

Hebben we gaan twijfelen over het hulpwerkwoord?

Een experimenteel en corpusgebaseerd onderzoek naar hulpwerkwoordselectie in clusters met een aspectueel werkwoord

Fien Croux

Masterproef aangeboden binnen de opleiding
Educatieve master in de talen

Promotor: prof. dr. Jeroen van Craenenbroeck
Co-promotor: prof. dr. Stef Grondelaers

Academiejaar 2024-2025

335.081 tekens



Ik verklaar me akkoord met de code of conduct van de faculteit Letteren voor geloofwaardig auteurschap.

Dankwoord

Eerst en vooral wil ik mijn promotor, prof. dr. Jeroen van Craenenbroeck, bedanken voor de interessante begeleiding en de waardevolle feedback. Ik bedank hem om te geloven in mijn kunnen wanneer ik dat zelf niet deed en om begripvol te zijn wanneer het even allemaal te veel werd. Ook bedank ik hem voor de kennismaking met prof. dr. Stef Grondelaers, die als co-promotor onmisbaar was voor deze masterproef. Hij gooide me in de diepe wateren van het experimentele onderzoek, en leerde me erin zwemmen. Zonder hen was het schrijven van deze masterproef nooit zo'n verrijkende ervaring geweest.

Mijn bijzondere dank gaat uit naar iedereen die heeft bijgedragen aan het experiment dat de basis vormt voor deze masterproef. Ik bedank graag prof. dr. Ton Van der Wouden om de zinnen kritisch na te lezen, en iedereen die hielp met het verspreiden van het experiment van West-Vlaanderen tot in Friesland, in het bijzonder prof. dr. Stef Grondelaers, mijn mama en mijn broer Kasper. Ook ben ik erg dankbaar voor alle familieleden, vrienden en kennissen die de tijd namen om aan het experiment deel te nemen.

Deze masterproef was ten slotte niet tot stand gekomen zonder de steun en hulp van de mensen die me omringen. Bedankt aan mijn thesisbuddy Lore voor de schrijfdates in de bib, en aan Tuur, met wie ik door de jaren heen de fascinatie voor het Nederlands kon delen. Bedankt aan alle KHB-vrienden, die me telkens vergaven wanneer ik het gesprek onderbrak om een van hun zinnen in mijn notities te noteren. Bedankt aan mijn liefste vriendinnen van Litt Gall, die tijdens feestjes en vakanties geïnteresseerd naar mijn verhalen over hulpwerkwoordselectie luisterden. Mijn broer Wout ben ik dankbaar voor de hulp met LaTeX en statistiek, mijn broer Kasper voor zijn oprechte interesse voor alles wat ik doe. Mijn moeke en opa wil ik bedanken om me onvoorwaardelijk te steunen in wat ik doe, ook al weten ze niet helemaal precies wat dat is. Daarnaast ben ik mijn mama en papa enorm dankbaar, omdat ik dankzij hen kon studeren wat ik echt graag doe, en ze me de tijd en ruimte gaven om dat op mijn eigen manier te doen. De allergrootste dank verdient mijn liefste Isaak, om geduldig, zorgend, inspirerend, humoristisch, relativerend, begripvol, en alles wat ik nog meer nodig had te zijn.

Tot slot, mijn liefste oma: ik ben dan niet de directrice van de normaalschool geworden, maar ik weet zeker dat je trots op me zou zijn. De liefde voor de taal kreeg ik van jou, en ik zal die altijd blijven koesteren.

Samenvatting

Deze masterproef onderzoekt hulpwerkwoordvariatie in werkwoordsclusters met een aspectueel werkwoord, zoals *Ze is/heeft blijven wachten*. Het onderzoek heeft tot doel de factoren te identificeren die de keuze voor het hulpwerkwoord van voltooid aspect (*hebben* of *zijn*) in zulke clusters beïnvloeden.

Op basis van onderzoek naar hulpwerkwoordvariatie in werkwoordsclusters met een modaal werkwoord, zoals *Ze is/heeft moeten vertrekken*, worden taalinterne en -externe factoren geselecteerd die mogelijk een invloed hebben op hulpwerkwoordselectie in clusters met een aspectueel werkwoord (Hofmans 1981, Van Eynde et al. 2016, Dhont 2024). De onderzochte factoren zijn het type hoofdwerkwoord, de afstand tussen het hulpwerkwoord en het aspectuele werkwoord, irrealis, geografische regio en register. Het onderzoek combineert een experiment met een corpusonderzoek. In het experimentele onderzoek zijn acceptabiliteitsoordelen van Vlaamse en Nederlandse respondenten bevestigd en geanalyseerd aan de hand van lineaire regressieanalyses. Het corpusonderzoek in het SoNaR-corpus en het Corpus Gesproken Nederlands biedt inzicht in de frequentie van de onderzochte factoren in clusters waarin *hebben* het hulpwerkwoord is. Op basis van de resultaten en eerdere analyses (Cardinaletti & Shlonsky 2004, Draye & Van der Horst 2006, Pots 2020) wordt tot slot een formele analyse voorgesteld voor hulpwerkwoordvariatie in clusters met een aspectueel werkwoord.

De resultaten tonen aan dat hulpwerkwoordselectie in clusters met een aspectueel werkwoord beïnvloed wordt door irrealis, geografische regio en register. Een irrealislezing verhoogt de voorkeur voor het hulpwerkwoord *hebben*, dat meer aanvaard wordt in Vlaanderen dan in Nederland en meer voorkomt in informeel dan in formeel taalgebruik. De invloed van het type hoofdwerkwoord en afstand wordt in de resultaten niet eenduidig aangetoond. De hulpwerkwoordvariatie wordt in de analyse verklaard door de status van het aspectuele werkwoord, dat zich in een grammaticalisatieproces van semi-lexicaliteit naar functionaliteit bevindt en zich daarin verschillend positioneert in Vlaanderen en Nederland.

