken eerst naar het Nederlands Nederlands. Voor het register met informaliteitsindex 2 (OR = 2.882 [2.068,4.0164]) bevestigen de gegevens in het logistische-regressiemodel de resultaten van de bivariate analyse (zie paragrafen 12.3.1 en 12.3.2): we stellen in vergelijking met het referentiepunt (‘nl.0’) een significante stijging van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  vast. In de lectische analyse hebben we betoogd dat dit geen informeel maar een formeel effect is, omdat het sterk gecontroleerde gesprekssituaties betreft die gepaard gaan met formeel en ambtelijk taalgebruik, met name parlementaire debatten in de Tweede Kamer. Het niet-significante verschil tussen het register met informaliteitsindex 1 en het referentiepunt bevestigt eveneens het resultaat op basis van de bivariate analyse (OR = 3.2352 [0.9678,10.814]).<sup>7</sup> De OR voor de meest informele gesprekssituaties (informaliteitsindex 3) verschilt tot slot ook significant van het formele referentiepunt (OR = 2.3731 [1.7366,3.2427]). Dit resultaat contrasteert met de bivariate analyse. Hiervoor kunnen we geen verklaring aanbieden, behalve dat de interviews met de leerkrachten Nederlands, die tot dit register behoren, enkele recurrente lexicale collocaties bevatten, zoals *middelbaar onderwijs*, *voortgezet onderwijs* en *lager onderwijs*.

Laten we nu de situatie in het Belgisch Nederlands bekijken. De OR die het formele Belgisch Nederlands (‘bel.0’) met het formele Nederlands Nederlands vergelijkt, wijst op een klein maar significant verschil (OR = 1.3033 [1.014,1.6751]). De *p*-waarde bevestigt deze vaststelling (*p* = 0.0385). Dit resultaat

strookt met de algemene neiging tot convergentie van het Belgisch Nederlands met het Nederlands Nederlands in formele gespreksituaties. In de informele gespreksituaties stellen we een progressieve stijging van de OR vast naarmate de informaliteitsindex stijgt ( $OR_{bel.1} = 2.6767$ ;  $OR_{bel.2} = 5.1851$ ;  $OR_{bel.3} = 9.3166$ ).

Met andere woorden: de resultaten van de logistische-regressieanalyse zijn conform met de drie lectische hypothesen:

- De onverbogen variant wordt vaker geselecteerd in het Belgisch Nederlands dan in het Nederlands Nederlands. Merk wel op dat het buigingsgedrag in de formele Belgische registers nauwelijks afwijkt van het formele Nederlands Nederlands. Het kleine maar significante onderscheid tussen het formele taalgebruik in beide nationale variëteiten sluit aan bij de sterke convergentie van het formele Belgisch Nederlands, het zogenaamde VRT-Nederlands, met het Nederlands Nederlands.
- In het Nederlands Nederlands valt de keuze van de onverbogen vorm samen met contexten die gekenmerkt worden door een zeer formeel en ambtelijk taalgebruik, zoals debatten in de Tweede Kamer. We constateren echter ook een significante stijging van de OR voor het register met informaliteitsindex 3, i.e. een effect dat niet samenhangt met formulaïsch en ambtelijk taalgebruik.
- In het Belgisch Nederlands wordt de exogene verbogen buigingsvorm in formele contexten gebruikt. In de informele gespreksituaties nemen we een progressieve stijging van de endogene onverbogen vorm waar. Dit stilistische effect hebben we gerelateerd aan de productiedruk: naarmate de productiedruk stijgt, selecteert de Vlaming progressief vaker de onverbogen variant.

#### *Lexicale collocabiliteit adjectief-nomenpaar*

Het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar dat we in de bivariate analyse (zie paragraaf 11.2.1) hebben waargenomen, wordt door de regressieanalyse bevestigd. De OR stijgt gestaag naarmate de collocationale associatiekracht tussen het adjectief en het substantief vergroot ( $OR_{kw.1} = 1.2266$ ;  $OR_{kw.2} = 1.8624$ ;  $OR_{kw.3} = 3.7210$ ;  $OR_{kw.4} = 9.0087$ ). De adjectief-nomenparen in het laagste collocabiliteitskwartiel verschillen niet significant ten overstaan van het referentiepunt, i.e. de afwezigheid van lexicale collocabiliteit ( $p = 0.2728$ ): dit hebben we verklaard als een gevolg van de lage lexicale specificiteit van de elementen die de adjectivische en substantivische posities bezetten in de observaties die tot het laagste collocabiliteitskwartiel ('lex.col<sub>kw.1</sub>') behoren. We concluderen dat het onverbogen adjectief een sterke lexicale relatie tussen het adjectief en het substantief markeert, waarbij het adjectief-nomenpaar als een gelexicaliseerde woordgroep geïdentificeerd wordt.

*Relationaliteit adjectief*

Naast de lexicale collocabiliteit van het adjectief-nomenpaar is het onderscheid tussen een relationeel en een kwalificerend adjectief de tweede lexicale/semantische parameter die de keuze tussen beide buigingsvarianten beïnvloedt (zie paragraaf 11.2.2). Het resultaat van de bivariate analyse wordt bevestigd: de selectie van het onverbogen buigingsmorfeem kan gebruikt worden als signaal dat het adjectief geen prototypische modifierende of kwalificerende functie heeft, maar dat het een relatie tussen twee substantivische concepten oproept, met name het concept aangeduid door het nominale hoofd en het concept aangeduid door het denominale relationele adjectief (OR = 2.4704 [2.0384,2.994]).

*Morfologische geleedheid adjectief*

De laatste factor die we in het regressiemodel hebben opgenomen, is de morfologische geleedheid van het adjectief. De OR (OR = 1.416 [1.0334,1.9404]) en de *p*-waarde (*p* = 0.0304) wijzen op een zwak maar significant effect. Deze vaststelling ondersteunt – net als de lage rangorde in het *forward stepwise*-model (zie paragraaf 14.2, figuur 14(1)) – onze stelling dat deze factor epifenomenaal is, meer bepaald ten overstaan van de prosodische structuur en de lengtegraad van het adjectief en de relationaliteit van het adjectief.

*Niet-significante factoren*

De volgende factoren zijn niet in het manuele logistische-regressiemodel in tabel 14(2) opgenomen, omdat ze geen significante bijdrage leveren tot de reductie van de onverklaarde buigingsvariatie:

- trappen van vergelijking van het adjectief: De multivariate analyse staft het resultaat van de bivariate analyse (zie paragraaf 8.2.2). Het is namelijk zo dat noch in de bivariate analyse noch in de multivariate analyse de corpusgegevens het frequent geciteerde deflecterende effect van de vormen van de vergrotende trap ondersteunen.
- diminuering substantief: Het regressiemodel ontkracht het deflecterende effect van het verkleinwoord dat we in het bivariaat onderzoek hebben vastgesteld (zie paragraaf 8.4.2). Net als bij de woordsoort van de determinator moeten we de tegenstrijdige resultaten van het bivariaat en het multivariaat onderzoek verklaren (cf. *infra*).
- sonoriteit adjectivische eindmedeklinker: In tegenstelling tot de resultaten van het bivariaat onderzoek (zie paragraaf 9.2.2) weerleggen de gegevens in het logistische-regressiemodel de hypothese dat een sonore adjectivische eindmedeklinker de keuze van de buigings-*e* faciliteert. Ook deze instantie van Simpsons paradox zullen we bij de algemene bespreking trachten te verklaren (cf. *infra*).

*Algemene bespreking*

Na de bespreking van de afzonderlijke factoren kijken we nu naar het globale beeld. Het model in tabel 14(2) toont aan dat de adjectivische buigingsalternantie bij neutra een complex maar modelleerbaar verschijnsel is. Het gebruik van beide adjectivische buigingsvarianten wordt bepaald door een reeks concurrerende en interagerende factoren die bovendien tot verschillende taalkundige domeinen behoren, zowel taalinterne als lectische, zowel formele als semantische.

Tabel 14(3) vergelijkt de resultaten van het bivariaat en het multivariaat onderzoek. De rijen onderscheiden de factoren in functie van de significantie van hun effect in het bivariaat onderzoek; de kolommen groeperen de factoren op basis van de significantie van hun effect in het logistische-regressiemodel in tabel 14(2). Zoals alle overzichtstabellen is ook het schema in deze tabel een generalisatie die niet alle nuances vat. De gecombineerde factoren, met name ‘a.len.pros’ en ‘reg.inf.index’, hebben geen directe tegenhanger in de bivariate analyses en de parameter voor de woordsoort van de determinator is opgesplitst omdat het effect van het aanwijzend voornaamwoord (‘det.cat=aanw.vnw’) en het effect van het bezittelijk voornaamwoord (‘det.cat=bez.vnw’) tegengesteld zijn.

			<b>Multivariate analyse</b>	
			<b>significant</b>	<b>niet-significant</b>
<b>Bivariate analyse</b>	<b>significant</b>		lex.col	det.cat=aanw.vnw
			a.rel	a.sonor
			reg.inf.index	n.dim
			a.len.pros	
			n.beginklank	
			n.inf	
			n.pros	
			n.genus	
			a.geleed	
			<b>niet-significant</b>	det.cat=bez.vnw
				det.cat=genitief

*Tabel 14(3): Overzicht van de significantie van de verklarende factoren in de bivariate en de multivariate analyses*

Voor het merendeel van de bestudeerde factoren bevestigt de logistische-regressieanalyse de resultaten van de bivariate analyses: factoren waarvan het afzonderlijke effect (niet) significant is, hebben eveneens in het regressiemodel een (niet-)significante impact op de responsvariabele. Voor een aantal factoren zijn de resultaten voor de bivariate analyse echter tegengesteld aan de resultaten voor de multivariate analyse. Ten eerste, verscheidene factoren die in de bivariate analyses een significant effect vertonen, verliezen in de logistische regressie hun significantie: het effect van het aanwijzend voornaamwoord, de diminuering van het

element dat de nominale kernpositie bezet en de sonoriteit van de adjectivische eindmedeklinker. Ten tweede, voor één factorwaarde is het effect enkel in de multivariate analyse significant, met name het effect van het bezittelijk voor-naamwoord.

Na de identificatie van de discrepanties tussen de bivariate analyses en de multivariate analyse gaan we op zoek naar een verklaring voor de waargenomen verschillen. We maken hierbij gebruik van een correspondentieanalyse. Dit is een exploratieve statistische techniek die factorwaarden groepeerd in functie van hun onderlinge relatie in een materiaalverzameling.<sup>8</sup> De correspondentieanalyse waarop deze bespreking gebaseerd is, groepeerd de factorwaarden in vier kwadranten. Factorwaarden die sterk gerelateerd zijn in de materiaalverzameling bevinden zich in hetzelfde kwadrant. Deze kwadranten worden in tabel 14(4) gevisualiseerd: factorwaarden die in de materiaalverzameling op basis van het CGN sterk gerelateerd zijn – dit wil zeggen vaker dan verwacht samen voorkomen –, staan in dezelfde cel in tabel 14(4).<sup>9</sup>

Dimensie 1		
	det.cat=bez.vnw	n.genus= <i>het</i> -woord
	det.cat=genitief	n.inf=ja
	a.comp=ja	n.beginklank=klinker
	n.dim=ja	a.len.pros=twee.s
	a.sonor=ja	a.len.pros=twee.sw
	n.beginklank=medeklinker	a.len.pros=meer.sw
	n.pros=s	reg.inf.index=nl.1
	reg.inf.index=nl.0	reg.inf.index=bel.0
	a.rel=neen	lex.col=-lex.col
		a.geleed=ja
Dimensie2	det.cat=aanw.vnw	det.cat= <i>het</i>
	a.comp=neen	n.dim=neen
	n.genus=bigenerisch	a.sonor=neen
	n.inf=neen	a.len.pros=meer.s
	a.len.pros=een.s	a.len.pros=meer.sww
	reg.inf.index=nl.3	n.pros=w
	reg.inf.index=bel.3	reg.inf.index=nl.2
	lex.col=kw.1	reg.inf.index=bel.1
	a.geleed=neen	reg.inf.index=bel.2
		lex.col=kw.2
		lex.col=kw.3
	lex.col=kw.4	
	a.rel=ja	

Tabel 14(4): Kwadranten voor de correspondentieanalyse op basis van de verklarende factoren

We verifiëren nu of de associaties tussen de waarden van de verklarende factoren kunnen worden aangewend ter verklaring van de voorbeelden van Simpsons paradox die we in onze materiaalverzameling hebben vastgesteld bij het buigingseffect van de woordsoort van de determinator, de sonoriteit van de adjectivische eindmedeklinker en de diminuering van het substantief. We kunnen hierover echter geen definitief uitsluitel geven bij gebrek aan een significantietest voor het resultaat van de correspondentieanalyse (zie ook noot 8).

	<b>Flecterende waarden</b>	<b>Niet-significante waarden</b>	<b>Deflecterende waarden</b>
<b>det.cat=het</b>	/	a.sonor=neen n.dim=neen	n.pros=w reg.inf.index=bel.1 reg.inf.index=bel.2 reg.inf.index=nl.2 lex.col=kw.2 lex.col=kw.3 lex.col=kw.4 a.rel=ja a.len.pros=meer.s a.len.pros=meer.sww
<b>det.cat=bez.vnw</b>	reg.inf.index=nl.0 a.rel=neen n.beginkl=medekl. n.pros=s	det.cat=genitief a.comp=ja a.sonor=ja n.dim=ja	/
<b>det.cat=aanw.vnw</b>	n.inf=neen a.len.pros=een.s a.geleed=neen	a.comp=neen lex.col=kw.1	n.genus=bigen reg.inf.ind=bel.3 reg.inf.ind=nl.3

*Tabel 14(5): Factorwaarden waarmee de verschillende determinatoren sterk geassocieerd zijn in de materiaalverzameling*

Bij het buigingseffect van de woordsoort van de determinator is Simpsons paradox een gevolg van de parameterwaardencombinaties waarin de drie determinatoren – *het*, aanwijzend voornaamwoord en bezittelijk voornaamwoord – frequent voorkomen (de genitief laten we buiten beschouwing). Doordat een bivariate analyse geen rekening houdt met andere factoren, zijn de bivariate buigingsprofielen voor deze observaties scheefgetrokken als gevolg van de interactie met andere factoren: in het bivariate buigingsprofiel van het bepaald lidwoord is de onverbogen buigingsvariant oververtegenwoordigd; in het bivariate buigingsprofiel van het bezittelijk voornaamwoord is de verbogen buigingsvariant oververtegenwoordigd. Tabel 14(5) geeft een overzicht van de factorwaarden waarmee deze determinatoren op basis van de correspondentieanalyse sterk geassocieerd zijn in de



materiaalverzameling. We kijken meer bepaald naar de factorwaarden die in tabel 14(4) in hetzelfde kwadrant als deze determinatoren voorkomen. We delen deze factorwaarden in drie klassen op: de waarden die het gebruik van de onverbogen vorm inhiberen ('flecterende waarden'), de waarden die het gebruik van de onverbogen vorm faciliteren ('deflecterende waarden') en de waarden die geen significant effect op de buigingsalternantie hebben in het logistische-regressiemodel in tabel 14(2) ('niet-significante waarden').

Het bepaald lidwoord, dat als ijkpunt fungeert voor de berekening van de impact van de woordsoort van de determinator, is hoofdzakelijk geassocieerd met factorwaarden die het gebruik van de onverbogen variant faciliteren. Hierdoor neigt het buigingsprofiel van dit element in de bivariate analyse te sterk naar de onverbogen buigingsvariant. Het bezittelijk voornaamwoord is dan weer sterk geassocieerd met flecterende factorwaarden: de sterke voorkeur van deze factorwaarden voor de buigings-*e* neutraliseert in de bivariate analyse het in essentie deflecterende effect van het bezittelijk voornaamwoord. Het beeld voor het aanwijzend voornaamwoord is minder duidelijk dan bij het bepaald lidwoord of het bezittelijk voornaamwoord. We kunnen niettemin het significant grotere aandeel verbogen vormen ten overstaan van het referentiepunt, i.e. het bepaald lidwoord, dat we in de bivariate analyse bij het aanwijzend voornaamwoord hebben vastgesteld in verband brengen met de te sterke deflecterende neiging van dit referentiepunt in de bivariate analyse. In de regressieanalyse worden de relaties tussen de verschillende verklarende variabelen gecontroleerd, waardoor de scheeftrekking, die voor een vertekend beeld zorgt in de bivariate analyse, gecorrigeerd wordt.

Twee factoren vertonen het tegengestelde beeld van het bezittelijk voornaamwoord in tabel 14(3), met name een significant bivariaat effect dat niet door de multivariate analyse gestaafd wordt: de diminuering van het substantief en de sonoriteit van de adjectivische eindlettergreep. Aangezien de waarden van deze parameters in dezelfde kwadranten in tabel 14(4) voorkomen, behandelen we ze samen in tabel 14(6). De contexten waar de waarden van deze factoren in de materiaalverzameling voorkomen, reiken een verklaring aan voor de discrepantie tussen de resultaten van de bivariate en de multivariate analyse. Tabel 14(6) vat de factorwaarden samen die sterk geassocieerd zijn met de waarden voor de diminuering van het element in de nominale kernpositie en met de waarden voor de sonoriteit van de adjectivische eindconsonant.

We kijken eerst naar de waarden voor de diminuering van het element in de kernpositie. De factorwaarde die in de bivariate analyse vaker de onverbogen variant selecteert ('n.dim=neen'), is sterk geassocieerd met factorwaarden die het gebruik van de onverbogen vorm faciliteren. De factorwaarde die in de bivariate analyse de verbogen variant verkiest ('n.dim=ja'), is daarentegen sterk gerelateerd aan factorwaarden die de buigings-*e* selecteren.

	<b>Flecterende waarden</b>	<b>Niet-significante waarden</b>	<b>Deflecterende waarden</b>
<b>n.dim=neen</b> <b>a.sonor=neen</b>	det.cat= <i>het</i>	/	n.pros=w reg.inf.index=bel.1 reg.inf.index=bel.2 reg.inf.index=nl.2 lex.col=kw.2 lex.col=kw.3 lex.col=kw.4 a.rel=ja a.len.pros=meer.s a.len.pros=meer.sww
<b>n.dim=ja</b> <b>a.sonor=ja</b>	n.beginkl=medekl. n.pros=s reg.inf.index=nl.0 a.rel=neen	det.cat=genitief a.comp=ja	det.cat=bez.vnw

*Tabel 14(6): Factorwaarden waarmee de verschillende waarden voor de diminuering van het element in de nominale kernpositie en de verschillende waarden voor de sonoriteit van de adjectivische eindmedeklinker sterk geassocieerd zijn in de materiaalverzameling*

Laten we nu de waarden voor de sonoriteit van de adjectivische eindconsonant bekijken, de tweede factor die een significant bivariaat effect vertoont dat de multivariate analyse niet ondersteunt. De flecterende parameterwaarde ('a.sonor=ja') is voornamelijk met factorwaarden geassocieerd die de buigings-*e* selecteren, terwijl de deflecterende parameterwaarde ('a.sonor=neen') vooral voorkomt met factorwaarden die de selectie van de buigings-*e* inhiberen.

De vastgestelde verschillen tussen de bivariate en de multivariate analyses ondersteunen onze stelling dat multivariaat onderzoek noodzakelijk is om greep te krijgen op complexe taalkundige verschijnselen, zoals de adjectivische buigingsalternantie bij neutra. Doordat bivariate analyses het effect van één enkele factor afzonderlijk beschouwen, worden interacties en de daaruit voortvloeiende conditioneringen van het buigingsprofiel niet in beschouwing genomen. Dit betekent echter niet dat we pleiten voor het opgeven van bivariate analyses in corpuslinguïstisch onderzoek. De bivariate analyses vormen een essentieel onderdeel van het kwantitatief-empirisch onderzoek, in eerste instantie bij het operationaliseren van de hypotheses, het parametriseren van de factoren en het interpreteren van de vastgestelde patronen, zoals de afzonderlijke bivariate analyses in deel II duidelijk illustreren. Het is echter een illusie te denken dat taalvariatie verklaard kan worden zonder rekening te houden met de competitie en de interactie tussen de verschillende verklarende factoren. De observaties in een corpus worden immers gedefinieerd door de combinatie van verschillende verklarende factor(waard)en die het effect van de afzonderlijke verklarende factor(waarde)en kan beïnvloeden: in de eerste plaats kan het effect van een factor verschillend zijn in functie van de factorwaarden-

combinatie waarin hij voorkomt, zoals aangetoond in de stratumanalyses in deel II; in de tweede plaats kan het reële effect van een factor versluierd zijn door de specifieke samenstelling van de materiaalverzameling. Dit is een gevolg van de ongecontroleerde samenstelling van corpusmateriaal: door zijn spontane natuur ontglipt het taalgebruik in een corpus aan de controle van de onderzoeker. Bijgevolg mag de onderzoeker zich niet tot de afzonderlijke effecten van de individuele factoren beperken, maar moet h/zij bij de uiteindelijke modellering rekening houden met de concrete interacties tussen de verklarende factoren in de materiaalverzameling, c.q. het taalgebruik, en met de eventuele partiële conditioneringen.

Laten we nu het schema in tabel 14(3) beschouwen vanuit het standpunt van de hypothesen die we uit de literatuur gedistilleerd hebben. Dit schema toont aan dat het taalgebruik in het CGN conform is met de meeste onderzoekshypothesen. Enkel het faciliterende effect van de comparatief en het inhiberende effect van de diminutief ten overstaan van de selectie van het onverbogen buigingsmorfeem worden niet door de corpusgegevens ondersteund. Van de vermoedens die we tijdens de empirische analyse geformuleerd hebben, wordt enkel de stijging van het aandeel onverbogen vormen bij een genominaliseerde infinitief door de multivariate analyse bevestigd. Het andere vermoeden, i.e. de stijging van het aandeel verbogen vormen bij de adjectieven met een sonore eindmedeklinker, wordt alleen door de bivariate analyse onderbouwd. Bij de evaluatie van de hypothesen met betrekking tot de conditionering van de buigingsalternantie mogen we niet vergeten dat de voorgestelde operationalisering soms afwijkt van de oorspronkelijke hypothese. De lexicale collocaties meten de lexicale associatiesterkte tussen het adjectief en het substantief. Deze notie is ruimer dan de idiomatische of de naamstatus van het adjectief-nomenpaar die in de introspectieve literatuur worden aangehaald, maar die een objectieerbare operationalisering in de weg staan (zie paragraaf 11.1.1). De prosodische analyse heeft de beklemtoning van de adjectivische eindlettergreep uitgebreid tot de meest rechtse metrische voet van het adjectief en het prosodische patroon op het niveau van de fonologische constituent (zie paragrafen 10.1.1 en 10.2.2).

Op basis van de regressiecoëfficiënten in tabel 14(2) groeperen we de significante verklarende factoren in drie groepen naargelang de impact op de buigingsalternantie. Daaraan voegen we een vierde groep toe voor de niet-significante factoren.

- *primaire factoren*: prosodisch patroon van het adjectief gecombineerd met de adjectivische lengte, lexicale collocabiliteit van het adjectief-nomenpaar, lectische conditionering
- *secundaire factoren*: relationaliteit van het adjectief, voorkomen van een genominaliseerde infinitief in de nominale kernpositie, woordsoort van de determinator, beginklank van het substantief, prosodisch patroon van het substantief

- *tertiaire factoren*: morfologische geleding van het adjectief, genus van het substantief
- *niet-significante factoren*: diminuering van het element in de nominale kernpositie, trappen van vergelijking van het adjectief, sonoriteit van de adjectivische eindmedeklinker

Deze vierledige classificatie aan de hand van de regressiecoëfficiënten bevestigt de rangschikking van de factoren op basis van de *forward stepwise*-regressie (zie paragraaf 14.2, figuur 14(1)). De primaire factoren realiseren niet alleen de grootste reductie van de onverklaarde variatie in de materiaalverzameling, maar definiëren bovendien de factorwaarden met de sterkste impact op de keuze van het buigingsmorfem. De primaire factoren zijn verschillend van aard: we hebben een syntagmatische factor, een lexicale/semantische factor en een lectische factor. Hoewel het effect van de secundaire factoren zeer significant is, vervullen ze een ondergeschikte rol in vergelijking met de primaire factoren: zowel de variatiereducerende kracht als de OR's zijn kleiner dan bij de primaire factoren. Twee van deze factoren zijn nauw verbonden met een primaire factor: de relationaliteit van het adjectief en het prosodische patroon van het substantief vullen respectievelijk de lexicale collocabiliteit van het adjectief-nomenpaar en de prosodische structuur van het adjectief aan. De tertiaire factoren definiëren zwakke condities op de buigingsalternantie, zowel op het vlak van de variatiereductie als op het vlak van de regressiecoëfficiënten. De zwakte van deze factoren wordt onderstreept door de relatief grote  $p$ -waarden ( $0.05 \geq p > 0.001$ ), die – met uitzondering van het genuseffect in het Nederlands Nederlands – niet meer significant zijn wanneer we de materiaalverzameling opdelen in twee regionaal gedefinieerde deelverzamelingen (zie paragraaf 14.3.4). Tot slot zijn er de niet-significante factoren die noch in de *forward stepwise*-analyse, noch in de manuele regressie een significante impact op de buigingsalternantie definiëren.

We concluderen dat het onverbogen adjectief – en bij uitbreiding de buigingsalternantie – in de gesproken standaardtaal drie primaire functies vervult:

- de realisatie van een optimale prosodische vorm: De onverbogen vorm wordt geselecteerd om een opeenvolging van onbeklemtoonde, voetloze lettergrepen te vermijden. Noteer dat het gebruik van de verbogen vorm bijna imperatief is bij een dreigende klemtoonbotsing.
- de identificatie van een lexicale collocatie: Het onverbogen buigingsmorfem wordt aangewend om het adjectief-nomenpaar als een lexicale collocatie te identificeren. Dit wil zeggen dat de onverbogen vorm een sterk lexicaal verband tussen het adjectief en het substantief aanduidt. De adjectivische positie in lexicale collocaties wordt vaak door een relationeel adjectief bezet. Het onverbogen adjectief signaleert dan dat het adjectief en het substantief niet alleen een lexicale maar ook een semantische eenheid vormen. Dit is: het adjectief modificeert niet langer de referent van het substantief maar roept een relatie tussen twee nominale concepten op. De

precieze aard van de lexicale/semantische conditionering van de buigingsalternantie vormt het onderwerp van deel IV.

- de markering van de gesprekssituatie: Deze functie is tegengesteld in Nederland en Vlaanderen. Terwijl de onverbogen vorm in Nederland kenmerkend is voor een zeer formeel en ambtelijk taalgebruik, komt deze buigingsvariant in Vlaanderen vooral voor in situaties die gekenmerkt worden door een grote productiedruk. Op het verschil tussen het Belgisch Nederlands en het Nederlands Nederlands komen we uitgebreid terug in paragraaf 14.3.4.

### ***14.3.3 Voorspellende kracht van de factorwaardencombinaties***

Tot nu toe hebben we de verklarende en de voorspellende kracht van het regressiemodel geëvalueerd (zie paragraaf 14.3.1) en hebben we de impact op de buigingsalternantie becommentarieerd van de verschillende factor(waard)en in het regressiemodel (zie paragraaf 14.3.2). De conclusie is dat de voorgestelde logistische regressie een adequate modellering is van de adjectivische buigingsalternantie bij neutra, zoals deze in het gesproken taalgebruik voorkomt.

We sluiten de logistische-regressieanalyse voor het volledige CGN af door de verschillende factorwaarden, die we tot nu toe afzonderlijk bekeken hebben, te combineren op basis van het logistische-regressiemodel. Dit doen we door de impact op de buigingsalternantie te meten van de verschillende logische combinaties van factorwaarden. Elke observatie in de materiaalverzameling wordt immers gekenmerkt door een bepaalde factorwaardencombinatie. Op basis van de coëfficiënten in het regressiemodel kunnen we voor elke logische combinatie van factorwaarden de door het model voorspelde kans op een onverbogen adjectief berekenen. Deze kans noemen we in navolging van Rietveld & Van Hout (1993) de  $p'$ -waarde van een factorwaardencombinatie. Dit is een score variërend tussen 0 en 1. Het scharnierpunt bedraagt 0.5: bij een  $p'$ -waarde kleiner dan of gelijk aan 0.5 voorspelt het model een verbogen adjectief; bij een  $p'$ -waarde groter dan 0.5 voorspelt het model een onverbogen adjectief.

De analyse bestaat uit twee fases. In de eerste fase berekenen we de  $p'$ -waarde voor elke logische combinatie van factorwaarden. Daarna vergelijken we de voorspelde buigingsvariant met het effectieve buigingsgedrag in de materiaalverzameling. Deze demarche komt neer op het evalueren van de voorspellende kracht van het regressiemodel (zie ook paragraaf 14.3.1): voor elke observatie in de materiaalverzameling berekenen we welke buigingsvariant het regressiemodel voorspelt en vervolgens kijken we welke buigingsvariant effectief in de materiaalverzameling, c.q. in het taalgebruik, voorkomt. Samengevat: op basis van het logistische-regressiemodel voor het taalgebruik in het CGN berekenen we voor iedere observatie in de materiaalverzameling de kans op een onverbogen adjectief. Door deze door het model voorspelde buigingsvariant te vergelijken met de effectieve buigingsvorm in de materiaalverzameling kunnen we nagaan hoe accuraat

het regressiemodel het gebruik van de buigingsmorfemen voorspelt en welke factorwaardencombinaties eventueel voor verbetering vatbaar zijn.

*De deflectiegraad van de verschillende factorwaardencombinaties*

Alvorens de berekeningswijze van de  $p'$ -waarden voor de factorwaardencombinaties toe te lichten, vermelden we dat we enkel de primaire en secundaire factoren geselecteerd hebben. De reden hiervoor is praktisch: de acht geselecteerde factoren zijn nog steeds goed voor 15,360 logische factorwaardencombinaties.<sup>10</sup>

Voor iedere factorwaardencombinatie  $w_i$  heeft de regressievergelijking een andere invulling en dus een ander resultaat, c.q. een andere waarde voor de logit  $g(w_i)$ . Denk hierbij aan de berekening van de  $y$ -waarde in de lineaire regressies in het vorige hoofdstuk, waar de waarde van de  $x$ 'en in de regressievergelijking de uiteindelijke  $y$ -waarde bepaalt (zie paragraaf 13.1.1). We moeten echter twee belangrijke verschillpunten tussen de lineaire en de logistische regressie in herinnering brengen. Ten eerste, doordat de verklarende variabelen nominale variabelen zijn, zijn de  $x$ -waarden ditmaal dummy-variabelen met 0 en 1 als mogelijke waarden. Ten tweede, het bekomen resultaat, de  $y$ -waarde, is de logit  $g(w_i)$ , i.e. een logaritmische waarde (zie paragraaf 13.2.1: vergelijking in (4)). Deze waarde wordt door middel van de volgende formule in een  $p'$ -waarde, i.e. een kans, omgezet (Rietveld & van Hout 1993: 330-332):

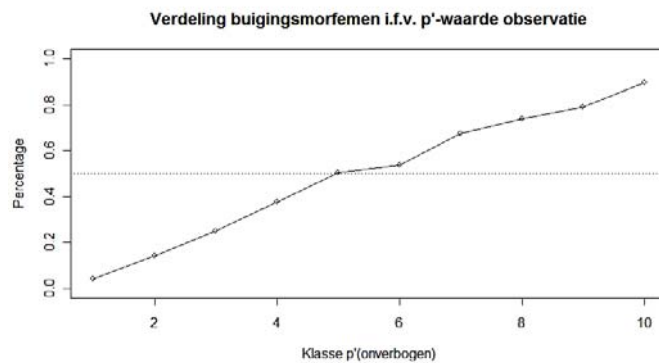
$$(2) \quad P'(\text{onverbogen} | w_i) = \frac{e^{g(w_i)}}{1 + e^{g(w_i)}}$$

De  $p'$ -waarden voor de parameterwaardencombinaties vormen een continuum van 0.0126 (c.q. een zeer kleine kans op een onverbogen adjectief) tot en met 0.9999 (c.q. een zeer grote kans op een onverbogen adjectief): bij een  $p'$ -waarde kleiner dan of gelijk aan 0.5 voorspelt het model het verbogen buigingsmorfeme; bij een  $p'$ -waarde groter dan 0.5 voorspelt het model de onverbogen pendant.

*Adequaatheid van het logistische-regressiemodel*

Figuur 14(2) visualiseert het aandeel verbogen/onverbogen vormen voor iedere  $p'$ -waarde. De X-as toont de door het model voorspelde  $p'$ -waarde van de verschillende parameterwaardencombinaties, zoals die op basis van het aangepaste logistische-regressiemodel berekend zijn. Om het overzicht te bewaren, hebben we de verschillende parameterwaardencombinaties in 10 klassen ingedeeld op basis van hun  $p'$ -waarde: klasse 1 verzamelt de parameterwaardencombinaties met een  $p'$ -waarde die tot het interval  $[0.0,0.1[$  behoort (c.q. de minst deflecerende factorwaardencombinaties volgens het regressiemodel); klasse 10 groepeerde de parameterwaardencombinaties met een  $p'$ -waarde groter dan 0.9 (c.q. de meest deflecerende parameterwaardencombinaties volgens het regressiemodel). De Y-as toont het reële percentage onverbogen vormen in de materiaalverzameling voor iedere klasse van factorwaardencombinaties die door de curve geïdentificeerd wordt. De horizontale stippellijn verdeelt de figuur in twee vlakken: de meetpunten

beneden de stippellijn identificeren de klassen met een voorkeur voor de verbogen buigingsvariant; de meetpunten boven de stippellijn wijzen op een voorkeur voor de onverbogen vorm van het adjectief.



Figuur 14(2): Evaluatie van de door het regressiemodel voorspelde buigingsvormen

Wat voor curve verwachten we nu in figuur 14(2)? Indien het model op adequate wijze de buigingsalternantie beschrijft, zoals de modelstatistieken in paragraaf 14.3.1 laten vermoeden, verwachten we een stijging van het aandeel onverbogen vormen naarmate de klasse voor de  $p'$ -waarden groter wordt.

De curve (met het aandeel onverbogen vormen) vertoont het verwachte patroon, met name een progressieve stijging naarmate de  $p'$ -waarde groter wordt. De X-as in figuur 14(2) kan in drie velden worden ingedeeld:<sup>11</sup>

- De vier klassen met de laagste  $p'$ -waarden vertonen een duidelijke voorkeur voor de verbogen buigingsvariant.
- De vier klassen met de hoogste  $p'$ -waarden vertonen een duidelijke voorkeur voor de onverbogen buigingsvariant.
- Bij de twee klassen centraal in de figuur (dit zijn: de klasse met een  $p'$ -waarde  $\in [0.4,0.5[$  en de klasse met een  $p'$ -waarde  $\in [0.5,0.6[$ ) liggen de meetpunten zeer dicht bij de horizontale stippellijn.

De grafische voorstelling van het door het logistische-regressiemodel voorspelde buigingsmorfem in figuur 14(2) illustreert de graduele aard van de buigingsalternantie: in functie van de parameterwaardencombinatie neemt de kans dat het (on)verbogen buigingsmorfem geselecteerd wordt progressief toe. Met betrekking tot de voorspellende kracht van het model toont figuur 14(2) duidelijk aan dat de zone met de meeste twijfel – en bijgevolg de grootste kans op een ‘foute voorspelling’ door het regressiemodel – zich centraal in de figuur bevindt bij de parameterwaardencombinaties met een  $p'$ -waarde die het scharnierpunt 0.5 benadert. Om kort te gaan: wanneer we van links naar rechts opschuiven in figuur 14(2) evolueren we van een duidelijke verbogen zone over een zone met twijfel naar een duidelijke onverbogen zone.

#### 14.3.4 Een apart model voor het Belgisch Nederlands?

Na de bespreking van het globale regressiemodel voor het volledige CGN in de vorige paragrafen concentreren we ons nu op de vraag of beide nationale variëteiten van het Nederlands een verschillend model vereisen. Deze vraag is ingegeven door de tegengestelde status van de buigingsvarianten in beide nationale variëteiten. Terwijl de verbogen vorm endogeen is in Nederland, is deze vorm exogeen in Vlaanderen. Bij de onverbogen vorm zien we het tegengestelde beeld: de zeer formele en ambtelijke status van deze vorm in Nederland contrasteert met de endogene en informele status in Vlaanderen (zie paragrafen 1.1.5, 2.3.1, 12.2.1 en 12.2.2).

We beantwoorden de vraag naar de noodzaak van twee afzonderlijke nationale modellen op de volgende wijze. We construeren twee aparte regressiemodellen, een eerste voor het Nederlands Nederlands en een tweede voor het Belgisch Nederlands. Vervolgens vergelijken we de impact van de verschillende factorwaarden. Hierbij letten we op de significantie van de verschillen tussen beide modellen. Na eerst twee manuele modellen te hebben geconstrueerd met alle verklarende variabelen die in tabel 14(1) worden voorgesteld, hebben we het model opnieuw berekend waarbij we alleen de variabelen hebben overgehouden die in minstens één nationale variëteit significant zijn.

We kijken eerst naar de modelstatistieken voor de logistische-regressiemodellen voor het Nederlands Nederlands en het Belgisch Nederlands:

	verklarende kracht	voorspellende kracht
<b>Nederlands Nederlands</b>	$2,160.5 - 1,365.6 = 794.86$ , $df = 20, p < 0.0000$	$c = 0.884$
<b>Belgisch Nederlands</b>	$3,169.2 - 2,346.9 = 822.3$ , $df = 20, p < 0.0000$	$c = 0.833$

Tabel 14(7): Modelstatistieken voor het regressiemodel voor het Nederlands Nederlands en het regressiemodel voor het Belgisch Nederlands

De modelstatistieken in tabel 14(7) tonen aan dat de adjectivische buigingsalternantie op basis van de behandelde factoren accurater gemodelleerd kan worden in het Nederlands Nederlands dan in het Belgisch Nederlands: zowel voor de verklarende als voor de voorspellende kracht scoort het Nederlandse model beter dan het Belgische model.<sup>12</sup> Een voor de hand liggende verklaring is de taalkloof in Vlaanderen. We hebben aangevoerd dat de taalkloof weerspiegeld wordt door de stilistische waarde van beide buigingsvarianten in het Belgisch Nederlands (zie paragraaf 12.3.2): het samen voorkomen van twee buigingsvormen met een tegengesteld statuut, een exogene vorm en een endogene vorm, leidt tot (een grotere) onzekerheid over de gebruikswaarde van beide varianten. De complexiteit van het onderliggende model is bovendien niet van die aard om de onzekerheid te verkleinen. We mogen de onzekerheid in het Belgisch Nederlands echter niet



overdrijven: de scores voor de verklarende kracht ( $p < 0.00001$ ) en de voorspellende kracht ( $c = 0.833$ ) wijzen op een zeer degelijk model. Een andere mogelijkheid die we niet zonder meer kunnen uitsluiten, is dat er voor Vlaanderen een nog niet-geïdentificeerde factor in het spel is.

Tabel 14(8) schematiseert de regressiemodellen voor het Nederlands Nederlands en het Belgisch Nederlands.<sup>13</sup> Om het overzicht te bewaren, beperken we ons tot de essentiële informatie: naast de factorwaarde rapporteren we voor elke nationale variëteit de regressiecoëfficiënt, de OR en het 95% BI voor de OR. De factorwaarden en de referentiewaarden zijn dezelfde als in tabel 14(1), met dit verschil dat de lectische variabele vereenvoudigd is tot een zuiver stilistische variabele ('inf.index'). Het meest formele register ('inf.index=0') doet telkens dienst als referentiepunt. De parameterwaarden die geen significante impact op de buigingsalternantie hebben in één van beide nationale variëteiten, zijn gecursiveerd; de parameterwaarden die een significant verschillend buigingseffect in beide nationale variëteiten definiëren, zijn vetgedrukt.