Inhoudsopgave

1	Inleiding	10
2	Literatuurstudie.....	12
2.1	Werkwoordsclusters	12
2.1.1	Drieledige werkwoordsclusters	12
2.1.2	Groepsvormende werkwoorden	13
2.2	Hulpwerkwoordselectie.....	13
2.2.1	Hulpwerkwoordselectie bij voltooid aspect	13
2.2.2	Hulpwerkwoordselectie bij drieledige werkwoordsclusters	15
2.3	Hulpwerkwoordvariatie in clusters met een modaal werkwoord (Mod-clusters).....	16
2.3.1	Verhouding tussen <i>hebben</i> en <i>zijn</i>	16
2.3.2	Geografische variatie.....	17
2.3.3	Taalinterne factoren	19
2.3.4	Conclusie: hulpwerkwoordvariatie in Mod-clusters	25
2.4	Hulpwerkwoordvariatie in clusters met een aspectueel werkwoord (Asp-clusters)	26
2.4.1	Verhouding tussen <i>zijn</i> en <i>hebben</i>	26
2.4.2	Geografische variatie.....	26
2.4.3	Taalinterne factoren	27
2.4.4	Conclusie: hulpwerkwoordvariatie in Asp-clusters	28
2.5	Analyse van hulpwerkwoordvariatie in drieledige werkwoordsclusters	28
2.5.1	Grammaticalisatie van het groepsvormende werkwoord	28
2.5.2	Coussé en Van de Velde (2014)	31
2.5.3	De status van het groepsvormende werkwoord.....	32
2.6	Conclusie	38
3	Onderzoeksvragen en hypotheses	39
3.1	Onderzoeksvragen	39
3.2	Hypotheses.....	40
3.2.1	Basishypothese.....	40
3.2.2	Hoofdwerkwoordhypothese.....	40
3.2.3	Afstandshypothese.....	41
3.2.4	Irrealishypothese.....	42
3.2.5	Geografische hypothese	43
3.2.6	Registerhypothese.....	43
3.3	Conclusie	43
4	Experimenteel onderzoek.....	44
4.1	Methodologie	44
4.1.1	Stimuli	44
4.1.2	Respondenten.....	56
4.1.3	Taak.....	58
4.1.4	Procedure	58
4.1.5	Analyse	60
4.1.6	Conclusie	61
4.2	Resultaten	62
4.2.1	Experiment 1	62

4.2.2	Experiment 2A	65
4.2.3	Experiment 2B	70
4.2.4	Experiment 3	75
4.2.5	Experiment 4	78
4.2.6	Conclusie: resultaten	84
4.3	Conclusie	85
5	Corpusonderzoek	86
5.1	Inleiding	86
5.2	Methodologie	86
5.2.1	Corpora	86
5.2.2	Dataverzameling	87
5.2.3	Annotatie	90
5.2.4	Analyse	91
5.3	Resultaten	93
5.3.1	Aspectuele werkwoorden	93
5.3.2	Hoofdwerkwoorden	93
5.3.3	Afstand	95
5.3.4	Modaliteit	96
5.3.5	Nationale variatie	97
5.3.6	Register	97
5.3.7	Conclusie	99
6	Discussie	100
6.1	Hulpwerkwoordvariatie in Asp-clusters	100
6.2	Taalinterne factoren	100
6.2.1	De invloed van het hoofdwerkwoord	100
6.2.2	De invloed van afstand	101
6.2.3	De invloed van irrealis	102
6.3	Taalexterne factoren	103
6.3.1	De invloed van geografische regio	103
6.3.2	De invloed van register	104
6.4	Analyse	105
6.4.1	Hulpwerkwoordvariatie bij Asp-clusters	105
6.4.2	Asp-clusters versus Mod-clusters	108
6.4.3	Irrealis	110
6.5	Beperkingen van het onderzoek en suggesties voor vervolgonderzoek	111
7	Conclusie	112
8	Referenties	113
	Appendix	118
	Appendix A. Overzicht stimuli	118
	Appendix B. Conditie experiment	121
	Appendix C. Demografische vragen	122
	Appendix D. Introductie experiment	123
	Appendix E. Corpusontwerp SoNaR-corpus	124
	Appendix F. Hoofdwerkwoorden corpusonderzoek	125

Figuren- en tabellenlijst

Figuur 2.1: Selectievoorkeur bij Mod-clusters in provincies uit Hofmans (1981) op zijn/hebben-as.....	18
Figuur 2.2: Selectievoorkeur van hoofdwkwoorden bij Mod-clusters op zijn/hebben-as (Hofmans 1981:86).....	20
Figuur 2.3: Voorkomen van hebben en zijn bij hoofdwkwoorden in Mod-clusters met moeten (Van Eynde et al. 2016:23).....	21
Figuur 2.4: Selectievoorkeur van modale wkwwoorden bij Mod-clusters op zijn/hebben-as (Hofmans 1981:88).....	23
Figuur 2.5: Selectievoorkeur in Mod-clusters bij vier taalinterne factoren op zijn/hebben-as.....	25
Figuur 3.1: Verwachte selectievoorkeur van hoofdwkwoorden in Asp-clusters	41
Figuur 4.2: Relatieve verdeling leeftijdsgroep per land	56
Figuur 4.1: Dichtheidsverdeling leeftijd van respondenten.....	56
Figuur 4.3: Absolute verdeling Regio_18j en Regio_nu.....	57
Figuur 4.4: Illustratie sliderschaal experiment.....	58
Figuur 4.5: Exp1 - Plot interactie Afstand*Hulpww.....	63
Figuur 4.6: Exp2A - Gemiddelde acceptabiliteitsscore als functie van Hulpww, HWW en Land	66
Figuur 4.7: Exp2A - Gemiddelde acceptabiliteitsscore als functie van Hulpww en Regio	67
Figuur 4.8: Exp2A - Plot interactie TypeHWW*Hulpww	68
Figuur 4.9: Exp2A - Plot driewegsinteractie TypeHWW*Hulpww*Land	69
Figuur 4.10: Exp2B - Gemiddelde acceptabiliteitsscore als functie van Hulpww, HWW en Land.....	71
Figuur 4.11: Exp2B - Gemiddelde acceptabiliteitsscore als functie van Hulpww en Regio	72
Figuur 4.12: Exp2B - Plot interactie TypeHWW*Hulpww.....	73
Figuur 4.13: Exp2B - Plot driewegsinteractie TypeHWW*Hulpww*Land.....	74
Figuur 4.14: Exp3 - Plot interactie Realis*Hulpww.....	76
Figuur 4.15: Exp3 - Plot driewegsinteractie Realis*Hulpww*Land	78
Figuur 4.16: Exp4 - Gemiddelde acceptabiliteitsscore als functie van Hulpww, Realis en Inversie	80
Figuur 4.17: Exp4 - Plot interactie Realis*Hulpww	81
Figuur 4.18: Exp4 - Plot driewegsinteractie Realis*Hulpww*Land	82
Figuur 4.19: Exp4 - Plot driewegsinteractie Realis*Hulpww*Leeftijdscategorie.....	83
Figuur 5.1: Absolute frequentie van afstand	96
Figuur 6.1: Positie aspectuele en modale wkwwoorden in grammaticalisatieproces per land (Vlaanderen vs. Nederland).....	109
Tabel 2.1: Kenmerken van functionele en lexicale groepsvormende wkwwoorden	33
Tabel 2.2: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende wkwwoorden in het Duits	36
Tabel 2.3: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende wkwwoorden in het Nederlands	36
Tabel 4.1: Overzicht van onderzoeksvragen en hypothesen per deelexperiment	45
Tabel 4.2: Types hoofdwkwoorden die in de experimenten gebruikt worden	46
Tabel 4.3: Type, aantal syllaben en frequentie van geselecteerde hoofdwkwoorden	47
Tabel 4.4: Afstandsmanipulaties gebruikt in Experiment 1	50
Tabel 4.5: Alle varianten van de eerste zin van Experiment 1 (E1_1)	50
Tabel 4.6: Hoofdwkwoorden die gebruikt worden in Experiment 2A.....	51
Tabel 4.7: Alle varianten van de eerste zin van Experiment 2A (E2A_1)	51
Tabel 4.8: Hoofdwkwoorden die gebruikt worden in Experiment 2B.....	52
Tabel 4.9: Alle varianten van de eerste zin van Experiment 2B (E2B_1).....	52
Tabel 4.10: Modaliteitsmanipulaties gebruikt in Experiment 3.....	52
Tabel 4.11: Varianten van de eerste zin van Experiment 3 (E3_1)	53
Tabel 4.12: Inversiemaniplulaties gebruikt in Experiment 4.....	54

Tabel 4.13: Alle varianten van de eerste zin van Experiment 4 (E4_1)	54
Tabel 4.14: Overzicht stimuli per deelexperiment	55
Tabel 4.15: Overzicht van verdeling provincies in regio's	57
Tabel 4.16: Schematische opbouw experiment voor conditie 1	59
Tabel 4.17: Overzicht opgenomen variabelen in verkennende analyse en regressieanalyse per deelexperiment	61
Tabel 4.18: Exp1 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Afstand en Land	62
Tabel 4.19: Exp1 – Regressieoutput basismodel (Afstand*Hulpww)	63
Tabel 4.20: Exp1 – Regressieoutput driewegsanalyse (Afstand*Hulpww*Land)	64
Tabel 4.21: Exp2A - Gemiddelde scores (met standaarddeviatie) als functie van Hulpww, TypeHWW en Land	65
Tabel 4.22: Exp2A - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww en HWW	66
Tabel 4.23: Exp2A – Regressieoutput basismodel (TypeHWW*Hulpww)	67
Tabel 4.24: Exp2A – Regressieoutput driewegsinteractie (TypeHWW*Hulpww*Land)	68
Tabel 4.25: Exp2B - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, TypeHWW en Land	70
Tabel 4.26: Exp2B - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww en HWW	71
Tabel 4.27: Exp2B – Regressieoutput basismodel (TypeHWW*Hulpww)	73
Tabel 4.28: Exp2B – Regressieoutput driewegsinteractie (TypeHWW*Hulpww*Land)	74
Tabel 4.29: Exp3 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Land	75
Tabel 4.30: Exp3 – Regressieoutput basismodel (Realis*Hulpww)	76
Tabel 4.31: Exp3 – Regressieoutput driewegsinteractie (Realis*Hulpww*Land)	77
Tabel 4.32: Exp4 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Land	79
Tabel 4.33: Exp4 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Inversie	79
Tabel 4.34: Exp4 – Regressieoutput basismodel (Realis*Hulpww)	80
Tabel 4.35: Exp4 – Regressieoutput driewegsinteractie (Realis*Hulpww*Land)	81
Tabel 4.36: Exp4 - Regressieoutput driewegsinteractie (Realis*Hulpww*Leeftijdscategorie)	82
Tabel 5.1: Indeling CGN-componenten op basis van aantal deelnemers en mate van spontaniteit (Van Eerten 2007:2)	92
Tabel 5.2: Absolute en relatieve frequentie van aspectuele werkwoorden als functie van land	93
Tabel 5.3: Absolute en relatieve frequentie van tien meest frequente hoofdwerkwoorden als functie van land	94
Tabel 5.4: Absolute en relatieve frequentie van slapen, lezen, volgen en winnen als functie van land ..	94
Tabel 5.5: Absolute en relatieve frequentie van afstand als functie van land	95
Tabel 5.6: Absolute en relatieve frequentie van tijd en modaliteit als functie van land	96
Tabel 5.7: Absolute en relatieve frequentie van Asp-clusters met hebben als functie van land	97
Tabel 5.8: Absolute en relatieve frequentie van voorkomende teksttypes van het SoNaR-corpus als functie van land	97
Tabel 5.9: Absolute en relatieve frequentie van voorkomende taalgebruikssituaties van het CGN als functie van land	98
Tabel 6.1: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende werkwoorden	105

Lijst met afkortingen

AIC	<i>Akaike Information Criterion</i>
ASP	Aspectueel werkwoord
AUX	Hulpwerkwoord van voltooid aspect
CGN	Corpus Gesproken Nederlands
H	Hebben
HWW	Hoofdwkwoord
irr	Irrealis
<i>M</i>	Gemiddelde
MOD	Modaal werkwoord
N	Nederland
OTT	Onvoltooid tegenwoordige tijd
OVT	Onvoltooid verleden tijd
rea	Realis
<i>SD</i>	Standaarddeviatie
V	Vlaanderen
Z	Zijn

1 Inleiding

In het VRT-programma *Het Huis* spreekt de Vlaamse televisiemaker Arnout Hauben in een interview de zin in (1a) uit. In de ondertiteling van het programma verschijnt de zin in (1b).

- (1) a. Ik **heb** blijven oefenen tot ik op een podiumpje kon staan.
- b. Ik **ben** blijven oefenen tot ik op een podiumpje kon staan. (*Het Huis*, 25 november 2024)

Het verschil tussen de zinnen in (1) is het hulpwerkwoord van voltooid aspect: in (1a) is dat *hebben*, in (1b) *zijn*. Die keuze van het hulpwerkwoord bij de vorming van de voltooid tijd in het Nederlands is meestal duidelijk. Een taalgebruiker weet dat het werkwoord *blijven* met *zijn* vervoegd wordt, zoals in (2a), en dat *oefenen* met *hebben* wordt gecombineerd, zoals in (2b).

- (2) a. Ik **ben**/*heb gisteren lang op het feestje **gebleven**.
- b. Ik **heb**/*ben van tevoren hard **geoefend**.

Wanneer de werkwoorden *blijven* en *oefenen* samen in een werkwoordscluster terechtkomen, zoals in (1), is de keuze van het hulpwerkwoord echter niet meer zo eenduidig, en ontstaat er hulpwerkwoordvariatie. Die variatie blijft beperkt in clusters waarbij het groepsvormende werkwoord een aspectueel werkwoord is, zoals *blijven* in (1). Het hulpwerkwoord van voltooid aspect *zijn* is standaardtaal in zulke clusters en wordt dus ook in de ondertiteling gebruikt. *Hebben* komt minder voor, maar wanneer het wordt gebruikt, is dat meestal door een Vlaming, zoals Arnout Hauben in (1). Hulpwerkwoordvariatie komt meer voor in werkwoordsclusters met een modaal werkwoord als *moeten*, zoals in (3). Dan wordt *hebben* (3a) vooral gebruikt in Nederland, en *zijn* (3b) vooral in Vlaanderen.