Factor	Nederlands Nederlands			Belgisch Nederlands		
	Coef.	OR	95% BI OR	Coef.	OR	95% BI OR
Intercept	-4.855			-3.641		
det.cat=aanw.vnw	-0.536	0.5849	[0.3705,0.9235]	<i>0.1092</i>	<i>1.1153</i>	<i>[0.7773,1.6003]</i>
det.cat=bez.vnw	<i>0.2676</i>	<i>1.3068</i>	<i>[0.846,2.0184]</i>	1.0515	2.862	[2.0179,4.0592]
det.cat=genitief	<i>1.1996</i>	<i>3.3187</i>	<i>[0.3183,34.596]</i>	<i>0.2519</i>	<i>1.2865</i>	<i>[0.2486,6.6565]</i>
n.inf=ja	2.0842	8.0381	[3.701,17.457]	1.0079	2.7398	[1.3301,5.6438]
n.genus=bigen	-1.213	0.2971	[0.1084,0.8144]	<i>-0.408</i>	<i>0.6647</i>	<i>[0.3841,1.1503]</i>
n.beginklank=klinker	0.5313	1.7011	[1.1546,2.5062]	0.5538	1.7398	[1.287,2.352]
a.len.pros= twee.s	1.0762	2.9335	[1.6071,5.3544]	<i>0.1634</i>	<i>1.1775</i>	<i>[0.7813,1.7747]</i>
<b>a.len.pros= twee.sw</b>	<b>1.9301</b>	<b>6.8901</b>	<b>[4.2518,11.165]</b>	<b>1.0779</b>	<b>2.9386</b>	<b>[2.11,4.0926]</b>
<b>a.len.pros= meer.s</b>	<b>1.9299</b>	<b>6.8888</b>	<b>[4.2526,11.159]</b>	<b>0.8362</b>	<b>2.3076</b>	<b>[1.695,3.1416]</b>
<b>a.len.pros= meer.sw</b>	<b>2.7412</b>	<b>15.505</b>	<b>[9.2239,26.065]</b>	<b>1.7995</b>	<b>6.0469</b>	<b>[4.2559,8.5916]</b>
a.len.pros= meer.sww	3.681	39.686	[21.982,71.646]	2.8745	17.7181	[11.693,26.845]
n.pros=w	0.6945	2.0027	[1.4911,2.6898]	0.2637	1.3018	[1.0462,1.6198]
inf.index=1	1.2676	3.5523	[1.0486,12.033]	0.7863	2.1953	[1.6441,2.9312]
inf.index=2	0.8698	2.3864	[1.7609,3.3287]	1.4208	4.1407	[3.1323,5.4738]
<b>inf.index=3</b>	<b>0.9305</b>	<b>2.5357</b>	<b>[1.7917,3.5887]</b>	<b>1.9243</b>	<b>6.8505</b>	<b>[4.9343,9.5108]</b>
lex.col=kw.1	<i>0.4025</i>	<i>1.4955</i>	<i>[0.7881,2.8377]</i>	<i>0.0909</i>	<i>1.0951</i>	<i>[0.7022,1.7078]</i>
lex.col=kw.2	0.877	2.4036	[1.4847,3.8913]	0.4146	1.5138	[1.0662,2.1491]
lex.col=kw.3	1.3944	4.0325	[2.4738,6.5733]	1.2812	3.6011	[2.566,5.0537]
<b>lex.col=kw.4</b>	<b>3.203</b>	<b>24.606</b>	<b>[14.939,40.528]</b>	<b>1.647</b>	<b>5.1914</b>	<b>[3.6913,7.3013]</b>
a.rel=ja	0.9997	2.7174	[1.9996,3.693]	1.0921	2.9805	[2.3804,3.7319]

Tabel 14(8): Regressiecoëfficiënten van de factorwaarden in het model voor het Nederlands Nederlands en het model voor het Belgisch Nederlands

We overlopen nu de factoren en factorwaarden die beide nationale variëteiten op significante wijze differentiëren. Concreet betekent dit dat we naar niet-overlappende BI'en uitkijken (cf. de vetgedrukte factorwaarden in tabel 14(8)).<sup>14</sup>

*Prosodische conditionering: adjectivische lengtegraad en prosodisch patroon  
adjectief*

Het buigingseffect van de prosodische conditionering van het adjectief is significant groter in het Nederlands Nederlands dan in het Belgisch Nederlands. Met uitzondering van de patronen 'a.len.pros=twee.s' en 'a.len.pros=meer.sw' zijn de 95% BI'en voor de OR significant groter in het Nederlandse deel van de materiaalverzameling:

- 'a.len.pros=twee.sw': 95%  $BI_{\text{NED}} = [4.2518, 11.165]$ ; 95%  $BI_{\text{BEL}} = [2.11, 4.0926]$
- 'a.len.pros=meer.s': 95%  $BI_{\text{NED}} = [4.2526, 11.159]$ ; 95%  $BI_{\text{BEL}} = [1.695, 3.1416]$
- 'a.len.pros=meer.sw': 95%  $BI_{\text{NED}} = [9.2239, 26.0651]$ ; 95%  $BI_{\text{BEL}} = [4.2559, 8.5916]$

Merk op dat de OR voor het prosodische patroon aan de linkerwoordrand van het substantief ('n.pros') niet-significant groter is in Nederland dan in Vlaanderen.

*Stilistische conditionering: informaliteitsindex*

De OR's bedragen bij twee informaliteitsindexen (c.q. 'inf.index=2' en 'inf.index=3') ongeveer het dubbele in het Belgisch Nederlands ten overstaan van het Nederlands Nederlands. Voor het meest informele register ('inf.index=3') is dit verschil significant (95%  $BI_{\text{NED}\&\text{inf.index=3}} = [1.7917, 3.5887]$ ; 95%  $BI_{\text{BEL}\&\text{inf.index=3}} = [4.9343, 9.5108]$ ). Voor het register met 'inf.index=2' is het stilistische effect tegengesteld voor beide nationale variëteiten: informeel in het Belgisch Nederlands tegenover zeer formeel en ambtelijk in het Nederlands Nederlands. Deze cijfers bevestigen de resultaten van de lectische analyse: het grotere aandeel onverbogen vormen in de informele registers weerspiegelt in het Belgisch Nederlands de taalkloof tussen het exogene Nederlandse superstraat (VRT-Nederlands) en het endogene Vlaamse substraat.

De buigingsalternantie is in het Belgisch Nederlands in ruime mate stilistisch geconditioneerd, zoals kan worden afgeleid uit de OR's.<sup>15</sup> Wat de buigingsalternantie aangaat, is Nederland stilistisch homogener dan Vlaanderen. Dit is in overeenstemming met het standaardiseringsniveau en de standaardiseringsgeschiedenis (zie ook paragraaf 12.2.1).

*Lexicale collocaties*

De lexicale collocabiliteit van het adjectief-nomenpaar vertoont het tegengestelde beeld van de stilistische variatie: ditmaal zijn de OR's in het Nederlands Nederlands groter dan hun tegenhangers in het Belgisch Nederlands, met een significant verschil

voor het hoogste collocabiliteitskwartiel (95%  $BI_{\text{NED}\&\text{lex.col}=\text{kw.4}} = [14.939,40.528]$ ; 95%  $BI_{\text{BEL}\&\text{lex.col}=\text{kw.4}} = [3.6913,7.3013]$ ).

Bij de interpretatie van deze verschillen moet worden opgemerkt dat de collocatie-indexen berekend zijn op basis van het volledige CGN. Om sluitende collocabiliteitsverschillen tussen het Belgisch Nederlands en het Nederlands Nederlands vast te stellen, moeten de collocatie-indexen voor beide nationale variëteiten afzonderlijk worden berekend. Rekening houdend met het geringe aantal AN-lemmatypes en AN-lemmatokens waarvoor in de volledige materiaalverzameling een significante collocatie-index (c.q.  $G^2$ -score) kan worden berekend (zie paragraaf 11.2.1), kan deze validatie niet op basis van de geraadpleegde 5<sup>e</sup> pre-release van het CGN worden uitgevoerd.

### *Bespreking*

De logistische-regressiemodellen die we voor de buigingsalternantie in beide nationale variëteiten hebben berekend, lopen op meerdere punten uit elkaar. Merk trouwens op dat de significante verschillen betrekking hebben op alle drie de primaire factoren:

- het prosodische patroon aan de rechterwoordrand van het adjectief gecombineerd met de adjectivische lengte
- de stilistische conditionering, c.q. de formaliteit van de gesprekssituatie
- de lexicale collocabiliteit van het adjectief-nomenpaar

Het buigingseffect van de prosodische patronen is groter in Nederland dan in Vlaanderen, zowel op het niveau van de metrische voet ('a.len.pros') als op het niveau van de prosodische constituent ('n.pros'). Voor drie contexten is dit verschil tussen de nationale variëteiten significant. Deze vaststelling suggereert dat Nederlanders gevoeliger zijn dan Vlamingen voor de constructie van een optimale prosodische vorm.

De andere factoren die significant verschillen in het Nederlands Nederlands en het Belgisch Nederlands, de stilistische en de lexicale/semantische conditionering, vertonen een analoog beeld: grotere OR's in het Belgisch Nederlands voor de stilistische conditionering en grotere OR's in het Nederlands Nederlands voor de lexicale collocabiliteit van het adjectief-nomenpaar. Deze verschillen cumuleren respectievelijk in de hoogste informaliteitsindex ('inf.index=3') en het hoogste collocabiliteitskwartiel ('lex.col=kw.4') in een significant verschil tussen beide nationale variëteiten.<sup>16</sup>

De volgende factoren hebben in geen van beide nationale variëteiten een significante invloed op de buigingsalternantie:

- de trappen van vergelijking van het adjectief
- de diminuering van het substantief
- de sonoriteit van de adjectivische eindmedeklinker
- de morfologische geleedheid van het adjectief

De eerste drie factoren hebben evenmin in het globale regressiemodel in tabel 14(2) een significant effect; de morfologische geleedheid van het adjectief definieert in het

globale model een zwak significant effect dat echter verdwijnt wanneer beide variëteiten afzonderlijk gemodelleerd worden.

We concluderen dat beide nationale variëteiten met betrekking tot de drie primaire factoren significant verschillen en dat de buigingsalternantie bij neutra in beide nationale variëteiten bijgevolg door een significant verschillend systeem gestuurd wordt. Een belangrijke oorzaak van deze significante verschillen tussen beide nationale variëteiten is de divergerende historische achtergrond van de adjectivische buiging en dito status van het onverbogen adjectief in beide nationale variëteiten. In het Belgisch Nederlands wordt het sterke stilistische effect veroorzaakt door het tegengestelde statuut van beide buigingsvarianten: het exogene verbogen adjectief is kenmerkend voor formele gespreksituaties, terwijl de endogene onverbogen vorm in informele gespreksituaties wordt geselecteerd. De significant grotere impact van de lexicale collocabiliteit van het adjectief-nomenpaar in het Nederlands Nederlands hangt – voor een deel – samen met het zeer formele en ambtelijke statuut van de onverbogen variant: tot de adjectief-nomenparen in het hoogste collocabiliteitskwartiel ('lex.col=4') in het Nederlandse deelcorpus behoren een aantal formulaische uitdrukkingen die allemaal in Tweede-Kamerdebatten voorkomen:

- (3) geregistreerd partnerschap, openbaar vervoer, landelijk pakket, voortgezet onderwijs, burgerlijk huwelijk, regionaal structuurplan, strafrechtelijk onderzoek, wetenschappelijk onderzoek.

De stilistische dimensie van het collocabiliteitsverschil wordt gestaafd door het vergelijkbare effect van de relationele adjectieven in beide nationale variëteiten (cf. de overlappende 95% BI'en in tabel 14(8)). De vergelijking van beide logistische-regressiemodellen in tabel 14(8) suggereert tot slot dat Nederlanders bij de keuze van het buigingsmorfeem gevoeliger zijn voor de realisatie van een optimale prosodische vorm.

Zonder de verschillen tussen beide modellen in tabel 14(8) te willen minimaliseren, wijzen we erop dat twee van de drie verschillen tussen het model voor het Nederlands Nederlands en het model voor het Belgisch Nederlands betrekking hebben op de grootte van het effect: bij de prosodische structuur van de adjectivische rechterwoordrand en bij de lexicale collocabiliteit van het adjectief-nomenpaar is het effect in het Nederlands Nederlands significant groter, zonder dat de zin van het effect omkeert. Dit ligt anders bij de stilistische conditionering: in de eerste plaats hebben de registers met informaliteitsindex 2 een tegengestelde stilistische betekenis in beide nationale variëteiten: informeel in het Belgisch Nederlands en zeer formeel en ambtelijk in het Nederlands Nederlands. Daarnaast vertonen de registers met de hoogste informaliteitsindex ('inf.index=3'), die in beide nationale variëteiten informeel zijn, een significant groter deflecterend effect in het Belgisch Nederlands.

Deze verschillen kunnen we tot slot beschouwen als een eerste, zeer voorzichtige indicatie ter bevestiging van Goossens' (2000) beduchtheid dat Vlaanderen door het gebruik van de tussentaal een nieuwe contextuele tweetaligheid ontwikkelt, te meer omdat het register één van de factoren is die significant verschillen. De falsificatie van Goossens' hypothese vormt echter het onderwerp van een afzonderlijk onderzoek naar de grammaticale kenmerken van de tussentaal en hoe deze kenmerken zich verhouden tot het VRT-Nederlands.

#### 14.4 Algemene bespreking

We ronden dit hoofdstuk af met een algemene bespreking van de adjectivische buigingsalternantie bij neutra in de gesproken standaardtaal. Eerst hernemen we de belangrijkste resultaten van de regressieanalyse. Vervolgens confronteren we deze resultaten met de algemene attitudes ten aanzien van de adjectivische buiging in het Nederlands en de traditionele syntactische visie op congruentie. Deze confrontatie brengt ons bij de slotvraag: waarom vertoont het attributieve adjectief in het Nederlands buigingsvariatie?

##### *Globaal model voor de adjectivische buigingsalternantie bij neutra*

Het onderzoek naar de adjectivische buigingsalternantie in het CGN heeft aan de ene kant aangetoond dat de selectie van het adjectivische buigingsmorfem bij neutra niet syntactisch voorspelbaar en redundant is: het aandeel gemarkeerde onverbogen vormen is namelijk te groot om de alternantie als een randverschijnsel te negeren ( $\frac{\text{verbogen}=76.75\%}{\text{onverbogen}=23.25\%}$ ; zie paragraaf 6.2). Aan de andere kant suggereert de logistische-regressieanalyse dat deze buigingsalternantie wel degelijk gemodelleerd kan worden: de keuze van het buigingsmorfem wordt bepaald door een netwerk van diverse factoren die met elkaar verstrengeld zijn. De modelleerbaarheid van de buigingsalternantie bij neutra blijkt overduidelijk uit de algemene modelstatistieken (zie ook paragraaf 14.3.1):

- verklarende kracht: De onverklaarde variatie wordt op significante wijze gereduceerd ( $p_{\text{model}} < 0.00001$ ).
- voorspellende kracht: Het resultaat voor de  $c$ -maat ( $c = 0.855$ ) bewijst dat het logistische-regressiemodel het gebruik van de buigingsvarianten op adequate wijze kan voorspellen. Bovendien is de voorspellende kracht het kleinst bij de factorwaardencombinaties die in het reële taalgebruik de meeste twijfel vertonen ten aanzien van de selectie van het adjectivische buigingsmorfem (zie paragraaf 14.3.3, figuur 14(2)).

Wanneer we de verschillende verklarende factoren beschouwen, stellen we vast dat een aantal factoren in het geheel van het model geen significante impact op de buigingsalternantie hebben:

- de trappen van vergelijking van het adjectief (met uitzondering van de superlatief die altijd verbogen is, zie paragraaf 6.2)

- de diminuering van het substantief
- de sonoriteit van de adjectivische eindmedeklinker

Twee factoren hebben een zwakke maar significante impact op de buigingsalternantie:

- het genus/de bigenericiteit van het substantief
- de morfologische geleedheid van het adjectief

De zwakte van deze factoren wordt bevestigd door het verdwijnen van de significantie in (minstens één van) de nationaal gedifferentieerde regressiemodellen (zie paragraaf 14.3.4, tabel 14(8)). Rekening houdend met de relatief grote  $p$ -waarden van deze factoren ( $0.05 \geq p > 0.001$ ) in vergelijking met de andere  $p$ -waarden in dit onderzoek ( $p \leq 0.001$ ), besluiten we deze factoren onder voorbehoud op te nemen.

De volgende factoren spelen een significante en prominente rol bij de keuze van de adjectivische buigingsvariant:

- het prosodische patroon aan de rechterrاند van het adjectief gecombineerd met de adjectivische lengte
- het prosodische patroon aan de linkerrاند van het substantief
- de lexicale collocabiliteit van het adjectief-nomenpaar
- het onderscheid tussen relationele/kwalificerende adjectieven
- de lectische kenmerken: de nationale variëteit gecombineerd met de (in)formaliteit van de gesprekssituatie
- de beginklank van het substantief
- het voorkomen van een genominalseerde infinitief in de nominale kernpositie
- de woordsoort van de determinator

Op basis van de OR's (zie tabel 14(2)) en de bijdrages tot de reductie van de onverklaarde variatie (zie figuur 14(1)) identificeren we drie primaire factoren:

- het prosodische patroon aan de rechterwoordrand van het adjectief gecombineerd met de adjectivische lengte
- de lexicale collocabiliteit van het adjectief-nomenpaar
- de lectische kenmerken, i.e. de nationale variëteit gecombineerd met het register

De eerste twee factoren zijn bovendien aan een secundaire factor gerelateerd, respectievelijk het prosodische patroon aan de linkerwoordrand van het substantief en de relationaliteit van het adjectief.

#### *Attitudes ten aanzien van de adjectivische buigingsalternantie bij neutra*

Laten we nu de verschillende attitudes die in de neerlandistische literatuur worden aangenomen ten overstaan van de adjectivische buigingsalternantie (zie paragraaf 1.2.2) met deze resultaten confronteren:

- *marginaliserende attitude*: De marginaliserende attitude, die de buigingsalternantie als een dialectisch of idiolectisch (rand)fenomeen beschouwt (De Schutter 1994, 1997; Kester 1996; Klooster 2001: 82-83;

Odiijk 1992; Rijpma & Schuringa 1978: 115-116; van den Toorn 1982: 165; Van Eynde 2003), wordt door de empirische gegevens ontkracht. Zoals we hierboven al vermeld hebben, is het gemarkeerde onverbogen adjectief te frequent in het reële taalgebruik, zowel in het Belgisch Nederlands als in het Nederlands Nederlands, om als een epifenomeen te worden afgedaan.

- *reducerende attitude*: De reducerende attitude herleidt de buigingsalternantie tot één of twee factoren, met name de ritmische of eufonische kenmerken van het adjectief en de semantische kenmerken van het adjectief-nomenpaar (Blom 1994, 1995; Booij 2002a, 2002c; Booij & van Santen 1998; Cockx 2000; de Haas & Trommelen 1993; Honselaar 1980; Klein & Visscher 1996; Penninckx & Buyse 1997; Permentier & Van Den Eynden 1998). Hoewel dit twee primaire factoren zijn, kunnen we deze attitude bezwaarlijk toereikend noemen. Ten eerste, verscheidene andere factoren – in de eerste plaats de lectische factoren – hebben eveneens een significante impact op de keuze van het adjectivische buigingsmorfeem. Wanneer we een regressiemodel construeren dat alleen de prosodische conditionering van het adjectief ('a.len.pros') en de lexicale collocabiliteit ('lex.col') omvat, vermindert zowel de verklarende ( $5,383.4 - 4,252.6 = 1,130.79$ ,  $df = 9$ ,  $p < 0.00001$ ) als de voorspellende kracht ( $c = 0.797$ ). Ten tweede, de prosodische en de lexicale/semantische conditionering van de buigingsalternantie die we hebben vastgesteld, is complexer dan doorgaans in de literatuur wordt aangenomen (zie paragrafen 10.1.1 en 11.1.1).<sup>17</sup>
- *globaliserende attitude*: De globaliserende attitude erkent de complexiteit van de buigingsalternantie maar slaagt er – bij ontstentenis van de gepaste methode en/of techniek – niet in een globaal model te construeren, waardoor sommige auteurs het buigingsgedrag als aleatoir bestempelen (ANS 1997; Degryse 1959; Lebrun & Schurmans-Swillen 1966; Raidt 1968; de Rooij 1980a, 1980b; Royen 1953). Op basis van het logistische-regressiemodel kunnen we concluderen dat de buigingsalternantie, hoewel zeer complex, niet aleatoir is. Bovendien is de impact van de diverse factoren niet gelijk, zodat we een rangorde aanbrenge tussen primaire, secundaire en tertiaire factoren enerzijds en niet-significante factoren anderzijds.

Het is opmerkelijk dat de lectische conditionering van de buigingsalternantie, dialectologisch onderzoek buiten beschouwing gelaten, nauwelijks aan bod komt in de literatuur. Bestaand empirisch onderzoek is tot één stilistisch stratum beperkt (Degryse 1959; Lebrun & Schurmans-Swillen 1966) of identificeert geen significant stilistisch verschil (de Rooij 1980a). De ANS (1997) verbindt de onverbogen variant zowel in het Nederlands Nederlands als in het Belgisch Nederlands met zeer formele gespreksituaties. Onderhavig onderzoek toont aan dat zowel de regionale als de stilistische dimensie op significante wijze de buigingsalternantie conditioneren: ten eerste, het gecombineerde effect van de regionale en stilistische kenmerken is een

primaire factor in het globale regressiemodel (zie paragrafen 14.2 en 14.3.2); ten tweede, het regressiemodel voor het Belgisch Nederlands verschilt op significante wijze van het regressiemodel voor het Nederlands Nederlands (zie paragraaf 14.3.4); ten derde, in tegenstelling tot de stellingname in de ANS (1997) is de stilistische waarde van het onverbogen buigingsmorfem tegengesteld in de gesproken taal in Nederland en Vlaanderen.

Kortom: de buigingsalternantie die optreedt bij neutra is enerzijds een wezenlijk kenmerk van het taalgebruik, maar is anderzijds wel degelijk modelleerbaar. Deze vaststelling impliceert dat theoretische modellen voor de adjectivische buiging de buigingsvariatie niet langer kunnen negeren. Deze demarche resulteert niet alleen in een adequatere beschrijving van de talige realiteit, maar opent daarnaast een nieuw spectrum aan niet-redundante functies voor de adjectivische buigingsmorfemen.

#### *Theoretische modellen voor congruentie*

De confrontatie met de resultaten van het empirisch onderzoek toont aan dat de traditionele visie op congruentie niet opgaat voor de adjectivische buiging(s-alternantie) bij neutra. De uiteindelijke buigingsvorm is noch syntactisch voorspelbaar noch redundant. Dit impliceert echter geen verwerping, maar een nuancering van de traditionele, syntactische visie op congruentie (zie paragraaf 4.2). Het basissysteem van de buiging van het attributieve adjectief is syntactisch: de kenmerken aanwezig in de syntactische context definiëren verschillende buigingscontexten (zie paragraaf 3.1, figuur 3(1)). In deze buigingscontexten komt naast het syntactisch voorspelde buigingsmorfem echter ook een gemarkeerde variant voor die bij neutra gemotiveerd is door de niet-syntactische factoren die in het tweede deel van dit proefschrift zijn voorgesteld.

De multivariate analyses in dit hoofdstuk tonen aan dat de keuze van de onverbogen vorm voor een aanzienlijk deel lexicaal/semantisch ('lex.col' en 'a.rel') en lectisch ('reg.inf.index') gemotiveerd is. Deze vaststelling sluit aan bij de discursieve congruentiemodellen (zie paragraaf 4.2.2). Volgens deze modellen drukken de buigingsmorfemen conceptuele informatie uit over een referent in het discours, over een gesprekspartner of over de gesprekscontext (Barlow 1991, 1999; Corbett 1979; Kiefer 1998; Pollard & Sag 1994). De adjectivische buigingsalternantie bij neutra is echter complexer: naast de lexicale/semantische en lectische factoren zijn er ook andere factoren – in de eerste plaats de prosodische structuur van het adjectief – die de keuze van het buigingsmorfem beïnvloeden en die niet duidelijk in de discursieve congruentiemodellen passen.<sup>18</sup>

Congruentie vertoont bijgevolg, net als alle andere taalkundige fenomenen, variatie waarbij het gebruik van de verschillende vormelijke varianten niet vrij maar gemotiveerd is. Dit weerlegt Haimans (1985) visie dat congruentie een overwinning van de indexicale op de iconische functie van taal is: het gemarkeerde onverbogen buigingsmorfem realiseert niet alleen prosodisch optimale patronen, maar wordt daarnaast aangewend om op iconische wijze gemarkeerde denotationele en lectische betekenisaspecten uit te drukken.<sup>19</sup> In het vervolg van deze paragraaf betogen we dat



het voorkomen van variatie bij congruentie bijna automatisch voortvloeit uit de (schijnbare) redundantie van deze elementen.

*Motivatie voor de adjectivische buigingsvariatie*

De erkenning van een systematische buigingsalternantie brengt ons bij de laatste vraag: waarom vertoont het attributieve adjectief buigingsvariatie? Op basis van Barlows (1991, 1999) discursieve congruentiemodel kunnen we antwoorden dat de alternantie tussen het verbogen adjectief en het onverbogen adjectief een conceptuele functie vervult, waarbij het onderscheid tussen beide vormen zowel op de denotationele als op de lectische betekenisaspecten betrekking heeft. Dit volstaat echter niet om de specifieke situatie van de adjectivische buigingsalternantie bij neutra te duiden. Er zitten immers verscheidene factoren in het globale logistische-regressiemodel in tabel 14(2) waarvoor Barlows congruentiemodel geen verklaring aanreikt, waardoor het aangepaste regressiemodel op basis van Barlows congruentiemodel een significant kleinere reductie van de onverklaarde variatie realiseert dan het globale regressiemodel in tabel 14(2) (cf. *supra* en noot 18). Om de vraag naar de motivatie van de adjectivische buigingsvariatie te beantwoorden, concentreren we ons bijgevolg op de specifieke situatie van de adjectivische buiging in het Nederlands.

Bij de bespreking van het basissysteem voor de verbuiging van het attributieve adjectief is de complexiteit van het onderliggende mechanisme opgevallen (zie paragraaf 3.1). De keuze tussen de buigings-*e* en het nulmorfeem is het resultaat van drie morfosyntactische factoren: het genus van het nominale hoofd, het getal van het nominale hoofd en de definitie van de NP, c.q. van de determinator. Hieraan moeten we toevoegen dat de buiging van het attributieve adjectief in het Nederlands niet langer gemotiveerd is (van Bree 1980; De Groot 1949; van Haeringen 1937), met uitzondering van de Vlaamse dialecten, waar de adnominale buiging een rol speelt in het drie-generasysteem (zie paragraaf 2.3.1). De voor de hand liggende hypothese, die trouwens ook door bovenvermelde auteurs geopperd wordt, is dat de adjectivische buiging uit het taalsysteem verdwijnt.

Deze stelling past in de functioneel gemotiveerde “hierarchy of [inflectional; JT] categories according to their relative strength” van Kiparsky (1982a), waarbij de sterkte van een inflectionele categorie moet worden verstaan als haar weerstand tegen klankwettige verdwijning. Zwakke inflectionele categorieën, i.e. categorieën die het makkelijkst fonetische uitholling ondergaan, “are those which register information that is relatively redundant” (Kiparsky 1982a: 99), zoals grammaticale casusmarkering en congruentie “[that] are specified by transformations on the basis of the tree configuration at a late stage in the derivation of a sentence”. Elders verwoordt Kiparsky (1982b: 67) de tendentie om semantisch gemotiveerde inflectionele categorieën te behouden op basis van dit principe:

“Morphological material which is predictable on the surface tends to be more susceptible to loss than morphological material which is not predictable on the surface”

Volgens Kiparsky's criteria is de adjectivische buiging in het Nederlands een zwakke categorie die onderhevig is aan verdwijning: de adjectivische buiging is niet alleen functioneel redundant maar bovendien is ze ook semantisch intransparant.

Deze argumentatie wordt bevestigd door het vergelijkende onderzoek uitgevoerd door Van Marle & Smits (1993) naar de evolutie van drie inflectionele categorieën in het overzeese Nederlands. De bestudeerde inflectionele categorieën zijn: de nominale meervoudsvorming, de werkwoordelijke vervoeging (tijd en persoon) en de verbuiging van het attributieve adjectief. Anders gezegd: een instantie van inherente inflectie (nominale meervoudsvorming), een instantie van contextuele inflectie (adjectivische buiging) en een mengvorm die beide inflectionele types combineert (werkwoordelijke vervoeging). De volgende variëteiten van het overzeese Nederlands zijn in het onderzoek opgenomen: het 17<sup>e</sup> eeuwse New Yorkse Nederlands, het Amerikaanse Nederlands en het Afrikaans. Van Marle & Smits (1993) bespreken de evoluties van de drie inflectionele categorieën aan de hand van twee criteria, met name de homogeniteit van de evoluties in de verschillende overzeese variëteiten en de stabiliteit van het systeem dat aan de basis van deze categorieën ligt. De homogeniteit van een categorie wordt bepaald door de gelijkenisgraad tussen de evoluties in de verschillende overzeese variëteiten. De stabiliteit van een inflectionele categorie definiëren we als de afstand tussen het continentale Nederlands en de variëteiten van het overzeese Nederlands. We situeren de door Van Marle & Smits bestudeerde inflectionele categorieën op een schaal die beide criteria combineert.

	Inherente inflectie	Contextuele inflectie
<b>Homogeniteit</b>	+	-
	N	V
<b>Stabiliteit</b>	+	-

Tabel 14(9): Homogeniteit en stabiliteit van de evoluties van de nominale meervoudsvorming, de werkwoordelijke vervoeging en de adjectivische buiging in het overzeese Nederlands

Tabel 14(9) toont aan dat er een relatie is tussen de homogeniteit en de stabiliteit van een inflectionele categorie enerzijds en de functionele redundantie en de semantische transparantie anderzijds. De nominale meervoudsvorming bezet de meest stabiele en homogene pool van het continuum. Deze categorie kent in de verschillende variëteiten van het overzeese Nederlands slechts marginale aanpassingen, zoals de extensie van één van beide meervoudsmorfemen (-en in het Amerikaanse Nederlands en -s in het Afrikaans). Aan de andere pool van het continuum in tabel 14(9) vinden we de adjectivische buiging, die het meest verwijderd is van het continentale systeem: de adjectivische buiging als zodanig is in

alle bestudeerde variëteiten verdwenen (zie ook Van Marle 1995a). Bovendien zijn er sterke verschillen tussen de evoluties in de verschillende variëteiten: ofwel wordt één van de twee buigingsvarianten geregulariseerd (de verbogen variant in het Amerikaanse Nederlands en de onverbogen variant in het 17<sup>e</sup> eeuwse New Yorkse Nederlands), ofwel wordt het systeem geherconditioneerd, zoals in het Afrikaans.<sup>20</sup> De verbale inflectie neemt een middenpositie in: aan de ene kant heeft de werkwoordelijke vervoeging verschillende veranderingen ondergaan in het overzeese Nederlands; aan de andere kant vertonen de evoluties in de verschillende overzeese variëteiten sterke gelijkenissen. De werkwoordelijke vervoeging kent meer bepaald een algemene reductie van de persoons- en tijdsdistinctie. Van Marle & Smits (1993) concluderen dat de kans dat een inflectionele categorie uitgehold wordt, samenhangt met de functionele motivatie en de semantische transparantie: de sterkst gemotiveerde categorie, i.e. de nominale meervoudsvorming, vertoont de grootste stabiliteit; de zwakst gemotiveerde categorie, i.e. de adjectivische buiging, is het minst stabiel. Dit verklaart waarom de adjectivische buiging als congruentiefenomeen in het overzeese Nederlands verdwenen is. In het Nederlands Nederlands dat in de migrantengemeenschap wordt gesproken, merkt Weerman (2003) de tendentie op om net als in sommige variëteiten van het overzeese Nederlands de verbogen vorm te generaliseren en bijgevolg het adjectief bij een onzijdig substantief in een indefiniëte NP te verbuigen, zoals *een mooie paard*.

De analyse van de evolutie van de adjectivische buiging in de overzeese variëteiten van het Nederlands bevestigt de functionele redundantie van de adjectivische buiging als een historisch relict. Het mag dan ook geen verbazing wekken dat er nieuwe opposities ontstaan wanneer dit abstracte en functieloze systeem behouden blijft, zoals dit het geval is in het continentale Nederlands, en dat taalgebruikers aan deze nieuwe vormelijke variatie functioneel gemotiveerde distincties relateren. Het is trouwens opmerkelijk dat een aantal van de factoren die we empirisch bestudeerd hebben ook in het historisch overzicht aan bod zijn gekomen: de lectische conditionering (nationale variëteit en tekstgenre), de prosodische conditionering (adjectieven op *-ig* en *-lijk*), de morfosyntactische conditionering (c.q. het gebruik van het onverbogen adjectief na een bezittelijk voor-naamwoord) en de lexicale/semantische conditionering (zij het bij *de*-woorden) beïnvloeden al sedert het Middel-Nederlands de keuze van de adjectivische buigingsvariant bij de neutra in de nominatief en de accusatief (zie hoofdstuk 2).

In toekomstig onderzoek moet het perspectief uiteraard verruimd worden van de buigingsalternantie bij neutra tot het globale model voor de buiging van het attributieve adjectief. De volgende vragen kunnen eventueel voor een eerste aanzet zorgen:

- Hoe verhoudt de buigingsalternantie bij neutra zich tot de buigingsalternantie bij *de*-woorden? Wordt de alternantie tussen de ‘regelmatige’ verbogen vorm en de gemarkeerde onverbogen vorm bij *de*-woorden en *het*-woorden door dezelfde factoren bepaald of is het gebruik van de

onverbogen vorm beperkt tot de iconische functie in gelexicaliseerde adjectief-nomenparen?

- Hoe evolueert de adjectivische buiging in beide nationale variëteiten van het Nederlands? Of nog: groeien Nederland en Vlaanderen verder uit elkaar?
- Waarom is de adjectivische buiging als morfosyntactisch mechanisme in het overzeese Nederlands verdwenen en niet in het continentale Nederlands, hoewel de keuze van het buigingsmorfeem sinds het einde van het Middel-Nederlands niet langer (duidelijk) gemotiveerd is?

## 14.5 Conclusie

We sluiten dit hoofdstuk af met een bondige samenvatting van de belangrijkste inhoudelijke en methodologische resultaten van het empirisch onderzoek op basis van het CGN.

Inhoudelijk hebben we aangetoond dat de adjectivische buigingsalternantie bij neutra geen aleatoir en onregelbaar fenomeen is dat uitsluitend twijfel en verwarring zaait bij de taalgebruiker. Het empirisch onderzoek heeft aangetoond dat de buigingsalternantie op een onderliggende regelmaat stoelt die echter op twee manieren afwijkt van de traditionele taalkundige benaderingen. Ten eerste, de waargenomen regelmatigheid is niet discreet maar gradueel en stochastisch. Ten tweede, de buigingsalternantie is niet reduceerbaar tot één of twee factoren, maar wordt bepaald door een complex netwerk van concurrerende en interagerende factoren die tot verschillende taalkundige domeinen behoren. In tabel 14(10) vatten we dit netwerk schematisch samen. De eerste kolom ('Type') identificeert het grammaticale domein waartoe de factor in kwestie behoort. De tweede kolom ('Rang') duidt de rang van de factor aan op basis van het logistische-regressiemodel: '1' staat voor primaire factor, '2' voor secundaire factor en '3' voor tertiaire factor (de niet-significante factoren zijn niet in het schema opgenomen). De laatste drie kolommen identificeren de verschillende significante factoren en de factorwaarden die het gebruik van beide buigingsvarianten faciliteren. De flecterende en deflecterende factorwaarden moeten uiteraard binnen de context van de parameter in kwestie geïnterpreteerd worden: voor iedere factor scheiden we de waarden die het gebruik van het verbogen adjectief vergemakkelijken en de waarden die de keuze van de onverbogen tegenhanger motiveren.

Bij de interpretatie van tabel 14(10) moeten we rekening houden met een bijkomende complicerende factor. Hoewel de meeste verklarende factoren een vergelijkbaar effect op de buigingsalternantie hebben in het Belgisch Nederlands en het Nederlands Nederlands, vertonen beide nationale variëteiten drie significante verschillen:

- het buigingseffect van de prosodische patronen is significant groter in het Nederlandse taalgebruik dan in het Vlaamse taalgebruik

14. MULTIVARIATE ANALYSE VAN DE ADJECTIVISCHE BUIGINGSALTERNANTIE BIJ NEUTRA

- het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar is significant groter in het Nederlands Nederlands
- de stilistische waarde van de onverbogen buigingsvariant is tegengesteld in beide nationale variëteiten

We onderscheiden bijgevolg één regionale variabele ('regio') en twee stilistische variabelen, een eerste voor het Nederlands Nederlands ('stijl<sub>NED</sub>') en een tweede voor het Belgisch Nederlands ('stijl<sub>BEL</sub>').

Type	Rang	Factorwaarde verbogen vorm	Factor	Factorwaarde onverbogen vorm
morf. synt	2	het, aanw.vnw, genitief	woordsoort determinator	bez.vnw
	2	substantief	woordsoort N	genominaliseerde inf.
	3	bigenerisch	genus N	het-woord
fonologisch	2	medeklinker	beginklank N	klinker
	1	'een.s', 'twee.s'	prosodische structuur A	'twee.sw', 'meer.s', 'meer.sw', 'meer.sww'
	2	's'	prosodische structuur N	'w'
morf.	3	niet geled	morf. geledheid A	geled
lexicaal / semantisch	1	geen lexicale collocatie, zeer zwakke lexicale collocatie	lexicale collocabiliteit AN-paar	lexicale collocatie
	2	kwalificerend A	semantische categorie A	relationeel A
lectisch	1	Nederlands Nederlands	regio	Belgisch Nederlands
	1	ongemarkeerd	stijl <sub>NED</sub>	zeer formeel, ambtelijk
	1	formeel, ongemarkeerd	stijl <sub>BEL</sub>	informeel

Tabel 14(10): Samenvattende tabel met de verklarende factoren die een significant effect op de buigingsalternantie in het CGN definiëren

Zoals het cijfer voor de rangorde aangeeft, zijn er drie primaire factoren (gesteld dat we de drie lectische factoren als één complexe factor beschouwen; zie paragraaf 14.1, tabel 14(1)). Deze primaire factoren onderscheiden zich van de overige factoren door een grotere bijdrage tot de reductie van de onverklaarde variatie én een grotere waarde voor de regressiecoëfficiënten, c.q. een grotere waarde voor de OR's. Merk op dat het zeer moeilijk – zo niet onmogelijk – is om de primaire factoren – laat staan alle verklarende parameters in tabel 14(10) – in één schematische conditie te bundelen. We kunnen evenwel enkele relaties tussen verklarende factoren identificeren. Voor Nederland is er een relatie tussen de lectische en de lexicale/semantische conditionering: de zeer formele en ambtelijke status van het onverbogen buigingsmorfeem in het Nederlands Nederlands is nauw verbonden met het formulaïsche taalgebruik in deze registers. Daarnaast zijn twee primaire factoren nauw verbonden met een secundaire factor. In de eerste plaats versterkt het prosodische patroon aan de linkerwoordrand van het substantief het

effect aan de rechterwoordrand van het adjectief. In de tweede plaats wordt de adjectivische positie in lexicale collocaties significant vaker dan verwacht door een relationeel adjectief bezet. We herinneren de lezer eraan dat de stratumanalyse geen significante partiële conditionering tussen beide lexicale/semantische factoren heeft vastgesteld (zie paragraaf 11.2.2).

De confrontatie van het logistische-regressiemodel met de neerlandistische literatuur leert ons dat geen enkele attitude ten overstaan van de buigingsalternantie conform is met het voorgestelde empirische model. Daarnaast hebben we gesuggereerd dat het voorkomen van buigingsvariatie normaal is in het (continentale) Nederlands, waar – in tegenstelling tot het overzeese Nederlands – het abstracte en functieloze congruentiesysteem niet verdwenen is: in een abstract en functieloos systeem ontstaan – al dan niet toevallig – nieuwe vormelijke opposities waaraan nieuwe functionele distincties verbonden worden.

Na de inhoudelijke samenvatting richten we nu onze blik op de methodologische aspecten van het onderzoek. We hopen de sceptische lezer overtuigd te hebben dat multivariaat onderzoek onontbeerlijk is bij corpuslinguïstisch variatieonderzoek. Ten eerste, door de constructie van een logistische-regressiemodel kan de verklarende en de voorspellende kracht gemeten worden, zodat de adequaatheid van het model én van het gevoerde onderzoek geëvalueerd kunnen worden. Ten tweede, de regressiecoëfficiënten – of de OR's zo men wil – worden berekend terwijl het effect van de andere factoren gecontroleerd wordt. Dit is de enige mogelijkheid om het werkelijke effect van de verschillende factor(waard)en in het taalgebruik te berekenen. De voorkomens van Simpsons paradox vormen bijkomende evidentie om bivariaat onderzoek aan te vullen met (verschillende types van) multivariaat onderzoek. Op basis van de resultaten van een correspondentieanalyse hebben we aangevoerd dat de tegengestelde waarden van bepaalde factoren in de bivariate en de multivariate analyses voortvloeien uit de ongecontroleerde samenstelling van een corpus: doordat bepaalde factorwaarden sterk gerelateerd zijn aan flecterende/deflecterende factorwaarden is hun reële deflecterende/flecterende effect verhoud in de bivariate analyse. Ten derde, de resultaten van de logistische-regressieanalyse kunnen gebruikt worden om een rangorde tussen de verschillende factoren aan te brengen: we hebben de variatiereducerende kracht van de factoren in de *forward stepwise*-regressie verbonden met de OR's van de verschillende factor(waard)en om primaire, secundaire, tertiaire en niet-significante factoren te onderscheiden.

Het verkregen regressiemodel kan voortaan als vertrekpunt dienen voor het falsificeren van specifieke onderzoekshypothesen in strikt gecontroleerde onderzoekscontexten. We denken in de eerste plaats aan enquêtes en psycholinguïstische experimenten die de informatie in het regressiemodel kunnen aanwenden bij de selectie van de testitems.

Met de logistische-regressieanalyse in dit hoofdstuk sluiten we de basisanalyse naar de adjectivische buigingsalternantie bij neutra af. Deze basisanalyse heeft de vorm van een tweeluik: in deel II hebben we de verschillende (potentiële) verklarende factoren geoperationaliseerd en hun (afzonderlijke) effect op de buigingsalternantie geïnterpreteerd; in deel III hebben we op basis van deze factoren een multivariaat model geconstrueerd, dat op significante wijze de onverklaarde buigingsvariatie reduceert en op een zeer adequate manier de selectie van het buigingsmorfeem beschrijft en verklaart.

In het vierde en laatste deel van dit proefschrift concentreren we ons op één specifieke factor, met name de lexicale collocabiliteit van het adjectief-nomenpaar. We stellen meer bepaald de vraag of het buigingseffect van de lexicale collocaties een louter lexicale tendens is, dan wel of er semantische patronen aan de basis van het deflecterende effect van de lexicale collocaties geponoerd kunnen worden. Tot nu toe hebben we deze vraag nog niet gesteld en hebben we het dan ook over de *lexicale/semantische* conditionering van de buigingsalternantie gehad.