- (3) a. Ik **heb** van mijn vriend op het feestje **moeten blijven**.
- b. Ik **ben** van mijn vriend op het feestje **moeten blijven**.

Hoewel de variatie in clusters als (1) en (3) een andere omvang en geografische variatie kent, lijkt er taalintern hetzelfde aan de hand: de hulpwerkwoordvariatie ontstaat wanneer er twijfel is over welk werkwoord van het cluster de keuze van het hulpwerkwoord mag bepalen. In (1a) en (3a) wordt het hulpwerkwoord geselecteerd door het groepsvormende werkwoord (*blijven* en *moeten*), terwijl in (1b) en (3b) de keuze bepaald lijkt te worden door het hoofdwkwoord (*oefenen* en *blijven*).

Naar hulpwerkwoordvariatie in clusters met een modaal werkwoord zoals (3) werd al uitgebreid onderzoek gevoerd (o.a. Hofmans 1981, Van Eynde et al. 2016, Dhont 2024). Daarbij werden verschillende taalinterne en -externe factoren geïdentificeerd die de hulpwerkwoordselectie in zulke clusters beïnvloeden. De hulpwerkwoordvariatie in clusters als (1) is in de onderzoeksliteratuur echter minder uitgebreid in kaart gebracht. Het doel van deze masterproef is om inzicht te verwerven in de factoren die de hulpwerkwoordselectie in clusters met een aspectueel werkwoord bepalen. Gebaseerd op het onderzoek naar clusters met modale werkwoorden wordt de invloed van vijf factoren onderzocht: het type hoofdwkwoord, afstand tussen het hulp- en aspectuele werkwoord, irrealis, geografische regio en register. Dat gebeurt aan de hand van een experimenteel onderzoek en een corpusonderzoek. In het experiment worden de acceptabiliteitsoordelen van Vlaamse en Nederlandse respondenten over clusters als (1) bevraagd. Het corpusonderzoek in het SoNaR-corpus en het Corpus Gesproken Nederlands biedt inzicht in het voorkomen van *hebben* in clusters met een aspectueel werkwoord in niet-

geëliciteerd taalgebruik. Op basis van de resultaten van het onderzoek stel ik tot slot een formele analyse van hulpwerkwoordselectie in werkwoordsclusters met aspectuele werkwoorden voor.

Deze masterproef is als volgt gestructureerd. In §2 wordt de literatuur over hulpwerkwoordselectie in clusters met aspectuele en modale werkwoorden gepresenteerd. Vervolgens baken ik de onderzoeksvragen af en formuleer ik de hypothesen voor het onderzoek (§3). Daarna wordt eerst het experimentele onderzoek (§4) toegelicht, en vervolgens het corpusonderzoek (§5). In de discussie (§6) breng ik de resultaten van het experiment en het corpusonderzoek samen, stel ik een analyse voor en bespreek ik de beperkingen van het onderzoek evenals perspectieven voor vervolgonderzoek. In de conclusie (§7) vat ik de belangrijkste bevindingen van het onderzoek samen.

2 Literatuurstudie

Dit hoofdstuk bevat een overzicht van de relevante literatuur met betrekking tot hulpwerkwoordselectie in clusters met een aspectueel werkwoord. Allereerst licht ik toe op welk type werkwoordsclusters dit onderzoek zich richt (§2.1). In §2.2 volgt een bespreking van hulpwerkwoordselectie bij voltooid aspect in het Nederlands. Daarna wordt de hulpwerkwoordvariatie in clusters met modale werkwoorden behandeld (§2.3), en in §2.4 komt de hulpwerkwoordvariatie in clusters met een aspectueel werkwoord aan bod. In §2.5 bespreek ik voorgestelde analyses voor hulpwerkwoordvariatie in clusters met een modaal of aspectueel werkwoord. Ik sluit af met een conclusie in §2.6.

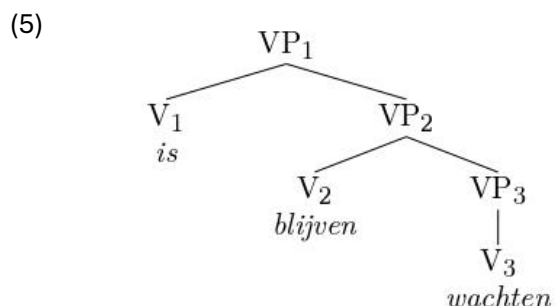
2.1 Werkwoordsclusters

In deze masterproef wordt hulpwerkwoordselectie in zinnen als (4) onderzocht. De zinnen bevatten allebei een drieledige werkwoordscluster, maar het soort groepsvormend werkwoord verschilt. In (4a) is dat het aspectuele werkwoord *blijven*, in (4b) het modale werkwoord *moeten*. In deze sectie definieer ik drieledige werkwoordsclusters (§2.1.1) en bespreek ik de groepsvormende werkwoorden die in het onderzoek aan bod komen (§2.1.2).

- (4) a. Ik denk dat Emma op hem **is blijven wachten**.
b. Ik denk dat Bram vroeg **heeft moeten vertrekken**.

2.1.1 Drieledige werkwoordsclusters

De werkwoordsclusters in deze masterproef hebben een structuur zoals in (5) (gebaseerd op Barbiers 2005:242).¹ Het hiërarchisch hoogste werkwoord (V_1) is steeds een finiet hulpwerkwoord van voltooid aspect (*hebben/zijn*). Dat hulpwerkwoord selecteert een groepsvormend werkwoord (V_2), dat op zijn beurt een hoofdwerkwoord (V_3) selecteert (Augustinus 2015:7, Broekhuis & Corver 2016:957-1066).



Ik spreek in deze masterproef over *drieledige werkwoordsclusters* wanneer drie werkwoorden in een hiërarchische relatie zoals in (5) staan. Of het finiete werkwoord deel uitmaakt van de werkwoordelijke eindgroep en dus in de tweede pool staat, zoals in (6a), of naar de eerste pool verplaatst, zoals in (6b), speelt voor mij geen rol.

(6)	voorveld	POOL 1	middenveld	POOL 2	achterveld
a.		dat	Emma op hem	is blijven wachten	
b.	Emma	is	op hem	blijven wachten	

¹ In het onderzoek naar drieledige werkwoordsclusters (o.a. Barbiers 2005) werd veel aandacht besteed aan woordvolgordevariatie. Ik neem hier de basisstructuur aan zonder aandacht te besteden aan de mogelijke invloed van de volgordevariatie op de structuur.

2.1.2 Groepsvormende werkwoorden

Er bestaan verschillende werkwoorden in het Nederlands die als groepsvormend werkwoord in een drieledige werkwoordscluster kunnen fungeren (Haeseryn et al. 1997:1663). In deze masterproef richt ik me tot twee types: de aspectuele werkwoorden *gaan*, *komen*, *blijven* en *beginnen* en de modale werkwoorden, waaronder *moeten*, *kunnen*, *mogen* en *willen* (Broekhuis & Corver 2016:150).