De motivatie voor dit afzonderlijke onderzoek is dat het CGN zowel vanuit een kwantitatief als vanuit een kwalitatief perspectief minder geschikt is om deze vraag te beantwoorden: kwantitatief kampen we met dataschaarste (zie ook paragraaf 11.2.1); kwalitatief verzamelt het CGN thematisch en lectisch zeer gediversifieerde gesprekken en gesprekssituaties. We hopen deze problemen op te lossen door een nieuw corpus te raadplegen, met name de edities van (januari tot en met september van) de jaargang 2003 van de Vlaamse kwaliteitskrant *De Standaard*.

## Noten

1. Een alternatieve oplossing is het opnemen van een interactiefactor in het regressiemodel, c.q.  $a.len * a.pros$  of  $a.len * an.pros$ . Deze optie wijzen we expliciet van de hand omdat ze de interpretatie van het regressiemodel aanzienlijk bemoeilijkt. De klemtoonbotsing is de enige informatie die niet direct aanwezig is in de huidige codering (voor de interpretatie verwijzen we naar paragraaf 14.3.2).
2. We spreken van een perfecte scheiding wanneer iedere waarde van een parameter consistent één waarde van de responsvariabele selecteert. Deze absolute voorkeur wordt vertaald in een zeer grote waarde voor de regressiecoëfficiënt en een zeer grote standaardfout, en bijgevolg een zeer ruim 95% BI voor de OR.
3. We houden eraan Hans Smessaert te bedanken voor zijn zeer waardevolle suggesties bij de analyse van de morfologische structuur van de adjectieven in onze materiaalverzameling.
4. Het voorgestelde model kan verbeterd worden door het verwijderen van de *influential observations*. Dit zijn observaties die ver van de andere observaties verwijderd zijn en op deze wijze de regressiecurve vervormen. Voor het globale regressiemodel telt de materiaalverzameling 25 *influential observations*. De afwijking van de curve wordt veroorzaakt door infrequente factorwaarden, namelijk ‘det.cat=genitief’, ‘n.inf=ja’, ‘n.genus=bigen’ en ‘reg.inf.index=nl.1’. De verwijdering van de *influential observations* uit de materiaalverzameling leidt tot een verbetering van de voorspellende kracht ( $c = 0.8600$ ) en tot een stijging van de verklaarde variatie ( $5,334.3 - 3,691.5 = 1,642.81$ ,  $df = 25$ ,  $p < 0.0000$ ). Het *Design*-pakket in R dat we voor de logistische-regressieanalyse gebruiken, voorziet speciale functies voor de identificatie van de *influential observations* (Harrell 2001).
5. Alvorens we de regressiecoëfficiënten interpreteren, testen we of er sprake is van multicollineariteit tussen de verklarende factoren. Hiervoor gebruiken we de *Variance Inflation Factor* (VIF), een diagnostische maat die in het *Design*-pakket in R voorzien is (Harrell 2001). In de statistiek spreekt men van multicollineariteit wanneer twee of meerdere verklarende variabelen sterk correleren. Het effect van deze variabelen op de responsvariabele wordt dan arbitrair omdat ze vrijwel dezelfde variatie verklaren. Voor een logistische-regressiemodel berekent de VIF-score het kwadraat van de toename van de standaarddeviatie van een regressiecoëfficiënt onder invloed van de andere verklarende factoren in het model. Er is geen absolute regel voor de interpretatie van de VIF-scores, maar een VIF-score groter dan 10 wordt algemeen als een sterke indicatie voor multicollineariteit beschouwd en een VIF-score groter dan 2.5 als een aanmaning tot voorzichtigheid.

Tabel 14(i) vat de VIF-scores voor het logistische-regressiemodel in tabel 14(2) samen. Bij slechts twee verklarende variabelen overschrijdt de VIF-score de waarde 2.5, met name voor ‘a.len.pros=meer.s’ (VIF = 2.877987) en voor ‘a.len.pros=meer.sw’ (VIF = 2.757547). We beschouwen deze lichte overschrijdingen van de drempelwaarde 2.5 niet



14. MULTIVARIATE ANALYSE VAN DE ADJECTIVISCHE BUIGINGSALTERNANTIE BIJ NEUTRA

als een gevaar voor de stabiliteit van deze regressiecoëfficiënten in het model in tabel 14(2). De oorzaak van deze hoge VIF-waarden ligt bij de sterke correlatie tussen de factoren ‘a.len.pros=meer.s’ en ‘a.len.pros=meer.sw’ aan de ene kant en de factor ‘a.geleed’ aan de andere kant: wanneer we de factor ‘a.geleed’ uit het model verwijderen, stellen we een sterke daling vast van de VIF-scores voor ‘a.len.pros=meer.s’ (VIF = 1.901557) en voor ‘a.len.pros=meer.sw’ (VIF = 1.822860). Dit bewijst andermaal de sterke correlatie tussen de morfologische geleedheid en de prosodische structuur van het adjectief.

Factorwaarde	VIF	Factorwaarde	VIF
det.cat=aanw.vnw	1.045648	reg.inf.index=nl.1	1.036326
det.cat=bez.vnw	1.098480	reg.inf.index=nl.2	1.546640
det.cat=genitief	1.012478	reg.inf.index=nl.3	1.678981
n.inf=ja	1.030555	reg.inf.index=bel.0	2.014794
n.genus=bigen	1.054592	reg.inf.index=bel.1	1.580276
n.beginklank=klink	1.050107	reg.inf.index=bel.2	1.632475
a.len.pros=twee.s	1.788134	reg.inf.index=bel.3	1.493734
a.len.pros=twee.sw	2.448413	lex.col=kw.1	1.115093
a.len.pros=meer.s	2.877987	lex.col=kw.2	1.079520
a.len.pros=meer.sw	2.757547	lex.col=kw.3	1.168172
a.len.pros=meer.sww	1.941290	lex.col=kw.4	1.236533
n.pros=w	1.055355	a.rel=ja	1.416471
		a.geleed=ja	2.421587

Tabel 14(i): VIF-scores voor de verklarende variabelen in het logistische-regressiemodel in tabel 14(2)

6. De lezer kan de regressievergelijking zelf reconstrueren aan de hand van de regressiecoëfficiënten en de factorwaarden in de eerste twee kolommen van tabel 14(2).
7. Het ruime 95% BI voor de OR wijst op dataschaarste, zoals duidelijk werd bij de analyse van de *influential observations* (zie noot 4 bij dit hoofdstuk). In de lectische analyse (zie paragraaf 12.3.2) hebben we geopperd dat het sterk deflecterende effect in het Nederlandse register met informaliteitsindex 1 het gevolg is van het grote aandeel lexicale collocaties gecombineerd met dataschaarste ( $n = 15$ ).
8. De correspondentieanalyse is een exploratieve statistische techniek die gebruikt wordt bij de analyse van 2x2-contingentietabellen en (zeer grote)  $rx$ c-contingentietabellen (in laatstgenoemd geval spreken we van meervoudige correspondentieanalyse) (zie o.a. Venables & Ripley 2002). Een correspondentieanalyse geeft de relaties tussen de verschillende variabelen weer aan de hand van een aantal latente dimensies (net zoals dat bij een *factoranalyse* het geval is voor numerische variabelen).

Een correspondentieanalyse verschaft twee soorten informatie. Ten eerste, de oorspronkelijke  $n$  dimensies worden tot één, twee of drie latente dimensies gereduceerd. Deze dimensies worden berekend in functie van een maximale bijdrage tot de globale  $\chi^2$ -statistiek voor de volledige contingentietabel. Vervolgens moeten deze dimensies geïnterpreteerd worden. Ten tweede, we kunnen ook naar de verschillende factorwaarden kijken

die doorgaans (in functie van twee latente dimensies) in vier kwadranten worden ingedeeld. Meetpunten die in hetzelfde kwadrant voorkomen, zijn sterk gerelateerd en leveren bijgevolg een grote positieve bijdrage tot de globale  $\chi^2$ -statistiek. Het is eigen aan (de meeste) exploratieve technieken dat er geen mogelijkheid voorzien is om de significantie van de vastgestelde patronen te testen.

Rekening houdend met onze doelstelling, met name het verklaren van Simpsons paradox bij een aantal factoren, zijn we niet zozeer in de latente dimensies geïnteresseerd, maar wel in de onderlinge positionering van de meetpunten, c.q. de factorwaarden. We voeren de (meervoudige) correspondentieanalyse uit op basis van de datamatrix die de invoer vormt van de logistische regressie in tabel 14(2), met dit verschil dat de responsvariabele eruit verwijderd is. We zijn namelijk geïnteresseerd in de onderlinge associaties tussen de verklarende factoren.

Voor het uitvoeren van de correspondentieanalyse hebben we gebruik gemaakt van het MASS-pakket in R (Venables & Ripley 2002).

9. Traditioneel wordt het resultaat van een correspondentieanalyse als een (twee-dimensionaal) *scatterplot* weergegeven. Een scatterplot is doorgaans moeilijk leesbaar doordat de labels bij de meetpunten elkaar overlappen. Bijgevolg hebben we ervoor geopteerd om de vier kwadranten van het scatterplot in tabelvorm weer te geven.

Sommige correspondenties in tabel 14(4) zijn gemotiveerd: de vaststelling dat 'reg.inf.ind=nl.2' in hetzelfde kwadrant staat als 'lex.col=kw.2', 'lex.col=kw.3' en 'lex.col=kw.4' hangt samen met het sterk formulaïsche taalgebruik in het Nederlandse register met informaliteitsindex 2 (zie paragraaf 12.3.2); 'a.geleed=ja' komt voor in hetzelfde kwadrant als één van de prosodische factoren waarmee het (een zwakke) multicollineariteit vertoont, met name 'a.len.pros=meer.sw'; 'a.rel' bevindt zich in hetzelfde kwadrant als 'lex.col=kw.2', 'lex.col=kw.3' en 'lex.col=kw.4', wat samenhangt met de vaststelling dat de adjectivische positie in lexicale collocaties significant vaker dan verwacht door een relationeel adjectief wordt bezet (zie paragraaf 11.2.2). Andere correspondenties zijn dan weer toevallig, zoals de constatering dat 'det.cat=bez.vnw' en 'n.beginklank=medeklinker' tot hetzelfde kwadrant behoren.

10. Op basis van het model in tabel 14(2) zijn er 64,440 logische factorwaardencombinaties. We hebben uiteraard een afzonderlijke regressieanalyse met de acht geselecteerde factoren uitgevoerd om de aangepaste regressiecoëfficiënten te berekenen. Dit is de aangepaste regressievergelijking:

$$\begin{aligned}
 g' = & -4.24852 * + (-0.15970 * \text{det.cat=aanw.vnw}) + (0.67422 * \text{det.cat=bez.vnw}) + \\
 & (0.34517 * \text{det.cat=genitief}) + (1.49372 * \text{n.inf=ja}) + (0.56087 * \\
 & \text{n.beginklank=klink}) + (0.49473 * \text{a.len.pros=twee.s}) + (1.40732 * \\
 & \text{a.len.pros=twee.sw}) + (1.26360 * \text{a.len.pros=meer.s}) + (2.14370 * \\
 & \text{a.len.pros=meer.sw}) + (3.11653 * \text{a.len.pros=meer.sww}) + (0.42776 * \\
 & \text{n.pros=w}) + (1.16640 * \text{reg.inf.index=nl.1}) + (1.08535 * \text{reg.inf.index=nl.2}) + \\
 & (0.86863 * \text{reg.inf.index=nl.3}) + (0.27105 * \text{reg.inf.index=bel.0}) + (1.01088 * \\
 & \text{reg.inf.index=bel.1}) + (1.66723 * \text{reg.inf.index=bel.2}) + (2.22303 * \\
 & \text{reg.inf.index=bel.3}) + (0.18698 * \text{lex.col=kw.1}) + (0.61080 * \text{lex.col=kw.2}) + \\
 & (1.30798 * \text{lex.col=kw.3}) + (2.21181 * \text{lex.col=kw.4}) + (0.99494 * \text{a.rel=ja})
 \end{aligned}$$

#### 14. MULTIVARIATE ANALYSE VAN DE ADJECTIVISCHE BUIGINGSALTERNANTIE BIJ NEUTRA

Hier moeten we opnieuw op de ongecontroleerde samenstelling van een corpus wijzen. Terwijl de samensteller van een enquête of een psycholinguïstisch experiment de items in de vragenlijst – en bijgevolg de gerealiseerde factorwaardencombinaties – controleert, is de corpuslinguïst aangewezen op de factorwaardencombinaties die in het corpus gerealiseerd worden. De materiaalverzameling realiseert nauwelijks 1,063 van de 15,360 logische parameterwaardencombinaties: terwijl het regressiemodel voor 70.29% ( $n = 10,797$ ) van de logische parameterwaardencombinaties een onverbogen adjectief voorspelt, vertoont slechts 15.97% van de observaties in de materiaalverzameling een parameterwaardencombinatie waarvoor het logistische-regressiemodel een onverbogen adjectief voorspelt. Dit hangt uiteraard samen met de gemarkeerde status van de onverbogen buigingsvariant.

11. Rekening houdend met de ongelijke verdeling van de beide buigingsvarianten in de materiaalverzameling ligt het voor de hand dat de klassen met de lage  $p$ -waarden oververtegenwoordigd zijn en dat de klassen met de hoge  $p$ -waarden ondervertegenwoordigd zijn. Tabel 14(ii) toont de frequentieverdeling van de verschillende klassen ('Kl.1' tot en met 'Kl.10').

Klasse $p$ -waarden									
Kl.1	Kl.2	Kl.3	Kl.4	Kl.5	Kl.6	Kl.7	Kl.8	Kl.9	Kl.10
2,071	999	477	381	263	251	148	169	101	104

Tabel 14(ii): Frequentieverdeling van de klassen voor de  $p$ -waarden

We stellen vast dat de klassen 1 en 2 samen goed zijn voor 61.85% van de observaties in de materiaalverzameling. De klassen met een  $p$ -waarde groter dan 0.5 bevatten samen slechts 16% van het totale aantal observaties.

12. Net als voor het globale regressiemodel in tabel 14(2) hebben we ook voor de nationaal gedefinieerde regressiemodellen de *influential observations* geïdentificeerd. De deelverzameling voor het Belgisch Nederlands bevat 27 *influential observations*. Wanneer we deze meetpunten uit de materiaalverzameling verwijderen, bekomen we de volgende modelstatistieken: verklarende kracht =  $3,110 - 2,245 = 865.45$ ,  $df = 25$ ,  $p < 0.00001$ ; voorspellende kracht:  $c = 0.845$ . Let wel: de genitief realiseert een perfecte scheiding, met een instabiele regressiecoëfficiënt en OR als gevolg: OR = 0.0023 [1.4e-26, 3.8e+20].

De Nederlands Nederlandse deelverzameling bevat ruim 68 *influential observations*. De verwijdering van deze observaties leidt tot een zeer sterke stijging van de verklarende ( $1,973 - 1,027 = 946.47$ ,  $df = 25$ ,  $p < 0.00001$ ) en de voorspellende kracht ( $c = 0.927$ ) van het model. Let wel: na de verwijdering van de *influential observations* realiseren twee factorwaarden een perfecte scheiding in het Nederlandse model, met instabiele regressiecoëfficiënten als gevolg: 'inf.index=1': OR = 0.0001 [4.6e-107, 6.1e+98]; 'n.genus=bigen': OR = 9.9e-05 [1.6e-29, 5.8e+20].

Rekening houdend met de instabiliteit van één of meerdere regressiecoëfficiënten in beide nationale modellen na de verwijdering van de *influential observations* beantwoorden we de vraag naar de noodzaak van een apart model voor het Belgisch Nederlands en het Nederlands Nederlands op basis van de logistische-regressiemodellen gerapporteerd in de tabellen 14(7) en 14(8), i.e. de regressiemodellen met inclusie van de *influential observations*.

13. We testen de verklarende variabelen in de logistische-regressiemodellen voor het Nederlands Nederlands en het Belgisch Nederlands eerst op multicollineariteit. Tabel 14(iii) rapporteert de VIF-scores voor de verklarende variabelen in beide nationaal gedefinieerde logistische-regressiemodellen. Alle VIF-scores in tabel 14(iii) zijn lager dan de drempelwaarde 2.5, zowel voor het Belgisch Nederlandse als voor het Nederlands Nederlandse regressiemodel. We besluiten bijgevolg dat er geen gevaar voor multicollineariteit is.

<b>Factorwaarde</b>	<b>VIF<sub>Nederlands.Nederlands</sub></b>	<b>VIF<sub>Belgisch.Nederlands</sub></b>
det.cat=aanw.vnw	1.065901	1.035088
det.cat=bez.vnw	1.064748	1.129698
det.cat=genitief	1.011943	1.016004
n.inf=ja	1.057436	1.021554
n.gen=bigen	1.082723	1.050051
n.beginklank=klink	1.081063	1.036965
a.len.pros=twee.s	1.576785	1.323141
a.len.pros=twee.sw	2.058521	1.501193
a.len.pros=meer.s	2.292686	1.777158
a.len.pros=meer.sw	2.324457	1.644310
a.len.pros=meer.sww	1.861265	1.374665
n.pros=w	1.086665	1.050306
inf.index=1	1.405375	1.215634
inf.index=2	1.051193	1.229399
inf.index=3	1.353393	1.274692
lex.col=kw.1	1.127068	1.102705
lex.col=kw.2	1.072810	1.084159
lex.col=kw.3	1.171450	1.175320
lex.col=kw.4	1.336715	1.200678
a.rel=ja	1.204364	1.240032

*Tabel 14(iii): VIF-scores voor de verklarende variabelen in de logistische-regressiemodellen voor het Nederlands Nederlands en het Belgisch Nederlands*

14. We benadrukken dat we in eerste instantie geïnteresseerd zijn in de factorwaarden die significant verschillen in beide nationale variëteiten. Factorwaarden die niet significant verschillen, zijn niet relevant, ook niet als de zin van de regressiecoëfficiënt in beide nationale variëteiten tegengesteld is. We verduidelijken dit aan de hand van een voorbeeld, met name het aanwijzend voornaamwoord ('det.cat=aanw.vnw'). De impact van dit element is tegengesteld in het Nederlands Nederlands (OR = 0.5849 [0.3705,0.9235]) en in het Belgisch Nederlands (OR = 1.1153 [0.7773,1.6003]). Dit verschil is echter niet significant, waardoor we enkel de volgende, vrijblijvende interpretatie kunnen voorstellen: in het Nederlands Nederlands komt de verbogen buigingsvariant significant vaker voor na een vorm van het aanwijzend voornaamwoord dan na het bepaald lidwoord; in het Belgisch Nederlands is dit verschil niet significant; het verschil tussen beide nationale variëteiten is evenmin significant, maar zo er een significant verschil zou zijn, dan zou het aanwijzend voornaamwoord in het Belgisch Nederlands waarschijnlijk een faciliterende context voor het gebruik van de onverbogen

buigingsvorm zijn. Het spreekt voor zich dat zulke vrijblijvende uitspraken niet stroken met de methodologische nauwkeurigheid die we nastreven.

15. Wanneer we een *forward stepwise*-regressie uitvoeren voor het Belgisch Nederlands is de informaliteitsindex de tweede factor die in het model wordt opgenomen, na de prosodische structuur aan de rechterwoordrand van het adjectief. Bij de *forward stepwise*-analyse voor het Nederlands Nederlands is dat de lexicale collocabiliteit van het adjectief-nomenpaar.
16. Voor de volledigheid vermelden we dat de verwijdering van de 68 *influential observations* in het Nederlands Nederlands tot de volgende verschillen in het regressiemodel leidt:
  - Het effect van de diminuering is na de verwijdering van de *influential observations* zwak significant (OR = 0.0825 [0.0102, 0.6690],  $p = 0.0195$ ). Bovendien is ook het verschil met het Belgisch Nederlands significant, zowel ten overstaan van het model met als ten overstaan van het model zonder de *influential observations*. Dit is een gevolg van het verwijderen van acht van de tien verkleinwoorden met een onverbogen adjectief uit de Nederlandse materiaalverzameling, waardoor het buigings-effect van de diminuering uitvergroet wordt.
  - De morfologische geleedheid van het adjectief definieert na de verwijdering van de *influential observations* een significant patroon (OR = 3.4459 [1.5128, 7.8491],  $p = 0.0032$ ). Dit is een gevolg van het verwijderen van (onverbogen) tweelettergrepige ongelede adjectieven met 'sw' als rechterwoordrand (bv. *eerder werk*, *ander doosje*, *ander woord*, *standaard verhaal*) en (onverbogen) eenlettergrepige gelede adjectieven (bv. *Zweeds parlement*, *Turks brood*). Dit toont andermaal aan dat de morfologische geleedheid van het adjectief sterk gecorreleerd is met de adjectivische lengte en de prosodische structuur aan de rechterwoordrand van het adjectief.
  - Andere verschillen als gevolg van de verwijdering van de *influential observations* leiden enkel tot een versterking van een bestaand significant effect: 'a.len.pros=meer.sww': OR = 38.5671 [15.6091, 95.2914]; 'lex.col=kw.4': OR = 44.1326 [24.2548, 80.3011].

In het Belgisch Nederlands noteren we één belangrijke verandering na de verwijdering van de *influential observations*: het genus van het element in de kernpositie definieert, net als in het Nederlands Nederlands, een significant effect (OR = 0.3887 [0.2035, 0.7422];  $p = 0.0042$ ).

17. We vergelijken beide logistische-regressiemodellen, c.q. het globale regressiemodel in tabel 14(2) en het regressiemodel voor de reducerende buigingsattitude, aan de hand van een ANOVA (Woods, Fletcher & Hughes 1986): we gebruiken beide regressiemodellen als verklarende variabelen; de verklaarde variatie is de responsvariabele (zie ook paragraaf 13.2.1, noot 8). Op deze wijze kunnen we het verschil in verklaarde variatie tussen het globale regressiemodel en het regressiemodel voor de reducerende buigingsattitude meten en testen of dit verschil significant is. Voor de ANOVA's die we gebruiken om twee logistische-regressiemodellen te vergelijken, rapporteren we de volgende informatie: de onverklaarde variatie van beide modellen (residu), het verschil in verklaarde variatie tussen het tweede model en het eerste model en de significantie van dit verschil ( $P(>|\chi^2|)$ ). Het verschil in verklaarde variatie tussen het logistische-regressiemodel voor de reducerende attitude en het globale model in tabel 14(2) is significant, zoals de gegevens

in tabel 14(iv) aantonen ( $p = 1.724e-90$ ). Het negatieve cijfer voor het verschil in verklaarde variatie (-472.9) duidt aan dat het regressiemodel voor de reducerende buigingsattitude significant minder variatie verklaart dan het globale regressiemodel.

<b>Logistische-regressiemodel</b>	<b>Residu</b>	<b>Vershil in</b>	<b>P(&gt; <math>\chi^2</math>)</b>
<b>verklaarde variatie</b>			
Globaal regressiemodel	3,779.7		
Regressiemodel voor de reducerende buigingsattitude	4,252.6	-472.9	1.724e-90

Tabel 14(iv): ANOVA voor de vergelijking van de verklaarde variatie in het globale regressiemodel en het regressiemodel voor de reducerende buigingsattitude

In feite zouden we het prosodische patroon van het adjectief moeten beperken tot de beklemtoning van de eindlettergreep. Deze aanpassing gaat gepaard met een bijkomende daling van zowel de verklarende kracht ( $5,383.4 - 4,580.0 = 803.39$ ,  $df = 5$ ,  $p < 0.00001$ ) als van de voorspellende kracht ( $c = 0.745$ ). Het spreekt voor zich dat ook dit model significant minder variatie verklaart dan het globale model in tabel 14(2), zoals kan worden afgelezen uit tabel 14(v).

<b>Logistische-regressiemodel</b>	<b>Residu</b>	<b>Vershil in</b>	<b>P(&gt; <math>\chi^2</math>)</b>
<b>verklaarde variatie</b>			
Globaal regressiemodel	3,779.7		
Regressiemodel voor de reducerende buigingsattitude(bis)	4,580.0	-800.3	217e-156

Tabel 14(v): ANOVA voor de vergelijking van de verklaarde variatie in het globale regressiemodel en het regressiemodel voor de reducerende buigingsattitude (met enkel de beklemtoning van de adjectivische eindlettergreep als prosodisch patroon)

18. Dit zijn de modelstatistieken voor het logistische-regressiemodel dat alleen de lexicale/semantische ('lex.col' en 'a.rel') en de lectische ('reg.inf.index') factoren bevat:

- verklarende kracht:  $5,383.4 - 4,391.5 = 991.94$ ,  $df = 12$ ,  $p < 0.00001$
- voorspellende kracht:  $c = 0.788$

We vergelijken nu aan de hand van een ANOVA de verklaarde variatie van het globale logistische-regressiemodel in tabel 14(2) met de variatie verklaard door bovenstaand logistische-regressiemodel op basis van het discursieve congruentiemodel. De gegevens in tabel 14(vi) tonen aan dat het discursieve regressiemodel significant minder variatie verklaart dan het globale regressiemodel:

<b>Logistische-regressiemodel</b>	<b>Residu</b>	<b>Vershil in</b>	<b>P(&gt; <math>\chi^2</math>)</b>
<b>verklaarde variatie</b>			
Globaal regressiemodel	3,779.7		
Regressiemodel voor het discursieve congruentiemodel	4,391.5	-611.8	2.38e-122

Tabel 14(vi): ANOVA voor de vergelijking van de verklaarde variatie in het globale regressiemodel en het regressiemodel voor het discursieve congruentiemodel (o.a. Barlow)

#### 14. MULTIVARIATE ANALYSE VAN DE ADJECTIVISCHE BUIGINGSALTERNANTIE BIJ NEUTRA

19. Merk op dat Haiman (1985) de lectische dimensie wellicht als indexicaal zou opvatten.
20. De adjectivische buiging in het Afrikaans is niet langer een morfosyntactisch maar een lexicaal fenomeen. De morfologische en fonologische structuur van het adjectivische lemma bepalen de uitgang. Een morfologisch complex adjectief eindigt altijd op *-e*. Bij een morfologisch enkelvoudig adjectief is de fonotactische structuur bepalend: adjectieven waarvan de stam eindigt op een consonantencluster, een fricatief of een stemloze occlusief nemen een 'buigings-*e*'; de andere adjectieven eindigen onveranderd op  $-\emptyset$  (Lass 1990: 91-98; van Marle 1995a; van Marle & Smits 1993; Raidt 1968: 102-114).

## DEEL IV

### **LEXICALE EN SEMANTISCHE ANALYSE: SEMANTISCHE CONDITIONERING VAN DE BUIGINGSALTERNANTIE**

In het laatste deel van dit proefschrift concentreren we ons op de lexicale conditionering van de buigingsalternantie. We stellen meer bepaald de vraag of de invloed van de lexicale collocaties op de buigingsalternantie, zoals vastgesteld in hoofdstuk 11, een louter lexicaal – en bijgevolg idiosyncratisch en idiomatisch – effect is, dan wel of er semantische categorieën aan de basis van het deflecterende effect van de lexicale collocaties geïdentificeerd kunnen worden.

Bij de lexicale/semantische analyse op basis van het CGN hebben we lexicale patronen geïdentificeerd die de selectie van het onverbogen adjectief motiveren, met name adjectief-nomenparen die een lexicale collocatie vormen. Daarnaast hebben we ook gezien dat de adjectivische positie in deze patronen significant vaker dan verwacht door een relationeel adjectief wordt bezet. In dit vierde deel diepen we de semantische interpretatie van de lexicale collocabiliteit tussen het adjectief en het substantief verder uit door de introductie van drie bijkomende semantische categorieën die in de literatuur over de buigingsalternantie en gefixeerde adjectief-nomenparen behandeld worden: categoriserende adjectieven, eigennamen en institutionele termen. De belangrijkste verschillen met het bestaande onderzoek zijn van methodologische aard. Ten eerste, we streven er expliciet naar om de semantische kenmerken door middel van een objectieerbare – of beter gezegd intersubjectieve – procedure te identificeren. Ten tweede, we testen de empirische validiteit van deze semantische kenmerken aan de hand van een kwantitatief-empirisch corpusonderzoek in plaats van een impressionistische analyse op basis van enkele zeer sterk gefixeerde adjectief-nomenparen die vooral in de neerlandistische vakliteratuur voorkomen.

De motivatie voor dit uitgebreide semantische onderzoek is van tweeërlei aard. Ten eerste, we hebben tijdens de lexicale/semantische analyse op basis van het CGN herhaaldelijk op het probleem van dataschaarste gewezen (zie hoofdstuk 11). Door het raadplegen van een ruimer corpus met krantenmateriaal hopen we zowel het aantal als het aandeel lexicale collocaties te vermeerderen. Het krantenmateriaal is niet alleen groter wat de omvang betreft, maar vertoont ook een grotere lectische en



thematische homogeniteit dan het CGN, waardoor de kans op recurrente adjectief-nomenparen stijgt. Deze materiaaluitbreiding is noodzakelijk indien we op zoek willen gaan naar een semantische interpretatie voor de AN-lemmatypes met een significante associatie tussen het adjectief en het substantief. Ten tweede, deze semantische analyse biedt de mogelijkheid om de empirische validiteit te testen van de semantische hypothesen voorgesteld door Blom (1994, 1995), Booij (2002a) en Honselaar (1980) (zie paragraaf 1.1.3.2).

Dit vierde en laatste deel van het onderzoek naar de adjectivische buigingsalternantie bij neutra omvat drie hoofdstukken. In hoofdstuk 15 presenteren we de samenstelling en de ontsluiting van het geraadpleegde *De Standaard*-corpus. Hoofdstuk 16 is gewijd aan de lexicale conditionering van de buigingsalternantie. Deze analyse is grotendeels een replicatie van het collocatieonderzoek op basis van het CGN (zie hoofdstuk 11), waarbij we de impact van de lexicale collocabiliteit op de buigingsalternantie en het communicatiemedium, met het onderscheid tussen gesproken en geschreven taal, van naderbij bekijken. De eigenlijke semantische analyse wordt gerapporteerd in hoofdstuk 17, dat de kern van dit vierde deel vormt. In dit hoofdstuk meten we eerst in welke mate de buigingsalternantie bij de lexicale collocaties semantisch gemotiveerd is. Daarna bespreken we de implicaties van het empirisch onderzoek voor de interpretatie van het deflecterende effect van de lexicale collocaties.

We concentreren ons in dit vierde deel hoofdzakelijk op de lexicale en semantische conditionering van de buigingsalternantie. De semantische analyse in hoofdstuk 17 heeft bijgevolg uitsluitend betrekking op de observaties waar het adjectief-nomenpaar een lexicale collocatie vormt. De andere verklarende factoren die we in de vorige twee delen op basis van het CGN hebben vastgesteld, worden niet in dit lexicale en semantische onderzoek opgenomen. Ten eerste, door de Vlaamse kwaliteitskrant *De Standaard* als corpus te gebruiken, zijn de lectische variabelen constant: de materiaalverzameling is uitsluitend opgebouwd uit algemeen geschreven Belgisch Nederlands. Ten tweede, het effect van de derde primaire factor, de prosodische conditionering, is wellicht minder groot in de geschreven taal dan in de gesproken taal, zodat het verdedigbaar is om deze parameter niet in het lexicale en semantische onderzoek op te nemen, te meer omdat dit onderzoek beperkt is tot de adjectief-nomenparen die een lexicale collocatie vormen.

## **Hoofdstuk 15:**

# **Voorstelling en analyse van het *De Standaard*-corpus**

Alvorens de lexicale en semantische analyse aan te vatten, stellen we bondig het gebruikte *De Standaard*-corpus voor (15.1) en lichten we de ontsluiting en bevraging van deze materiaalverzameling toe (15.2). We beschrijven meer bepaald de *ad hoc* procedure die we ontwikkeld hebben om dit corpus te analyseren. We sluiten dit hoofdstuk af met de presentatie van de resulterende materiaalverzameling voor de analyse van de buigingsalternantie (15.3).

### **15.1 Voorstelling van het *De Standaard*-corpus**

Het *De Standaard*-corpus (voortaan VDDS) omvat de edities van januari tot en met september van de jaargang 2003 van deze Vlaamse kwaliteitskrant.<sup>1</sup> Dit corpus bevat 17,911,865 woorden, bijna het vijfvoud (4.83) van het geannoteerde deel van de 5<sup>e</sup> pre-release van het CGN (zie paragraaf 6.1). Een materiaalverzameling van deze omvang moet ons in staat stellen om het probleem van de dataschaarste op te lossen waarmee we bij de analyse van de lexicale effecten in het CGN geconfronteerd werden (zie paragraaf 11.2.1).

Daarnaast is er ook een kwalitatieve motivatie voor de keuze van dit materiaal. Door het gebruik van deze Vlaamse kwaliteitskrant controleren we de lectische factoren: de nationale variëteit en het register zijn constant. Het betreft met name de formele variëteit van het geschreven standaard Belgisch Nederlands. We kunnen bovendien aannemen dat het aantal recurrente adjectief-nomenparen groter is in een corpus bestaande uit één jaargang van één krant dan in een corpus met gesproken materiaal dat in uiteenlopende gesprekssituaties verzameld is. Niet alleen de dubbele redactionele controle (c.q. door de journalist en door de eindredactie) maar ook het verslaan van de dagelijkse actualiteit leidt tot een grotere thematische consistentie en verhoogt de kans op repetitief taalgebruik. De bijzondere aandacht die het politieke en maatschappelijke leven in een kwaliteitskrant genieten, resulteert normaliter in de aanwezigheid van thema's die het gebruik van adjectief-nomenparen met naam-status, i.e. van lexicale collocaties, bevorderen.

Kortom: het VDDS-corpus is zowel om kwantitatieve als om kwalitatieve redenen geschikt voor het onderzoek naar de lexicale en semantische conditionering van de buigingsalternantie.

## 15.2 Ontsluiting en analyse van het *De Standaard*-corpus

Om een maximale replicateerbaarheid van het onderzoek te garanderen, beschrijven we in deze paragraaf de stapsgewijze procedure die we gebruikt hebben om het VDDS-corpus te bevragen met het oog op de analyse van de buigingsalternantie.

Vergeleken met het CGN heeft het VDDS-corpus één groot nadeel: het materiaal is niet verrijkt met taalkundige informatie. Dit is hoogst problematisch aangezien we zowel de woordsoortannotatie als de lemmatisering nodig hebben bij het onderzoek naar de impact van lexicale collocaties op de buigingsalternantie. Bij ontstentenis van een stilistisch vergelijkbare geannoteerde materiaalverzameling kunnen we geen beroep doen op een ‘lerende’ *part of speech*-tagger. We hebben er bijgevolg voor geopteerd om het materiaal zelf te annoteren en te lemmatiseren. Hiervoor hebben we een *ad hoc* procedure opgesteld vanuit het standpunt van onze onderzoeksdoelstellingen, met name de analyse van lexicale preferentiepatronen bij de keuze van de adjectivische buigingsvariant en de mogelijke identificatie van een onderliggende semantische regelmaat. Voor de realisatie van deze doelstellingen hebben we de volgende gegevens nodig:

- een materiaalverzameling met de instanties van  $[\text{DET}_{[\text{DEF}]} \text{A N}_{[\text{ONZ,ENK}]}]_{\text{NP}}$
- een deelcorpus bestaande uit de NP's met de structuur  $[\text{DET? A N}]_{\text{NP}}$ , verrijkt met woordsoortannotatie en lemmatisering

De samenstelling van de eerste materiaalverzameling vereist nauwelijks enige toelichting: deze materiaalverzameling bevat de instanties van de syntactische context waar de buigingsalternantie optreedt en vormt het equivalent van de CGN-materiaalverzameling (zie paragraaf 6.2). De tweede materiaalverzameling is een deelcorpus uit het VDDS-corpus dat enkel NP's met de structuur  $[\text{DET A N}]_{\text{NP}}$  of met de structuur  $[\text{A N}]_{\text{NP}}$  bevat. De motivatie voor de compilatie van dit deelcorpus is zowel praktisch als inhoudelijk van aard. Vanuit een praktisch standpunt kunnen we op deze wijze de annotatietaak beperken tot specifieke contexten en (combinaties van) woordsoorten. Inhoudelijk gezien willen we de associatie tussen het adjectief en het substantief zo nauwkeurig mogelijk berekenen. Hiervoor beperken we de zoekruimte tot de NP's. Dit is: we verwijderen de contexten die irrelevant zijn voor adjectief-nomensequenties en die bijgevolg voor ruis zorgen (voor een uitgebreide verantwoording verwijzen we naar paragraaf 11.1.1.2.). We beperken het aantal adjectieven in de NP's in dit deelcorpus tot één omdat dit het aantal adjectieven in de onderzoekscontext is. De conditie dat de NP's altijd een adjectief moeten bevatten, is technisch: de gebruikte *log likelihood ratio*-statistiek ( $G^2$ ; zie paragraaf 11.1.1.2) berekent de associatie tussen twee elementen, zodat NP's die enkel een N bevatten, uitgesloten zijn. Bij uitbreiding sluiten we ook de pendant met determinator van een NP zonder adjectief uit, i.e. de sequentie  $[\text{DET N}]_{\text{NP}}$ , zodat we een evenwichtige materiaalverzameling bekomen voor de analyse van de impact op de buigingsalternantie van de collocatieve kracht tussen het adjectief en het substantief.

In de volgende paragrafen presenteren we de compilatie van het deelcorpus met de twee- en drieplaatsige NP's (15.2.1) en van de materiaalverzameling met de adjectivische buigingsalternantie (15.2.2).

### 15.2.1 *Compilatie van het deelcorpus met twee- en drieplaatsige NP's*

We beginnen met de compilatie van het deelcorpus met de twee- en drieplaatsige NP's, voortaan het  $VDDS_{NP}$ -corpus. Tabel 15(1) stelt de invulling voor van de drie syntagmatische posities in het NP-deelcorpus. Het is duidelijk dat de aanwezigheid van open woordklassen in de adjectivische positie en de nominale kernpositie het belangrijkste probleem vormt bij de taalkundige verrijking van de gegevens.

determinatorpositie	adjectiefpositie	kernpositie
lidwoorden	inflectionele vormen van	inflectionele vormen van
voornaamwoorden	het adjectief	het substantief
(zero) <sub>2</sub> -plaatsige NP	inflectionele vormen van het voltooid deelwoord	genominaliseerde infinitief
	inflectionele vormen van het onvoltooid deelwoord	
GESLOTEN WOORDSOORT	OPEN WOORDSOORT	OPEN WOORDSOORT

Tabel 15(1): Maximale invulling van de NP bij de compilatie van het  $VDDS_{NP}$ -corpus

Voor het compileren en het annoteren van de materiaalverzameling met de twee- en drieplaatsige NP's doen we een beroep op de woordsoortbestanden van *Van Dale*: deze bestanden bevatten de verschillende lemmata die een woordsoort in het *Groot Van Dale Woordenboek* instantiëren samen met hun inflectionele paradigma. Het bestand met de substantieven bevat bijvoorbeeld het lemma *boek* en de verschillende inflectionele realisatievormen van dit lemma: *boek*, *boeken*, *boekje* en *boekjes*. Voor de verschillende open woordklassen die de adjectivische positie en de kernpositie in het schema in tabel 15(1) kunnen bezetten, beschikken we over een woordsoortbestand van *Van Dale*.

We beperken het NP-deelcorpus tot de meest courante invulling van de drie posities binnen de NP. Hiervoor baseren we ons op de materiaalverzameling die we uit het CGN geëxtraheerd hebben. Geanteponeerde nominale genitieven in determinatorpositie worden niet opgenomen: primo, deze invulling van de determinatorpositie bleek zeldzaam te zijn in het CGN; secundo, deze mogelijkheid zou de zoekruimte voor de compilatie van het NP-deelcorpus aanzienlijk vergroten door de inclusie van een open woordsoort in de determinatorpositie die nu enkel door gesloten woordsoorten gerealiseerd wordt. Nominalisaties hebben we, met uitzondering van de infinitief, om dezelfde redenen van de nominale kernpositie uitgesloten.

Voor de compilatie en de verrijking van het corpus met de twee- en drieplaatsige NP's hebben we deze stapsgewijze procedure toegepast:

- stap 1: tokenizeren
- stap 2: woordsoortannotatie
- stap 3: compilatie VDD<sub>NP</sub>
- stap 4: lemmatisering

We lichten deze stappen kort toe.

#### *Stap 1: Tokenizeren*

Het tokenizeren vormt de eerste fase van de morfosyntactische verrijking van een corpus. Tijdens deze voorbereidende fase worden de woordgrenzen – en dus ook de woorden – in het corpus geïdentificeerd. De volgende karakters en karakter-sequenties worden als een apart woord beschouwd:

- meta-informatie tussen vishaken
- leestekens: '.', '!', '?', ';', ':', ',', '}', '{', '}', '"""', ''''
- sequenties bestaande uit alfanumerieke karakters, inclusief het koppelteken en de ampersand

Deze fase is volledig automatisch uitgevoerd.

#### *Stap 2: Woordsoortannotatie*

Na het tokenizeren worden de verschillende tokens van hun woordsoortinformatie voorzien. Dit is een *match and generate*-procedure op basis van deze vragen:

- Stap 2.a: Staat het betreffende element in de tokenlijst van één van de woordsoorten in tabel 15(1)?
- Stap 2.b: Ken het gepaste woordsoortlabel aan het element in kwestie toe (indien de vraag in stap 2.a positief beantwoord is).