In groepsvormende hoedanigheid selecteren die werkwoorden een infinitief (Broekhuis & Corver 2016:150). Het aspectuele werkwoord *blijven* selecteert in (7a) de infinitief *wachten*; het modale *moeten* selecteert in (7c) *vertrekken*. Wanneer die werkwoorden in groepsvormende hoedanigheid voltooid worden gemaakt, verschijnen ze als gevolg van het IPP-effect als infinitief, zoals *blijven* in (7b) en *moeten* in (7d) (Van Craenenbroeck 2019:43).

- (7)
- a. Emma **blijft** op hem **wachten**.
 - b. Emma **is** op hem **blijven wachten**.
 - c. Bram **moet** vroeg **vertrekken**.
 - d. Bram **heeft** vroeg **moeten vertrekken**.

De drieledige werkwoordsclusters die aan bod komen in deze masterproef zijn dus het resultaat van het in voltooide tijd plaatsen van de aspectuele en modale werkwoorden in groepsvormende hoedanigheid. Naar de twee types werkwoordsclusters wordt verder in deze masterproef verwezen met ‘Asp-clusters’ en ‘Mod-clusters’, zoals gedefinieerd in (8).

- (8)
- Asp-cluster:** drieledige werkwoordscluster met een aspectueel werkwoord
 - Mod-cluster:** drieledige werkwoordscluster met een modaal werkwoord

Ik vat de geïntroduceerde begrippen samen aan de hand van de twee voorbeelden in (4), hieronder herhaald als (9). De zin in (9a) bevat een Asp-cluster. Die drieledige werkwoordscluster bestaat uit het hulpwerkwoord van voltooid aspect *zijn* (AUX), het aspectuele werkwoord *blijven* (ASP) en het hoofdwerkwoord *wachten* (HWW). De zin in (9b) bevat een Mod-cluster. Die drieledige werkwoordscluster bestaat uit het hulpwerkwoord van voltooid aspect *hebben* (AUX), het modale werkwoord *moeten* (MOD) *moeten* en het hoofdwerkwoord *vertrekken* (HWW). Ik geef de werkwoorden van het cluster weer naast de voorbeeldzin.

- (9)
- | | |
|---|-----------------|
| a. Emma is op hem blijven wachten . | AUX – ASP – HWW |
| b. Bram heeft vroeg moeten vertrekken . | AUX – MOD – HWW |

2.2 Hulpwerkwoordselectie

In deze sectie bespreek ik hulpwerkwoordselectie in het Nederlands. In de eerste plaats licht ik hulpwerkwoordselectie bij voltooid aspect over het algemeen toe (§2.2.1), vervolgens de hulpwerkwoordselectie bij drieledige werkwoordsclusters (§2.2.2).

2.2.1 Hulpwerkwoordselectie bij voltooid aspect

Het Nederlands beschikt over twee hulpwerkwoorden van voltooid aspect: *hebben* en *zijn*. De keuze tussen die twee hulpwerkwoorden bij de vorming van de voltooide tijden is niet vrij, maar wordt bepaald door het hoofdwerkwoord (Ackema & Sorace 2017:2). Zo selecteert het hoofdwerkwoord *wachten* het hulpwerkwoord van voltooid aspect *hebben* (10a), terwijl het hoofdwerkwoord *vertrekken* het

hulpwerkwoord *zijn* selecteert (10b). *Hebben* door *zijn* vervangen in (10a) en *zijn* door *hebben* vervangen in (10b) leidt tot ongrammaticaliteit.

- (10) a. Emma **heeft**/*is op hem **gewacht**.
b. Bram **is**/*heeft vroeg **vertrokken**.

De generalisatie voor hulpwerkwoordselectie is dat alle transitieve en onergatieve werkwoorden steeds *hebben* selecteren, terwijl de ergatieve werkwoorden soms *hebben* en soms *zijn* selecteren (Broekhuis & Corver 2016:211). In de zinnen in (11) wordt *hebben* geselecteerd omdat *lezen* (11a) een transitief werkwoord en *lachen* (11b) een onergatief werkwoord is. De zinnen in (12) bevatten de ergatieve werkwoorden *bloeden* en *arriveren*. Aangezien de atelische ergatieve werkwoorden *hebben* selecteren en de telische ergatieve werkwoorden *zijn*, wordt *bloeden* (12a) over het algemeen vervoegd met *hebben* en *arriveren* (12b) met *zijn* (Broekhuis & Corver 2016:951-952).

- (11) a. Jan **heeft**/*is het boek **gelezen**.
b. Jan **heeft**/*is **gelachen**. (Broekhuis & Corver 2016:951-952)
- (12) a. De jongen **heeft**/*is hevig **gebloed**.
b. De jongen **is**/*heeft op tijd **gearriveerd**. (Broekhuis & Corver 2016:205-207)

Dat betekent niet dat het zomaar mogelijk is om een lijst samen te stellen van werkwoorden die *hebben* selecteren en van werkwoorden die *zijn* selecteren. Bepaalde categorieën werkwoorden worden namelijk met *hebben* of *zijn* gecombineerd afhankelijk van de syntactische context. Het atelische ergatieve werkwoord *bloeden* in (12a) kan bijvoorbeeld telisch worden gemaakt door het toevoegen van *dood*, zoals in (13). Dan wordt het hulpwerkwoord *zijn* geselecteerd in plaats van *hebben* (Broekhuis & Corver 2016:205-206). Bewegingswerkwoorden als *fietsen* selecteren daarnaast normaliter het hulpwerkwoord *hebben*, zoals in (14a), maar worden met *zijn* vervoegd wanneer er een richtingsobject aanwezig is, zoals in (14b) (Ackema & Sorace 2017:6). Tot slot zijn er werkwoorden die zowel transitief als intransitief gebruikt kunnen worden, zoals *genezen* in (15). Bij transitief gebruik, zoals in (15a), selecteren ze *hebben*, bij intransitief gebruik selecteren ze *zijn* (15b) (Haeseryn et al. 1997:75-78).

- (13) De jongen **is**/*heeft *dood* **gebloed**. (Broekhuis & Corver 2016:205-206)
- (14) a. Marie **heeft**/*is **gefietst**.
b. Marie **is**/*heeft *naar Den Haag* **gefietst**. (Ackema & Sorace 2017:6)
- (15) a. De dokter **heeft**/*is *hem* **genezen**.
b. Hij **is**/*heeft **genezen**. (Haeseryn et al. 1997:76)

Daarnaast zijn er in het Nederlands enkele werkwoorden die in bepaalde betekenissen zowel *hebben* als *zijn* kunnen selecteren. De werkwoorden die in dat verband het vaakst vermeld worden zijn *vergeten* in de betekenis ‘er niet aan denken te’, zoals geïllustreerd in (16a), en *verliezen* in de betekenis ‘iets kwijtraken’ (16b) (Haeseryn et al. 1997:87-80, Ackema & Sorace 2017:10). Andere werkwoorden die met beide hulpwerkwoorden gecombineerd worden zijn bijvoorbeeld *opvolgen* in de betekenis ‘komen na iemand’ (16c), *volgen* in de betekenis ‘achter iemand aan gaan’ (16d) en *optreden* in de betekenis ‘actie ondernemen’ (16e) (Haeseryn et al. 1997:78-80, Taaladvies 2024, Onze Taal 2024).