Tabel 15(2) toont de onderscheiden woordsoortlabels:

Positie	Label	Elementen
DETERMINATOR	<DET-onz-enk-def>	definiete, enkelvoudige onzijdige of bigenerische determinator
	<DET-rest>	andere determinator
ADJECTIEF	<ADJ-adj>	adjectief
	<ADJ-vd>	voltooid deelwoord
	<ADJ-od>	onvoltooid deelwoord
KERN	<NOM-onz-enk >	enkelvoudig <i>het</i> -woord of bigenerisch substantief
	<NOM-dim-enk>	enkelvoudig diminutief
	<NOM-inf>	infinietief
	<NOM-rest>	andere substantivische vormen (c.q. <i>de</i> -woorden en meervoudsvormen)

*Tabel 15(2): Labels voor de woordsoortannotatie in het VDD<sub>NP</sub>-corpus*

Bij de labels in tabel 15(2) moeten enkele bemerkingen worden geformuleerd. Ten eerste, de labels zijn specifiek ontworpen voor de compilatie van de twee materiaalverzamelingen die we nodig hebben bij het onderzoek naar het buigingseffect van de lexicale collocaties. In het kader van deze gespecialiseerde taak gebruiken we speciale labels voor de identificatie van de syntactische context waarin de buigingsvariatie optreedt:

- (1) a. determinatorpositie: <DET-onz-enk-def>  
b. kernpositie: <NOM-onz-enk>, <NOM-dim-enk> en <NOM-inf>

De instanties van deze contexten worden in een aparte gegevensbank ondergebracht (zie paragraaf 15.3, tabel 15(5)). Voor de andere NP's is een verdere morfosyntactische specificatie overbodig binnen het kader van dit onderzoek, vanwaar het gebruik van algemene codes, zoals <DET-rest> of <NOM-rest>.

Ten tweede, iedere substantivische vorm waarmee twee genera geassocieerd kunnen worden, beschouwen we voorlopig als een *het*-woord. In een latere fase van de annotatieprocedure zal het genus van deze vormen manueel gecontroleerd worden (zie stap 3).

Ten derde, de vormen van het verkleinwoord worden als een apart inflectioneel paradigma beschouwd en krijgen later ook een apart lemma (zie stap 4). Deze analyse strookt niet met de traditionele visie dat het substantivische inflectionele paradigma vier woordvormen bevat die de categorieën getal (enkelvoud vs. meervoud) en gradatie ('-diminutief' vs. 'diminutief') uitdrukken, zoals geïllustreerd in tabel 15(3) voor de substantieven *boek* en *trui* (zie ook paragraaf 4.1, tabel 4(1)):

	enkelvoud	meervoud
-diminutief	<i>boek; trui</i>	<i>boeken; truien</i>
diminutief	<i>boekje; truitje</i>	<i>boekjes; truitjes</i>

Tabel 15(3): Traditioneel inflectioneel paradigma van een substantief

Het traditionele paradigma zoals voorgesteld in tabel 15(3) is problematisch voor het collocatieonderzoek. We mogen niet uit het oog verliezen dat de lexicale associatiekracht tussen het adjectief en het substantief als een verklarende factor voor de buigingsalternantie fungeert. Doordat het onverbogen adjectief taalstructureel enkel bij *het*-woorden kan voorkomen, is er een disproportie tussen het lemma van een *de*-woord en het lemma van een *het*-woord: het aandeel onverbogen adjectieven in het buigingsprofiel van het lemma van een *de*-woord, dat ook het onzijdige diminutief bevat, is bijgevolg nooit groter dan het aandeel onverbogen adjectieven in het lemma van een *het*-woord, dat eveneens de vormen van het diminutief bevat. Met andere woorden: het aandeel onverbogen adjectieven bij de woordvormen van het lemma *trui* (dat de woordvorm *truitje* bevat) is per definitie niet groter dan het aandeel onverbogen adjectieven bij de woordvormen van het lemma *boek* (dat de

woordvorm *boekje* bevat). Om dit euvel op te lossen, splitsen wij de lemmata van *boek* en *trui*, zoals voorgesteld in tabel 15(3), in twee aparte lemmata op: een eerste lemma voor de niet-gediminueerde vormen (*boek*<sub><NOM-onz-enk></sub> en *trui*<sub><NOM-rest></sub>) en een tweede lemma voor de vormen van het diminutief (*boekje*<sub><NOM-dim-enk></sub> en *truitje*<sub><NOM-dim-enk></sub>). Het paradigma voor de niet-gediminueerde vormen bevat de woordvormen voor het enkelvoud, *boek* en *trui*, en het meervoud, *boeken* en *truien*; het paradigma voor het verkleinwoord bestaat eveneens uit de woordvormen voor het enkelvoud, *boekje* en *truitje*, en het meervoud, *boekjes* en *truitjes*. Op deze wijze – het *de*-woord *trui* en het *het*-woord *truitje* vormen twee aparte lemmata – verwijderen we de contexten waar het onverbogen adjectief om structurele redenen nagenoeg uitgesloten is uit het onderzoek naar de invloed van de lexicale collocabiliteit op de buigingsalternantie. Bij deze keuze weten we ons gesteund door taalkundige argumenten. De diminuering is een voorbeeld van inherente inflectie en vertoont naast morfosyntactische kenmerken ook derivationale kenmerken (zie paragraaf 4.1 noot 7): primo, de diminutiefvorm is altijd onzijdig, ongeacht wat het genus van de basisvorm is; secundo, het verkleinwoord modificeert de lexicale betekenis van het substantief (Bakema 1998); tertio, het diminueringsmechanisme ligt aan de basis van gelexicaliseerde vormen, zoals *meisje*, *biertje* ('glas bier'), *jasje* ('colbertjas, blazer') of *lijfje* ('deel van een kledingstuk').

Het resultaat van deze annotatiefase is enigszins teleurstellend door het grote aandeel ambigue tokens, i.e. woordvormen waaraan meer dan één woordsoortcode is toegekend. Dit is een gevolg van de natuurlijke ambiguïteit van taal: *blind* en *blinde* kunnen allebei zowel een adjectief als een substantief zijn; *klinken* kan zowel een substantief als een infinitief zijn. Daarnaast wordt deze ambiguïteit versterkt door de volledigheid van de *Van Dale*-lijsten, met de inclusie van onder andere archaïsmen.<sup>2</sup> Een aantal determinatoren kunnen tot slot ook als substantief of werkwoord gebruikt worden, zoals *mijn*, *zijn*, *haar*, *ons*, *dat* of *het*.<sup>3</sup> De instanties van ambiguïteit waarin een determinatorvorm betrokken is, hebben we met een betrouwbaarheid van minstens 95% automatisch gedesambigüeerd.<sup>4</sup> De 24 contexten waar de ambiguïteit adjectivische en/of substantivische codes betreft, hebben we manueel gecontroleerd (n = 425,330).<sup>5</sup>

### Stap 3: Compilatie van het VDDS<sub>NP</sub>-corpus

Op basis van het VDDS-corpus met gedesambigüeerde woordsoortannotatie wordt vervolgens het NP-deelcorpus (VDDS<sub>NP</sub>) gecompileerd. We hebben alle twee- en drieplaatsige NP's die in een krantenkop voorkomen uit het VDDS<sub>NP</sub>-corpus verwijderd omdat er in krantenkoppen sprake is van een aparte stijl. Bij tweeplaatsige NP's stelt zich een bijkomend probleem: de NP moet namelijk effectief tweeplaatsig zijn. Een tournure wordt niet als een tweeplaatsige NP beschouwd – en bijgevolg uit de materiaalverzameling verwijderd – wanneer de adjectief-nomensequentie tot een grotere NP behoort (2), wanneer het adjectief door een bijwoord wordt gemodificeerd (3)<sup>6</sup>, of wanneer er sprake is van een conjunctie van adjectieven (4) of nominale kernen (5).

- (2) [een voor ons zeer *herkenbaar landschap*]<sub>NP</sub>  
 (3) hij heeft [ronduit *spectaculaire ideeën*]<sub>NP</sub>  
 (4) [een [aangenaam en *stemmig*]<sub>AP</sub> *arboretum*]<sub>NP</sub>  
 (5) [een [knappe actie] en [*strak schot*]]<sub>NP</sub>

Doordat we elke woordvorm die een onzijdig substantief kan zijn voorlopig als *het*-woord geannoteerd hebben (zie stap 2), moeten de bigenerische vormen afzonderlijk gecontroleerd worden (hiervoor hebben we dezelfde procedure toegepast als bij het CGN, zie paragraaf 6.2.2).<sup>7</sup>

De toepassing van deze criteria resulteert in een VDDSNP-corpus dat 1,495,347 woorden omvat. Tabel 15(4) toont de specifieke samenstelling van deze materiaalverzameling (de aantallen verwijzen naar de NP's).

Determinatorpositie <sup>8</sup>	AN-sequentie	NP <sub>3-plaatsig</sub>	NP <sub>2-plaatsig</sub>
<DET-onz-enk-def>	<ADJ-adj><NOM-onz-enk>	48,384	39,269
<DET-rest>		29,000	
<DET-onz-enk-def>	<ADJ-vd><NOM-onz-enk>	2,775	1,683
<DET-rest>		2,211	
<DET-onz-enk-def>	<ADJ-od><NOM-onz-enk>	743	274
<DET-rest>		591	
<DET-onz-enk-def>	<ADJ-adj><NOM-dim-enk>	1,216	206
<DET-rest>		2,093	
<DET-onz-enk-def>	<ADJ-vd><NOM-dim-enk>	64	0
<DET-rest>		92	
<DET-onz-enk-def>	<ADJ-od><NOM-dim-enk>	33	1
<DET-rest>		46	
<DET-onz-enk-def>	<ADJ-adj><NOM-inf>	820	1,026
<DET-rest>		135	
<DET-onz-enk-def>	<ADJ-vd><NOM-inf>	72	25
<DET-rest>		12	
<DET-onz-enk-def>	<ADJ-od><NOM-inf>	17	26
<DET-rest>		0	
<DET-rest>	<ADJ-adj><NOM-rest>	259,345	142,717
<DET-rest>	<ADJ-vd><NOM-rest>	17,805	5,963
<DET-rest>	<ADJ-od><NOM-rest>	4,080	2,043
Σ	<b>562,860</b> <b>(100.00%)</b>	<b>369,627</b> <b>(65.67%)</b>	<b>193,233</b> <b>(34.33%)</b>

Tabel 15(4): Samenstelling van het VDDSNP-corpus (aantal NP's)

#### Stap 4: Lemmatisering

De laatste fase van de taalkundige verrijking van de VDDSNP-materiaalverzameling is het lemmatiseren van de woordvormen. Op basis van de combinatie van de woordvorm en de woordsoortannotatie wordt het correcte lemma in de *Van Dale*-



woordsoortbestanden opgezocht. Ondanks het gebruik van de gedesambigueerde woordsoortannotatie (zie stap 2) zijn er nog enkele resterende instanties van ambiguïteit, waarbij één woordvorm met meerdere lemmata van dezelfde woordsoort kan worden geassocieerd, zoals *schepen* (enkelvoudsvorm vs. meervoudsvorm van *ship*) en *jongen* (enkelvoudsvorm vs. meervoudsvorm van *jong*). Deze gevallen zijn allen manueel gedesambigueerd.

### 15.2.2 *Compilatie van de materiaalverzameling met de buigingsalternantie*

Uit het VDDS<sub>NP</sub>-corpus, waarvan de structuur wordt beschreven in tabel 15(4), hebben we de materiaalverzameling geëxtraheerd die de definiete NP's met een enkelvoudig *het*-woord als hoofd bevat: [DET<sub>[DEF]</sub> A N<sub>[ONZ,ENK]</sub>]<sub>NP</sub>. De toepassing van de criteria die we hebben opgesteld voor het exciperen van het CGN (zie paragraaf 6.2.1) levert een materiaalverzameling met 54,124 observaties op. Deze observaties zijn vervolgens geverifieerd met betrekking tot de restricties die we voor de validatie van de observaties in de CGN-gegevensbank geformuleerd hebben (zie paragraaf 6.2.2). Omdat deze restricties ook in onderhavige materiaalverzameling strikt gerespecteerd worden, komen we hier niet op terug, met uitzondering van de superlatieven. Tegen de regel in worden namelijk 8 van de 6,861 superlatieven niet verbogen. Zeven van deze onverbogen superlatieven modifieren het substantief *aantal*. De vaststelling dat het substantief *aantal* geleidelijk naar een adjectivisch statuut (meer bepaald een kwantificerende modifier) evolueert, reikt een mogelijke verklaring aan voor het gebruik van de onverbogen superlatief in deze context: de afwezigheid van de buigings-*e* identificeert in deze context het adjectief als een bijwoord (7 van de 104 superlatieven bij *aantal* zijn onverbogen).

- (6)
  - a. Terwijl het grootst aantal rechtszaken burgerlijke zaken zijn
  - b. de achtervolging van Mario Cipollini op Alfredo Binda voor het grootst aantal Giro-ritzege
  - c. Laminaat is het belangrijkste product geworden van de concurrent Unilin.

We hebben alle superlatieven, inclusief de acht attestaties met een onverbogen vorm, uit de materiaalverzameling verwijderd.

Na de controle van de elementen die de drie posities in de NP bezetten, bekomen we een materiaalverzameling die 42,235 observaties telt.

### 15.3 **Resulterende materiaalverzameling**

Tabel 15(5) toont de verdeling van de adjectivische buigingsvarianten in de materiaalverzameling die we aan de hand van bovenvermelde criteria uit het VDDS-corpus geëxcerpeerd hebben. Deze frequentieverdeling bevestigt de vaststellingen

die we op basis van het CGN geformuleerd hebben (zie paragraaf 6.3, tabel 6(4)): aan de ene kant is de verbogen vorm onmiskenbaar de dominante buigingsvariant; aan de andere kant kan de onverbogen vorm niet als een epifenomeen worden afgedaan. Zelfs voor een Vlaamse kwaliteitskrant, waar de algemene – noordelijke – norm als streefdoel wordt gehanteerd (Permentier & Van Den Eynden 1998: 114), kunnen we met 95% betrouwbaarheid stellen dat tussen 30.33% en 31.21% van de adjectieven onverbogen blijft in een definitieve NP met een enkelvoudig onzijdig substantief als hoofd.

	n	%
<b>verbogen</b>	29,241	69.23
<b>onverbogen</b>	12,994	30.77
<b>Σ</b>	<b>42,235</b>	<b>100.0</b>

Tabel 15(5): Verdeling van de verbogen en de onverbogen adjectieven in het  $VDDS_{NP}$ -corpus

Het gebruik van krantenmateriaal creëert een nieuwe stilistische dimensie op basis van het communicatiemedium met het onderscheid tussen gesproken taal (CGN) en geschreven taal (*De Standaard*). Tabel 15(6) vat de buigingsprofielen samen voor het  $VDDS$ -corpus, het CGN en het Belgische deel van het CGN ( $CGN_{BEL}$ ). De tabelstatistieken tonen aan dat het communicatiemedium een significante invloed op de buigingsalternantie heeft, zowel voor de distinctie tussen het  $VDDS$ -corpus en het volledige CGN ( $\chi^2 = 118.3247$ ,  $df = 1$ ,  $p < 2.2e-16$ ) als voor de distinctie tussen het  $VDDS$ -corpus en het  $CGN_{BEL}$  ( $\chi^2 = 14.7703$ ,  $df = 1$ ,  $p = 0.0001214$ ).

	VDDS		CGN		CGN <sub>BEL</sub>	
<b>verbogen</b>	29,241	69.23	3,810	76.75	1,969	72.76
<b>onverbogen</b>	12,994	30.77	1,154	23.25	737	27.24
<b>Σ</b>	<b>42,235</b>	<b>100.0</b>	<b>4,964</b>	<b>100.0</b>	<b>2,706</b>	<b>100.0</b>

Tabel 15(6): Buigingsvarianten in  $VDDS$ , CGN en  $CGN_{BEL}$

Wanneer we de OR's beschouwen, stellen we vast dat het buigingsprofiel voor de geschreven taal in het  $VDDS$ -corpus significant meer onverbogen vormen bevat dan de buigingsprofielen voor de gesproken taal in het volledige CGN (OR = 1.4671 [1.3693,1.572]) of in het  $CGN_{BEL}$  (OR = 1.1872 [1.0882,1.2953]).

Een mogelijke verklaring voor het grotere aandeel onverbogen adjectieven in het  $VDDS$ -corpus vergeleken met het CGN en het  $CGN_{BEL}$  is dat de Vlaamse kwaliteitskrant *De Standaard* zich een Nederlands Nederlandse buigingsattitude aanmeet: het onverbogen buigingsmorfeem wordt aangewend om formele en ambtelijke situaties te markeren, zoals dit het geval is in het Nederlands Nederlands (zie paragraaf 12.2.2). Toegepast op een krant kunnen we dit als volgt herformuleren: het

#### DEEL IV

significant grotere aandeel onverbogen adjectieven in het VDDS-corpus houdt verband met de behandelde thema's, die een prominente positie verlenen aan de politieke actualiteit (zie ook paragraaf 15.1). Laat dit nu net de contexten zijn waar lexicale collocaties frequent voorkomen. Deze hypothese wordt in het volgende hoofdstuk gefalsificeerd (zie paragraaf 16.3).

## Noten

1. Het krantenmateriaal is ons via Dirk Geeraerts ter beschikking gesteld door *Van Dale Lexicografie*. Wij wensen hierbij *Van Dale Lexicografie* en Dirk Geeraerts te bedanken.
2. Wat te denken van werkwoorden als *kinderen* ‘kramen’ en *beurzen* ‘naar de beurs gaan, beurs houden’ of van het gewestelijke adjectivische gebruik van *vlag* als ‘zacht, zoel’.
3. De EGVD leert ons dat het gebruik als zelfstandig naamwoord van *dat* en *het* vrij zeldzaam is: *dat* komt meestal als diminutiefvorm voor in de tournure *een ditje en een datje*, of wordt ter vervanging van vloekwoorden gebruikt, zoals in *loop naar de dit en dat*; het substantivische gebruik van *het* is beperkt tot de psychologie om het ‘es’ aan te duiden.
4. De betrouwbaarheid van de automatische desambigueringsregels is telkens op een willekeurig staal van 250 observaties getest. Indien het aantal voorkomens van een ambigue context minder dan 250 bedraagt, hebben we voor een manuele desambiguering geopteerd.
5. Deze 24 contexten zijn: <ADJ-adj><NOM-onz-enk>, <ADJ-adj><NOM-rest>, <NOM-rest><NOM-inf>, <NOM-onz-enk><NOM-inf>, <NOM-onz-enk><NOM-dim-enk>, <NOM-dim-enk><NOM-rest>, <ADJ-adj><NOM-rest><NOM-inf>, <ADJ-vd><NOM-rest>, <ADJ-vd><NOM-onz-enk><NOM-inf>, <ADJ-od><NOM-onz-enk>, <ADJ-vd><NOM-inf>, <ADJ-vd><ADJ-od>, <ADJ-od><NOM-rest>, <ADJ-adj><NOM-rest><NOM-inf>, <ADJ-adj><NOM-onz-enk><NOM-inf>, <ADJ-adj><NOM-inf>, <ADJ-adj><ADJ-vd><NOM-rest><NOM-inf>, <ADJ-adj><ADJ-vd><NOM-rest>, <ADJ-adj><ADJ-vd><NOM-onz-enk>, <ADJ-adj><ADJ-vd><NOM-inf>, <ADJ-adj><ADJ-od><NOM-rest>, <ADJ-adj><ADJ-od><NOM-onz-enk>, <ADJ-adj><ADJ-vd>, <ADJ-adj><ADJ-od>.
6. De tests die we gebruikt hebben om de adverbiale modificatie te analyseren, zijn de mogelijkheid tot extrapositie en de mogelijkheid tot *clefting*. Een zin als (i) wordt wel opgenomen in het  $VDDS_{NP}$ -corpus omdat het bijwoord *bijna* het werkwoord en niet het adjectief modificeert.
  - (i) ze zijn intussen bijna *opgejaagd wild* geworden  
 → [extrapositie] bijna zijn ze intussen opgejaagd wild geworden  
 → [*clefting*] het is opgejaagd wild dat ze intussen bijna geworden zijn

Het voorbeeld in (ii) wordt uit het  $VDDS_{NP}$ -corpus verwijderd omdat het bijwoord *ronduit* het adjectief *spectaculaire* modificeert en bijgevolg tot de NP behoort:

  - (ii) hij heeft ronduit *spectaculaire ideeën*  
 → [extrapositie] \* ronduit heeft hij spectaculaire ideeën  
 → [*clefting*] \* het zijn spectaculaire ideeën die hij ronduit heeft vs. het zijn ronduit spectaculaire ideeën die hij heeft

#### DEEL IV

Indien we voorbeelden als (ii) in het  $VDDS_{NP}$ -corpus zouden opnemen, moeten we ook ‘drieplaatsige’ NP’s opnemen waar het adjectief door een bijwoord gemodificeerd wordt, zoals *het heel grote kind* of *de ronduit spectaculaire ideeën*.

7. Ook hier zorgt de rijkdom van het *Van Dale*-materiaal voor onverwachte ambiguïteiten: als *de*-woord duidt *president* een staatshoofd aan, als *het*-woord een weefsel.
8. Bij tweeplaatsige NP’s is de determinatorpositie leeg.

## Hoofdstuk 16:

### Lexicale conditionering

In dit hoofdstuk bestuderen we de invloed van lexicale collocaties op de buigingsalternantie in het VDDS-corpus. Deze analyse is in ruime mate een replicatie van de lexicale analyse op basis van het CGN (zie paragraaf 11.2.1). We vatten deze analyse aan met het rapporteren van enkele overzichtsstatistieken voor de lexicale collocabiliteit (16.1). Daarna bekijken we de impact van de lexicale associatiesterkte tussen het adjectief en het substantief op de buigingsalternantie (16.2). We sluiten dit hoofdstuk af met een stilistische zijsprong (16.3). We meten meer bepaald in welke mate het significant grotere aandeel onverbogen vormen in het VDDS-corpus ten aanzien van het CGN (zie paragraaf 15.3, tabel 15(6)) samenhangt met de aanwezigheid van lexicale collocaties.

Deze lexicale analyse wil de volgende hypothesen falsificeren, die we op basis van de literatuurstudie en de CGN-analyse naar voor kunnen schuiven:

- Adjectief-nomenparen die een lexicale collocatie vormen, selecteren significant vaker het onverbogen buigingsmorfeem dan adjectief-nomenparen die geen (significante) lexicale collocatie vormen.
- De kans op een onverbogen adjectief stijgt op significante wijze naarmate de associatiesterkte tussen het adjectief en het substantief groter wordt.
- Het significant grotere aandeel onverbogen adjectieven in de krantentaal is een gevolg van het grotere aandeel lexicale collocaties in het VDDS-corpus ten aanzien van het CGN en het CGN<sub>BEL</sub> (zie paragraaf 15.3).

#### 16.1 Overzichtsstatistieken voor de lexicale collocabiliteit van de adjectief-nomenparen in het *De Standaard*-corpus

We kijken eerst naar de token/type-ratio voor de adjectief-nomenparen. Deze maat bedraagt in het VDDS-corpus bijna het dubbele van deze in het CGN: 2.37 voor het VDDS-corpus tegen 1.42 voor het CGN. Dit leidt tot een forse stijging van het aandeel AN-tokens waarvoor een significante  $G^2$ -score kan worden berekend: terwijl het aandeel van de AN-lemmatypes met een significante  $G^2$ -score gelijk is in het krantenmateriaal en de gesproken taal (ongeveer één op acht), heeft meer dan de helft van de AN-lemmatokens in het VDDS-corpus een significante  $G^2$ -score (tegen één op drie in het CGN). Deze gegevens worden samengevat in tabel 16(1). We beschouwen de uitbreiding van het CGN-materiaal met het VDDS-materiaal bijgevolg als geslaagd: als gevolg van het repetitieve taalgebruik in een

kwaliteitskrant stellen we een stijging vast van het aantal adjectief-nomenparen die een significante lexicale collocatie vormen.<sup>1</sup>

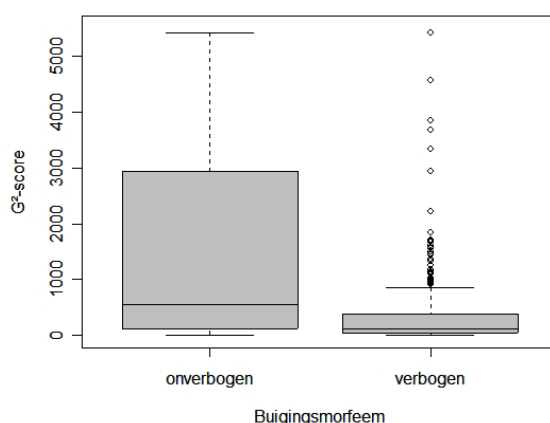
	VDDS	CGN
$n(\text{AN}_{\text{type}})$	2,241/17,829 (12.57%)	482/3,475 (13.87%)
$n(\text{AN}_{\text{token}})$	23,953/42,235 (56.71%)	1,668/4,964 (33.60%)

Tabel 16(1): Adjectief-nomenparen met een significante  $G^2$ -score in het VDDS-corpus en het CGN

Bovendien is een kwaliteitskrant in eerste instantie in de politieke en maatschappelijke actualiteit geïnteresseerd, een domein dat gekenmerkt wordt door lexicale collocaties. Dit blijkt enerzijds uit het grote aandeel lexicale collocaties in de Tweede-Kamerdebatten in het Nederlandse deel van het CGN. Anderzijds citeert de taalzuiveringsliteratuur – die journalisten tot haar publiek rekent – onder de rubriek adjectief-nomenparen met naamstatus veel voorbeelden uit het domein van de politiek (*inter al.* Cockx 2000: 30; Klein & Visscher 1996: 69; Penninckx & Buyse 1997: 40; Permentier & Van Den Eynde 1998: 182-183):

- (1) het kort geding, het openbaar kunstbezit, het Koninklijk Besluit, het Algemeen Nederlands Verbond, het burgerlijk wetboek, het administratief recht, het Europees Parlement, het Vlaams Parlement

We richten onze blik nu op de verdeling van de  $G^2$ -scores over beide buigingsvarianten. Deze verdeling wordt samengevat door de twee boxplots in figuur 16(1). Let wel: deze boxplots visualiseren enkel de verdeling van de AN-lemmatokens met een significante  $G^2$ -score. De plots voor de gehele materiaalverzameling zijn te onduidelijk doordat het merendeel van de observaties bij de verbogen buigingsvariant in de zone tussen 0 en 100 zijn samengepakt.



Figuur 16(1): Verdeling van de scores van de  $G^2$ -statistiek over beide buigingsvarianten in de observaties met een significante  $G^2$ -score in het VDDS-corpus

Deze boxplots tonen hetzelfde beeld als de boxplots voor het CGN (zie paragraaf 11.2.1, figuur 11(2)). Ten eerste, de boxplot voor de meetpunten met een onverbogen adjectief (linker boxplot) bezet een bredere zone op het  $G^2$ -continuüm dan de boxplot voor de meetpunten met een verbogen adjectief (rechter boxplot). Ten tweede, terwijl de adjectief-nomenparen met een verbogen adjectief de lagere  $G^2$ -scores realiseren, bezetten de observaties met een onverbogen adjectief de zone met de hogere collocabiliteitsgraden op het continuüm met de  $G^2$ -scores. Deze gegevens leveren een eerste bevestiging van de hypothese dat het onverbogen adjectief een sterke lexicale associatie tussen het adjectief en het substantief markeert. In de volgende paragraaf gaan we dieper in op de relatie tussen de lexicale collocabiliteit van het adjectief-nomenpaar en de buigingsalternantie in het VDDS-corpus.

## 16.2 Invloed van de lexicale collocabiliteit van het adjectief-nomenpaar op de buigingsalternantie

De bespreking van de invloed van de lexicale collocabiliteit op de buigingsalternantie bestaat uit drie delen. We starten de discussie met de binaire impact van de lexicale collocabiliteit: we vergelijken de buigingsprofielen van de adjectief-nomenparen die een significante lexicale associatie vormen met deze van de adjectief-nomenparen die geen significante associatie vertonen. Vervolgens richten we onze aandacht op de observaties met een lexicale collocatie: eerst kijken we naar de verdeling van beide buigingsvarianten over de collocabiliteitskwartielen; daarna verifiëren we of er een lineair verband bestaat tussen de lexicale associatiesterkte tussen een adjectief en een substantief en de kans op een onverbogen buigingsmorfeem.

De criteria om een adjectief-nomenpaar als een significante lexicale collocatie te beschouwen zijn dezelfde als bij de CGN-analyse (zie ook paragraaf 11.1.1.2 voor een uitgebreide discussie):

- $G^2(A_{\text{lemma}}, N_{\text{lemma}}) \geq 3.84$  (i.e. de kritische waarde voor het significantieniveau gelijk aan 0.05 in een 2x2-contingentietabel)
- $f_{\text{geobserveerd}}(A_{\text{lemma}}, N_{\text{lemma}}) > f_{\text{verwacht}}(A_{\text{lemma}}, N_{\text{lemma}})$
- in de vier cellen van de contingentietabel die de basis vormt voor de  $G^2$ -berekening is de geobserveerde frequentie groter of gelijk aan drie

We bespreken nu de resultaten van het empirisch onderzoek. Tabel 16(2) toont de buigingsprofielen van de deelverzameling met een significante  $G^2$ -score en het complement zonder significante  $G^2$ -score. Deze frequentieverdeling is significant ( $\chi^2 = 2,217.753$ ,  $df = 1$ ,  $p < 2.2e-16$ ).



	¬lex.col		lex.col		Σ	
<b>verbogen</b>	14,871	81.34	14,370	59.99	29,241	69.23
<b>onverbogen</b>	3,411	18.66	9,583	40.01	12,994	30.77
<b>Σ</b>	<b>18,282</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>	<b>42,235</b>	<b>100.0</b>

Tabel 16(2): Invloed van de lexicale collocabiliteit van het adjectief-nomenpaar op de adjectivische buigingsalternantie

In de krantentaal van het VDDS-corpus is de kans op een onverbogen adjectief in een adjectief-nomenpaar dat een lexicale collocatie vormt dubbel zo groot als in een adjectief-nomenpaar dat een toevallige combinatie is:  $RR_{\text{onverbogen}} = 2.1443$  [2.07259,2.2185]. De odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  bedraagt bijna het drievoud bij de lexicale collocaties: 2.9074 [2.7786,3.0422].

Laten we nu naar de invloed van de lexicale associatiesterkte tussen het adjectief en het substantief kijken. Hiervoor beperken we het materiaal tot de adjectief-nomenparen met een significante  $G^2$ -score ( $n = 23,953$ ). Voor het CGN hebben we vastgesteld dat de kans op een onverbogen adjectief progressief toeneemt naarmate de lexicale associatiekracht tussen het adjectief en het substantief groter is (zie paragraaf 11.2.1, tabel 11(6)). De gegevens voor de krantentaal vertonen dezelfde tendentie wanneer we de observaties met een significante adjectief-nomencollocatie in kwartielen indelen op basis van de  $G^2$ -score ( $\chi^2 = 3,125.364$ ,  $df = 3$ ,  $p < 2.2e-16$ ).

	lex.col <sub>kw1</sub>		lex.col <sub>kw2</sub>		lex.col <sub>kw3</sub>		lex.col <sub>kw4</sub>		Σ <sub>lex.col</sub>	
<b>verb</b>	4,742	79.23	4,258	71.28	3,407	56.85	1,963	32.71	14,370	69.23
<b>onverb</b>	1,243	20.77	1,716	28.72	2,586	43.15	4,038	67.29	9,583	30.77
<b>Σ</b>	<b>5,985</b>	<b>100.0</b>	<b>5,974</b>	<b>100.0</b>	<b>5,993</b>	<b>100.0</b>	<b>6,001</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>

Tabel 16(3): Invloed van de collocabiliteitsgraad van het adjectief-nomenpaar op de adjectivische buigingsalternantie

De kans op een onverbogen vorm stijgt progressief naarmate de lexicale associatie sterker wordt. Tabel 16(4) vat de  $RR_{\text{onverbogen}}$ 's en de OR's samen die de buigingsprofielen voor de collocabiliteitskwartielen vergelijken met de observaties waar het adjectief en het substantief geen significante associatie vormen ('¬lex.col', cf. tabel 16(2)). De cijfers tonen een graduele stijging van de kans op een onverbogen adjectief ( $RR_{\text{onv}}$ ) en van de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ .

	$RR_{\text{onv}}$	95% BI	OR	95% BI
<b>lex.col<sub>kw.1</sub> vs. ¬lex.col</b>	1.11	[1.05,1.18]	1.14	[1.06,1.23]
<b>lex.col<sub>kw.2</sub> vs. ¬lex.col</b>	1.54	[1.46,1.62]	1.76	[1.64,1.88]
<b>lex.col<sub>kw.3</sub> vs. ¬lex.col</b>	2.31	[2.22,2.41]	3.31	[3.11,3.53]
<b>lex.col<sub>kw.4</sub> vs. ¬lex.col</b>	3.61	[3.48,3.74]	8.97	[8.40,9.58]

Tabel 16(4):  $RR$ 's en  $OR$ 's die het buigingsprofiel van de collocabiliteitskwartielen vergelijken met het buigingsprofiel van de observaties zonder lexicale collocatie

Het progressief stijgende aandeel onverbogen vormen doet een lineair verband vermoeden tussen de kans op een onverbogen adjectief en de lexicale associatiesterkte. Om dit lineaire verband statistisch aan te tonen, zetten we de responsvariabele om in een ordinale variabele.<sup>2</sup> Hiervoor maken we gebruik van de deflectiegraad van het AN-lemmatype, i.e. het aandeel onverbogen vormen in het buigingsprofiel van het AN-lemmatype. Op basis van het aandeel onverbogen vormen onderscheiden we de volgende vijf deflectiegraden voor de AN-lemmatypes:

Deflectiegraad	P(onv.adj)
1 altijd verbogen	$P(\text{onv.adj}) = 0.0$
2 meestal verbogen	$0.0 < P(\text{onv.adj}) < 0.4$
3 twijfel	$0.4 \leq P(\text{onv.adj}) < 0.7$
4 meestal onverbogen	$0.7 \leq P(\text{onv.adj}) < 1.0$
5 altijd onverbogen	$P(\text{onv.adj}) = 1.0$

Tabel 16(5): Deflectiegraden van het AN-lemmatype

De MH- $\chi^2$  (type I) identificeert een significante lineaire associatie tussen de deflectiegraad van een AN-lemmatype en de collocabiliteitscore, ingedeeld in kwartielen (MH- $\chi^2 = 2,969.5435$ ,  $df = 1$ ,  $p < 2.2e-16$ ). Voor de zin en de omvang van deze lineaire associatie gebruiken we de  $\gamma$ -index:  $\gamma = 0.3834$  [0.3772,0.3836]. Deze score wijst op een positieve en significante associatie. Dit betekent dat de deflectiegraad van een AN-lemmatype volgens een lineair verband stijgt naarmate de associatiegraad tussen het adjectief en het substantief (uitgedrukt in kwartielen) toeneemt. De grootte van de  $\gamma$ -index moeten we als volgt interpreteren: wanneer we de rang van de onafhankelijke variabele kennen (c.q. het lexicale collocabiliteitskwartiel), reduceren we de fouten tegen het voorspellen van de rang van de responsvariabele (c.q. de deflectiegraad) met 38.34%, wat een zeer bevredigend resultaat is.

Tot slot kijken we naar de interne variatie: bij ruim 66.35% van de AN-lemmatypes (1,487/2,241) selecteren alle instanties dezelfde buigingsvariant. Dit wijst op een grote consistentie bij de keuze van de buigingsvariant in de geschreven taal.<sup>3</sup>

### 16.3 Lexicale collocabiliteit en medium: het buigingsverschil tussen gesproken en geschreven taal als een lexicaal neveneffect

Bij de presentatie van het VDDS-corpus hebben we vastgesteld dat de onverbogen buigingsvariant significant vaker geselecteerd wordt in het geschreven materiaal in het VDDS-corpus dan in het gesproken materiaal in het CGN en het CGN<sub>BEL</sub>, i.e. het Belgische deel van het CGN (zie paragraaf 15.3, tabel 15(6)). We ronden de analyse naar het buigingseffect van de lexicale collocabiliteit in het VDDS-corpus af door na

te gaan in welke mate het stilistische verschil tussen gesproken en geschreven taal samenhangt met het buigingseffect van de lexicale collocabiliteit in beide media.

Rekening houdend met de kenmerken van de krantentaal hebben we in paragraaf 15.3 de volgende hypothese naar voor geschoven:

- Het significant grotere aandeel onverbogen vormen in het krantenmateriaal is een neveneffect van het grotere aandeel lexicale collocaties in het VDDS-corpus.

Voor de falsificatie van deze hypothese gaan we met behulp van de partiële  $\chi^2$ -statistiek na of het aandeel lexicale collocaties significant groter is in het VDDS-corpus dan in het CGN of het CGN<sub>BEL</sub>.

We bekijken de verdeling van de lexicale collocaties in het VDDS-corpus enerzijds en in het CGN (tabel 16(6)) en het CGN<sub>BEL</sub> (tabel 16(7)) anderzijds (tussen ronde haken rapporteren we de absolute frequenties van iedere cel). Tabel 16(6) bevat de 95% BI'en voor de partiële  $\chi^2$ -statistiek van de frequentieverdeling die ontstaat wanneer we de aan-/afwezigheid van de lexicale collocabiliteit ('lex.col' vs. '-lex.col') uitzetten tegen het medium (VDDS vs. CGN).

	<b>-lex.col</b>	<b>lex.col</b>
<b>CGN</b>	[1.4854,1.5865] (3,296)	[0.5149,0.5716] (1,668)
<b>VDDS</b>	[0.9302,0.9515] (18,282)	[1.0411,1.0597] (23,953)

Tabel 16(6): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de lexicale collocaties in het CGN en het VDDS-corpus

Tabel 16(7) presenteert dezelfde informatie, maar dan voor de oppositie tussen VDDS en CGN<sub>BEL</sub>:

	<b>-lex.col</b>	<b>lex.col</b>
<b>CGN<sub>BEL</sub></b>	[1.4064,1.5406] (1,780)	[0.5787,0.6576] (926)
<b>VDDS</b>	[0.9588,0.9805] (18,282)	[1.0156,1.0333] (23,953)

Tabel 16(7): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de lexicale collocaties in het CGN<sub>BEL</sub> en het VDDS-corpus

De cijfers in de tabellen 16(6) en 16(7) tonen aan dat het aantal observaties met een lexicale collocatie significant groter dan verwacht is in het geschreven krantenmateriaal en significant kleiner dan verwacht in het gesproken materiaal, zowel voor het volledige CGN als voor het Belgische deelcorpus (CGN<sub>BEL</sub>). De observaties zonder (significante) lexicale collocatie vertonen het tegengestelde beeld: een significante ondervertegenwoordiging in het krantenmateriaal en een significante oververtegenwoordiging in het gesproken materiaal.<sup>4</sup>

De cijfers in de tabellen 16(6) en 16(7) tonen aan dat de adjectief-nomenparen die een lexicale collocatie realiseren significant oververtegenwoordigd zijn in het VDDS-corpus. Deze vaststelling geeft echter geen definitief uitsluitsel dat het grotere aandeel onverbogen vormen in het VDDS-corpus gerelateerd is aan het grotere aandeel lexicale collocaties. Daarom vergelijken we de buigingsprofielen in het VDDS-corpus enerzijds en het CGN en het CGN<sub>BEL</sub> anderzijds na exclusie van de lexicale collocaties: indien de significante oververtegenwoordiging van lexicale collocaties het grotere aandeel onverbogen vormen in het geschreven materiaal veroorzaakt, verwachten we dat het significante verschil tussen de materiaalverzamelingen met het geschreven en het gesproken materiaal verdwijnt na verwijdering van de observaties met een lexicale collocatie. Tabel 16(8) vat de buigingsprofielen samen voor het VDDS-corpus, het CGN en het CGN<sub>BEL</sub> na exclusie van de lexicale collocaties:

	VDDS		CGN		CGN <sub>BEL</sub>	
<b>verbogen</b>	14,871	81.34	2,722	82.58	1,406	78.99
<b>onverbogen</b>	3,411	18.66	574	17.42	374	21.01
<b>Σ</b>	<b>18,282</b>	<b>100.0</b>	<b>3,296</b>	<b>100.0</b>	<b>1,780</b>	<b>100.0</b>

Tabel 16(8): Buigingsprofielen voor het VDDS-corpus, het CGN en het CGN<sub>BEL</sub> na exclusie van de lexicale collocaties

Het verschil tussen de buigingsprofielen voor het VDDS-corpus en het volledige CGN is na exclusie van de lexicale collocaties niet langer significant ( $\chi^2 = 2.7819$ ,  $df = 1$ ,  $p = 0.09533$ ). Het VDDS-corpus en het CGN<sub>BEL</sub> daarentegen vertonen wel een significant verschil ( $\chi^2 = 5.717$ ,  $df = 1$ ,  $p = 0.0168$ ). Merk op dat dit significante verschil de aangereikte verklaring niet ondermijnt: de OR wijst op een lichte maar significante daling van het aandeel onverbogen vormen in het geschreven Belgisch Nederlands vergeleken met het gesproken Belgisch Nederlands (OR = 0.8623 [0.7648,0.9722]). Met andere woorden: na exclusie van de lexicale collocaties bevat het buigingsprofiel voor het geschreven Belgisch Nederlands significant minder onverbogen adjectieven dan het buigingsprofiel voor het gesproken Belgisch Nederlands. Deze constatering strookt met de formele(re) status van de verbogen variant in het Belgisch Nederlands.<sup>5</sup>

We concluderen dat het grotere aandeel onverbogen adjectieven in het krantenmateriaal een gevolg is van het formele(re) register en de door dit medium behandelde thema's. Ten eerste, het CGN verzamelt uiteenlopende gespreks-situaties: dit corpus is een amalgaam van gespreksonderwerpen met weinig thematische consistentie en weinig repetitiviteit tot gevolg. Een corpus dat de edities van negen opeenvolgende maanden van één jaargang van een krant bundelt, zoals het VDDS-corpus, behandelt daarentegen een kleiner aantal thema's en kent een grote(re) herhaling van de behandelde thema's. De grotere token/type-ratio voor adjectief-nomenparen in het VDDS-corpus onderbouwt dit argument (2.37 in het

VDDS-corpus tegen 1.42 in het CGN; zie paragraaf 16.1).<sup>6</sup> Ten tweede, doordat een (kwaliteits)krant in eerste instantie de politieke en maatschappelijke actualiteit verslaat, wordt het taalgebruik gekenmerkt door een groter aandeel gespecialiseerde termen – en lexicale collocaties –, zoals de gegevens in de tabellen 16(6) en 16(7) aantonen: bij neutra selecteert *De Standaard* het onverbogen buigingsmorfeem ter identificatie van adjectief-nomenparen die een lexicale eenheid vormen. Als gevolg vertoont het taalgebruik in het VDDS-corpus ten overstaan van het CGN en het CGN<sub>BEL</sub> een groter aandeel onverbogen adjectieven. Dit sluit aan bij het deflecterende effect dat we voor het Nederlands Nederlands in de Tweede-Kamerdebatten hebben vastgesteld en bij de visie van de ANS (1997), die één identieke stilistische functie voor beide nationale variëteiten naar voren schuift. Na exclusie van de lexicale collocaties bezet het VDDS-corpus binnen het Belgisch Nederlands de flecterende pool van het buigingscontinuum. Deze ambivalente positie van het VDDS-corpus ten aanzien van het CGN – en het CGN<sub>BEL</sub> in het bijzonder – onderstreept andermaal de complexiteit van de buigingsalternantie: het register van het VDDS-corpus bezit tegelijk flecterende (c.q. het formele/gecontroleerde taalgebruik) en deflecterende (c.q. het relatief grote aandeel lexicale collocaties) eigenschappen.