- | | | |
|------|--|---------------------------|
| (16) | a. Ik heb/ben mijn paraplu vergeten . | (Haeseryn et al. 1997:79) |
| | b. Ik heb/ben mijn horloge verloren . | (Haeseryn et al. 1997:78) |
| | c. Wie heeft/is Willem III opgevolgd ? | (Haeseryn et al. 1997:80) |
| | d. Die hond heeft/is ons tot aan het hek gevolgd . | (Taaladvies 2024) |
| | e. De politie heeft/is direct opgetreden . | (Onze Taal 2024) |

Samengevat selecteren sommige werkwoorden steevast *hebben*, andere steevast *zijn*, en andere laten, soms afhankelijk van de context, beide hulpwerkwoorden toe. Ik geef in de rest van de voorbeeldzinnen met een codering aan wat de selectievoordeur van het hoofdwerkwoord (HWW) is en welk hulpwerkwoord van voltooid aspect (AUX) in de zin geselecteerd is. Dat doe ik aan de hand van een ‘H’ voor *hebben* en/of een ‘Z’ voor *zijn*, zoals in (17). Een werkwoord als *lezen* (17a) noem ik in het vervolg van de masterproef een ‘H-werkwoord’, omdat het *hebben* selecteert. Het werkwoord *vertrekken* (17b) noem ik naar analogie een ‘Z-werkwoord’, net als *genezen* in (17c), dat in de betreffende syntactische context enkel *zijn* kan selecteren (zie (15)). Een werkwoord als *verliezen* in (17d) noem ik tot slot een ‘HZ-werkwoord’, omdat het zowel *hebben* als *zijn* kan selecteren. De ‘H’ bij het hulpwerkwoord (AUX) geeft in dat geval aan dat in de voorbeeldzin *hebben*, en niet *zijn* werd geselecteerd.

- | | | |
|------|---|--------------------------------------|
| (17) | a. Ik heb het boek gelezen . | AUX _H – HWW _H |
| | b. Ik ben vertrokken . | AUX _Z – HWW _Z |
| | c. Ik ben genezen . | AUX _Z – HWW _Z |
| | d. Ik heb mijn handtas verloren . | AUX _H – HWW _{HZ} |

2.2.2 Hulpwerkwoordselectie bij drieledige werkwoordsclusters

In drieledige werkwoordsclusters wordt het hulpwerkwoord van voltooid aspect over het algemeen geselecteerd door het groepsvormende werkwoord (Haeseryn et al. 1997:80). Het werkwoord *blijven* is een Z-werkwoord: het selecteert *zijn* bij voltooid aspect, zoals geïllustreerd in (18a). Wanneer *blijven* als groepsvormend werkwoord fungeert, zoals in (18b), is *zijn* het hulpwerkwoord van het Asp-cluster. *Moeten* daarentegen is een H-werkwoord: het selecteert *hebben*, zoals geïllustreerd in (19a). In het Mod-cluster in (19b) is *hebben* het hulpwerkwoord van voltooid aspect (Haeseryn et al. 1997:80).

- | | | |
|------|---|---------------------------|
| (18) | a. Zijn /*hebben jullie de hele middag in de kantine gebleven ? | |
| | b. Zijn /*hebben jullie de hele middag in de kantine blijven zitten ? (Haeseryn et al. 1997:80) | |
| (19) | a. Rekenen? Dat heb /*ben ik nooit goed gekund . | |
| | b. Ik heb /*ben nooit goed kunnen rekenen . | (Haeseryn et al. 1997:80) |

Het verschil tussen voorbeelden (18) en (19) van Haeseryn et al. (1997:80) is dat in (18b) de selectievoordeur van het groepsvormende werkwoord en het hoofdwerkwoord niet gelijk zijn, maar in (19b) wel. In (18b) selecteert *blijven* het hulpwerkwoord *zijn*, maar het hoofdwerkwoord *zitten* selecteert *hebben* (*ik heb gezeten*). Het hulpwerkwoord van het Asp-cluster is dus bepaald door het groepsvormende werkwoord. In (19b) daarentegen selecteren zowel *kunnen* als *rekenen* het hulpwerkwoord *hebben* (*ik heb gerekend*). In principe is in (19b) dus niet vast te stellen door welk werkwoord de keuze voor het hulpwerkwoord *hebben* is bepaald (Haeseryn et al. 1997:80).

Wanneer in Mod-clusters als (19b) het groepsvormende werkwoord en het hoofdwkwoord elk een ander hulpwerkwoord selecteren, kan er volgens Haeseryn et al. (1997:80-81) hulpwerkwoordvariatie optreden. In (20) is het groepsvormende werkwoord *kunnen* een H-werkwoord, maar het hoofdwkwoord *komen* een Z-werkwoord. In (20a) bepaalt *kunnen* de keuze van het hulpwerkwoord, terwijl in (20b) het hulpwerkwoord *zijn* door het hoofdwkwoord *komen* wordt geselecteerd.

- (20) a. Door de gladheid **hebben** ze niet **kunnen** komen.
 b. Door de gladheid **zijn** ze niet kunnen **komen**. (Haeseryn et al. 1997:81)

Haeseryn et al. (1997:81) stellen dat hulpwerkwoordvariatie zoals in (20) zich enkel kan voordoen bij de groepsvormende werkwoorden *durven*, *kunnen*, *moeten*, *hoeven*, *mogen*, *proberen*, *trachten*, *weten* en *willen*. Enkel bij die werkwoorden, waaronder de modale werkwoorden, kan de keuze van het hulpwerkwoord bepaald worden door het hoofdwkwoord in plaats van door het groepsvormende werkwoord. Die opvatting wordt algemeen ondersteund in de literatuur (o.a. Paardekooper 1978:101, Broekhuis & Corver 2016:952-953) en de hulpwerkwoordvariatie in Mod-clusters is reeds uitgebreid onderzocht (o.a. Hofmans 1981, Draye & Van der Horst 2006, Van Eynde et al. 2016, Dhont 2024).

Onderzoek toont echter aan dat hulpwerkwoordvariatie niet beperkt blijft tot de werkwoorden die Haeseryn et al. (1997:80-81) vermelden. Ook in Asp-clusters zoals in (18), waarbij volgens Haeseryn et al. (1997:80) enkel het hulpwerkwoord *zijn* mogelijk is, blijken beide hulpwerkwoordvarianten naast elkaar te kunnen bestaan (Hofmans 1982b, Draye & Van der Horst 2006, Van Eynde et al. 2016, Dhont 2024). In het Asp-cluster in (21) kan de keuze van hulpwerkwoord dan ook bepaald worden door ofwel het groepsvormende werkwoord, zoals in (21a), ofwel door het hoofdwkwoord, zoals in (21b).

- (21) a. Emma **is** op hem **blijven** wachten. $AUX_Z - ASP_Z - HWW_H$
 b. Emma **heeft** op hem blijven **wachten**. $AUX_H - ASP_Z - HWW_H$

Zoals blijkt uit het voorgaande, is hulpwerkwoordvariatie in clusters met een modaal werkwoord in de onderzoeksliteratuur uitgebreider in kaart gebracht dan hulpwerkwoordvariatie in clusters met een aspectueel werkwoord. In wat volgt bespreek ik het beschikbare onderzoek voor beide types. Ik begin met de Mod-clusters (§2.3), daarna volgen de Asp-clusters (§2.4).

2.3 Hulpwerkwoordvariatie in clusters met een modaal werkwoord (Mod-clusters)

In deze sectie beschrijf ik de hulpwerkwoordvariatie in clusters met een modaal werkwoord (Mod-clusters). Eerst wordt in §2.3.1 de verhouding tussen de hulpwerkwoorden *hebben* en *zijn* in Mod-clusters besproken. Vervolgens licht ik toe hoe die algemene verhouding geografische variatie vertoont (§2.3.2). In §2.3.3 komen de taalinterne factoren aan bod die hulpwerkwoordselectie in Mod-clusters beïnvloeden. Ik vat tot slot de belangrijkste bevindingen samen in §2.3.4.

2.3.1 Verhouding tussen *hebben* en *zijn*

De hulpwerkwoordvariatie in Mod-clusters kan in kaart gebracht worden door de verhouding tussen de hulpwerkwoorden *hebben* en *zijn*. Die verhouding kan enkel berekend worden in Mod-clusters waarin variatie mogelijk is en waarin de selectievoorkeur van het modale werkwoord en het hoofdwkwoord dus verschillen (zie §2.2.2). De modale werkwoorden (o.a. *moeten*, *kunnen*, *mogen*, *willen*) selecteren steeds *hebben*, dus enkel wanneer het hoofdwkwoord *zijn* selecteert, is hulpwerkwoordvariatie mogelijk. In het Mod-cluster in (22a) kan dus variatie optreden, want het hoofdwkwoord is het Z-

werkwoord *vertrekken*. In (22b) daarentegen is het hoofdwkwoord *lezen* net als *moeten* een H-werkwoord, dus is er in principe geen hulpwerkwoordvariatie mogelijk. Dan is het hulpwerkwoord van voltooid aspect steeds *hebben* (zie §2.2.2).

- | | | |
|------|--|--|
| (22) | a. Ze hebben/zijn eerder moeten vertrekken . | AUX _{H/Z} – MOD _H – HWW _Z |
| | b. Ze hebben/*zijn het boek moeten lezen . | AUX _H – MOD _H – HWW _H |

Van Eynde et al. (2016:21) onderzoeken de verhouding tussen *hebben* en *zijn* in Mod-clusters met de modale werkwoorden *moeten* en *kunnen* aan de hand van een corpusonderzoek in de componenten Wikipedia, boeken en autocues van de SoNaR-treebank (Oostdijk et al. 2013).² Ze vinden 2 158 Mod-clusters met een Z-hoofdwkwoord zoals die in (22a). In die clusters is het hulpwerkwoord van voltooid aspect in 78.2% van de gevallen *hebben*, en slechts in 21.8% van de gevallen *zijn*. Op die manier tonen ze aan dat *zijn* de ‘alternatieve’ hulpwerkwoordvariant is, die minder voorkomt dan de ‘normale’ variant *hebben* (Van Eynde et al. 2016:22). Die algemene verhouding wordt echter sterk genuanceerd door een belangrijke geografische variabele, die wordt toegelicht in de volgende sectie.

2.3.2 Geografische variatie

De hulpwerkwoordselectie in Mod-clusters kent een grote geografische variatie. Er is sprake van nationale variatie tussen Vlaanderen en Nederland (§2.3.2.1), maar ook van regionale variatie binnen beide gebieden (§2.3.2.2).

2.3.2.1 Nationale variatie

De algemene tendens is dat *hebben* in Nederland meer voorkomt dan *zijn*, terwijl *zijn* in Vlaanderen vaker voorkomt dan *hebben*, zoals geïllustreerd in (23) (Haeseryn et al. 1997:81, Draye & Van der Horst 2006).

- | | | | |
|------|--|------------|--|
| (23) | a. Kasper heeft niet naar school kunnen gaan . | NEDERLAND | AUX _H – MOD _H – HWW _Z |
| | b. Kasper is niet naar school kunnen gaan . | VLAANDEREN | AUX _Z – MOD _H – HWW _Z |

Die nationale variatie in Mod-clusters werd voor het eerst onderzocht in het enquêteonderzoek van Hofmans (1981). Daarbij vulden 750 Vlaamse en Nederlandse respondenten in 40 zinnen zoals (24) een vorm van *hebben* of *zijn* in op basis van hun eigen taalgebruik. Als beide hulpwerkwoorden mogelijk waren gaven ze dat aan, maar duiden ze een voorkeursvariant aan (Hofmans 1981:106).

- | | | |
|------|--|-----------------------------|
| (24) | Waarom _____ ik moeten komen en hij niet? | (Hofmans 1981:112, zin 15). |
|------|--|-----------------------------|

Uit de resultaten blijkt dat 80% van de 250 Nederlandse respondenten de voorkeur geeft aan *hebben*, terwijl 67% van de 500 Vlaamse respondenten *zijn* verkiest (Hofmans 1981:101). Die nationale tendens wordt bevestigd in het corpusonderzoek van Dhont (2024), die nationale variatie in hulpwerkwoordselectie bij Mod-clusters onderzoekt in de krantencomponent van het SoNaR-500-corpus (Oostdijk et al. 2013).³ De resultaten tonen dat *hebben* in Nederland in 95.3% van de gevallen

² De zinnen in de SoNaR-treebank, het met syntactische structuren verrijkte SoNaR-corpus, zijn ontleend met behulp van de Alpino-parser, die enkel voor de werkwoorden *kunnen* en *moeten* de mogelijkheid tot hulpwerkwoordvariatie voorziet. Voor de andere modale werkwoorden die alternantie toestaan (bv. *mogen*) laat de parser geen variatie toe, waardoor aan die constructies een andere syntactische structuur is toegekend (Van Eynde et al. 2016:18). Daarom kunnen Van Eynde et al. (2016:22) enkel voor Mod-clusters met *moeten* en *kunnen* relevante gegevens extraheren over de *hebben/zijn*-verhouding.