#### 16.4 Conclusie

De basisanalyse naar de impact van lexicale collocaties op de buigingsalternantie in het VDDS-corpus bevestigt de resultaten voor het CGN (paragrafen 16.1 en 16.2). Ten eerste, in adjectief-nomenparen die een lexicale collocatie vormen, selecteert het adjectief significant vaker het onverbogen buigingsmorfeem dan in adjectief-nomenparen die geen significante lexicale collocatie vormen. Ten tweede, de kans op een onverbogen adjectief stijgt progressief naarmate de lexicale associatiesterkte tussen het adjectief en het substantief groter wordt. We hebben bovendien aangetoond dat er een lineair verband bestaat tussen de collocabiliteitsgraad tussen het adjectief en het substantief enerzijds en de deflectiegraad van het AN-lemmatype, i.e. de kans dat een instantie van het AN-lemmatype een onverbogen adjectief selecteert, anderzijds. Ten derde, de verschillende instanties van de AN-lemmatypes vertonen een relatief grote consistentie bij de selectie van het adjectivische buigingsmorfeem, zoals de het percentage AN-lemmatypes zonder interne variatie aantoont. Tot slot, het stilistische verschil tussen gesproken en geschreven taal, met een significant groter aandeel onverbogen vormen in het krantenmateriaal, is een neveneffect van het significant grotere aandeel lexicale collocaties in het VDDS-corpus ten aanzien van het CGN en het CGN<sub>BEL</sub> (paragraaf 16.3).

Samengevat: de lexicale collocabiliteit van het adjectief-nomenpaar is een sterke indicator voor het gebruik van het onverbogen adjectief in de geschreven taal, zoals vertegenwoordigd door de Vlaamse kwaliteitskrant *De Standaard*.

In het volgende hoofdstuk concentreren we ons op de observaties met een significante  $G^2$ -score en onderzoeken we in hoeverre de lexicale collocaties en hun deflecterende effect op de buigingsalternantie een louter lexicaal fenomeen zijn. Concreet betekent dit dat we testen of er semantische tendensen aan de basis liggen van de adjectivische buigingsalternantie bij de lexicale collocaties.

## Noten

1. Het valt nog te bezien hoe de situatie zou zijn in een gesproken corpus dat (veel) groter is dan de door ons geraadpleegde 5<sup>e</sup> pre-release van het CGN.
2. Om statistisch een lineair verband te kunnen vaststellen, moeten beide variabelen minimaal tot het ordinale meetniveau behoren. Dit impliceert dat geen van beide variabelen tot het nominale meetniveau behoort.
3. We signaleren dat verscheidene AN-lemmatypes een marginale interne variatie vertonen. Hieronder verstaan we AN-lemmatypes waarvan slechts een zeer klein aantal attestaties niet de dominante buigingsvariant selecteren, zoals de volgende voorbeelden illustreren:
  - (i) openbaar vervoer ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{3}{299}$ ), openbaar ministerie ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{1}{176}$ ), Verenigd Koninkrijk ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{2}{280}$ ), gerechtelijk verlof ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{1}{50}$ ), gerechtelijk akkoord ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{1}{43}$ ), slecht nieuws ( $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{40}{1}$ ), Vlaams Blok (naam politieke partij;  $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{1}{1180}$ ), Liberaal Appel (naam politieke partij;  $\frac{n(\text{verbogen})}{n(\text{onverbogen})} = \frac{1}{186}$ )
4. De vergelijkingen tussen het aandeel lexicale collocaties in het CGN en het CGN<sub>BEL</sub> enerzijds en het VDDS-corpus anderzijds in de tabellen 16(6) en 16(7) zijn oneerlijk door de zeer verschillende corpusomvang. In een grotere dataset zijn patronen immers sneller significant. De OR's die de odds  $\frac{\text{lex.col}}{\text{-lex.col}}$  in het VDDS-corpus vergelijken met het CGN (OR = 2.5889 [2.4334, 2.7544]) en het CGN<sub>BEL</sub> (OR = 2.5185 [2.3209, 2.7329]) bevestigen de resultaten van de partiële  $\chi^2$ -statistiek in de tabellen 16(6) en 16(7).
5. We hebben twee afzonderlijke stratumanalyses uitgevoerd om het buigingseffect van de lexicale collocaties in de gesproken en de geschreven taal te vergelijken: een eerste stratumanalyse om het effect van de lexicale collocabiliteit te vergelijken in het VDDS-corpus en het volledige CGN; een tweede stratumanalyse om het effect van de lexicale collocabiliteit te vergelijken in het VDDS-corpus en het CGN<sub>BEL</sub>. Geen van beide stratumanalyses wijst op een significant verschillend buigingseffect van de lexicale collocaties in de gesproken en de geschreven taal.
 

We rapporteren eerst de stratumanalyse die de invloed van de lexicale collocabiliteit in het VDDS-corpus en het volledige CGN vergelijkt. Wanneer we het communicatiemedium, c.q. geschreven taal tegenover gesproken taal, controleren, blijft het deflecterende effect van de lexicale collocabiliteit behouden (MH- $\chi^2 = 2,405.5043$ ,  $df = 1$ ,  $p < 2.2e-12$ ; CMH-OR = 2.8666 [2.7461, 2.9924]). De *Breslow-Day*-test is niet significant ( $p = 0.0544$ ). Dit wijst op homogene OR's in beide stilistische strata.

We doen deze vergelijking nu over met de onderscheiding tussen VDDS en CGN<sub>BEL</sub> als stratumvariabele. Het buigingseffect van de aan-/afwezigheid van lexicale collocabiliteit blijft behouden wanneer we het communicatiemedium binnen het Belgische materiaal controleren (MH- $\chi^2 = 2,319.5355$ ,  $df = 1$ ,  $p < 2.2e-16$ ; CMH-OR = 2.8741 [2.7508, 3.003]). De zwak significante  $p$ -waarde voor de *Breslow-Day*-test ( $p = 0.0477$ ) wordt niet onderbouwd door de OR's voor de verschillende strata, zoals de overlappende

BI'en voor de strata gedefinieerd door  $CGN_{BEL}$  en het VDDS-corpus in tabel 16(i) aantonen:

<b>Stratum</b>	<b>OR</b>	<b>95% BI OR</b>
$CGN_{BEL}$	2.4239	[2.0336,2.8857]
VDDS	2.9074	[2.7786,3.0422]
CMH-OR	2.8741	[2.7508,3.003]

*Tabel 16(i): Samenvatting van de stratumanalyse voor de invloed op de buigingsalternantie van de aan-/afwezigheid van lexicale collocabiliteit in het  $CGN_{BEL}$  en het VDDS-corpus*

Samenvattend kunnen we stellen dat geen van beide stratumanalyses een uitgesproken significant verschillend effect van de lexicale collocabiliteit aanduidt in het algemeen gesproken (Belgisch) Nederlands aan de ene kant en het geschreven Nederlands dat in een Vlaamse kwaliteitskrant wordt gehanteerd aan de andere kant. Binnen het Belgisch Nederlands identificeert de *Breslow-Day*-statistiek een zwak significant effect dat niet bevestigd wordt door de 95% BI'en voor de OR's: het deflecterende effect van de lexicale collocaties is niet-significant groter in het geschreven Belgisch Nederlands.

- Ook hier moeten we erop wijzen dat de hogere token/type-ratio voor een deel te maken heeft met de omvang van het VDDS-corpus.



## **Hoofdstuk 17:**

# **Semantische conditionering van de adjectivische buigingsalternantie bij lexicale collocaties**

Tot nu toe hebben we het *De Standaard*-corpus (VDDS) gebruikt om een louter lexicale vraagstelling te behandelen. We hebben namelijk de invloed bestudeerd van de lexicale collocaties op de buigingsalternantie (hoofdstuk 16). In voorliggend hoofdstuk concentreren we ons op de semantische aspecten die verbonden zijn met de keuze van het buigingsmorfeem in de observaties die een lexicale collocatie vormen. In eerste instantie identificeren we de semantische restricties op de buigingsalternantie bij de observaties die een lexicale collocatie vormen (17.1). Daarna gaan we na in hoeverre de lexicale collocabiliteit een betrouwbare indicator is voor de identificatie van de semantische condities op het gebruik van de adjectivische buigingsvarianten (17.2). We ronden de semantische analyse af met een algemene discussie waarbij we dieper ingaan op de status van de lexicale collocaties (17.3). We trachten meer bepaald een antwoord te formuleren op de vraag of de invloed van de lexicale collocabiliteit op de buigingsalternantie een louter lexicaal – en derhalve idiomatisch en idiosyncratisch – effect is, of dat er semantische patronen aan de basis van het deflecterende effect van de lexicale collocaties geponeerd kunnen worden.

We onderstrepen dat deze semantische analyse beperkt is tot de deelverzameling met de lexicale collocaties ( $n_{\text{type}} = 2,241$ ;  $n_{\text{token}} = 23,953$ ; zie tabel 16(1)).

### **17.1 Semantische restricties op de buigingsalternantie in lexicale collocaties**

In het eerste deel van de semantische analyse richten we onze aandacht op de semantische conditionering van de buigingsalternantie bij lexicale collocaties. We lichten eerst de hypothesen en de semantische codering van de materiaalverzameling toe (17.1.1). Daarna rapporteren we de resultaten van het empirisch onderzoek (17.1.2). Deze analyse is een uitdieping van het onderzoek naar de invloed van de relationele adjectieven op de buigingsalternantie (zie paragraaf 11.2.2).

### 17.1.1 *Hypotheses en operationalisering van de semantische factoren*

Het onderzoek naar de semantische conditionering van de buigingsalternantie vertrekt van de volgende vier hypothesen die we op basis van het literatuuroverzicht geformuleerd hebben (zie paragraaf 1.1.3):

- Adjectief-nomenparen met een relationeel adjectief in de adjectivische positie vormen een faciliterende context voor het gebruik van het onverbogen buigingsmorfeem (Booij 2002a; Heynderickx 2001; zie ook paragraaf 11.2.2).
- In adjectief-nomenparen met een categoriserend adjectief in de adjectivische positie wordt de onverbogen buigingsvariant vaker geselecteerd dan in adjectief-nomenparen waar een ander element de adjectivische positie bezet (Blom 1994, 1995; Booij 2002a; Honselaar 1980).
- Wanneer het adjectief-nomenpaar als geheel een eigennaam is, wordt bij voorkeur het onverbogen adjectief geselecteerd (Broekhuis 1999: 210; Lebrun & Schurmans-Swillen 1966: 177; de Rooij 1980a: 12-13).
- Adjectief-nomenparen die een institutionele term vormen, vertonen een duidelijke voorkeur voor het onverbogen adjectief (ANS 1997: 408; Broekhuis 1999: 210; Cockx 2000: 30; Klein & Visscher 1996: 69; Lebrun & Schurmans-Swillen 1966: 177; de Rooij 1980a: 12-13).

Merk op dat deze hypothesen in twee groepen uiteenvallen: de eerste twee hypothesen verifiëren de invloed op de buigingsalternantie van adjectivische kenmerken, met name de relationaliteit en de categorialiteit van het adjectief; de laatste twee hypothesen testen semantische kenmerken die betrekking hebben op het adjectief-nomenpaar als geheel, met name de status van de adjectief-nomensequentie als eigennaam of als institutionele term.

Om een maximale intersubjectiviteit te garanderen bij de semantische analyse definiëren we voor ieder semantisch kenmerk een aantal operationaliseerbare criteria. Maar zelfs dan is de semantische annotatie van de observaties in de materiaalverzameling geen sinecure: bij iedere semantische factor worden we geconfronteerd met een restgroep van twijfelgevallen waarvoor het niet is uitgesloten dat verschillende onderzoekers een afwijkende codering voorstellen.<sup>1</sup> We zullen de criteria die we bij de semantische annotatie van de materiaalverzameling hebben toegepast, expliciteren en illustreren, zodat de lezer zich een duidelijk beeld kan vormen van de inhoud van de verschillende semantische labels. Bij de codering van de semantische eigenschappen maken we gebruik van een co-text van 20 woorden links en 20 woorden rechts van het AN-lemmatype. Daarnaast hebben we alle informatie gemaskeerd die de codering van de semantische kenmerken kan beïnvloeden, zoals informatie over het buigingsgedrag of de collocabiliteitsgraad van het adjectief-nomenpaar in kwestie. In geval van twijfel hebben we het AN-lemmatype in kwestie in drie afzonderlijke sessies

geanalyseerd, waarbij het alternatief geselecteerd wordt dat minstens tweemaal het beste lijkt.

We hebben de semantische kenmerken op het niveau van het AN-lemmatype gecodeerd, meer bepaald op basis van de eerste drie instanties van het AN-lemmatype in kwestie.<sup>2</sup> Op deze wijze reduceren we de annotatie-inspanning van 23,953 tokens tot 2,241 AN-lemmatypes, of preciezer tot 6,723 tokens. Alvorens we tot deze reductie van de gegevens zijn overgegaan, hebben we de semantische homogeniteit van de AN-lemmatypes getest op basis van een willekeurig staal van 150 AN-lemmatypes ( $n_{\text{token}} = 1,208$ ). Onder de semantische homogeniteit van een AN-lemmatype verstaan we de mate waarin de verschillende instanties van een AN-lemmatype dezelfde betekenis uitdrukken. We beschouwen een AN-lemmatype enkel als semantisch homogeen indien alle instanties aan de volgende twee voorwaarden voldoen:

- de verschillende instanties drukken dezelfde betekenis uit
- de verschillende instanties hebben een identieke codering voor de semantische parameters

In (1) bespreken we enkele semantisch heterogene AN-lemmatypes:

- (1) a. *Brits team*: het Britse nationale team vs. een team uit Groot-Brittannië
- b. *gezamenlijk vermogen*: het elektrische vermogen vs. het financiële vermogen
- c. *kloppend hart*: letterlijke, fysische betekenis vs. metaforische betekenis
- d. *Vlaams onderwijs*: het onderwijs in Vlaanderen vs. het onderwijs in Vlaanderen als institutie

In de staalname vertonen slechts 11 van de 150 AN-lemmatypes (7.33%) een heterogeen betekenisprofiel. De omvang van de semantische heterogeniteit is bovendien zeer beperkt: de 11 AN-lemmatypes zijn samen goed voor 126 tokens, waarvan er slechts 19 de niet-dominante betekenis van het AN-lemmatype instantiëren. Dit betekent dat we aan slechts 1.57% van de AN-tokens in de staalname een foutief semantisch label hebben toegekend. We concluderen bijgevolg dat de AN-lemmatypes die een lexicale collocatie vormen een grote semantische stabiliteit vertonen in het VDDS-corpus, wat de semantische codering op het typeniveau rechtvaardigt.

Na deze inleidende opmerkingen lichten we nu de operationalisering en de codering van de verschillende semantische parameters toe.

#### *Relationele adjectieven*

Voor de codering van de relationele adjectieven maken we gebruik van dezelfde procedure die we voor de CGN-analyse hebben voorgesteld (zie paragraaf 11.1.2). We komen hier dan ook niet op terug.

### *Categoriserende adjectieven*

Bij de analyse van de relationele adjectieven in het CGN (zie paragraaf 11.2.2) hebben we reeds gewezen op de categoriserende betekenis van sommige relationele adjectieven, zoals in *technisch onderwijs*, *mondeling examen* of *neuraal netwerk*. De combinatie van een categoriserend adjectief met een substantief definieert een subtype van het concept of de categorie aangeduid door het substantief in de nominale kernpositie: een mondeling examen is een specifiek soort examen (naast bijvoorbeeld een schriftelijk examen).<sup>3</sup> In de Angelsaksische literatuur spreekt men van “classifying adjectives” (Halliday 1994).

De categoriserende adjectieven hebben twee typische eigenschappen. Ten eerste, het adjectief-nomenpaar vormt een habituele combinatie. We wijzen erop dat lexicale collocaties uitermate geschikt zijn om habituele adjectief-nomenparen te identificeren. Ten tweede, het adjectief identificeert een kenmerk of een dimensie die saillant is voor het substantief. Dit verklaart waarom we in *bruin brood* en *Frans brood* (Belgisch Nederlandse variant voor *stokbrood*) wel van categoriserende adjectieven spreken en bij *vers brood* niet: versheid wordt als een minder essentiële eigenschap van brood ervaren. Dit blijkt bijvoorbeeld uit de labels in een bakkerij, die doorgaans de adjectieven *bruin*, *wit* en *Frans* vermelden, maar zelden gewag maken van *vers* of *oudbakken*. Het is dan ook geen toeval dat de categoriserende adjectieven de positie het dichtst bij de nominale kernpositie bezetten (Halliday 1994; Heynderickx 2001): *een vers bruin brood* klinkt aannemelijker dan *een bruin vers brood*, contrastieve contexten niet te na gesproken.

Voor de identificatie van de categoriserende adjectieven gebruiken we de volgende test: indien het adjectief een categoriserend adjectief is, vormt het adjectief-nomenpaar een antwoord op de vraag *wat voor (een) N?* of op de vraag *wat voor een soort N?* De geschiktheid van deze test blijkt uit de volgende beschrijving in de ANS (1997: 321): “De woordcombinaties *wat voor* en *wat voor een* worden gebruikt om een classificering te vragen voor de zelfstandigheid genoemd door het kernwoord”. Het vragend voornaamwoord *welk* is daarentegen ambigu tussen een identificerende en een classificerende functie. We verduidelijken deze test aan de hand van een voorbeeld. De sequentie *bruin brood* vormt een antwoord op de *wat voor*-vraag (2a), terwijl dat voor *vers brood* niet het geval is (2b):

- (2)      Wat voor een brood wil je?
- a. een bruin brood, een wit brood; een Frans brood, een stokbrood, een boerenbrood
- b. \* een vers brood, \* een oudbakken brood

We zijn ons terdege bewust dat er contexten zijn waarin het adjectief *vers* in *vers brood* als een categoriserend adjectief kan worden gebruikt. Daarom evalueren we

de *wat voor*-vraag voor ieder AN-lemmatype in de effectieve gebruikscontext van het corpus, c.q. in de contexten gedefinieerd door de eerste drie instanties van het AN-lemmatype (cf. *supra*). Bij de evaluatie van de *wat voor*-vraag hebben we ons bovendien laten leiden door de vaststelling dat adjectief-nomenparen met een categoriserend adjectief een subklasse definiëren op een lager niveau in een taxonomie (zie o.a. Cruse 1986: 136-155): op een bepaalde dimensie in een taxonomie bevinden zich meestal verscheidene concurrerende subtypes van een substantivisch concept (bv. een dimensie met *bruin brood* naast *wit brood*; een dimensie met *Frans brood/stokbrood* naast *boerenbrood*).<sup>4</sup>

De geciteerde voorbeelden tonen aan dat de relationele en de categoriserende adjectieven geen één op één relatie onderhouden: niet alle relationele adjectieven zijn categoriserende adjectieven en niet alle categoriserende adjectieven zijn relationele adjectieven, zoals de voorbeelden in tabel 17(1) illustreren:

		relationeel adjectief	
		ja	neen
categoriserend adjectief	ja	<i>ministerieel besluit,</i> <i>koninklijk besluit,</i> <i>economisch beleid, medisch</i> <i>dossier</i>	<i>bruin brood, groot</i> <i>seminarie, algemeen</i> <i>belang</i>
	neen	<i>Amerikaans leger, Brits</i> <i>parlement</i> <sup>5</sup> , <i>negentiende-</i> <i>eeuws België</i>	<i>bruin paard, groot huis,</i> <i>mooi gewaad</i>

Tabel 17(1): Interacties tussen de relationaliteit en de categorialiteit van het adjectief

Merk op dat de niet-categoriserende relationele adjectieven hoofdzakelijk een (minder essentiële) locatieve en temporele relatie met het concept van het nominale hoofd onderhouden.<sup>6</sup>

#### *Eigennamen met een woordgroepstructuur*

Dit zijn de adjectief-nomenparen die in hun geheel als eigennaam fungeren. We hebben het dan over woordgroepen zoals *Romeinse Rijk*, *Witte Huis*<sup>7</sup> of *Etnografisch Museum*.

Op basis van de analyse van enkele prototypische gevallen, zoals bovengenoemde voorbeelden, hebben we drie noodzakelijke en voldoende voorwaarden naar voren geschoven om van een eigennaam met woordgroepstructuur te kunnen spreken. Ten eerste, zowel het adjectief als het substantief beginnen met een kapitaal. Hierbij moeten we enkele minimale paren vermelden die nu eens met en dan weer zonder hoofdletter worden geschreven: *Burgerlijk Wetboek/burgerlijk wetboek*, *Stedelijk Zwembad/stedelijk zwembad*. Hoewel het onderscheid tussen beide schrijfwijzen niet altijd semantisch gemotiveerd is, zoals bij *Burgerlijk Wetboek/burgerlijk wetboek*, beschouwen we enkel de AN-lemmatypes waar beide vormen met een hoofdletter beginnen als een eigennaam. Ten tweede, de

aanwezigheid van twee kapitalen is een noodzakelijke maar geen voldoende voorwaarde, zoals blijkt uit *Amerikaans Hooggerechtshof* of *Frans Hof* (van *Cassatie*). In beide gevallen heeft het adjectief een specificerende functie: het adjectief relateert de referent van het nominale hoofd aan een bepaalde natie. Het tweede criterium stelt bijgevolg dat het adjectief in een eigennaam incidenteel niet-specificerend is, zoals dat in *Rode Kruis* of *Witte Huis* het geval is. Ten derde, in een eigennaam kunnen het adjectief en het substantief niet gescheiden worden door een ander adjectief: *het Belgische Rode Kruis*/\**het Rode Belgische Kruis*, *het Aalsterse Stedelijk Zwembad*/\**het Stedelijk(e) Aalsters(e) Zwembad*.

Ondanks het poneren van deze drie noodzakelijke en voldoende voorwaarden resten er een aantal twijfelgevallen. Dit zijn hoofdzakelijk instanties waar het adjectief van een geografische naam is afgeleid, zoals *Vlaams Artsensyndicaat* of *Europees Parlement*. In dergelijke gevallen hebben we op het internet gezocht naar evidentie voor het bestaan van een organisatie met het adjectief-nomenpaar als naam. Het bestaan van een website met het adjectief-nomenpaar als naam, zoals dit het geval is voor *Europees Parlement*, of het voorkomen van een acroniem dat zowel het adjectief als het substantief bevat, zoals VAS voor *Vlaams Artsensyndicaat*, beschouwen we als voldoende evidentie voor een eigennaamstatus (bovenop bovenvermelde condities).

#### *Institutionele termen*

We gebruiken de notie ‘institutionele term’ als verzamelnaam voor drie soorten adjectief-nomencombinaties: namen van instituties, standaardfuncties in een institutie of een organisatie, en officiële en ambtelijke termen. We benadrukken dat de adjectief-nomencombinatie in de drie subtypes als geheel een institutionele term vormt.

Onder namen van instituties verstaan we de adjectief-nomenparen die een institutie aanduiden. Dit zijn openbare organisaties en instellingen die binnen de verschillende maatschappelijke domeinen een taak vervullen:

- politiek: *Vlaams Parlement*, *Brussels Gewest*, *Witte Huis*
- administratief: *administratief centrum*, *federaal secretariaat*
- juridisch: *openbaar ministerie*, *grondwettelijk hof*
- militair: *militair hospitaal*, *militair domein*, *Centraal Commando*
- medisch: *Rode Kruis*, *universitair ziekenhuis*, *algemeen ziekenhuis*
- cultureel: *cultureel centrum*, *Etnografisch Museum*
- sportief: *stedelijk zwembad*
- educatief: *gewoon onderwijs*, *katholiek onderwijs*, *vrij net*

Deze sterk gefixeerde adjectief-nomencombinaties hebben meestal een collectieve betekenis: ze identificeren een groep van mensen en middelen als een eenheid. De aangehaalde voorbeelden illustreren eveneens de overlapping met de categorie der eigennamen.

De standaardfuncties identificeren wettelijk bepaalde functies binnen een institutie of organisatie, zoals *medisch personeel*, *technisch personeel* of

*onderwijzend personeel*. Tot deze groep rekenen we ook de namen van functies waar het collectieve betekenismoment niet aanwezig is, zoals *Vlaams parlementslid* of *Europees parlementslid*.

De laatste subklasse institutionele termen omvat de namen van officiële en ambtelijke concepten die niet in de twee vorige categorieën kunnen worden ondergebracht. Het zijn standaardbenamingen die meestal wettelijk zijn vastgelegd en een welomschreven betekenis hebben in een bepaald institutioneel kader:

- medisch: *medisch dossier*
- economisch: *forfaitair bedrag, primair saldo*
- politiek: *buitenlands beleid, Europees voorzitterschap*
- juridisch: *fiscaal recht, kort geding*
- administratief: *administratief dossier*

### 17.1.2 Resultaten en bespreking

In deze paragraaf testen we de empirische validiteit van de semantische hypothesen die we geformuleerd hebben met betrekking tot de selectie van de onverbogen buigingsvariant in adjectief-nomencollocaties. Het empirisch onderzoek bestaat uit twee delen: eerst kijken we naar de invloed op de buigingsalternantie van de afzonderlijke semantische factoren (17.1.2.1); daarna modelleren we de gecombineerde impact van deze factoren aan de hand van een logistische regressieanalyse (17.1.2.2). Op basis van het regressiemodel kunnen we bepalen in welke mate de buigingsalternantie bij de lexicale collocaties semantisch gemotiveerd is. Tot slot becommentariëren we de resultaten van het empirisch onderzoek naar de semantische conditionering van de buigingsalternantie bij lexicale collocaties (17.1.2.3).

#### 17.1.2.1 Bivariate analyses

##### Relationele adjectieven

Tabel 17(2) vat de impact van de relationele adjectieven op de buigingsalternantie samen. Deze frequentieverdeling is significant ( $\chi^2 = 2,841.902$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	-rel.adj		rel.adj		$\Sigma$	
<b>verbogen</b>	6,952	77.37	2,631	43.60	14,370	69.23
<b>onverbogen</b>	5,374	22.63	8,996	56.40	9,583	30.77
<b><math>\Sigma</math></b>	<b>12,326</b>	<b>100.0</b>	<b>11,627</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>

Tabel 17(2): Invloed van de relationaliteit van het adjectief op de adjectivische buigingsalternantie

De gegevens in tabel 17(2) ondersteunen de hypothese met betrekking tot het buigingsgedrag van de relationele adjectieven: bij een relationeel adjectief bedraagt de kans op een onverbogen vorm meer dan het dubbele van die bij een kwalificerend

adjectief ( $RR_{\text{onverbogen}} = 2.4924 [2.4019, 2.5865]$ ); voor de odds  $\text{onverbogen}/\text{verbogen}$  stellen we meer dan een verviervoudiging vast ( $OR = 4.4232 [4.1817, 4.6788]$ ). Deze resultaten bevestigen de tendentie die we in het CGN hebben waargenomen om het onverbogen buigingsmorfeem te selecteren in NP's waar het adjectief geen kwalificerende functie vervult, maar een relatie tussen twee substantivische concepten oproept (zie paragraaf 11.2.2).

#### *Categoriserende adjectieven*

De categorialiteit van het adjectief heeft een significante invloed op de buigingsalternantie ( $\chi^2 = 3,435.955$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	$\neg\text{cat.adj}$		$\text{cat.adj}$		$\Sigma$	
<b>verbogen</b>	13,220	69.41	1,150	23.43	14,370	69.23
<b>onverbogen</b>	5,825	30.59	3,758	76.57	9,583	30.77
<b><math>\Sigma</math></b>	<b>19,045</b>	<b>100.0</b>	<b>4,908</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>

Tabel 17(3): *Invloed van de categorialiteit van het adjectief op de adjectivische buigingsalternantie*

Het RR en de OR zijn conform met de hypothese die het categoriserend adjectief als een faciliterende factor voor de selectie van het onverbogen buigingsmorfeem bestempelt ( $RR_{\text{onverbogen}} = 2.5034 [2.4382, 2.5704]$ ;  $OR = 7.4164 [6.8951, 7.9772]$ ).

Relationele adjectieven zijn significant vaker dan verwacht ook categoriserende adjectieven (voor voorbeelden verwijzen we naar tabel 17(1)). Tabel 17(4) toont de resultaten van de partiële  $\chi^2$ -statistiek op basis van de kruistabel die de categoriserende adjectieven tegen de relationele adjectieven uitzet (tussen ronde haken rapporteren we de absolute tokenfrequentie voor iedere cel).

	$\neg\text{cat.adj}$	$\text{cat.adj}$
$\neg\text{rel.adj}$	[1.0784, 1.1108] (10,120)	[0.8957, 0.9256] (1,507)
$\text{rel.adj}$	[0.6016, 0.6634] (8,925)	[1.3046, 1.3885] (3,401)

Tabel 17(4): *95% BI en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de categoriserende adjectieven over de relationele adjectieven*

#### *Eigennamen met woordgroepstructuur*

Bij de adjectief-nomenparen die als eigenaam fungeren ('eig.naam'), stellen we een sterke stijging vast van de kans op een onverbogen adjectief ( $RR_{\text{onverbogen}} = 2.6191 [2.5496, 2.6905]$ ) en van de odds  $\text{onverbogen}/\text{verbogen}$  ( $OR = 7.8217 [7.2957, 8.3856]$ ) in vergelijking met de andere adjectief-nomenparen (' $\neg$ eig.naam').



	$\neg$ eig.naam		eig.naam		$\Sigma$	
<b>verbogen</b>	13,057	76.27	1,313	29.12	14,370	69.23
<b>onverbogen</b>	5,364	23.73	4,219	70.88	9,583	30.77
<b><math>\Sigma</math></b>	<b>18,421</b>	<b>100.0</b>	<b>5,532</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>

Tabel 17(5): Invloed van de eigennamen met woordgroepstructuur op de adjectivische buigingsalternantie

Wanneer we de elementen beschouwen die de adjectivische positie in een eigen-naam bezetten, stellen we vast dat de relationele adjectieven significant oververtegenwoordigd zijn. In tabel 17(6) rapporteren we de 95% BI'en van de partiële  $\chi^2$ -statistiek voor de distributie die ontstaat wanneer we de eigen-naamstatus van het adjectief-nomenpaar (' $\neg$ eig.naam' tegenover 'eig.naam') tegen de relationaliteit van het adjectief uitzetten (' $\neg$ rel.adj' tegenover 'rel.adj').

	$\neg$ eig.naam	eig.naam
<b><math>\neg</math>rel.adj</b>	[1.1487,1.1823] (10,422)	[0.4241,0.4734] (1,205)
<b>rel.adj</b>	[0.8287,0.8589] (7,999)	[1.4790,1.5609] (4,327)

Tabel 17(6): 95% BI'en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de relationaliteit van het adjectief over de eigennamen

De categoriserende adjectieven vertonen echter het tegengestelde beeld: op basis van de cijfers in tabel 17(7) stellen we een significante ondervertegenwoordiging vast van de categoriserende adjectieven bij de adjectief-nomenparen die een eigen-naam vormen en een significante oververtegenwoordiging bij de adjectief-nomenparen die geen eigen-naam vormen.

	$\neg$ eig.naam	eig.naam
<b><math>\neg</math>cat.adj</b>	[0.9535,0.9738] (14,115)	[1.0929,1.1487] (4,930)
<b>cat.adj</b>	[1.1099,1.1716] (4,306)	[0.4892,0.5729] (602)

Tabel 17(7): 95% BI'en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de categorialiteit van het adjectief over de eigennamen

De reden voor de ondervertegenwoordiging van de categoriserende adjectieven in de observaties waar het adjectief-nomenpaar een eigen-naam vormt, moeten we zoeken bij enkele hoogfrequente eigennamen waar het adjectief geen categoriserende

functie heeft, zoals *Vlaams Blok* (n = 1,181), *Verenigd Koninkrijk* (n = 282), *Witte Huis* (n = 292) of *Rode Kruis* (n = 200).

	<b>-eig.naam</b>	<b>eig.naam</b>
<b>¬cat.adj</b>	[0.9961,1.0148]	[0.9092,0.9989]
<b>cat.adj</b>	[0.9563,1.0085]	[1.0566,1.2371]

Tabel 17(8): 95% BI'en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de categorialiteit van het adjectief over de eigennamen (na exclusie van de eigennamen zonder categoriserend adjectief met een tokenfrequentie van minstens 200)

Wanneer we de attestaties van de eigennamen zonder categoriserend adjectief met een frequentie groter dan of gelijk aan 200 uit de materiaalverzameling verwijderen, treffen we in de adjectivische positie van de eigennamen met een woordgroepstructuur significant minder vaak dan verwacht een niet-categoriserend adjectief aan, zoals de 95% BI'en voor de partiële  $\chi^2$ -statistiek in tabel 17(8) aantonen: voor de andere cellen in deze tabel verschilt de geobserveerde frequentie niet significant van de verwachte frequentie.<sup>8</sup>

We besluiten dat de adjectivische positie bij eigennamen bij voorkeur door een relationeel adjectief of – zij het in mindere mate – een categoriserend adjectief wordt bezet.

#### *Institutionele termen*

De adjectief-nomenparen die een institutionele term vormen ('inst.term'), selecteren significant vaker de onverbogen buigingsvariant dan de adjectief-nomenparen die niet als institutionele term fungeren ( $\chi^2 = 8,406.895$ ,  $df = 1$ ,  $p < 2.2e-16$ ).

	<b>-inst.term</b>		<b>inst.term</b>		<b><math>\Sigma</math></b>	
<b>verbogen</b>	13,139	83.39	1,231	20.58	14,370	69.23
<b>onverbogen</b>	3,405	16.61	6,178	89.42	9,583	30.77
<b><math>\Sigma</math></b>	<b>16,544</b>	<b>100.0</b>	<b>7,409</b>	<b>100.0</b>	<b>23,953</b>	<b>100.0</b>

Tabel 17(9): Invloed van institutionele termen op de adjectivische buigingsalternantie

Een blik op het RR en de OR leert ons dat dit een zeer sterk effect is:  $RR_{\text{onverbogen}} = 4.0515$  [3.9254,4.1816];  $OR = 19.3658$  [18.0231,20.8085].

Laten we nu de elementen bekijken die de adjectivische positie in de institutionele termen bezetten. Tabel 17(10) toont de BI'en van de partiële  $\chi^2$ -statistiek voor de verdeling van de waarden voor de relationaliteit van het adjectief ('¬rel.adj' vs. 'rel.adj') over de adjectief-nomenparen die een institutionele term vormen ('inst.term') en de adjectief-nomenparen die geen institutionele term vormen ('¬inst.term'): bij de institutionele termen treffen we significant meer relationele adjectieven aan dan verwacht; bij de andere adjectief-nomenparen zijn de niet-relationele adjectieven significant oververtegenwoordigd.

	<b>-inst.term</b>	<b>inst.term</b>
<b>¬rel.adj</b>	[1.2034,1.2406] (9,814)	[0.4818,0.5264] (1,813)
<b>rel.adj</b>	[0.7745,0.8065] (6,730)	[1.4341,1.5014] (5,596)

Tabel 17(10): 95% BI'en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de relationaliteit van het adjectief over de institutionele termen

Tabel 17(11) bekijkt de verdeling van de waarden voor de categorialiteit van het adjectief over de institutionele termen en de complementaire verzameling zonder institutionele termen. De gegevens in deze tabel duiden op een significante oververtegenwoordiging van de categoriserende adjectieven in de adjectief-nomenparen die een institutionele term vormen.

	<b>-inst.term</b>	<b>inst.term</b>
<b>¬cat.adj</b>	[1.0891,1.1116] (14,474)	[0.7557,0.7961] (4,571)
<b>cat.adj</b>	[0.5854,0.6357] (2,070)	[1.804,1.934] (2,838)

Tabel 17(11): 95% BI'en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de categorialiteit van het adjectief over de institutionele termen

Net als bij een eigenaam wordt de adjectivische positie in een institutionele term significant vaker dan verwacht door een relationeel of een categoriserend adjectief bezet.

#### 17.1.2.2 Multivariate analyse

We sluiten de semantische analyse van de buigingsalternantie in lexicale collocaties af met een logistische-regressieanalyse om de semantische restricties op de keuze van het adjectivische buigingsmorfeem te modelleren. Ten eerste, de modelstatistieken geven een globaal beeld van de adequaatheid van de semantische factoren om de buigingsalternantie in lexicale collocaties te conditioneren. Ten tweede, rekening houdend met de sterke interactie tussen de semantische factoren is een regressieanalyse de enige mogelijkheid om de reële invloed van de verschillende factoren op de buigingsalternantie te meten. Een (meervoudig) regressiemodel berekent immers het effect op de responsvariabele van een verklarende variabele terwijl het effect van de andere verklarende variabelen gecontroleerd wordt. (Voor een algemene inleiding in de logistische-regressietechniek verwijzen we naar hoofdstuk 13.)

De responsvariabele is de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$ , of de logit van deze odds om precies te zijn. De vier semantische factoren die we in de vorige paragraaf aan de hand van bivariate analyses bestudeerd hebben, zijn de verklarende variabelen. Doordat de verklarende variabelen nominale variabelen zijn, maken we gebruik van dummy-variabelen. Tabel 17(12) vat de codering van de semantische factoren in het logistische-regressiemodel samen. We hebben telkens de factorwaarde die in de bivariate analyses het gebruik van het verbogen adjectief vergemakkelijkt als ijkwaarde gekozen.

Factor	Referentie	Waarde
a.rel (relationaliteit A)	neen	ja
a.cat (categorialiteit A)	neen	ja
an.eig naam (eigenaamstatus AN)	neen	ja
an.inst.term (institutionaliteit AN)	neen	ja

Tabel 17(12): Codering van de verklarende variabelen in het logistische-regressiemodel met de semantische factoren

Aangezien we geïnteresseerd zijn in de gecombineerde impact van de semantische factoren, hebben we voor de constructie van het logistische-regressiemodel de vier semantische factoren manueel ingevoerd.<sup>9,10</sup>

Laten we nu de resultaten van de regressieanalyse bekijken. We beginnen met de modelstatistieken:

- verklarende kracht:  $\text{variatie}_{\text{VERKLAARD}} = \text{variatie}_{\text{TOTAAL}} - \text{variatie}_{\text{RESIDU}}$   
 $= 32,243 - 20,881$   
 $= 11,362.14$

(gecorrigeerd a.h.v. AIC; zie paragraaf 13.2.1 noot 6)

$df = 4; p < 0.0000$

- voorspellende kracht:  $c\text{-score} = 0.859$

Zowel de verklarende als de voorspellende kracht bewijzen de deugdelijkheid van het voorgestelde model: het regressiemodel realiseert een significante reductie van de onverklaarde buigingsvariatie en de  $c$ -score wijst op een zeer degelijk model.<sup>11</sup> Dit betekent dat de keuze van de adjectivische buigingsvariant in de observaties waar het adjectief-nomenpaar een lexicale collocatie vormt op adequate wijze gemodelleerd kan worden op basis van de semantische kenmerken van het adjectief en het adjectief-nomenpaar.

Na de bespreking van de modelstatistieken richten we onze aandacht nu op de verklarende variabelen. Tabel 17(13) stelt de regressiecoëfficiënten en de OR's van de semantische factoren voor. De resultaten van de logistische-regressieanalyse bevestigen de resultaten van de bivariate analyses: de behandelde semantische kenmerken definiëren contexten waar de odds  $\frac{\text{onverbogen}}{\text{verbogen}}$  significant stijgen in vergelijking met de ijkwaarde, ook wanneer het effect van de andere semantische factoren gecontroleerd wordt.

Factor	Coëf	S.F	Wald Z	OR	95% BI OR	p
Intercept	-2.1393	0.0301	-71.02			0.000
a.rel=ja	0.8259	0.0361	22.84	2.2839	[2.1279,2.4513]	0.000
a.cat=ja	1.8174	0.0455	39.90	6.1558	[5.6306,6.73]	0.000
an.eig naam=ja	1.0336	0.057	18.10	2.8112	[2.514,3.1434]	0.000
an.inst.term=ja	1.9064	0.0519	36.69	6.7288	[6.078,7.4493]	0.000

Tabel 17(13): Regressiecoëfficiënten van de semantische factoren in het logistische-regressiemodel op basis van de lexicale collocaties in het VDDS-corporus

Een nadere inspectie van de scores voor de OR leert ons dat de semantische parameters in primaire en secundaire factoren kunnen worden opgedeeld: het deflectorende effect is significant groter bij de categoriserende adjectieven ('a.cat') en de institutionele adjectief-nomenparen ('an.inst.term') dan bij de relationele adjectieven ('a.rel') en de eigennamen ('an.eig naam').<sup>12</sup>

Een verklaring voor de lagere waarde van de OR van de relationele adjectieven in vergelijking met de categoriserende adjectieven moeten we zoeken bij de relationele adjectieven die afgeleid zijn van een geografische naam of van een tijdsbepalend substantief: deze adjectieven selecteren significant minder vaak het onverbogen buigingsmorfeem dan de andere relationele adjectieven – die bovendien vaak een categoriserende functie vervullen ( $\chi^2 = 243.3057$ ,  $df = 1$ ,  $p < 2.2e-16$ ;  $RR_{\text{verbogen}} = 1.3882$  [1.3309,1.4479];  $OR = 0.5627$  [0.5234,0.605]). De relationele adjectieven die van een geografische naam of van een temporeel substantief zijn afgeleid, roepen vaak een minder essentiële (locatieve of temporele) relatie tussen het adjectief en het nominale hoofd op dan de andere relationele adjectieven (zie paragraaf 17.1.1).

Bij de eigennamen is de relatief kleine OR een gevolg van de namen van geografische entiteiten die een sterke tot absolute voorkeur voor de buigings-*e* vertonen, zoals *Britse Imperium*, *Beloofde Land*<sup>13</sup>, *Verre Oosten*, *Nabije Oosten*, *Rode Plein*, *Britse Rijk*, *Ottomaanse Rijk*, *Romeinse Rijk*, *Arabisch(e) Schiereiland* of *Zwarte Woud*. Daarnaast zijn er vier andere eigennamen die consistent de buigings-*e* kiezen: *Rode Kruis*, *Witte Huis*, *Oude Testament* en *Nieuwe Testament*.

Bij een aantal AN-lemmatypes met een sterke voorkeur voor het verbogen adjectief lijkt de significante  $G^2$ -score het gevolg te zijn van een thematische scheefftrekking van de materiaalverzameling. Dit wil zeggen: bij een deel van de lexicale collocaties die niet het onverbogen adjectief selecteren, lijkt de collocatonele status samen te hangen met een thema dat 'toevallig' actueel was in de periode waarop de geraadpleegde edities van *De Standaard* betrekking hebben: *Amerikaans leger* (oorlog in Irak), *Iraaks leger* (oorlog in Irak), *Palestijns probleem/leiderschap* (Palestijnse kwestie), *arm zuiden* (mondiale politiek; globalisering), *wisselvallig weer* (weer), *voormalig personeel* (economie), *voormalig*

*management* (economie), *voorbij seizoen* (sport), *voorbij kwartaal* (economie), *verwacht resultaat* (economie).<sup>14</sup>

### 17.1.2.3 *Bespreking van de semantische conditionering van de buigingsalternantie*

De resultaten van het empirisch onderzoek zijn conform met de semantische hypotheses. De vier bestudeerde semantische factoren zijn significante indicatoren voor de buigingsalternantie bij lexicale collocaties: met betrekking tot de semantische kenmerken van het adjectief-nomenpaar identificeert het onverbogen adjectief een status als eigennaam of als institutionele term; met betrekking tot de semantische eigenschappen van het adjectief identificeert het onverbogen buigingsmorfem een relationeel of een categoriserend verband tussen het adjectief en het nominale hoofd.

Het onverbogen adjectief wordt dus niet uitsluitend geselecteerd om adjectief-nomenparen met een gelexicaliseerde betekenis te markeren, zoals dit prototypisch het geval is bij de eigennamen en de institutionele termen, maar ook om een relationele of categoriserende relatie tussen het adjectief en het nominale hoofd aan te duiden. Bovendien wordt de adjectivische positie in eigennamen en institutionele termen significant vaker dan verwacht door een relationeel of een categoriserend adjectief bezet. De ondervertegenwoordiging van de categoriserende adjectieven bij de adjectief-nomenparen die een eigennaam vormen, hebben we verklaard als een neveneffect van een tiental hoogfrequente eigennamen waar de adjectivische positie door een niet-categoriserend adjectief wordt bezet.

De modelstatistieken voor de logistische regressie tonen aan dat de keuze van het buigingsmorfem bij de adjectief-nomenparen die een lexicale collocatie vormen op adequate wijze gemodelleerd kan worden aan de hand van de semantische kenmerken van het adjectief en het adjectief-nomenpaar. Aangezien de vier semantische factoren in het logistische-regressiemodel een significant effect op de buigingsalternantie definiëren, kunnen we geen enkele van deze factoren verwijderen zonder een significante daling van de verklaarde variatie te veroorzaken. Wanneer we naar de omvang van het effect van deze semantische eigenschappen kijken (cf. de OR's in tabel 17(13)), stellen we vast dat er twee primaire en twee secundaire semantische condities op het gebruik van de buigingsvarianten bij de lexicale collocaties onderscheiden kunnen worden. Tabel 17(14) schematiseert de semantische restricties op het gebruik van het onverbogen adjectief, i.e. de waarden van de semantische factoren die het gebruik van het onverbogen buigingsmorfem motiveren.

	<b>Eigenschappen A</b>	<b>Eigenschappen AN-paar</b>
<b>primaire factoren</b>	categoriserend A	institutionele term
<b>secundaire factoren</b>	relationeel A	eigennaam

*Tabel 17(14): Overzicht van de semantische conditionering van het onverbogen adjectief in lexicale collocaties*

De primaire en de secundaire factoren zijn gelijk over de eigenschappen van het adjectief en het adjectief-nomenpaar verdeeld. Wat de kenmerken van het element in de adjectivische positie aangaat, stellen we vast dat het effect van een categoriserend adjectief groter is dan het effect van een relationeel adjectief. Dit onderbouwt de semantische stelling van Blom (1994, 1995), Booij (2002a) en Honselaar (1980), die het onverbogen adjectief met een categoriserende betekenis verbinden en het verbogen adjectief met een individualiserende betekenis, die de referent van het nominale hoofd modificeert (zie paragraaf 1.1.3.2). Wat de kenmerken van het adjectief-nomenpaar betreft, bedraagt het deflecterende effect bij een institutionele term bijna het drievoud van bij een eigenaam. We hebben aangevoerd dat dit het gevolg is van een aantal eigennamen met een sterke – tot absolute – voorkeur voor de buigings-*e*, waaronder veel namen van geografische entiteiten.

De vaststelling dat de buigingsalternantie bij de lexicale collocaties aan de hand van uitsluitend semantische parameters op adequate wijze gemodelleerd kan worden, vormt tot slot een eerste indicatie dat de lexicale collocaties een geschikte heuristiek zijn voor de identificatie van de semantische restricties op de buigingsalternantie. Deze problematiek vormt het onderwerp van de volgende paragraaf.

## **17.2 Relatie tussen de semantische condities en de lexicale collocaties**

Na de analyse van de semantische conditionering van de adjectivische buigingsalternantie bij de lexicale collocaties testen we nu in hoeverre deze semantische condities kenmerkend zijn voor de adjectief-nomenparen die een lexicale collocatie vormen. Het is immers niet uitgesloten dat er geen relatie is tussen de aanwezigheid van de deflecterende gemarkeerde waarden van de semantische factoren die we in de vorige paragraaf geanalyseerd hebben en de lexicale collocabiliteit van het adjectief-nomenpaar. De vraag die we in deze paragraaf trachten te beantwoorden, luidt als volgt: zijn de adjectief-nomencollocaties een betrouwbare indicator voor de identificatie van de relationele adjectieven, de categoriserende adjectieven en de adjectief-nomenparen die een eigenaam of een institutionele term vormen? Pas als deze vraag positief beantwoord wordt, kunnen we stellen dat deze semantische eigenschappen kenmerkend zijn voor (het deflecterende effect bij) de lexicale collocaties en dat de lexicale collocaties bijgevolg niet alleen lexicale patronen identificeren.

Eerst becommentariëren we de hypothesen en de operationalisering (17.2.1). Daarna presenteren we de resultaten van de empirische analyse (17.2.2).

### **17.2.1 Hypothesen en operationalisering**

Van Sterkenburg (1993) presenteert een semantische categorisatie voor de “gelexicaliseerde woordgroepen van het type A+N”. Met uitzondering van de relationele adjectieven, die Booij (2002a) en Heynderickx (2001) aan adjectief-nomenparen met naamstatus relateren, vinden we de vier onderscheiden semantische condities op de buigingsalternantie terug in de door Van Sterkenburg voorgestelde categorisatie: de categoriserende adjectieven, de eigennamen en de institutionele termen.<sup>15</sup> We schuiven bijgevolg de volgende algemene hypothese naar voren met betrekking tot de relatie tussen de bestudeerde semantische kenmerken en de lexicale collocabiliteit van het adjectief-nomenpaar:

- De lexicale collocabiliteit van het adjectief-nomenpaar vormt een betrouwbare indicator voor de identificatie van eigennamen, institutionele termen, relationele adjectieven en categoriserende adjectieven.

Voor de verificatie van deze hypothese gebruiken we twee diagnostische tests: een interne test en een externe test. Aan de hand van de externe test vergelijken we de realisatie van de semantische kenmerken in de materiaalverzameling met de lexicale collocaties en in een controlegroep zonder lexicale collocaties. De interne test meet binnen de materiaalverzameling met de lexicale collocaties de relatie tussen de semantische kenmerken enerzijds en de lexicale associatiesterkte tussen het adjectief en het substantief anderzijds. We lichten achtereenvolgens de externe en de interne test toe.

#### *Externe controletest*

Bij de externe test vergelijken we de verdeling van de waarden van de semantische factoren over de observaties die een significante collocatie vormen en een controlegroep met observaties die geen significante collocatie vormen. Tabel 17(15) presenteert de samenstelling van de controlegroep. Deze materiaalverzameling bevat 423 AN-lemmatypes die goed zijn voor 1,266 tokens. We hebben ons bij de samenstelling van deze controlegroep geconcentreerd op de AN-lemmatypes die geen significante collocatie vormen ( $p(G^2) > 0.05$ ) en de (significante) collocaties die significant minder vaak dan verwacht voorkomen ( $f_{\text{geobserveerd}}(A_{\text{lemma}}N_{\text{lemma}}) \leq f_{\text{verwacht}}(A_{\text{lemma}}N_{\text{lemma}})$ ).<sup>16</sup>

Selectie criterium	$n_{\text{type}}$	$n_{\text{token}}$
$f(A_{\text{lemma}}N_{\text{lemma}}) \geq 3 \ \& \ p(G^2) \leq 0.05 \ \&$	135	598
$f_{\text{geobserveerd}}(A_{\text{lemma}}N_{\text{lemma}}) \leq f_{\text{verwacht}}(A_{\text{lemma}}N_{\text{lemma}})$		
$f(A_{\text{lemma}}N_{\text{lemma}}) \geq 3 \ \& \ p(G^2) > 0.05$	17	126
$f(A_{\text{lemma}}N_{\text{lemma}}) = 2 \ \& \ p(G^2) > 0.05$	271	542
<b><math>\Sigma</math></b>	<b>423</b>	<b>1,266</b>

Tabel 17(15): Samenstelling van de controlegroep zonder lexicale collocaties

We testen de verdeling van de waarden van de semantische factoren over de lexicale collocaties en de controlegroep aan de hand van de partiële  $\chi^2$ -statistiek. De algemene hypothese kan derhalve als volgt geherformuleerd worden:

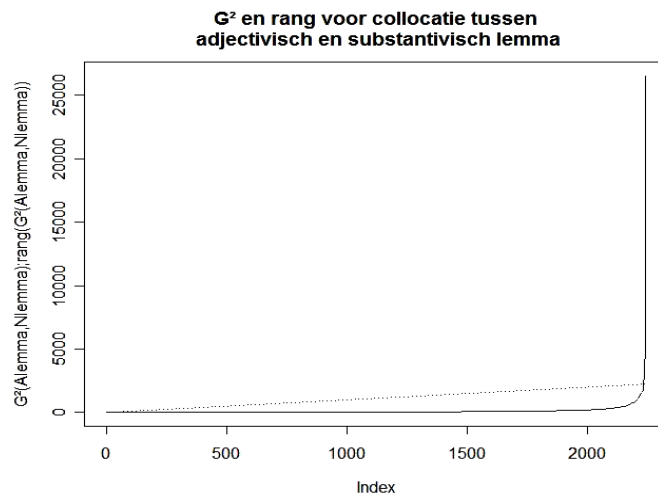


- De gemarkeerde waarden van de semantische factoren, c.q. de waarden die het gebruik van de onverbogen vorm motiveren (cf. tabel 17(12)), komen significant vaker voor dan verwacht in de materiaalverzameling met de lexicale collocaties en significant minder vaak dan verwacht in de controlegroep.
- De ongemarkeerde waarden van de semantische factoren treffen we significant vaker aan dan verwacht in de controlegroep en significant minder vaak dan verwacht in de groep met de lexicale collocaties.

#### *Interne controletest*

Zoals aangekondigd in de inleiding vullen we de externe test aan met een interne test. De doelstelling van deze interne controle is na te gaan wat het verband is tussen de semantische eigenschappen van het adjectief en het adjectief-nomenpaar aan de ene kant en de lexicale collocatiesterkte van het adjectief-nomenpaar aan de andere kant.

Voor het kwantificeren van de collocatiesterkte transformeren we de  $G^2$ -score in een collocabiliteitsrang. De  $G^2$ -score is immers geen lineaire maat; dit betekent dat de  $G^2$ -score niet volgens een rechte lijn stijgt wanneer we de verschillende AN-lemmatypes ordenen van de laagste  $G^2$ -score naar de hoogste  $G^2$ -score.<sup>17</sup> Dit wordt gevisualiseerd in figuur 17(1): op de X-as in figuur 17(1) zijn de 2,241 AN-lemmatypes gerangschikt in functie van hun oplopende  $G^2$ -score; de Y-as stelt de  $G^2$ -score en het rangordnummer voor de  $G^2$ -score van de AN-lemmatypes voor. De volle lijn in figuur 17(1) stelt de  $G^2$ -score voor tussen het adjectivische en het substantivische lemma, geordend van de laagste naar de hoogste waarde: in plaats van een gelijkmatige lineaire stijging van de curve stellen we eerst een zeer zachte en geleidelijke stijging vast; aan het rechteruiteinde van de curve zien we daarentegen een zeer sterke stijging van de  $G^2$ -score.



*Figuur 17(1): Stijging van de G<sup>2</sup>-score en van het rangordnummer voor de G<sup>2</sup>-score*

Deze curve toont duidelijk aan dat het verschil tussen de twee laagste G<sup>2</sup>-scores kleiner is dan het verschil tussen de twee hoogste G<sup>2</sup>-scores. Om deze ongerijmdheid op te lossen, transformeren we de G<sup>2</sup>-score in een lineaire maat: we rangschikken de AN-lemmatypes in functie van de grootte van de G<sup>2</sup>-score, vertrekkend bij de laagste en eindigend bij de hoogste score, en vervolgens kennen we aan ieder AN-lemmatype een oplopend rangordnummer toe, beginnend bij 1 en eindigend bij 2,241. Dit rangordnummer wordt in figuur 17(1) voorgesteld door de stippellijn, die een lineaire stijging vertoont.

Voor de interne controletest gebruiken we een ANOVA om de significantie en de zin van het verband te kwantificeren tussen de collocabiliteitsrang van een adjectief-nomenpaar enerzijds en de verschillende semantische kenmerken anderzijds. De ANOVA (Woods, Fletcher & Hughes 1986: 194-221) is de statistische techniek die standaard gebruikt wordt om het effect te meten van één of meerdere categorische onafhankelijke variabelen (c.q. de semantische eigenschappen van het adjectief of van het adjectief-nomenpaar) op een numerische responsvariabele (c.q. de collocabiliteitsrang van het adjectief-nomenpaar). In die hoedanigheid kan de ANOVA als een variant op de lineaire regressie worden beschouwd, met als specifieke eigenschap dat de verklarende variabelen categorisch in plaats van numerisch zijn (voor een algemene beschrijving van de lineaire regressie verwijzen we naar paragraaf 13.1). Een ANOVA test de H<sub>0</sub> dat de gemiddelde waarde van de responsvariabele (c.q. de gemiddelde collocabiliteitsrang van het adjectief-nomenpaar) in de verschillende groepen gedefinieerd door de verklarende variabele (bv. 'a.rel=neen' tegenover 'a.rel=ja' voor de relationaliteit van het adjectief) gelijk is. Net als bij de lineaire regressie wordt hiervoor de *F*-statistiek gebruikt.

Voor iedere ANOVA rapporteren we twee resultaten: ten eerste, op basis van de significantie van de associatie tussen een semantische factor en de collocatiesterkte van het adjectief-nomenpaar gaan we na of de collocabiliteitsrang significant verschilt voor de waarden van de semantische factor in kwestie; ten tweede, de regressiecoëfficiënt van de (deflecerende) gemarkeerde waarde van de verklarende variabele toont de zin van de associatie. We maken de lezer erop attent dat, net als bij de logistische-regressieanalyses (zie paragrafen 13.2.1 en 17.1.2), de nominale verklarende variabelen als dummy-variabelen gecodeerd worden. Hierbij gebruiken we de waarde die de selectie van de buigings-*e* faciliteert als referentiepunt (zie tabel 17(12)). Bij de relationele adjectieven fungeert bijgevolg de waarde 'a.rel=neen' als referentiewaarde (dummy-variabele = 0). De significantie van de afwijking van de parameterwaarden ten overstaan van de dummy-waarde wordt berekend door middel van de *t*-test, zodat de rapportering van de regressiecoëfficiënten vergelijkbaar is met de lineaire regressie.<sup>18</sup>

We voeren vier *one way* ANOVA's uit, waarbij telkens de invloed van één semantische factor op de collocabiliteitsrang wordt berekend.<sup>19</sup> De keuze om vier

*one way* ANOVA's uit te voeren, i.e. ANOVA's met één verklarende variabele, wordt gemotiveerd door de onderzoeksvraag: we zijn enkel geïnteresseerd in de associatie tussen de afzonderlijke semantische kenmerken enerzijds en de collocatiesterkte anderzijds. Anders gezegd: is de lexicale collocatiesterkte van het adjectief-nomenpaar een goede indicator van de afzonderlijke semantische kenmerken? Het modelleren van de collocabiliteitsrang in functie van de verschillende combinaties van semantische kenmerken overstijgt de doelstelling van dit onderzoek.

Voor de interne controletest passen we de globale hypothese als volgt aan:

- De lexicale collocabiliteitsrang tussen het adjectivische en het substantivische lemma vertoont een positieve significante associatie met de gemarkeerde semantische kenmerken, c.q. de aanwezigheid van een relationeel of een categoriserend adjectief en de aanwezigheid van een eigenaam of een institutionele term.

### 17.2.2 Resultaten en bespreking

We stellen achtereenvolgens de resultaten van de externe (17.2.2.1) en de interne controletests (17.2.2.2) voor.

#### 17.2.2.1 Externe controletest

##### *Relationele adjectieven*

De 95% BI'en voor de partiële  $\chi^2$ -statistiek in tabel 17(16) tonen aan dat de frequentie van de relationele adjectieven in de controlegroep significant kleiner is dan verwacht en dat de niet-relationele adjectieven significant vaker dan verwacht in deze materiaalverzameling voorkomen (tussen ronde haken rapporteren we de absolute frequenties voor iedere cel). Merk op dat de cijfers voor de groep met de lexicale collocaties ('lex.col') op een oververtegenwoordiging van de relationele adjectieven en een ondervertegenwoordiging van de kwalificerende adjectieven duiden, maar dat deze effecten net niet significant zijn (cf. de scharnierwaarde 1 behoort telkens nipt tot het 95% BI).<sup>20</sup>

	controlegroep	lex.col
<b>¬rel.adj</b>	[1.1512,1.3233] (770)	[0.9743,1.0006] (11,627)
<b>rel.adj</b>	[0.7034,0.8377] (496)	[0.9994,1.0249] (12,326)

Tabel 17(16): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de relationele adjectieven over de lexicale collocaties en de controlegroep

Deze cijfers staven de resultaten op basis van het CGN (zie paragraaf 11.2.2, tabel 11(10)).

*Categoriserende adjectieven*

De 95% BI'en voor de partiële  $\chi^2$ -test in tabel 17(17) tonen aan dat de adjectivische positie in een lexicale collocatie significant vaker dan verwacht door een categoriserend adjectief wordt bezet. NP's zonder significante associatie tussen het adjectief en het substantief bevatten daarentegen significant meer niet-categoriserende adjectieven en significant minder categoriserende adjectieven dan verwacht.<sup>21</sup>

	<b>controlegroep</b>	<b>lex.col</b>
<b>-cat.adj</b>	[1.0856,1.2144]	[0.9852,0.999]
	(1,169)	(19,045)
<b>cat.adj</b>	[0.3115,0.4659]	[1.0061,1.058]
	(97)	(4,908)

Tabel 17(17): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de categoriserende adjectieven over de lexicale collocaties en de controlegroep

*Eigennamen met woordgroepstructuur*

De lexicale collocaties vormen een goede indicator voor de identificatie van eigennamen met woordgroepstructuur. De gegevens in tabel 17(18) bewijzen dat de eigennamen significant ondervertegenwoordigd zijn in de controlegroep en significant oververtegenwoordigd zijn in de groep met de lexicale collocaties. De adjectief-nomenparen zonder eigenaamstatus vertonen het tegengestelde beeld.<sup>22</sup>

	<b>controlegroep</b>	<b>lex.col</b>
<b>-eig.naam</b>	[1.1922,1.3287]	[0.9788,0.9936]
	(1,245)	(18,430)
<b>eig.naam</b>	[0.0431,0.1077]	[1.0244,1.0733]
	(21)	(5,523)

Tabel 17(18): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de eigennamen over de lexicale collocaties en de controlegroep

*Institutionele termen*

Uit de cijfers in tabel 17(19) leiden we af dat we bij de lexicale collocaties significant meer institutionele termen aantreffen dan verwacht en dat in de controlegroep significant minder institutionele termen dan verwacht voorkomen.<sup>23</sup>

	<b>controlegroep</b>	<b>lex.col</b>
<b>-inst.term</b>	[1.3202,1.4197]	[0.9499,0.9679]
	(1,231)	(16,544)
<b>inst.term</b>	[0.0314,0.0626]	[1.0843,1.1271]
	(35)	(7,409)

Tabel 17(19): 95% BI'en partiële  $\chi^2$ -statistiek voor de verdeling van de institutionele termen over de lexicale collocaties en de controlegroep

### Bespreking

De empirische gegevens zijn conform met de hypothesen die we voor de externe controletest geformuleerd hebben: we treffen in de groep met lexicale collocaties significant meer categoriserende adjectieven, eigennamen en institutionele termen aan dan verwacht; in de controlegroep zonder lexicale collocaties komen deze elementen, samen met de relationele adjectieven, significant minder vaak voor dan verwacht. We concluderen dat de verschillende semantische eigenschappen typerend zijn voor de adjectief-nomenclollocaties.

In de volgende paragraaf gaan we na wat de omvang en de zin is van het verband tussen deze semantische kenmerken en de collocatiesterkte tussen het adjectief en het substantief.

#### 17.2.2.2 Interne controletest

Tabel 17(20) vat de resultaten van de ANOVA's samen. Om het overzicht te bewaren, rapporteren we voor ieder ANOVA-model (linker deel tabel) enkel het resultaat van de  $F$ -statistiek en de daarbijbehorende  $p$ -waarde (i.e.  $P(>F)$ ). Voor de gemarkeerde waarde van de verklarende factor (rechter deel tabel) rapporteren we de regressiecoëfficiënt en de bijbehorende  $p$ -waarde (i.e.  $P(>|t|)$ ). We benadrukken dat we voor iedere semantische factor een aparte *one way* ANOVA hebben uitgevoerd (Dit wordt weergegeven door de stippellijnen tussen de rijen in tabel 17(20)).

ANOVA			Regressiecoëfficiënt		
Factor	$F$ -statistiek	$p(>F)$	Factorwaarde	Coëf	$P(> t )$
a.rel	111.87	< 2.2e-16	a.rel=ja	10.58	< 2.2e-16
a.cat	105.14	< 2.2e-16	a.cat=ja	10.25	< 2.2e-16
an.eig naam	3,454.3	< 2.2e-16	an.eig naam=ja	58.77	< 2.2e-16
an.inst.term	4,364.8	< 2.2e-16	an.inst.term=ja	66.07	< 2.2e-16

Tabel 17(20): Samenvatting van de vier one way ANOVA's met de collocabiliteitsrang als responsvariabele en de semantische factoren als verklarende variabele

De gegevens in tabel 17(20) stemmen overeen met de hypothese. Ten eerste, de resultaten van de  $F$ -statistiek tonen aan dat voor iedere semantische parameter afzonderlijk de gemiddelde collocabiliteitsrang tussen het adjectief en het substantief significant verschilt in de (twee) groepen gedefinieerd door de waarden van deze semantische parameters. Ten tweede, de gemarkeerde waarde van iedere semantische factor heeft een positieve en significante regressiecoëfficiënt. Dit betekent dat de collocabiliteitsrang tussen het adjectivische lemma en het substantivische lemma stijgt in de volgende contexten:

- het element in adjectivische positie is een relationeel adjectief

- het element in adjectivische positie is een categoriserend adjectief
- het adjectief-nomenpaar is een eigennaam
- het adjectief-nomenpaar is een institutionele term

Hierbij moeten we signaleren dat de associatie het sterkst is voor de semantische kenmerken van het adjectief-nomenpaar als geheel.

### *17.2.2.3 Bespreking van de semantische tendensen aan de basis van de lexicale collocaties*

De resultaten van deze twee diagnostische tests ondersteunen de hypothesen die we geformuleerd hebben met betrekking tot de semantische basis van de lexicale collocaties. De empirische analyse heeft aangetoond dat de lexicale collocabiliteit een sterke indicator is voor de vier onderscheiden semantische condities op de buigingsalternantie bij de lexicale collocaties. Ten eerste, we treffen de semantische kenmerken die het gebruik van de onverbogen buigingsvariant motiveren significant vaker dan verwacht aan bij lexicale collocaties – met uitzondering van de relationele adjectieven (zie tabel 17(16)) – en significant minder vaak dan verwacht in de controlegroep zonder lexicale collocaties (paragraaf 17.2.2.1). Ten tweede, de waarden van de semantische categorieën definiëren significante verschillen ten overstaan van de lexicale collocabiliteitsgraad van het adjectief-nomenpaar: naarmate de collocatiesterkte tussen het adjectief en het substantief toeneemt, wordt de kans dat het adjectief-nomenpaar een gemarkeerde semantische eigenschap uitdrukt significant groter, zowel op het niveau van het adjectief (c.q. een relationeel en een categoriserend adjectief) als op het niveau van het adjectief-nomenpaar (c.q. een eigennaam en een geïnstitutionaliseerde term) (paragraaf 17.2.2.2). Omgekeerd betekent dit dat de lexicale collocaties niet beperkt zijn tot idiomatische uitdrukkingen. In dit verband moeten we echter in herinnering brengen dat de eigenamen en de institutionele termen in de ANOVA's de hoogste regressiecoëfficiënten hebben.

## **17.3 Algemene bespreking**

We sluiten de semantische analyse af met een algemene bespreking van de lexicale en semantische conditionering van de adjectivische buigingsalternantie bij lexicale collocaties. Hierbij trachten we de volgende vragen te beantwoorden:

- Wat zijn de lexicale en semantische restricties op het gebruik van beide buigingsvarianten bij lexicale collocaties?
- Wat is de status van de adjectief-nomencollocaties en van het onverbogen adjectief in deze tournures? Met andere woorden: zijn het idiomatische uitdrukkingen of definiëren ze regelmatige semantische patronen?

Vanuit een lexicaal perspectief heeft het onverbogen adjectief een iconische functie: het gebruik van het onverbogen buigingsmorfeem identificeert het adjectief-

nomenpaar als een lexicale collocatie. Dit wil zeggen: het adjectief-nomenpaar is een habituele combinatie die een hechte lexicale eenheid vormt. In die hoedanigheid kan de adjectief-nomencollocatie als een eenheid geactiveerd worden. Het verbogen adjectief signaleert daarentegen de – vanuit taaltheoretisch standpunt – prototypische situatie waar het adjectief en het substantief vrij tot een NP geassembleerd kunnen worden.

Deze bespreking van de lexicale collocabiliteit vereist een aanvullende semantische interpretatie. We hebben aangetoond dat lexicale collocaties significant geassocieerd zijn met semantische eigenschappen van het adjectief en van het adjectief-nomenpaar, meer bepaald met de waarden die het gebruik van het onverbogen buigingsmorfeem motiveren. Op het niveau van het adjectief identificeert het onverbogen buigingsmorfeem een niet-prototypisch verband tussen het adjectief en het substantief, hetzij een relationeel hetzij een categoriserend verband. In het eerste geval roept het adjectief een relatie tussen twee nominale concepten op: in *prinselijk huwelijk* identificeert het adjectief een bezitsrelatie tussen het nominale concept aan de basis van het relationele adjectief ('prins(es)') en het concept aangeduid door het nominale hoofd ('huwelijk'). In het tweede geval identificeert het adjectief een saillante subcategorie van het concept aangeduid door het substantief: in *algemeen reglement* verwijst het adjectief-nomenpaar naar een specifiek soort *reglement* dat van toepassing is in alle geledingen van een organisatie. Het is duidelijk dat het adjectief noch in *prinselijk huwelijk* noch in *algemeen reglement* de referent aangeduid door het substantief kwalificeert: een *prinselijk huwelijk* is niet noodzakelijk weelderig en groots; de voorschriften in een *algemeen reglement* zijn niet noodzakelijk universele principes. Merk op dat dit twee productieve semantische categorieën zijn die niet gelimiteerd zijn tot idiomatische uitdrukkingen.

De sterke voorkeur van de categoriserende adjectieven voor het onverbogen buigingsmorfeem vormt evidentie ter ondersteuning van de semantische modellen voorgesteld door Blom (1994, 1995), Booij (2002a) en Honselaar (1980): de empirische gegevens in het VDDS-corpus onderbouwen de hypothese dat het onverbogen adjectief een categoriserende relatie identificeert, waarbij het kenmerk uitgedrukt door het adjectief niet de individuele referent maar (een aspect van) het substantivische concept specificiert. Ondanks de status als primaire factor van de categorialiteit van het adjectief (zie tabel 17(14)) is een logistische-regressiemodel met enkel deze semantische factor als verklarende variabele onvoldoende om de buigingsalternantie bij lexicale collocaties adequaat te modelleren, zoals kan worden afgeleid uit de modelstatistieken voor de verklarende kracht ( $32,243 - 28,798 = 3,444.93$ ,  $df = 1$ ,  $p = 0.00001$ ) en de voorspellende kracht ( $c = 0.656$ ) van het enkelvoudige logistische-regressiemodel met de categorialiteit van het adjectief ('a.cat') als verklarende variabele.<sup>24</sup>

Wanneer we naar de semantische eigenschappen op het niveau van het adjectief-nomenpaar kijken, stellen we vast dat de lexicale collocaties vaak een eigenaam of

een institutionele term identificeren. Dit blijkt duidelijk uit de hoge regressie-coëfficiënten in de ANOVA's (zie tabel 17(20)). In het eerste geval vormt het adjectief-nomenpaar een lexicale en semantische eenheid, waarbij het adjectief (incidenteel) zijn specificerende functie verliest. In het tweede geval, wat in feite een afgezwakte versie van het eerste is, fungeert de sequentie van het adjectief en het substantief als een officiële en institutionele benaming. Hoewel eigennamen en institutionele termen typische instanties van idiomaticiteit zijn, is het opmerkelijk dat deze adjectief-nomenparen significant vaker dan verwacht gebruik maken van relationele en categoriserende adjectieven (zie paragraaf 17.1.2.1), zoals in *medisch dossier*, *algemeen stemrecht*, *S/stedelijk Z/zwembad*, *militair domein* of *technisch personeel*. Deze vaststelling suggereert dat de semantische structuur van een aantal eigennamen en institutionele termen transparant(er) is dan op basis van hun idiomatische karakter vermoed kan worden.

Kortom: het onverbogen adjectief wordt geselecteerd in contexten waar het adjectief en het substantief een lexicale eenheid vormen. Dit zijn meestal adjectief-nomenparen waar het adjectief geen prototypische kwalificerende functie heeft. Dit sluit aan bij Barlows (1991, 1999) discursieve congruentiemodel (zie paragraaf 4.2.2): de ongemarkeerde buigings-*e* duidt aan dat het adjectief informatie verschaft over een referent in het discours; het gemarkeerde onverbogen buigingsmorfeem signaleert op iconische wijze de situatie waar het adjectief niet de referent aangeduid door het nominale hoofd kwalificeert, en bijgevolg een afwijking van de prototypische situatie, hetzij door een relatie tussen twee nominale concepten op te roepen, hetzij door een subklasse van het nominale hoofd te identificeren, hetzij door het adjectief-nomenpaar als een eigenaam of een institutionele term te identificeren, hetzij door een combinatie van deze functies.

We benadrukken dat de verschillende semantische condities op de selectie van het onverbogen buigingsmorfeem heterogeen van aard zijn, zodat ze niet tot één schematische restrictie gereduceerd kunnen worden. Het gemeenschappelijke kenmerk van de contexten waar het onverbogen adjectief geselecteerd wordt, is dat het een afwijking van de prototypische NP betreft. Het onverbogen adjectief signaleert op iconische wijze een semantische specialisatie die samenhangt met de collocatieve status van het adjectief-nomenpaar:

- *relationeel adjectief*: Dirven (1999: 60-61) wijst erop dat het relationele adjectief en het nominale hoofd een “close conceptual unity” vormen “through a predication link underlying the relational adjective and the noun”. Booi (2002a) en Heynderickx (2001) verwoorden een analoge visie. Dit verklaart de associatie tussen de relationele adjectieven en de lexicale collocaties.
- *categoriserend adjectief*: Doordat de combinatie van een categoriserend adjectief en een substantief een subtype, c.q. subklasse, van het substantief in kwestie identificeert, opereert het geheel als een benoemingsprocédé. In dit opzicht is het interessant dat er twijfel bestaat over de status van



adjectief-nomenparen met een categoriserend adjectief: *grootseminarie* (EGVD) tegenover *groot seminarie* (CGN), *bruinbrood* (EGVD) tegenover *bruin brood* (VDDS).<sup>25</sup> De lexicale collocabiliteit is bovendien uitermate geschikt om het habituele karakter van de combinatie van een categoriserend adjectief en een substantief te vatten.

- *eigenaam*: De relatie tussen de eigennamen en de lexicale collocabiliteit ligt voor de hand, aangezien het adjectief in een adjectief-nomenpaar met eigenaamstatus (incidenteel) zijn specificerende betekenis verliest.
- *institutionele term*: Typerend voor de institutionele termen is dat het adjectief-nomenpaar als een eenheid opereert bij de identificatie van een ambtelijk of institutioneel concept.

Dit brengt ons bij de tweede vraag, met name de vraag naar de status van de lexicale collocaties, en bij uitbreiding van het onverbogen buigingsmorfeem in de adjectief-nomencollocaties. We brengen in herinnering dat we bij deze discussie de andere restricties op de buigingsalternantie, die we op basis van het CGN geïdentificeerd hebben (zie hoofdstuk 14), buiten beschouwing laten.<sup>26</sup>

Binnen bepaalde taalkundige tradities wordt het onverbogen adjectief doorgaans afgedaan als een markeerder van idiomatieiteit (*inter al.* de Haas & Trommelen 1993; Odijk 1992; Van Eynde 2003). We hebben dit de marginaliserende attitude ten overstaan van de buigingsalternantie genoemd (zie paragraaf 1.2.2). Het voorkomen van het onverbogen adjectief wordt op deze wijze gelimiteerd tot een aantal idiosyncratische gevallen. Kenmerkend voor deze attitude is de analyse van Odijk (1992), die een “Formal Lexicalisation”-regel voorstelt: deze regel houdt in dat er een lexicale knoop N boven de syntagmatische knoop N' wordt geplaatst, zodat de sequentie van het adjectief en het substantief ‘onzichtbaar’ is voor het toekennen van het buigingsmorfeem. Met andere woorden: de sequentie van adjectief en substantief wordt als een lexicale eenheid beschouwd die als zodanig in het lexicon wordt opgenomen. De redenering aan de basis van Odijks analyse en de marginaliserende attitude in het algemeen kunnen we op de volgende wijze parafraseren:

De congruentieregels van de Nederlandse grammatica voorzien dat het adjectief in een definiete NP met een enkelvoudig *het*-woord als hoofd een buigings-*e* krijgt. Wanneer de buigings-*e* in deze context ontbreekt, is de NP ofwel agrammaticaal ofwel niet door middel van de regels van de Nederlandse grammatica afgeleid. Bijgevolg is deze tournure een idiomatische uitdrukking die in het lexicon moet worden opgeslagen.<sup>27</sup>

Deze redenering vertrekt van een dichotomische visie op de taalkundige kennis die een grammaticale en een lexicale component onderscheidt: de grammaticale component bevat de abstracte regels om de regelmatige en productieve patronen van een taal te beschrijven; het lexicon is de restklasse die alle elementen en combinaties

van elementen bevat die niet door het regelsysteem van de grammatica kunnen worden verklaard. Deze visie, die zowel in het structuralisme als de Generatieve Grammatica wordt aangehangen, heeft de twintigste-eeuwse manier van denken over taal beheerst.

Op basis van de resultaten en de methodologie van voorliggend onderzoek kunnen wij noch deze dichotomische visie op taalkundige kennis noch de idiomatische visie op het onverbogen adjectief verdedigen.

Ten eerste, de resultaten van het lexicaal onderzoek en het semantisch onderzoek bewijzen dat de keuze van het onverbogen adjectief niet tot idiosyncratische instanties kan worden gereduceerd, maar dat er een lexicale en semantische regelmaat ten grondslag ligt aan de buigingsalternantie: de verklarende en de voorspellende kracht van het logistische-regressiemodel geven aan dat de buigingsalternantie bij de lexicale collocaties op adequate wijze gemodelleerd kan worden op basis van de bestudeerde semantische condities (zie paragraaf 17.1.2.2). Ten eerste, de kans op een onverbogen vorm neemt progressief toe naarmate de lexicale associatie tussen het adjectivische en het substantivische lemma sterker wordt. Dit is bovendien een lineair verband (zie paragraaf 16.2). Ten tweede, de lexicale collocabiliteit is een significante indicator voor de semantische kenmerken die het gebruik van het onverbogen adjectief motiveren (zie paragraaf 17.2). Het deflecterende effect van de lexicale collocabiliteit vertoont bijgevolg regelmatige patronen en valt samen met (gemarkeerde) semantische categorieën. In een ruimere onderzoekscontext naar de grammaticale status van gelexicaliseerde adjectief-nomenparen wijzen Booij (2002a), De Caluwe (1990) en Hüning (2004) op de productiviteit van dit mechanisme als benoemingsprocédé. Hierbij valt op dat het syntagmatische proces productiever is dan het concurrerende morfologische proces, c.q. de AN-samenstellingen (*kleinkind, grootouder, roodrok*).

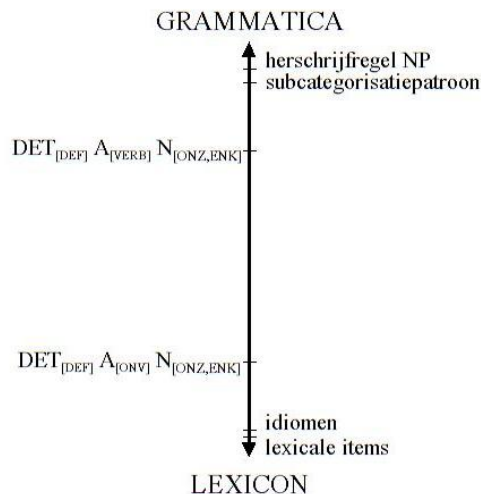
Ten tweede, de dichotomische visie op taalkundige kennis is problematisch vanuit het perspectief van de corpuslinguïstische methodologie die we hebben toegepast. Eén van de belangrijkste descriptieve verwezenlijkingen van de corpuslinguïstiek is het inzicht dat talige uitingen ten dele uit recurrende lexicale patronen zijn opgebouwd. Corpusonderzoek heeft aangetoond dat lexicale restricties niet alleen optreden tussen lexicale items op de syntagmatische as (Sinclair 1991; Stubbs 1995, 2001) maar ook tussen lexicale items en abstracte patronen, c.q. grammaticale constructies, op de paradigmatische as (Stefanowitsch & Gries 2003). De lexicale restricties op de combinatoriek van talige items zijn niet beperkt tot louter formele en formulaïsche sequenties, maar kunnen ook semantisch betekenisvolle structuren blootleggen, zoals dit ook in onderhavig onderzoek het geval is (zie paragrafen 17.1 en 17.2). Sinclair (1991: 109-121) stelt dan ook dat het “open-choice principle” moet worden aangevuld met het “idiom principle”:<sup>28</sup>

“The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute

single choices, even though they might appear to be analysable into segments” (Sinclair 1991: 110).

Daarnaast vormt de kwantitatief-empirische corpusanalyse de belangrijkste motivatie voor de collocatieanalyse: het is net omdat de noties van idiomaticiteit en lexicale compositionaliteit een objectiveerbare empirische operationalisering in de weg staan, dat we als alternatief voor het collocatieonderzoek hebben geopteerd (zie paragraaf 11.1.1).

Het afwijzen van de dichotomische visie veronderstelt uiteraard dat we een alternatief voorstellen dat in overeenstemming is met de resultaten en de methodologie van het gevoerde onderzoek. Hiervoor gaan we te rade bij de zogenaamde *usage-based* modellen die in het ruimere theoretische kader van de Cognitieve Grammatica en de Constructiegrammatica ontwikkeld worden (Barlow & Kemmer 2000; Bybee & Hopper 2001; Croft & Cruse 2004; Tomasello 2000). Eén van de centrale noties aan de basis van de *usage-based* modellen is de assumptie dat taalkundige kennis niet in twee dichotomische componenten kan worden opgedeeld, maar een continuum vormt variërend van zeer schematische eenheden (grammaticale pool) tot zeer concrete eenheden (lexicale pool). De grammaticale pool wordt bezet door abstracte patronen, zoals de herschrijffregel voor een NP (bv. NP → DET A N) of het subcategorisatiepatroon van een werkwoordelijke klasse (bv. een transitief patroon voorziet een onderwerp en een lijdend voorwerp); de lexicale pool wordt bezet door lexicale items en idiomen (bv. *de kat de bel aanbinden*). Tussen deze twee extremen treffen we verschillende schematiciteitsgraden aan, zoals grammaticale patronen die deels lexicaal zijn ingevuld.



Figuur 17(2): Positionering op het continuum tussen grammatica en lexicon van de NP's met een verbogen adjectief en de NP's met een onverbogen adjectief

Figuur 17(2) visualiseert het continuüm voor de taalkundige kennis, met daarin de twee buigingspatronen, met name het verbogen patroon ( $\text{Det}_{[\text{DEF}]} \text{A}_{[\text{VERB}]} \text{N}_{[\text{ONZ,ENK}]}$ ) en het onverbogen patroon ( $\text{Det}_{[\text{DEF}]} \text{A}_{[\text{ONV}]} \text{N}_{[\text{ONZ,ENK}]}$ ). Merk op dat de afstanden op het continuüm in figuur 17(2) relatief zijn: het is ons vooral te doen om de onderlinge positionering van beide buigingspatronen op dit continuüm en niet om de exacte posities.

Het verbogen patroon neemt een positie in aan de grammaticale pool van het continuüm: de adjectivische en de substantivische positie kunnen (relatief) vrij en los van elkaar worden ingevuld, aangezien dit het buigingspatroon is dat we aantreffen wanneer het adjectief en het substantief geen lexicale collocatie vormen. Het onverbogen patroon daarentegen bevindt zich aan de lexicale pool van het continuüm, aangezien het adjectief en het substantief een lexicale collocatie vormen. Dit betekent dat de adjectivische en substantivische posities niet onafhankelijk van elkaar kunnen worden ingevuld, maar dat de lexicale realisatie van beide posities sterk gerelateerd is. Deze positie aan de lexicale pool van het continuüm weerspiegelt de semantische tendenties aan de basis van de lexicale collocaties: de eigennamen en de institutionele termen hebben een sterk idiomatisch karakter; de relationele en de categoriserende adjectieven drukken zeer saillante eigenschappen van het substantivische concept uit, wat gepaard gaat met een grotere lexicale selectiviteit tussen adjectief en substantief (cf. *supra*).

Dit voorstel sluit aan bij Booij's (2002a) analyse van de gelexicaliseerde adjectief-nomenparen. Hij grijpt eveneens terug naar het theoretische kader van de Constructiegrammatica om deze items te beschrijven. Hij spreekt van "constructional idioms" om de taakverdeling tussen syntaxis (c.q. grammatica) en morfologie (c.q. lexicon) te modelleren:

"It is appropriate to assume a constructional idiom in the lexicon with the form  $[\text{A N}]_{\text{NP}}$ , with two open positions, and no terminal element fixed. This template is a specific instantiation of the general syntactic template for NP's. [...] Most properties of the AN phrasal names are inherited from NPs in general as defined by the syntactic module. It is only the name function that needs to be specified for the AN phrases. By establishing a formal relation of inheritance between constructional idioms and the canonical syntactic templates as defined by the syntax, we express the generalization that most constructional idioms are not completely arbitrary units of words, but form a subset of the types of word combinations that are defined as well formed by the syntax" (Booij 2002a: 315-316)

Booij's stelling dat *constructional idioms* geen volledig vrije woordcombinaties vormen, is conform met de empirische vaststelling dat deze items lexicale collocaties vormen.

De voorgestelde oplossing is eveneens in overeenstemming met de stellingname van onder andere De Caluwe (1990), Van Sterkenburg (1993) en Hüning (2004) dat gelexicaliseerde woordgroepen een overgangsgebied tussen morfologie, c.q. lexicon, en syntaxis, c.q. grammatica, bezetten. Het model in figuur 17(2) erkent enerzijds de hechte lexicale relatie tussen het adjectief en het substantief, zonder daarom echter een productief mechanisme uit te sluiten. Hüning (2004: 170) stelt in dit verband:

“De notie ‘gelexicaliseerd’ suggereert een historisch proces en wordt vaak juist als een soort tegenhanger van ‘productief’ gebruikt, waarbij productiviteit dan met (semantische) doorzichtigheid correspondeert. Booijs terminologie [*constructional idiom*; JT] benadrukt nu de uitbreidbaarheid van het patroon om nieuwe benamingen, nieuwe uitdrukkingen met een classificerende functie te creëren.”

Een continuum, zoals in figuur 17(2), is tot slot een maximalistische voorstellingswijze van taalkundige kennis. Dit wil zeggen dat een abstract patroon en concrete lexicale instanties van dit patroon elkaar niet wederzijds uitsluiten, maar samen kunnen voorkomen (Barlow & Kemmer 2000; Bybee & Hopper 2001; Croft & Cruse 2004). Anders gezegd: redundante informatie is niet *a priori* uitgesloten, zoals dit wel het geval is bij de dichotomische voorstellingswijze, waar een patroon ofwel regelmatig is – en in die hoedanigheid door de regels van de grammatica wordt gegenereerd – ofwel als een uitzondering naar het lexicon wordt verwezen. Bijgevolg kunnen in de *usage-based* voorstellingswijze adjectief-nomenclotaties die een productief semantisch mechanisme identificeren als eenheid worden opgeslagen, zonder dat dit impliceert dat het om idiosyncratische uitzonderingen gaat. We maken de lezer erop attent dat meerdere adjectief-nomenparen die een productieve semantische eigenschap uitdrukken net door de habitualiteit en de saillantheid van de (lexicale) combinatie een lexicale collocatie vormen, i.e. significant vaker samen voorkomen dan verwacht, en om die reden als een eenheid (kunnen) worden opgeslagen.

Samenvattend kunnen we stellen dat de Constructiegrammatica door de expliciete afwijzing van een dichotomische visie op taalkundige kennis een geschikt theoretisch kader aanreikt voor een graduele modellering van de taakverdeling tussen lexicon, c.q. morfologie, en grammatica, c.q. syntaxis. Toegepast op de adjectivische buigingsalternantie betekent dit dat de adjectief-nomenparen met een onverbogen adjectief een positie in het lexicale deel van het continuum bezetten, terwijl de adjectief-nomenparen met een verbogen adjectief een positie in het grammaticale deel van het continuum innemen. Deze onderlinge positionering van beide buigingspatronen verzoent aan de ene kant de vaststelling dat de onverbogen vorm kenmerkend is voor NP's met sterke lexicale restricties op de co-occurentie van de elementen in de adjectivische en de substantivische posities terwijl de verbogen vorm NP's identificeert waar de adjectivische en substantivische posities (relatief) onafhankelijk van elkaar kunnen worden ingevuld, en aan de andere kant de vaststelling dat het gebruik van het onverbogen buigingsmorfem niet beperkt is

tot idiosyncratische en idiomatische uitdrukkingen, zoals voorliggend semantisch onderzoek heeft aangetoond.

#### 17.4 Conclusie

In dit hoofdstuk zijn we dieper ingegaan op de precieze aard van het buigingseffect van de lexicale collocaties. We hebben meer bepaald de vraag gesteld of er semantische patronen geïdentificeerd kunnen worden aan de basis van het deflecterende effect van de adjectief-nomenparen die een lexicale collocatie vormen in het VDDS-corpus (zie paragraaf 16.2). Het antwoord op deze vraag bestaat uit twee delen: in eerste instantie hebben we gezocht naar een semantische motivatie voor de keuze van het (onverbogen) buigingsmorfeem in de lexicale collocaties (paragraaf 17.1); daarna zijn we nagegaan of de vastgestelde semantische effecten kenmerkend zijn voor de adjectief-nomencollocaties (paragraaf 17.2). We hebben dit hoofdstuk afgesloten met een algemene bespreking van de status van het onverbogen buigingsmorfeem in de adjectief-nomencollocaties (paragraaf 17.3).

Op basis van de verkennende literatuurstudie in hoofdstuk 1 hebben we vier hypothesen naar voren geschoven voor de semantische conditionering van de adjectivische buigingsalternantie bij lexicale collocaties. Zowel de bivariate analyses als de multivariate analyse onderbouwen deze hypothesen. In lexicale collocaties kan het onverbogen buigingsmorfeem aangewend worden om (één of meerdere van) de volgende vier betekenisaspecten op het niveau van het adjectief of op het niveau van het adjectief-nomenpaar uit te drukken:

- het onverbogen buigingsmorfeem identificeert het element in de adjectivische positie als een relationeel adjectief
- het onverbogen buigingsmorfeem identificeert het element in de adjectivische positie als een categoriserend adjectief
- het onverbogen buigingsmorfeem identificeert het adjectief-nomenpaar als een eigennaam
- het onverbogen buigingsmorfeem identificeert het adjectief-nomenpaar als een institutionele term

We wensen te onderstrepen dat het zeer moeilijk is om deze vier semantische functies in één schematische conditie samen te vatten. Wat volgens ons gemeenschappelijk is aan deze vier semantische motivaties voor het gebruik van het onverbogen adjectief is dat ze een situatie markeren die afwijkt van het prototypische adjectief-nomenpaar waar het adjectief de referent aangeduid door het substantief kwalificeert door een eigenschap van deze referent te noemen.

De logistische-regressieanalyse op basis van de semantische eigenschappen bewijst in de eerste plaats dat de keuze van de gemarkeerde onverbogen buigingsvorm in adjectief-nomencollocaties semantisch gemotiveerd is. De resultaten voor de modelstatistieken wijzen op een zeer degelijk model, zowel wat de verklarende kracht als wat de voorspellende kracht betreft. Daarnaast kunnen we aan de hand

van de regressiecoëfficiënten (c.q. de OR's) twee primaire en twee secundaire semantische condities onderscheiden: de primaire semantische functies zijn het adjectief-nomenpaar als een institutionele term markeren en het element in de adjectivische positie als een categoriserend adjectief identificeren; de secundaire semantische factoren zijn het adjectief-nomenpaar als een eigennaam identificeren en signaleren dat het element in de adjectivische positie een relationeel adjectief is. De secundaire status van de eigennamen hangt samen met een aantal (zeer frequente) eigennamen die consistent de buigings-*e* kiezen.

Na de analyse van de semantische motivatie van de onverbogen vorm bij adjectief-nomencollocaties zijn we nagegaan in welke mate de geïdentificeerde restricties kenmerkend zijn voor lexicale collocaties. Hiervoor hebben we een externe en een interne controletest voorgesteld. De externe test heeft aangetoond dat de gemarkeerde semantische kenmerken, die het gebruik van het onverbogen adjectief motiveren, significant vaker bij de adjectief-nomencollocaties voorkomen dan bij de observaties waar het adjectief en het substantief geen significante lexicale associatie onderhouden. Aan de hand van de interne controletest hebben we berekend dat de gemiddelde collocabiliteitsrang significant hoger is bij de deflecterende gemarkeerde waarden van de semantische kenmerken dan bij de flecterende ongemarkeerde waarden.

We hebben dit hoofdstuk afgesloten met een algemene discussie over de status van de buigingsalternantie en meer bepaald over de status van het onverbogen buigingsmorfeem in lexicale collocaties. Op grond van de resultaten van het empirisch onderzoek en de gehanteerde corpuslinguïstische methodologie hebben we aangevoerd dat de marginaliserende visie onhoudbaar is. Deze marginaliserende visie gaat ervan uit dat adjectief-nomenparen met een buigings-*e* via de regels van de grammatica gegenereerd worden en dat de adjectief-nomenparen met de onverbogen vorm van het adjectief idiosyncratische uitzonderingen zijn die in het lexicon moeten worden opgesomd. We hebben deze dichotomische visie genuanceerd: aan de ene kant wordt het onverbogen buigingsmorfeem geselecteerd om het adjectief-nomenpaar als een lexicale collocatie te markeren, i.e. als een lexicale eenheid die als geheel in het lexicon geactiveerd kan worden; aan de andere kant toont het logistische-regressiemodel voor de semantische factoren dat de keuze van het onverbogen adjectief in adjectief-nomencollocaties semantisch gemotiveerd is.

Als alternatief voor de dichotomische visie die de taalkundige kennis in een grammaticale en een lexicale component opdeelt, zijn we te rade gegaan bij de *usage-based* modellen, meer bepaald bij de Constructiegrammatica. Deze taalkundige stroming gaat ervan uit dat de taalkundige kennis een continuum vormt dat varieert van zeer regelmatige en abstracte (grammaticale) patronen tot zeer specifieke en idiosyncratische (lexicale) uitdrukkingen en items. Op basis van de resultaten voor het VDDS-corpus hebben we het NP-patroon met de gemarkeerde onverbogen adjectivische variant dichter bij de lexicale pool van het continuum

gesitueerd dan het 'regelmatige' NP-patroon met de verbogen pendant. Deze voorstellingswijze erkent enerzijds de sterke lexicale associatie tussen het adjectief en het substantief in NP's die het onverbogen adjectief selecteren, zonder dat, zoals in de dichotomische modellen, de mogelijkheid van een productief semantisch mechanisme wordt uitgesloten. Het gebruik van een continuum voorziet bovendien de mogelijkheid dat naast het abstracte NP-patroon ook habituele en recurrente adjectief-nomenparen worden opgeslagen, zonder dat dit impliceert dat deze specifieke adjectief-nomenparen een idioom of een idiosyncratische uitzondering zijn.



## Noten

1. Rekening houdend met de inbreng van de persoonlijke appreciatie en de encyclopedische kennis van de onderzoeker tijdens de semantische codering spreken we niet van een objectiveerbare maar van een intersubjectieve procedure. Voor een uitgebreide discussie verwijzen we naar Geeraerts (1989: 57-59). In paragraaf 11.1.1.2 hebben we al gewezen op de methodologische problemen die voortvloeien uit het gebrek aan een objectiveerbare semantische testprocedure voor corpuslinguïstisch onderzoek.
2. We hebben de volgorde van de verschillende attestaties van een AN-lemmatype uiteraard eerst gerandomiseerd.
3. Vanuit een formeel-semantisch perspectief definieert iedere adjectief-nomencombinatie een deelverzameling van de entiteiten waaraan het nominale hoofd refereert.
4. We vestigen de aandacht op de alternantie tussen AN-samenstellingen, zoals *stokbrood* en *boerenbrood*, en adjectief-nomenparen met een categoriserend adjectief, zoals *Frans brood*. Dit onderstreept de hechte associatie tussen een categoriserend adjectief en het nominale hoofd (zie ook paragraaf 11.2.2). We hebben er al eerder op gewezen dat de EGVD naast de samenstelling *bruinbrood* ook de woordgroep *bruin brood* vermeldt onder het lemma *brood*.
5. Het adjectief-nomenpaar *Brits parlement* is een voorbeeld van een sequentie waar het adjectief afhankelijk van de context al dan niet als een categoriserend adjectief geïnterpreteerd moet worden: wanneer deze sequentie het parlement van Groot-Brittannië identificeert, is *Brits* een niet-categoriserend relationeel adjectief; wanneer deze tournure daarentegen een type parlement aanduidt dat niet de vorm van een halfmond heeft, maar waar de regering en de oppositie tegenover elkaar zitten, is er wel sprake van een categoriserend relationeel adjectief, zoals in *het Waals parlement is een Brits parlement*.
6. Voor een rangschikking van de mogelijke relaties tussen het relationeel adjectief en het nominale hoofd in functie van de saillantheid van deze relatie verwijzen we naar Dirven (1999).
7. Met het gebruik van de buigings-*e* bij adjectief-nomenparen zoals *Witte Huis* of *Romeinse Rijk* signaleren we dat deze adjectief-nomenparen consistent de verbogen variant selecteren in het VDDS-corpus.
8. Wanneer we dit criterium versoepelen tot alle AN-lemmatypes met een frequentie groter dan of gelijk aan 100, is het aantal categorische adjectieven in de adjectivische positie groter dan verwacht, zoals de cijfers in tabel 17(i) aantonen.

	$\neg$ eig.naam	eig.naam
$\neg$ cat.adj	[1.1119,1.1335]	[0.6206,0.6647]
cat.adj	[0.5434,0.5899]	[2.1832,2.3386]

Tabel 17(i): 95% BI en partiële  $\chi^2$ -statistiek voor de frequentieverdeling van de categorialiteit van het adjectief over de eigennamen (na exclusie van de eigennamen zonder categoriserend adjectief met een tokenfrequentie van minstens 100)

9. Voor de constructie van het logistische-regressiemodel hebben we net als bij de logistische-regressieanalyses in hoofdstuk 14 gebruik gemaakt van het *Design*-pakket (Harrell 2001).
10. Alvorens we de resultaten van de regressieanalyse rapporteren, testen we de eventuele multicollineariteit van de verklarende factoren aan de hand van de *Variance Inflation Factor* (VIF) (zie ook paragraaf 14.3.2 noot 5). De VIF-score berekent in welke mate de collineariteit tussen de verschillende verklarende factoren de precisie van de regressie-coëfficiënten aantast. De interpretatie van de VIF-scores stoelt op de volgende vuistregel: een VIF-score groter dan 10 wijst op multicollineariteit; een VIF-score groter dan 2.5 wordt in logistische-regressiemodellen algemeen als een reden tot voorzichtigheid beschouwd. Tabel 17(ii) vat de VIF-indexen voor de semantische factoren samen:

Factorwaarde	VIF
a.rel=ja	1.033053
a.cat=ja	1.186446
an.eignaam=ja	1.984011
an.inst.term=ja	1.838259

Tabel 17(ii): VIF-scores voor de verklarende variabelen in het logistische-regressiemodel

Aangezien de VIF-score voor geen enkele verklarende variabele de waarde 2.5 overschrijdt, is er geen gevaar voor multicollineariteit.

11. Het voorgestelde logistische-regressiemodel bevat, in tegenstelling tot de modellen op basis van het CGN (zie paragraaf 14.3.1 noot 4 en paragraaf 14.3.4 noten 12 en 16), geen *influential observations*. De reden hiervoor is dat het aantal verklarende factor(waard)en in het VDDS-corpus kleiner is dan in het CGN en het aantal observaties groter. Bijgevolg bevat de materiaalverzameling op basis van het VDDS-corpus geen parameterwaardencombinaties die ondervertegenwoordigd zijn.
12. Deze classificatie op basis van de manuele regressieanalyse wordt bevestigd door de resultaten van de *forward stepwise*-regressie, met dit verschil dat de relationaliteit van het adjectief (derde parameter die door het *forward stepwise*-algoritme geselecteerd wordt) de onverklaarde variatie sterker reduceert dan de eigenaamstatus van het adjectief-nomenpaar (vierde en laatste parameter die door het *forward stepwise*-algoritme geselecteerd wordt). De institutionele status van het adjectief-nomenpaar en de categorialiteit van het adjectief worden respectievelijk tijdens de eerste en de tweede stap van het *forward stepwise*-algoritme in het regressiemodel opgenomen.

13. Merk op dat *Heilig Land* in tegenstelling tot *Beloofde Land* altijd onverbogen is in het VDDS-corpus.
14. De relatie tussen de thematische scheeftekening van de materiaalverzameling en de selectie van de buigings-*e* is enkel een veronderstelling die stoelt op hoofdzakelijk anekdotische observaties tijdens de semantische codering van de observaties in de materiaalverzameling. Het falsificeren van deze stelling – en in het bijzonder het operationaliseren van de thematische scheeftekening – vormt het onderwerp van een afzonderlijk onderzoek en overstijgt de doelstellingen van voorliggende analyse.
15. De institutionele termen worden niet als zodanig door Van Sterkenburg (1993) onderscheiden, maar kunnen in de ruimere klasse van de “technische termen” worden ondergebracht. Rekening houdend met de thema’s die in een kwaliteitskrant behandeld worden, lijkt het ons verdedigbaar om institutionele termen als ambtelijk en politiek jargon te bestempelen.
16. We herinneren de lezer eraan dat de  $G^2$ -statistiek die we gebruikt hebben voor het kwantificeren van de lexicale associatiesterkte tussen het adjectief en het substantief enkel de omvang en de significantie van de associatie tussen het adjectief en het substantief berekent en niets zegt over de richting van deze associatie. De richting van deze associatie – komen beide elementen significant vaker of significant minder vaak voor dan verwacht? – moeten we zelf bepalen door de geobserveerde frequentie met de verwachte frequentie te vergelijken. Daarnaast hebben we bij het opstellen van de selectiecriteria voor de controlegroep erover gewaakt om geen *hapax legomena* op te nemen. Voor een uitgebreide discussie van de restricties op het gebruik en de interpretatie van de  $G^2$ -statistiek verwijzen we naar paragraaf 11.1.1.2.
17. We berekenen in feite de *omgekeerde* collocabiliteitsrang. In de statistiek werkt men meestal met aflopende rangordes, terwijl in dit onderzoek de collocabiliteitsrang wordt toegekend op basis van de oplopende rangorde van de  $G^2$ -scores.
18. Aangezien de semantische factoren binomiale variabelen zijn, gaat een significante *p*-waarde voor de *F*-statistiek bijna automatisch gepaard met een significante *p*-waarde voor de *t*-test.
19. Anders gezegd: een *one way* ANOVA is een enkelvoudige lineaire regressie, maar dan met een nominale in plaats van een numerische verklarende variabele.
20. Omwille van het verschil in omvang tussen de controlegroep ( $n = 1,266$ ) en de groep met de lexicale collocaties ( $n = 23,953$ ) berekenen we als validatie voor de partiële  $\chi^2$ -statistiek de OR die de odds  $^{\text{rel.adj}}/_{\text{rel.adj}}$  bij de lexicale collocaties vergelijkt met de controlegroep: OR = 1.6457 [1.466,1.8475]. Deze gegevens bevestigen het resultaat van de partiële  $\chi^2$ -statistiek: de odds  $^{\text{rel.adj}}/_{\text{rel.adj}}$  is meer dan de helft groter bij de lexicale collocaties dan bij de controlegroep. Het 95% BI, dat de scharnierwaarde 1 niet bevat, staft de significantie van de verschillen aangeduid door de partiële  $\chi^2$ -statistiek.
- Deze validatietest op basis van de OR herhalen we bij de andere semantische factoren (zie noten 21, 22 en 23).

21. Dit resultaat wordt bevestigd door de OR die de odds  $^{cat.adj}/_{-cat.adj}$  voor de lexicale collocaties met de controlegroep vergelijkt: OR = 3.1057 [2.5188,3.8294].
22. De controle aan de hand van de OR die de odds  $^{eig.naam}/_{-eig.naam}$  bij de lexicale collocaties met de controlegroep vergelijkt, bevestigt het resultaat van de partiële  $\chi^2$ -statistiek: OR = 17.7664 [11.5301,27.3757].
23. Ook voor deze semantische eigenschap wordt het resultaat van de partiële  $\chi^2$ -statistiek gestaafd door de OR: OR = 15.751 [11.2437,22.0651]. De odds  $^{inst.term}/_{-inst.term}$  zijn met andere woorden vijftienmaal groter bij de lexicale collocaties dan in de controlegroep.
24. Een enkelvoudig logistische-regressiemodel met de categorialiteit van het adjectief ('a.cat') als verklarende variabele en de buigingsalternantie als responsvariabele gaat gepaard met een significante daling van de verklaarde variatie in vergelijking met het globale logistische-regressiemodel dat de vier semantische factoren omvat (zie paragraaf 17.1.2.2, tabel 17(13)). Tabel 17(iii) vat de resultaten samen van de ANOVA die we hebben uitgevoerd om de twee logistische-regressiemodellen te vergelijken. (Voor een nadere toelichting bij deze toepassing van de ANOVA verwijzen we naar paragraaf 14.3.2 noot 17.) De combinatie van het negatieve cijfer voor het verschil in verklaarde variatie (-7,917.2) en de zeer kleine  $p$ -waarde (0.0) duidt op een significante en zeer sterke daling van de verklaarde variatie in het enkelvoudige model ten opzichte van het globale, meer-voudige model.

Logistische-regressiemodel	Residu	Vershil in verklaarde variatie	P(>  $\chi^2$ )
Globaal regressiemodel	20,880.6		
Enkelvoudig regressiemodel met 'a.cat' als verklarende variabele	28,797.8	-7,917.2	0.0

Tabel 17(iii): ANOVA voor de vergelijking van de verklaarde variatie in het globale regressiemodel met de vier semantische factoren en in het enkelvoudige regressiemodel met de categorialiteit van het adjectief als verklarende variabele

25. We wijzen erop dat we niet teveel belang aan de orthografie mogen hechten. Deze voorbeelden hebben bijgevolgd een louter illustratieve status.
26. Als gevolg van het gebruik van *De Standaard* als corpus zijn de lectische variabelen constant: de materiaalverzameling omvat uitsluitend algemeen geschreven Belgisch Nederlands. Daarnaast is het effect van de prosodische conditionering wellicht minder groot in de geschreven taal dan in de gesproken taal, zodat het verdedigbaar is om deze parameter niet in het semantische onderzoek op te nemen. Tot slot brengen we in herinnering dat deze semantische analyse beperkt is tot de adjectief-nomenparen waar het adjectief en het substantief een lexicale collocatie vormen, waardoor het effect van de andere factoren wellicht minder groot is.
27. Een derde mogelijkheid is dat het gebruik van het onverbogen adjectief dialectisch is. Deze mogelijkheid lijkt niet van toepassing op *De Standaard*, waar de algemene – noordelijke – norm als streefdoel wordt gehanteerd (Permentier & Van Den Eynden 1998: 114) (zie ook paragraaf 15.3).

28. Het *open-choice principle* is kenmerkend voor de dichotomische visie op taalkundige kennis: een zin wordt beschouwd als het resultaat van een grammaticaal derivatieproces waarbij regels een abstracte (boom)structuur genereren waar de (terminale) posities ingevuld worden door lexicale items. De enige restrictie op de selectie van de lexicale items is de conformiteit met de restricties op de verschillende posities in de (boom)structuur. Met andere woorden: het eindresultaat van de derivatie moet een grammaticale zin zijn. Corpuslinguïstisch onderzoek heeft echter aangetoond dat deze visie ontoereikend is om het reële taalgebruik te modelleren, met de formulering van het *idiom principle* als gevolg.

## CONCLUSIE

Ter afsluiting hernemen we de belangrijkste descriptieve en methodologische bevindingen van voorliggend onderzoek naar de adjectivische buigingsalternantie in een definiëte NP met een enkelvoudig *het*-woord als hoofd. Dit proefschrift bestaat uit vier delen. Het eerste deel is een verkennende literatuurstudie waar we het onderzoeksobject en de onderzoeksmethodologie afbakenen en in een ruimer kader situeren. Deze literatuurstudie mondt uit in de onderzoekshypotheses die we voor de verschillende condities op de buigingsalternantie geformuleerd hebben, en in algemene onderzoeksvragen naar de aard van de adjectivische buigingsalternantie bij neutra.

De overige drie delen van deze dissertatie rapporteren de resultaten van het empirisch onderzoek. De delen II en III vormen een methodologisch tweeluik voor het basisonderzoek naar het gebruik van de adjectivische buigingsvormen bij neutra, zoals die voorkomen in het *Corpus Gesproken Nederlands* (CGN). Dit basisonderzoek wordt afgesloten met een multivariaat model, meer bepaald een logistische-regressiemodel, dat de diverse condities integreert die de keuze van het adjectivische buigingsmorfeem in het gesproken standaard Nederlands bepalen. In deel IV concentreren we ons op één specifieke conditie, met name de lexicale en semantische conditionering van de keuze van het buigingsmorfeem.

Laten we nu de condities op de adjectivische buigingsalternantie bij neutra overlopen. We bespreken eerst de algemene restricties op de selectie van het buigingsmorfeem in het gesproken standaard Nederlands, zoals we die in het CGN geïdentificeerd hebben. Daarna behandelen we de semantische conditionering van het buigingseffect bij de lexicale collocaties op basis van de Vlaamse kwaliteitskrant *De Standaard*.

In de delen II en III hebben we op basis van het CGN een globaal model geconstrueerd voor de buigingsalternantie in de gesproken standaardtaal. Tabel 1 stelt de factoren voor die de buigingsalternantie in het CGN conditioneren. Deze tabel bevat alleen de restricties met een significante impact op de keuze van het adjectivische buigingsmorfeem. Op basis van de bijdrage van deze factoren tot de reductie van de onverklaarde (buigings)variatie en op basis van hun impact op de selectie van de buigingsvariant, hebben we ze geordend in primaire, secundaire en tertiaire factoren ('Rang' in tabel 1). Wat opvalt in deze tabel, is de heterogeniteit van de verklarende factoren: in de eerste plaats behoren deze factoren tot verschillende taalkundige domeinen die traditioneel niet samen behandeld worden; daarnaast is het nagenoeg onmogelijk om deze factoren, zelfs de primaire factoren, in één schematische conditie samen te vatten. In deze algemene conclusie gaan we

CONCLUSIE

niet dieper in op het buigingseffect van de afzonderlijke factoren. Hiervoor verwijzen we de lezer naar de uitgebreide conclusies bij de hoofdstukken 8 tot en met 12.

Type	Rang	Factorwaarde verbogen vorm	Factor	Factorwaarde onverbogen vorm
morf./synt	2	<i>het</i> , aanw.vnw, genitief	woordsoort determinator	bez.vnw
	2	substantief	woordsoort N	genominaliseerde inf
	3	bigenerisch	genus N	<i>het</i> -woord
fonologisch	2	medeklinker	beginklank N	klinker
	1	beklemtoonde eindlettergreep	prosodische structuur rechterwoordrand A	minstens één onbeklemtoonde lettergreep
		beklemtoonde beginlettergreep	prosodische structuur linkerwoordrand N	onbeklemtoonde beginlettergreep
	2			
morf.	3	niet geled	morf. geledheid A	geled
lexicaal / semantisch	1	geen lexicale collocatie, zeer zwakke lexicale collocatie	lexicale collocabiliteit AN-paar	lexicale collocatie
	2	kwalificerend A	semantische categorie A	relationeel A
lectisch	1	Nederlands Nederlands	regio	Belgisch Nederlands
	1	ongemarkeerd	stijl <sub>NED</sub>	zeer formeel, ambtelijk
	1	formeel, ongemarkeerd	stijl <sub>BEL</sub>	informeel

Tabel 1: Overzicht van de conditionering van de adjectivische buigingsalternantie in het standaard gesproken Nederlands op basis van het Corpus Gesproken Nederlands

Ondanks de heterogeniteit van de factoren in tabel 1 kunnen toch enkele relaties tussen factoren geïdentificeerd worden. Ten eerste, de sterkere voorkeur voor een onverbogen adjectief vóór een substantief dat op een klinker begint, is beperkt tot de informele gesprekscontexten. Ten tweede, het zeer formele en ambtelijke karakter van de onverbogen vorm in het Nederlands Nederlands hangt nauw samen met het gebruik van formulaische adjectief-nomencollocaties in de formele gespreks-situaties, met name in de Tweede-Kamerdebatten. Ten derde, twee primaire factoren zijn gerelateerd aan een secundaire factor: het buigingseffect van het prosodische patroon aan de rechterwoordrand van het adjectief kan versterkt worden door het prosodische patroon aan de linkerwoordrand van het substantief; in de adjectief-nomencollocaties wordt de adjectivische positie significant vaker dan verwacht door een relationeel adjectief bezet.

Bij de interpretatie van het schema voor de buigingsalternantie in tabel 1 moeten we er rekening mee houden dat de afzonderlijke modellen die we voor het Nederlands Nederlands en voor het Belgisch Nederlands hebben voorgesteld

significant verschillen. Concreet definiëren de drie primaire factoren een significant verschillend effect in beide nationale variëteiten:

- Het buigingseffect van de restricties op de prosodische structuur van het adjectief is voor enkele patronen significant groter in het Nederlands Nederlands.
- Het buigingseffect van de lexicale collocabiliteit van het adjectief-nomenpaar is over de hele lijn groter in het Nederlands Nederlands, met een significant verschil in het hoogste collocabiliteitskwartiel.
- De stilistische betekenis van het onverbogen buigingsmorfeem is wat de interpretatie betreft tegengesteld in beide nationale variëteiten: zeer formeel en ambtelijk in Nederland tegenover informeel in Vlaanderen. Deze tegengestelde betekenis verklaart wellicht het significant grotere effect in de meest informele registers: in de registers met de hoogste informaliteits-index, die zowel in Nederland als in Vlaanderen informeel zijn, is het deflecterende effect significant groter in Vlaanderen dan in Nederland.

Wat de eerste twee verschillen aangaat, is de zin van het buigingseffect identiek in beide nationale variëteiten: het verschil heeft alleen betrekking op de grootte van het effect, dat telkens significant groter is in Nederland. Het stilistische verschil, daarentegen, is niet alleen significant groter in België voor de meest informele registers, maar is op het vlak van de interpretatie tegengesteld in beide nationale variëteiten.

In deel IV hebben we het globale buigingsmodel op basis van het CGN aangevuld met een apart model voor de semantische restricties op de keuze van het buigingsmorfeem in adjectief-nomenclotaties. In dit laatste deel van het onderzoek stellen we meer bepaald de vraag naar de semantische interpretatie van het deflecterende effect van de lexicale collocaties dat we op basis van het CGN hebben vastgesteld. Hiervoor hebben we de Vlaamse kwaliteitskrant *De Standaard* geraadpleegd, omdat het CGN zowel kwantitatief als kwalitatief minder geschikt is voor het beantwoorden van de semantische vraagstelling: kwantitatief is het aantal observaties waar het adjectief en het substantief een lexicale collocatie vormen relatief klein; kwalitatief verzamelt het CGN materiaal uit zowel thematisch als lectisch heterogene gesprekscontexten.

Op de vraag naar de semantische interpretatie van het deflecterende buigings-effect bij de lexicale collocaties hebben we een genuanceerd antwoord geformuleerd: aan de ene kant wordt het onverbogen buigingsmorfeem aangewend om op iconische wijze adjectief-nomenparen te markeren waar het adjectief en het substantief een sterke lexicale associatie vertonen; aan de andere kant kunnen aan de basis van dit deflecterende effect verschillende semantische tendensen geïdentificeerd worden. Deze semantische tendensen worden samengevat in tabel 2:



Type	Rang	Factorwaarde verbogen vorm	Factor	Factorwaarde onverbogen vorm
kenmerken A	2	kwalificerend A	relationaliteit A	relationeel A
	1	niet-categoriserend A	categorialiteit A	categoriserend A
kenmerken AN	2	neen	eigennaam	ja
	1	neen	institutionele term	ja

Tabel 2: Overzicht van de semantische conditionering van de adjectivische buigingsalternantie in lexicale collocaties op basis van het De Standaard-corpus (VDDS)

De resultaten van het empirisch onderzoek, samengevat in de tabellen 1 en 2, dwingen ons de bestaande neerlandistische attitudes met betrekking tot de buigingsalternantie te weerleggen of minstens te nuanceren:

- De adjectivische buigingsalternantie bij neutra kan in het reële taalgebruik niet worden afgedaan als een randverschijnsel. Daarvoor is de gemarkeerde onverbogen variant te frequent, zowel in het standaard gesproken Nederlands als in de Vlaamse krantentaal. De verklarende en voorspellende kracht van het logistische-regressiemodel voor de lexicale collocaties in *De Standaard* (deel IV) toont bovendien aan dat het gebruik van de onverbogen vorm in adjectief-nomenclotaties niet tot een louter lexicaal en idiosyncratisch verschijnsel gereduceerd kan worden, maar – ten dele – gemotiveerd is op basis van de semantische eigenschappen van het adjectief en het adjectief-nomenpaar.
- Hoewel de reducerende attitude door de erkenning van de buigingsalternantie een stap voorwaarts is in vergelijking met de marginaliserende attitude, bewijst het globale logistische-regressiemodel op basis van het CGN dat de prosodische en/of de lexicale/semantische restricties onvoldoende zijn om de buigingsalternantie te beschrijven. De semantische modellen die de keuze van het onverbogen adjectief in de adjectief-nomenidiomen, c.q. de adjectief-nomenclotaties volgens onze operationalisering, met categoriserende adjectieven relateren, doorstaan de toetsing met de empirische realiteit van *De Standaard*, maar volstaan niet om de keuze van het buigingsmorfeem bij de lexicale collocaties adequaat te modelleren.
- Tot slot hebben we de bestaande empirische analyses aangevuld door een globaal model te ontwikkelen voor de buigingsalternantie in het gesproken standaard Nederlands (deel III). Dit model toont aan dat de keuze van het adjectivische buigingsmorfeem bij neutra niet aleatoir is, zoals sommige analyses opperen, maar gemodelleerd kan worden, mits het gebruik van de

gepaste methode, c.q. empirisch corpusonderzoek, en het gebruik van de gepaste techniek, c.q. multivariate technieken.

Na de samenvatting van de inhoudelijke resultaten richten we ons nu op de methodologische bevindingen van dit onderzoek. We hebben de toegepaste methode als kwantitatief-empirisch corpusonderzoek gekwalificeerd. Dit is de meest geschikte methode voor de realisatie van onze onderzoeksdoelstelling, met name de beschrijving van de condities op de adjectivische buigingsalternantie bij neutra. Bij het empirisch onderzoek hebben we bij de analyse van de verschillende condities, c.q. hypothesen, een maximale objectiveerbaarheid en repliceerbaarheid nagestreefd (deel II).

Wat de kwantitatieve aspecten van het onderzoek aangaat, hebben we geargumenteed dat bivariate analyses naar het effect van afzonderlijke condities ontoereikend zijn om een complex verschijnsel als de adjectivische buigingsalternantie bij neutra op adequate wijze in kaart te brengen. De basisanalyse in de delen II en III vormt dan ook een methodologisch tweeluik, waarbij de bivariate analyses van de afzonderlijke factoren worden aangevuld met een multivariate analyse om het cumulatieve effect van de verschillende factoren te berekenen. Dit is echter geen betoog voor het opgeven van bivariaat onderzoek. De afzonderlijke bivariate analyses in deel II van dit proefschrift bewijzen het belang van bivariaat onderzoek bij de operationalisering, de parametrisering en de interpretatie van de afzonderlijke condities op het gebruik van beide adjectivische buigingsmorfemen. Deze bivariate analyses kunnen echter onmogelijk het einddoel van het empirisch onderzoek vormen: de complexiteit inherent aan het spontaan gerealiseerde taalgebruik vraagt om verdere en complexere methodologische stappen, waarbij de competitie en de interactie tussen de verschillende verklarende factoren in acht worden genomen. Door de ongecontroleerde samenstelling van een corpus, i.e. een verzameling van spontaan gerealiseerd taal materiaal, kan het effect van een verklarende factor immers geconditioneerd worden door andere factoren, zoals de stratumanalyses herhaaldelijk hebben aangetoond, of kan het reële effect van een factor(waarde) versluierd zijn door toedoen van de associatie(s) met andere, sterkere factoren, zoals we bij de bespreking van de voorbeelden van Simpsons paradox geïllustreerd hebben. Om deze concurrentie en interactie tussen de verklarende factoren in beeld te brengen hebben we de bivariate analyses (deel II) aangevuld met een multivariate analyse (deel III): aan de hand van een logistische-regressiemodel hebben we de gecombineerde impact van de verschillende verklarende factoren op de adjectivische buigingsalternantie bij neutra gemodelleerd. In een regressieanalyse wordt het effect op de responsvariabele van een factor(waarde) berekend, terwijl het effect van de andere factor(waarden) gecontroleerd wordt. Een multivariaat model, c.q. een meervoudig logistische-regressiemodel, biedt verschillende descriptieve voordelen. In de eerste plaats kunnen we de verschillende (significante) verklarende factoren rangschikken op basis van hun bijdrage tot de reductie van de onverklaarde

## CONCLUSIE

variatie en op basis van hun impact op de keuze van het adjectivische buigingsmorfeem (c.q. de grootte van de OR's voor de factorwaarden). Daarnaast biedt de constructie van een statistisch model de mogelijkheid om de descriptieve en verklarende kracht van het model en – bij uitbreiding – van het gevoerde onderzoek te evalueren.

We zijn van oordeel dat we de vooropgestelde doelstellingen gerealiseerd hebben. Vanuit een descriptief perspectief hebben we een globaal model voorgesteld dat niet alleen de significante condities op de buigingsalternantie samenbrengt, maar daarnaast ook een rangorde voorstelt op basis van de impact van deze condities op de keuze van het buigingsmorfeem. Dit model kan als vertrekpunt dienen bij toekomstig onderzoek naar de adjectivische buigingsalternantie bij neutra. De algemene onderzoeksvragen die aan het einde van deel I (zie paragraaf 5.4) geformuleerd zijn, kunnen we op de volgende wijze beantwoorden:

- De adjectivische buigingsalternantie bij neutra is geen marginaal verschijnsel.
- De keuze van het buigingsmorfeem is niet aleatoir, maar kan op adequate wijze gemodelleerd worden, zoals de modelstatistieken voor het logistische-regressiemodel op basis van het gesproken Nederlands aantonen. De concrete restricties op de selectie van beide buigingsvarianten worden samengevat in tabel 1.
- Het Belgisch Nederlands en het Nederlands Nederlands vertonen significante verschillen met betrekking tot de keuze van het buigingsmorfeem.

Vanuit een methodologisch standpunt hebben we voor de analyse van iedere conditie een objectieerbare procedure voorgesteld indien dit mogelijk bleek, en in het andere geval een intersubjectieve procedure. Daarnaast toont voorliggend onderzoek de descriptieve meerwaarde en de noodzaak aan van het gebruik van multivariate technieken om een complex fenomeen als de adjectivische buigingsalternantie bij neutra te modelleren.

Dit proefschrift beantwoordt uiteraard niet alle vragen over de adjectivische buigingsalternantie bij neutra. In eerste instantie moet het semantische onderzoek uit deel IV (hoofdstuk 17) gerepliceerd worden op basis van een ruimer corpus met gesproken materiaal om de semantische conditionering van de adjectivische buiging bij lexicale collocaties in de gesproken taal te bestuderen. Daarnaast moet verder onderzoek uitwijzen wat de relaties tussen de verschillende semantische condities zijn. Dit uitgebreide lexicale en semantische onderzoek kan als vertrekpunt dienen voor de vergelijking van het gebruik als benoemingsprocédé van adjectief-nomencollocaties en adjectief-nomensamenstellingen. De relatie tussen de buigingsalternantie in Nederland en Vlaanderen moet eveneens aan een diepgaand onderzoek onderworpen worden, waarbij expliciet de vraag wordt gesteld naar de tussentalige functie van de adjectivische buiging en van de adnominale flexie in het algemeen.

De voorgestelde regressiemodellen kunnen als uitgangspunt dienen voor een ruimer empirisch onderzoek naar de adjectivische buiging waarin ook de buiging van het adjectief bij *de*-woorden wordt betrokken. In de literatuur wordt er vanuit gegaan dat de buigingsalternantie strikter geconditioneerd is bij *de*-woorden dan bij *het*-woorden, hoewel hier – voor zover wij weten – geen empirisch onderzoek naar verricht is. Anders gezegd: welke factoren in de tabellen 1 en 2 zijn ook van toepassing op de keuze van het adjectivische buigingsmorfeem bij *de*-woorden? Uiteindelijk moet ook de vraag worden gesteld waarom de adjectivische buiging in het continentale Nederlands niet verdwenen is, zoals dat in verschillende variëteiten van het overzeese Nederlands wel het geval is. Deze vraag kan niet losgekoppeld worden van de vaststelling dat het adjectief bij neutra al sinds het Middelnederlands buigingsvariatie vertoont, zonder dat één van beide varianten (ten nadele van de andere) geregulariseerd is.

Methodologisch kunnen verfijndere methodes gebruikt worden om deelaspecten van voorliggend onderzoek te verfijnen. We denken in de eerste plaats aan experimenteel onderzoek naar de semantische patronen die aan de basis liggen van de lexicale collocaties, waarbij de semantische condities zeer gericht gemanipuleerd en geanalyseerd kunnen worden (terwijl de andere condities constant blijven).

Tot slot moeten complexe empirische modellen, zoals het model voorgesteld in voorliggend proefschrift, met theoretische taalkundige modellen geconfronteerd worden. Voor zover wij weten, bestaat er geen algemeen taalkundig model dat de diversiteit aan verklarende factoren – die zowel taalstructureel als semantisch zijn en die zowel taalintern als lectisch zijn – die in dit proefschrift naar boven zijn gekomen, integreert.

# Literatuur

- Agresti, Alan. 1996. *An Introduction to Categorical Data Analysis*. New York: Wiley.
- Anderson, Stephen R. 1982. "Where's Morphology?". *Linguistic Inquiry* 13. 571-612.
- Anderson, Stephen R. 1988. "Inflection". In Michael Hammond & Michael Noonan (red.) *Theoretical Morphology: Approaches to Modern Linguistics*. California: Academic Press. 23-44.
- Anderson, Stephen R. 1991. *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- ANS 1997: Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Tweede, geheel herziene druk. Groningen: Martinus Nijhoff Uitgevers – Deurne: Wolters Plantyn.
- Baayen, R.H., R. Piepenbrock & H. van Rijn. 1993. *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bakema, Peter. 1998. Het verkleinwoord verklaard: een morfosemantische studie over dimunitieven in het Nederlands. Doctorale dissertatie, Katholieke Universiteit Leuven.
- Barlow, Michael. 1991. "The Agreement Hierarchy and Grammatical Theory". *Berkeley Linguistic Society* 17. 30-40.
- Barlow, Michael. 1999. "Agreement as a Discourse Phenomenon". *Folia Linguistica* 33(2). 187-210.
- Barlow, Michael & Suzanne Kemmer (red.). 2000. *Usage-Based Models of Language*. Stanford: CSLI Publications.
- Bauer, Laurie. 2003. *Introducing Linguistic Morphology*. 2<sup>e</sup> herziene uitgave. Edinburgh: Edinburgh University Press.
- Bell, Alan. 1984. "Language style as audience design". *Language in Society* 13(2). 145-204.
- Bergen, Benjamin. Te verschijnen. "Social Variability and Probabilistic Language Processing." International Computer Science Institute. Technical Report.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

LITERATUUR

- Blom, Alied. 1994. "Het ondoorgrondelijk bijvoeglijk naamwoord". *Forum der Letteren* 35(2). 81-94.
- Blom, Alied. 1995. "Het ondoorgrondelijk(e) bijvoeglijk naamwoord". *Onze Taal* 64(4). 82-84.
- Bolinger, Dwight. 1967. "Adjectives in English: Attribution and prediction". *Lingua* 18. 1-34.
- Booij, Geert E. 1992. "Congruentie in Nederlandse NP's". *Spektator* 21(2). 119-135.
- Booij, Geert E. 1993. "Against Split Morphology". In Geert Booij & Jaap van Marle (red.) *Yearbook of Morphology 1993*. Dordrecht: Kluwer. 27-49.
- Booij, Geert E. 1995. *The Phonology of Dutch*. Oxford: Clarendon Press.
- Booij, Geert E. 1996. "Inherent versus contextual inflection and the split morphology hypothesis". In Geert E. Booij & Jaap van Marle (red.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer. 1-15.
- Booij, Geert E. 1998. "Phonological output constraints in morphology". In Wolfgang Kehrein & Richard Wiese (red.) *Phonology and Morphology of the Germanic Languages*. Tübingen: Niemeyer. 143-163.
- Booij, Geert E. 1999. "Morfologie". In Willy Smedts & Piet W. Paardekooper (red.) *De Nederlandse taalkunde in kaart*. Leuven: Acco. 121-128
- Booij, Geert E. 2000. "Inflection and derivation". In Geert E. Booij, Christian Lehman & Joachim Mugdan (red.) *Morphology. An International Handbook on Inflection and Word-Formation*. Berlijn: Mouton de Gruyter. 360-369.
- Booij, Geert E. 2002a. "Constructional Idioms, Morphology, and the Dutch Lexicon". *Journal of Germanic Linguistics* 14(4). 301-329.
- Booij, Geert E. 2002b. "Language variation and phonological theory: inflected adjectives in Dutch and related languages". In Jan Berns & Jaap van Marle (red.) *Trends in Linguistics. Present-day Dialectology*. Berlijn: Mouton de Gruyter. 35-56.
- Booij, Geert E. 2002c. *The Morphology of Dutch*. Oxford: Oxford University Press.
- Booij, Geert E. & Arianne van Santen. 1998. *Morfologie. De woordstructuur van het Nederlands*. 2<sup>e</sup> herziene druk. Amsterdam: Amsterdam University Press.
- van Bree, Cor. 1980. *Historische grammatica van het Nederlands*. Dordrecht: Foris.
- Broekhuis, Hans. 1999. "Adjectives and Adjective Phrases". *Modern Grammar of Dutch Working Papers 2*. Tilburg: Universiteit Tilburg.
- Bybee, Joan L. 2000. "Units of morphological structure". In Geert E. Booij, Christian Lehman & Joachim Mugdan (red.) *Morphology. An International Handbook on Inflection and Word-Formation*. Berlijn: Mouton de Gruyter. 370-377.

- Bybee, Joan L. Te verschijnen. "Diachronic Linguistics". In Dirk Geeraerts & Hubert Cuyckens (red.) *Handbook of Cognitive Linguistics*. Oxford: Oxford University Press.
- Bybee, Joan & Paul Hopper (red.). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Carstairs-McCarthy, Andrew. 1998. "Paradigmatic Structure: Inflectional Paradigms and Morphological Classes". In Andrew Spencer & Arnold M. Zwicky (red.) *Handbook of Morphology*. Oxford: Blackwell Publishers. 322-334.
- Chandan, Mukherjee, Howard White & Marc Whyte. 1998. *Econometrics and Data Analysis for Developing Countries*. London: Routledge.
- Church, Kenneth Ward & Patrick Hanks. 1990. "Word association norms, mutual information, and lexicography". *Computational Linguistics* 16(1). 22-29.
- Clarke, Sandra. 1997. "English Verbal -s Revisited: The Evidence from Newfoundland". *American Speech* 72(3). 227-259.
- Clyne, Michael (red.) 1992. *Pluricentric Languages. Differing Norms in Different Nations*. Berlijn: Mouton de Gruyter.
- Cockx, Paul. 2000. *Taalwijzer*. 3<sup>e</sup> druk. Leuven: Davidsfonds.
- Corbett, Greville G. 1979. "The Agreement Hierarchy". *Journal of Linguistics* 15. 203-224.
- Corbett, Greville G. 1998. "Morphology and Agreement". In Andrew Spencer & Arnold M. Zwicky (red.) *Handbook of Morphology*. Oxford: Blackwell Publishers. 191-205.
- Coupland, Nikolas. 1980. "Style-shifting in a Cardiff work setting". *Language in Society* 9(1). 1-12.
- Croft, William & D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Cruse, D. Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Daan, Jo. 1969. "Ons oude huis". *Taal en Tongval* 21. 112-114.
- De Caluwe, Johan. 1990. "Complementariteit tussen morfologische en in oorsprong syntactische benoemingsprocédés". In Johan De Caluwe (red.) *Betekenis en productiviteit*. Gent: Seminarie voor Duitse Taalkunde. 9-23.
- De Caluwe, Johan. 2003. "Tien stellingen over functie en status van tussentaal in Vlaanderen". In Johan De Caluwe, Dirk Geeraerts, Sjaak Kroon, Virginie Mamadouh, Ronald Soetaert, Luc Top & Ton Vallen (red.) *Taalvariatie en taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen – Apeldoorn: Garant. 57-67.
- De Groot, A. W. 1949. *Structurele Syntaxis*. Den Haag: Servire.

LITERATUUR

- De Schutter, Georges. 1968. "Semantische subcategorieën van het Nederlandse adjectief". *Handelingen der Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* XXII. 175-191.
- De Schutter, Georges. 1993. "Klemtoonpatronen in de Nederlandse woordenschat". *Leuvense Bijdragen* 82(1). 61-82.
- De Schutter, Georges. 1994. "De Interne Structuur van Nominale Constituenten, en hun Formele Kenmerken". *Gramma/TTT* 3(1). 13-31.
- De Schutter, Georges. 1997. "The Noun Phrase in Dutch". *Leuvense Bijdragen* 86(3). 309-356.
- De Sutter, Gert. 2005. Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. Doctorale dissertatie, Katholieke Universiteit Leuven.
- De Vooy, C. G. N. 1967. *Nederlandse Spraakkunst*. 7<sup>e</sup> druk. Groningen: Wolters.
- Debrock, Mark & Piet Mertens. 1993. *Phonétique générale et française*. Leuven: Universitaire Pers Leuven.
- Degryse, Zr. Lydie. 1959. Statistisch en beschrijvend onderzoek van de verbuiging van het bijvoeglijk naamwoord in het Modern Nederlands Proza (ca. 1950). Licentiaatsverhandeling, Katholieke Universiteit Leuven.
- Dirven, René. 1999. "The cognitive motivation for adjective sequences in attribution". *Journal of English Studies* 1. 57-67.
- Dubois, Sylvie & Barbara M. Horvath. 2003. "The English Vernacular of the Creoles of Louisiana". *Language Variation and Change* 15. 255-288.
- Dunning, Ted. 1993. "Accurate methods for the statistics of surprise and coincidence". *Computational Linguistics* 19(1). 61-74.
- Ebert, Robert P. 1992. "Internal and external factors in syntactic change in an historical speech community". In Marinel Gerritsen & Dieter Stein (red.) *Internal and External Factors in Syntactic Change*. Berlijn: Mouton de Gruyter. 201-228.
- Eckert, Penelope. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell Publishers.
- Everaert, Martin. 1993. "Vaste verbindingen (in woordenboeken)". *Spektator* 22(1). 3-27.
- Evert, Stefan & Brigitte Krenn. 2001. "Methods for the Qualitative Evaluation of Lexical Association Measures". *Proceedings of the 39<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Toulouse. 188-195.
- Evert, Stefan, Ulrich Heid & Kristina Spanger. 2004. "Identifying Morphosyntactic Preferences in Collocations". *Proceedings of the 4<sup>th</sup> International Conference on Language Resources and Evaluation*, Lissabon. 907-910.



- Ferguson, Charles A. & Michael Barlow. 1988. "Introduction". In Michael Barlow & Charles A. Ferguson (red.) *Agreement in Natural Language. Approaches, Theories, Descriptions*. Stanford: CSLI Publications. 1-22.
- Geeraerts, Dirk. 1986. *Woordbetekenis: een overzicht van de lexicale semantiek*. Leuven: Acco.
- Geeraerts, Dirk. 1989. *Wat er in een woord zit. Facetten van de lexicale semantiek*. Leuven: Peeters.
- Geeraerts, Dirk. 1995. "Specialization and Reinterpretation in Idioms". In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (red.) *Idioms: Structural and Psychological Perspectives*. Hillsdale: Lawrence Erlbaum. 57-74.
- Geeraerts, Dirk. 2000. *Van Dale Groot woordenboek der Nederlandse taal*. CD-rom: Versie 1.0 Plus op basis van de dertiende uitgave. Utrecht – Antwerpen: Van Dale Lexicografie.
- Geeraerts, Dirk. 2001. "Het Nederlands in Vlaanderen: gedrag, beleid, attitudes". *Ons Erfdeel* 44. 337-344.
- Geeraerts, Dirk. 2003a. "Cultural models of linguistic standardization". In René Dirven, Roslyn Frank & Martin Pütz (red.) *Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings*. Berlijn: Mouton de Gruyter. 25-68.
- Geeraerts, Dirk. 2003b. "Decontextualizing and recontextualizing tendencies in 20<sup>th</sup>-century linguistics and literary theory". In Ewald Mengel, Hans-Jörg Smidt & Michael Steppat (red.) *Anglistentag 2002 Bayreuth*. Trier: Wissenschaftlicher Verlag. 369-379.
- Geeraerts, Dirk. 2005. "Lectal Variation and Empirical Data in Cognitive Linguistics". Te verschijnen in Francisco Ruiz de Mendoza (red.) *Cognitive Linguistics, Functionalism, Discourse Studies: Common Ground and New Directions*. Berlijn: Mouton de Gruyter.
- Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlijn: Mouton de Gruyter.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P. J. Meertens Instituut.
- Geerts, Guido. 1966. *Genus en geslacht in de gouden eeuw. Een bijdrage tot de studie van de nominale classificatie en daarmee samenhangende adnominale flexievormen en pronominale verschijnselen in Hollands taalgebruik van de zeventiende eeuw*. Brussel: Belgisch Interuniversitair Centrum voor Neerlandistiek.

LITERATUUR

- Geerts, Guido. 1992. "Is Dutch a pluricentric language?". In Michael Clyne (red.) *Pluricentric Languages. Differing Norms in Different Nations*. Berlijn: Mouton de Gruyter. 71-91.
- Gibbs, Raymond W. 1995. "Idiomaticity and Human Cognition". In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (red.) *Idioms: Structural and Psychological Perspectives*. Hillsdale: Lawrence Erlbaum. 97-116
- van Ginneken, Jacques. 1934. "De geschiedenis der drie geslachten in Nederland". *Onze Taaltuin* 3(2). 33-42.
- van Ginneken, Jacques. 1937. "Het onbepaald lidwoord en het geslacht". *Onze Taaltuin* 5(12). 353-356.
- Goossens, Jan. 2000. "De toekomst van het Nederlands in Vlaanderen". *Ons Erfdeel* 43. 2-13.
- Gries, Stefan. 2001. "A Multifactorial Analysis of Syntactic Variation: Particle Movement Revisited." *Journal of Quantitative Linguistics* 8(1). 33-50.
- Gries, Stefan. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London: Continuum Press.
- Grondelaers, Stefan. 2000. De distributie van niet-anaforsch *er* buiten de eerste zinsplaats. Sociolectische, functionele en psycholinguïstische aspecten van *er*'s status als presentatief signaal. Doctorale dissertatie, Katholieke Universiteit Leuven.
- Grondelaers, Stefan, Marc Brysbaert, Dirk Speelman & Dirk Geeraerts. 2001. "*Er* als accessibility marker: on- en offline evidentie voor een procedurele interpretatie van presentatieve zinnen". *Gramma/TTT* 9(1). 1-22.
- Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van Den Heede & Dirk Speelman. 2000. "Het CONDIV-corpus geschreven Nederlands". *Nederlandse Taalkunde* 5(4). 356-363.
- Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts. 2002. "Regressing on *er*. Statistical analysis of texts and language variation". In A. Morin & P. Sébillot (red.) *6<sup>th</sup> International Conference on the Statistical Analysis of Textual Data*. Rennes: Institut National de Recherche en Informatique et en Automatique. 335-346.
- Gussenhoven, Carlos. 1983a. "Stress shift and the nucleus". *Linguistics* 21. 303-339.
- Gussenhoven, Carlos. 1983b. "Stress shift in Dutch as a rhetorical device". *Linguistics* 21. 603-619.
- Guy, Gregory R. 1994. "The Phonology of Variation". In Katherine Beals, Jeannette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki & Erica Zeinfeld (red.) *CLS 30. Papers from the 30<sup>th</sup> Regional Meeting of the Chicago Linguistic Society*. Volume 2. *Parasession on Variation in Linguistic Theory*. 133-149.

- de Haas, Wim & Mieke Trommelen. 1993. *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming*. Den Haag: Sdu Uitgevers.
- van Haeringen, Coenraad B. 1937. "Opmerkingen bij de apocope van *-e*". *De Nieuwe Taalgids* 31. 241-250, 322-331.
- van Haeringen, Coenraad B. 1962. "Franciscaner, Benedictijner, Karmelieter". In Coenraad B. van Haeringen (red.) *Grammarie*. Assen: Van Gorcum & Comp. 197-202.
- Haeseryn, Walter. 1996. "Grammaticale verschillen tussen het Nederlands in België en het Nederlands in Nederland: een poging tot inventarisatie". In Roeland van Hout, Joep Kruijssen & Toon Hagen (red.) *Taalvariaties. Toonzettingen en modulaties op een thema*. Dordrecht: Foris. 109-126.
- Haiman, John. 1985. *Natural Syntax*. Cambridge: Cambridge University Press.
- Halliday, Michael A.K. 1994. *An Introduction to Functional Grammar*. 2<sup>e</sup> uitgave. London: Arnold.
- Hampe, Beate. 2000. "Facing up to the Meaning of 'face up to': A Cognitive Semantico-Pragmatic Analysis of an English Verb-Particle Construction". In Ad Foolen & Frederike van der Leek (red.) *Constructions in Cognitive Linguistics*. Amsterdam: John Benjamins. 81-101.
- Harrell, Frank E. 2001. *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression*. New York: Springer.
- Heemskerk, José. 2001. "Het VDD-klemtoonsysteem. Documentatie, aanbevelingen en regels". Interne nota Van Dale Lexicografie.
- den Hertog, C.H. 1897. *De Nederlandsche taal. Praktische Spraakkunst van het hedendaagsche Nederlandsch*. Amsterdam.
- Heylen, Kris. 2005. Zur Abfolge (pro)nominaler Satzglieder im Deutschen. Eine korpusbasierte Analyse. Doctorale dissertatie, Katholieke Universiteit Leuven. [te verschijnen]
- Heylen, Kris, Jose Tummers & Dirk Geeraerts. Te verschijnen. "Methodological issues in corpus-based Cognitive Linguistics". In Gitte Kristiansen & René Dirven (red.) *Cognitive sociolinguistics: Language variation, cultural models, social systems*. Berlijn: Mouton de Gruyter.
- Heynderickx, Prescilla. 2001. *Relationele adjectieven in het Nederlands*. Berchem: EPO Drukkerij.
- Hinskens, Frans, Roeland van Hout & W. Leo Wetzels 1997. "Balancing Data and Theory in the Study of Phonological Variation". In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (red.) *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins. 1-34.

LITERATUUR

- Honselaar, Wim. 1980. "On the semantics of the adjective-noun combinations". In A.A. Barentsen, B.M. Groen & R. Sprenger (red.) *Studies in Slavic and General Linguistics*. Volume 1. 187-206.
- Hoppenbrouwers, Cor & Geer Hoppenbrouwers. 2001. *De indeling van de Nederlandse streektalen*. Assen: Van Gorcum.
- van der Horst, Johannes M. 1992. "Tets over *veel* en *vele*." In E.C. Schermer-Vermeer, W.G. Klooster & A.F. Florijn (red.) *De kunst van de grammatica. Artikelen aangeboden aan Frida Balk-Smit Duyzentkunst*. Amsterdam: Vakgroep Nederlandse Taalkunde Universiteit van Amsterdam. 111-118.
- van der Horst, Johannes M. 1995. *Analytische taalkunde*. Groningen: Martinus Nijhoff.
- van der Horst, Joop & Kees van der Horst. 1999. *Geschiedenis van het Nederlands in de twintigste eeuw*. Den Haag: Sdu Uitgevers.
- Hudson, Richard A. 1996. *Sociolinguistics*. 2<sup>e</sup> uitgave. Cambridge: Cambridge University Press.
- Hüning, Matthias. 2004. "Over woorden en woordgroepen: A+N-verbindingen in het Nederlands en het Duits". In Stefan Kiedron & Agata Kowalska-Szubert (red.) *Thesaurus polyglottus et flores quadrilingues. Festschrift für Stanislaw Predota zum 60. Geburtstag*. Wrocław: Oficyna Wydawnicza ATUT. 159-171.
- van der Hulst, Harry & Jan Kooij. 1998. "Prosodic choices and the Dutch nominal plural". In Wolfgang Kehrein & Richard Wiese (red.) *Phonology and Morphology of the Germanic Languages*. Tübingen: Niemeyer. 187-197.
- Hunston, Susan. 2001. "Colligation, lexis, pattern, and text". In Mike Scott & Geoff Thompson (red.) *Patterns of Text: In Honour of Michael Hoey*. Amsterdam: John Benjamins. 13-34.
- Kager, René & Ellis Visch. 1988. "Metrical constituency and rhythmic adjustment". *Phonology* 5. 21-71.
- van Kerckvoorde, Colette M. 1993. *An Introduction to Middle Dutch*. Berlijn: Mouton de Gruyter.
- Kester, Ellen-Petra. 1996. *The Nature of Adjectival Inflection*. Doctorale dissertatie, Universiteit Utrecht.
- Kiefer, Ferenc. 1998. "Morphology and Pragmatics". In Andrew Spencer & Arnold M. Zwicky (red.) *Handbook of Morphology*. Oxford: Blackwell Publishers. 272-279.
- Kiparsky, Paul. 1982a. "Explanation in phonology". In Paul Kiparsky (red.) *Explanation in phonology*. Dordrecht: Foris Publications. 81-118.
- Kiparsky, Paul. 1982b. "Historical linguistics". In Paul Kiparsky (red.) *Explanation in phonology*. Dordrecht: Foris Publications. 45-80.

- Klein, M. & M. Visscher. 1996. *Handboek Verzorgd Nederlands. Spellingsregels. Schrijfdeviezen*. 2<sup>e</sup> herziene druk. Groningen: Nijhoff.
- Klooster, Wim. 2001. *Grammatica van het hedendaags Nederlands. Een volledig overzicht*. Den Haag: Sdu Uitgevers.
- Kooij, Jan & Marc van Oostendorp. 2003. *Fonologie. Uitnodiging tot de klankleer van het Nederlands*. Amsterdam: Amsterdam University Press.
- Krahe, Hans & W. Meid. 1969. *Germanische Sprachwissenschaft. II. Formenlehre*. Berlijn: Walter de Gruyter.
- Kristiansen, Gitte. 2003. "How to do things with allophones: Linguistic stereotypes as cognitive reference points in social cognition". In René Dirven, Roslyn Frank & Martin Pütz (red.) *Cognitive Models in Language and Thought: Ideology, Metaphors, and Meanings*. Berlijn: Mouton de Gruyter.
- Labov, William. 1969. "Contraction, deletion, and inherent variation in English copula". *Language* 45. 725-762.
- Labov, William. 1972a. *Sociolinguistic Patterns*. Oxford: Blackwell Publishers.
- Labov, William. 1972b. "Some principles of linguistic methodology". *Language in Society* 1. 97-120.
- Labov, William (red.) 1980. *Locating Language in Time and Space*. New York: Academic Press.
- Labov, William. 1994. *Principles of Linguistic Change. Volume 1. Internal Factors*. Oxford: Blackwell Publishers.
- Lass, Roger. 1990. "How to do things with junk: exaptation in language evolution". *Journal of Linguistics* 26. 79-102.
- Lebrun, Yvan & Gerlinda Schurmans-Swillen. 1966. "Verbogen tegenover onverbogen adjectieven in de taal van de Zuidnederlandse dagbladpers". *Taal en Tongval* 18(1). 175-187.
- Leech, Geoffrey, Brian Francis & Xunfeng Xu. 1994. "The use of computer corpora in the textual demonstrability of gradience in linguistic categories". In Catherine Fuchs & Bernard Victorri (red.) *Continuity in Linguistic Semantics*. Amsterdam: John Benjamins. 57-78.
- Lehmann, Christian. 1988. "On the function of agreement". In Michael Barlow & Charles A. Ferguson (red.) *Agreement in Natural Language. Approaches, Theories, Descriptions*. Stanford: CSLI Publications. 55-65.
- Levelt, Willem J.M. 1989. *Speaking. From Intention to Articulation*. Cambridge: The MIT Press.
- Levinson, Stephen & Sérgio Meira. 2003. "'Natural concepts' in the spatial topological domain – adpositional meanings in cross-linguistic perspective: An exercise in semantic typology". *Language* 79(3). 485-516.
- van Loon, Jozef. 1986. *Historische fonologie van het Nederlands*. Leuven: Acco.

LITERATUUR

- van Loon, Jozef. 1988. "Zwakke, sterke en pronominale adjectiefflexie in het Oudnederlands". *Naamkunde* 20. 37-41.
- Manning, Christopher D. & Hinrich Schütze. 2002. *Foundations of Statistical Natural Language Processing*. Cambridge: The MIT Press.
- van Marle, Jaap. 1995a. "On the fate of adjectival declension in Overseas Dutch (with some notes on the history of Dutch)". In Henning Andersen (red.) *Historical Linguistics 1993*. Amsterdam: John Benjamins. 183-294.
- van Marle, Jaap. 1995b. "The unity of morphology: on the interwovenness of the derivational and inflectional dimension of the word". In Geert E. Booij & Jaap van Marle (red.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer. 67-82.
- van Marle, Jaap. 1997. "Dialect versus standard language: nature versus culture". In Jenny Cheshire & Dieter Stein (red.) *Taming the Vernacular. From Dialect to Written Language*. London – New York: Longman. 13-34.
- van Marle, Jaap & Caroline Smits. 1993. "The Inflectional Systems of Overseas Dutch". In Henk Aertsen & Robert J. Jeffers (red.) *Historical Linguistics 1989*. Amsterdam: John Benjamins. 313-328.
- Marynissen, Ann. 2004. "De conditionering van de sjwa-apocope bij zijn ontstaan in het Middelnederlands". In Johan De Caluwe, Georges De Schutter, Magda Devos & Jacques Van Keymeulen (red.) *Taeldeman, man van de taal, schatbewaarder van de taal*. Gent: Academia Press. 609-620.
- Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press.
- Meesters, Gert. 2002. Marginale morfologie in het Nederlands: paradigmatische samenstellingen, neoklassieke composita en splintercomposita. Doctorale dissertatie, Katholieke Universiteit Leuven.
- Mondorf, Britta. 2003. "Support for *more*-support". In Günter Rohdenburg & Britta Mondorf (red.) *Determinants of Grammatical Variation*. Berlijn: Mouton de Gruyter. 251-304.
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.
- Moravcik, Edith A. 1978. "Agreement". In Joseph H. Greenberg (red.) *Universals of Human Language*. Volume 4. *Syntax*. Stanford: Stanford University Press. 331-373.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Longman.
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. "Idioms". *Language* 70(3). 491-538.

- Odiijk, Jan. 1992. "Uninflected adjectives in Dutch". In Reineke Bok-Bennema & Roeland van Hout (red.) *Linguistics in the Netherlands 9*. Amsterdam: John Benjamins. 197-208.
- Oostdijk, Nelleke. 2000. "Het Corpus Gesproken Nederlands". *Nederlandse Taalkunde* 5(3). 280-284.
- van Oostendorp, Marc. 2000. *Phonological Projection. A Theory of Feature Content and Prosodic Structure*. Berlijn: Mouton de Gruyter.
- Overdiep, G.S. 1937. *Stilistische grammatica van het modern Nederlandsch*. Zwolle: W.E.J. Tjeenk Willink.
- Pander Maat, Henk & Ted Sanders. 2001. "Subjectivity in causal connectives: An empirical study of language in use". *Cognitive Linguistics* 12(3). 247-273.
- Paolillo, John C. 2001. *Analyzing Linguistic Variation. Statistical Models and Methods*. Stanford: CSLI Publications.
- Pedersen, Ted. 1996. "Fishing for Exactness". *Proceedings of the South-Central SAS Users Group Conference (SCSUG-96)*, Austin, TX, October 27-29. 188-200.
- Penninckx, Willy & Paul Buyse. 1997. *Correct taalgebruik*. 6<sup>e</sup> druk. Kortrijk: UGA.
- Permentier, Ludo & Ludo Van Den Eynden. 1998. *Stijlboek*. 2<sup>e</sup> druk. Groot-Bijgaarden: Scoop.
- Peters, P. R. 1937. "De geslachtsvormen van het adjectief in de Nederlandse dialecten". *Onze Taaltuin* 5(12). 357-379.
- Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Poplack, Shana. 2001. "Variability, frequency and productivity in the irrealis domain of French". In Joan Bybee & Paul Hopper (red.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins. 405-428.
- Prokosch, E. 1939. *A Comparative Germanic Grammar*. Philadelphia: Linguistic Society of America.
- Quak, A. & Joop M. van der Horst. 2002. *Inleiding Oudnederlands*. Leuven: Universitaire Pers Leuven.
- R Development Core Team. 2004. *R: A language and environment for statistical computing*. Wenen: R Foundation for Statistical Computing. Url: <http://www.r-project.org>.
- Raidt, Edith Hildegard. 1968. *Geskiedenis van die byvoeglike verbuiging in Nederlands en Afrikaans*. Nasou: Beperk.
- Rand, David & David Sankoff. 1990. *GoldVarb: A Variable Rule Application for Macintosh*. Montreal: Centre de Recherches Mathématiques, Université de Montréal.
- Rietveld, Toni & Roeland van Hout. 1993. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlijn: Mouton de Gruyter.

LITERATUUR

- Rietveld, Toni, Roeland van Hout & Mirjam Ernestus. 2004. "Pitfalls in Corpus Research". *Computers and the Humanities* 38. 343-362.
- Rijpma, E. & F.G. Schuringa. 1978. *Nederlandse Spraakkunst*. 25<sup>e</sup> druk, bewerkt door Jan van Bakel. Groningen: Wolters-Noordhoff.
- de Rooij, Jaap. 1972. "Algemeen Zuidnederlands?". *Bijdragen en Mededelingen der dialectcommissie van de Koninklijke Nederlandse Academie van Wetenschappen te Amsterdam* XLIII. 5-18.
- de Rooij, Jaap. 1980a. "Ons bruin(e) paard I". *Taal en Tongval* 32. 3-25.
- de Rooij, Jaap. 1980b. "Ons bruin(e) paard II". *Taal en Tongval* 32. 109-129.
- Rousseau, Pascale & David Sankoff. 1978. "Advances in variable rule methodology". In David Sankoff (red.) *Linguistic Variation: Models and Methods*. New York: Academic Press. 57-69.
- Royen, P. Gerlach. 1953. *Buigingsverschijnselen in het Nederlands*. Deel III. Amsterdam: North-Holland Publishing Company.
- Schenk, André. 1995. "The Syntactic Behavior of Idioms". In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (red.) *Idioms: Structural and Psychological Perspectives*. Hillsdale: Lawrence Erlbaum. 253-272.
- Schultink, Hendrik. 1962. *De morfologische valentie van het ongelede adjectief in modern Nederlands*. Den Haag: Van Goor Zonen. [doctorale dissertatie, Rijksuniversiteit Leiden]
- Schultink, Hendrik. 1977. "Reacties op 'stress clash': de accentuering van samenstellende afleidingen, afleidingen van composita, en composita in het Nederlands". *Spektator* 8. 195-208.
- Schultink, Hendrik. 1980. "Boundaries, wordclasses, and the accentuation of derived words in Dutch". In Wim Zonneveld, Frans van Coetsem & Orwin W. Robinson (red.) *Studies in Dutch Phonology*. Den Haag: Martinus Nijhoff. 205-221.
- Schultink, Hendrik. 2000. "The Netherlands". In Geert E. Booij, Christian Lehman & Joachim Mugdan (red.) *Morphology. An International Handbook on Inflection and Word-Formation*. Berlijn: Mouton de Gruyter. 162-170.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John. 1995. *Collins Cobuild English Dictionary*. London: Harper Collins.
- Speelman, Dirk. 1997. *Abundantia Verborum: A computer tool for carrying out corpus-based linguistic case studies*. Doctorale dissertatie, Katholieke Universiteit Leuven. Url: <http://www.ling.arts.kuleuven.ac.be/genling/abundant/>.
- Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts. 2003. "Profile-Based Linguistic Uniformity as a Generic Method for Comparing Language Varieties." *Computers and the Humanities* 37(3). 317-337.



- Stefanowitsch, Anatol. 2003. "Constructional semantics as a limit to grammatical alternation: The two genitives of English". In Günter Rohdenburg & Britta Mondorf (red.) *Determinants of Grammatical Variation*. Berlijn: Mouton de Gruyter. 413-444.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. "Collostructions: Investigating the interaction between words and constructions." *International Journal of Corpus Linguistics* 8(2). 209-243.
- Stefanowitsch, Anatol & Stefan Th. Gries. Te verschijnen. "Register and constructional meaning: A collostructional case study". Te verschijnen in René Dirven & Gitte Kristiansen (red.) *Cognitive Sociolinguistics: Language variation, Cultural Models, Social Systems*. Berlijn: Mouton de Gruyter.
- van Sterkenburg, Piet G.J. 1993. "Gelexicaliseerde woordgroepen van het type A+N". *Tabu* 23(1-2). 131-142.
- Stubbs, Michael. 1995. "Collocations and semantic profiles: On the cause of the trouble with quantitative studies". *Functions of Language* 2(1). 23-56.
- Stubbs, Michael. 2001. *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Stump, Gregory. 1998. "Inflection". In Andrew Spencer & Arnold M. Zwicky (red.) *Handbook of Morphology*. Oxford: Blackwell Publishers. 13-43.
- Stuurman, Frits. 1989. "Van X-bar structuren voor NPs naar deflectie bij prenominale adjectieven in het Nederlands". *Gramma* 13. 161-185.
- Szmrecsanyi, Benedikt. 2005. "Language users as creatures of habit: A corpus-based analysis of persistence in spoken English". *Corpus Linguistics and Linguistic Theory* 1(1). 113-150.
- Taeldeman, Johan. 1978. "De vokaalstructuur van de 'Oostvlaamse' dialecten. Een poging tot historische en geografische situering in het Zuidnederlandse dialektlandschap". *Bijdragen en Mededelingen der Koninklijke Nederlandse Academie voor Wetenschappen in Amsterdam* LI. 3-68.
- Taeldeman, Johan. 1980. "Inflectional aspects of adjectives in the dialects of Dutch-speaking Belgium". In Wim Zonneveld, Frans van Coetsem & Orrin W. Robinson (red.) *Studies in Dutch Phonology*. Groningen: Martinus Nijhoff. 223-246.
- Taeldeman, Johan. 1992. "Welk Nederlands voor de Vlamingen?". *Nederlands van Nu* 40(2). 33-52.
- Tagliamonte, Sali A. 2003. "'Every place has a different toll': Determinants of grammatical variation in cross variety perspective". In Günter Rohdenburg & Britta Mondorf (red.) *Determinants of Grammatical Variation*. Berlijn: Mouton de Gruyter. 531-554.

LITERATUUR

- Te Winkel, J. 1901. *Geschiedenis der Nederlandsche Taal*. Vertaald door F.C. Wieder. Culemborg: Blom & Olivierse.
- Tomasello, Michael. 2000. "First Steps toward a Usage-based Theory of Language Acquisition". *Cognitive Linguistics* 11(1-2). 61-82.
- van den Toorn, M.C. 1982. *Nederlandse Grammatica*. Groningen: Wolters Nijhoff.
- van den Toorn, M.C., J.A. van Pijnenburg, W.J.J. Leuvensteijn & J.M. van der Horst. 1997. *Geschiedenis van de Nederlandse taal*. Amsterdam: Amsterdam University Press.
- Trommelen, Mieke. 1983. *The syllable in Dutch. With special reference to diminutive formation*. Dordrecht: Foris.
- Trommelen, Mieke & Wim Zonneveld. 1989. *Klemtoon en metrische fonologie*. Muiderberg: Coutinho.
- Trommelen, Mieke & Wim Zonneveld. 1999. "Word stress in West-Germanic and North-Germanic languages: Dutch". In Harry van der Hulst (red.) *Word Prosodic Systems in the Languages of Europe*. Berlijn: Mouton de Gruyter. 492-515.
- Tummers, Jose, Kris Heylen & Dirk Geeraerts. 2005. "Usage-based approaches in Cognitive Linguistics. A technical state of the art". *Corpus Linguistics and Linguistic Theory* 1(2).
- Tummers, Jose, Dirk Speelman & Dirk Geeraerts. 2005. "Inflectional variation in Belgian and Netherlandic Dutch: A usage-based account of the adjectival inflection". In Nicole Delbecque, Johan Van Der Auwera & Dirk Geeraerts (red.) *Perspectives on Variation. Sociolinguistic, Historical, Comparative*. Berlijn: Mouton de Gruyter. 93-110.
- Van de Velde, Hans. 1996. "Verkavelingsvlaams. Wat is me dat nu?". In Roeland van Hout, Joep Kruijsen & Toon Hagen (red.) *Taalvariaties. Toonzettingen en modulaties op een thema*. Dordrecht: Foris. 261-271.
- Van Eynde, Frank. 2001. "Part of Speech Tagging en Lemmatisering". Interne publicatie CGN-project.
- Van Eynde, Frank. 2003. "Morpho-Syntactic Agreement and Index Agreement in Dutch NPs". In Tanja Gaustad (red.) *Computational Linguistics in the Netherlands 2002*. Amsterdam: Rodopi. 111-127.
- Van Helten, W. L. 1973 [1889<sup>1</sup>]. *Middelnederlandsche Spraakkunst*. Groningen: Wolters.
- Van Loey, A. 1949. *Middelnederlandse spraakkunst*. Antwerpen: De Sikkel.
- Van Loey, A. 1970 [1959<sup>1</sup>]. *Schönfelds historische grammatica van het Nederlands*. 8<sup>e</sup> druk. Zupthen: W.J. Thieme & Cie.
- Venables, William N. & Brian D. Ripley. 2002. *Modern Applied Statistics with S*. 4<sup>e</sup> uitgave. Berlijn: Springer Verlag.

- Verhagen, Arie. 1993. "The Inflection of Adjectives with Nominal Infinitives". In Robert S. Kirsner (red.) *The Low Countries and Beyond*. New York: University Press of America. 97-109.
- Verhagen, Arie & Jeroen van de Weijer (red.). 2003. *Usage-Based Approaches to Dutch*. Utrecht: LOT.
- Verhasselt, J. 1974. "Adjectieven met of zonder e". *Nu Nog* 22. 140-141.
- Verschuieren, Jef. 1987. *Pragmatics as a theory of linguistic adaptation*. Antwerpen: IPRA.
- Weeber, Marc, Rein Vos & R. Harald Baayen. 2000. "Extracting the Lowest-Frequency Words: Pitfalls and Possibilities". *Computational Linguistics* 26(3). 301-317.
- Weerman, Fred. 2003 "Een mooie verhaal; veranderingen in uitgangen". In Jan Stroop (red.) *Waar gaat het Nederlands naartoe?* Amsterdam: Bert Bakker. 249-260.
- Weijnen, Antonius. 1991. *Vergelijkende klankleer van de Nederlandse dialecten*. 's Gravenhage: Sdu Uitgevers.
- Wonnacott, Thomas H. & Ronald J. Wonnacott. 1990. *Introducing Statistics*. 5<sup>e</sup> uitgave. New York: John Wiley & Sons.
- Woods, Anthony, Paul Fletcher & Arthur Hughes. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press.
- van der Wouden, Ton 1992. "Beperkingen op het opreden van lexicale elementen". *De Nieuwe Taalgids* 85. 513-538.
- van der Wouden, Ton 2001. "Collocational behaviour in non-content words". In Béatrice Daille & Geoffrey Williams (red.) *Collocation. Computational Extraction, Analysis and Exploitation*. Toulouse: Institut de Recherche en Informatique de Toulouse. 16-23.
- Zonneveld, Wim. 2003. "Zijn we er op  $\varnothing$  gegaan in de fonologie?". In Georges De Schutter & Steven Gillis (red.) *Fonologische kruispunten*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde. 21-52.
- Zwicky, Arnold M. 1986. "German adjective agreement in GPSG". *Linguistics* 24. 957-990.