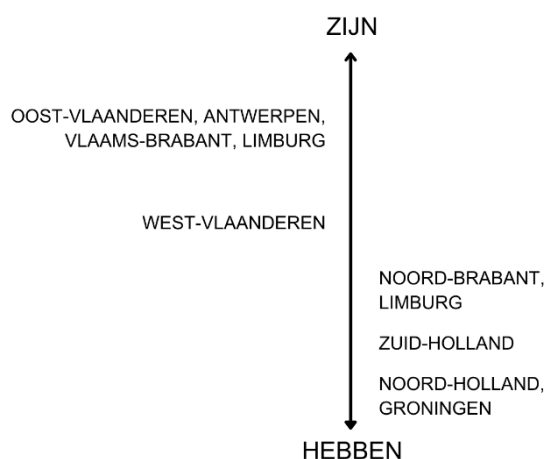
³ Dhont (2024) rekent tot wat hij ‘Type A-clusters’ noemt zowel clusters met modale werkwoorden als clusters met de groepsvormende werkwoorden *trachten*, *proberen*, *lopen*, *weten* en *durven*. Aangezien die laatste werkwoorden bijna niet

voorkomt, terwijl in Vlaanderen *zijn* in 78.5% en *hebben* slechts in 21.5% van de gevallen gebruikt wordt (Dhont 2024:190). Het *hebben* van Nederland staat dus tegenover het *zijn* van Vlaanderen, maar de voorkeur voor *zijn* in Vlaanderen is minder sterk dan de voorkeur voor *hebben* in Nederland (Draye & Van der Horst 2006:230, Haeseryn et al. 1997:81).

2.3.2.2 Regionale variatie

Hofmans (1981) toont aan dat de geografische variatie verder gaat dan de dichotomie Nederland-Vlaanderen: er is ook sprake van regionale variatie binnen Nederland en Vlaanderen. De Vlaamse en Nederlandse deelnemers aan het enquêteonderzoek zijn verdeeld over de provincies West-Vlaanderen, Oost-Vlaanderen, Antwerpen, Vlaams-Brabant en Limburg enerzijds, en Groningen, Noord-Holland, Zuid-Holland, Noord-Brabant en Nederlands-Limburg anderzijds (Hofmans 1981:68).⁴ De resultaten tonen aan dat er in Nederland een duidelijk verschil in voorkeur bestaat tussen de verschillende provincies. In de zuidelijke provincies (Noord-Brabant en Nederlands-Limburg) wordt *zijn* meer verkozen (respectievelijk in 26% en 33% van de gevallen) dan in de noordelijke provincies (Noord-Holland en Groningen, respectievelijk in 4% en 6%). In Zuid-Holland wordt *zijn* in 11% van de zinnen verkozen, waardoor het een positie tussen het noorden en zuiden inneemt (Hofmans 1981:98). Zo schetst Hofmans (1981:102) een geleidelijke overgang van noord naar zuid wat de voorkeur voor *zijn* betreft: *zijn* komt in het noorden van Nederland nauwelijks voor, maar wordt richting het zuiden steeds vaker gebruikt.

In Vlaanderen blijken er minder grote verschillen tussen de provincies te bestaan. Enkel West-Vlaanderen onderscheidt zich van de andere Vlaamse provincies. In West-Vlaanderen worden *hebben* en *zijn* immers evenveel geselecteerd, terwijl in de overige vier provincies in 66% van de gevallen *zijn* verkozen wordt. Het hulpwerkwoord *zijn* komt dus meer voor in Limburg, Vlaams-Brabant, Antwerpen en Oost-Vlaanderen dan in West-Vlaanderen (Hofmans 1981:98-99). De regionale variatie die blijkt uit het onderzoek van Hofmans (1981) wordt gevisualiseerd in Figuur 2.1. De provincies zijn in de figuur op een as tussen *zijn* en *hebben* geplaatst afhankelijk van de vastgestelde selectievoorkeur in Mod-clusters.⁵



Figuur 2.1: Selectievoorkeur bij Mod-clusters in provincies uit Hofmans (1981) op *zijn/hebben*-as

voorkomen in het onderzoek en de resultaten dus niet beïnvloeden, maak ik daar geen onderscheid in bespreek ik de globale resultaten voor de Type A-clusters (Dhont 2024:191).

⁴ Niet alle Nederlandse provincies zijn vertegenwoordigd in het enquêteonderzoek (Hofmans 1981:86).

⁵ De plaatsing van de verschillende provincies op de as is indicatief en niet gebaseerd op exacte metingen. Dat geldt ook voor de visualisaties op de *zijn/hebben*-as in de rest van dit hoofdstuk.

Samengevat bestaat er in Nederland een voorkeur voor *hebben* als hulpwerkwoord van voltooid aspect in Mod-clusters, terwijl de voorkeur in Vlaanderen - zij het minder uitgesproken – uitgaat naar *zijn*. Binnen Nederland is de voorkeur voor *hebben* het sterkst in de noordelijke provincies en het minst sterk in de zuidelijke provincies, waar *zijn* meer aanvaard wordt. In Vlaanderen is de voorkeur voor *zijn* het minst sterk in West-Vlaanderen, terwijl die in de andere Vlaamse provincies even sterk is (Hofmans 1981:101).

2.3.3 Taalinterne factoren

2.3.3.1 Inleiding

Het enquêteonderzoek van Hofmans (1981) is ook het vertrekpunt van het onderzoek naar taalinterne factoren die hulpwerkwoordselectie in Mod-clusters beïnvloeden. De Mod-clusters uit de 40 enquêtezinnen (zie §2.3.2.1) zijn opgebouwd uit een variatie aan hoofdwkwoorden en modale werkwoorden, en komen zowel in hoofd- als in bijzinnen voor. Op basis van de resultaten identificeert Hofmans (1981:73-74) vier taalinterne factoren die volgens hem voor zowel Vlamingen als Nederlanders een invloed hebben op de voorkeur voor *hebben* of *zijn*, met name de aard van het hoofdwkwoord, de afstand tussen het hulp- en het modale werkwoord, de aard van het modale werkwoord en irrealis.

Latere studies naar de invloed van taalinterne factoren op hulpwerkwoordselectie in Mod-clusters (Van Eynde et al. 2016, Dhont 2024) zijn expliciet voortgebouwd op de resultaten van Hofmans (1981). In wat volgt worden de bevindingen per taalinterne factor uiteengezet: het hoofdwkwoord (§2.3.3.2), de afstand (§2.3.3.3), het modale werkwoord (§2.3.3.4) en irrealis (§2.3.3.5).

2.3.3.2 Hoofdwerkwoord

Uit het onderzoek van zowel Hofmans (1981) als Van Eynde et al. (2016) blijkt dat de aard van het hoofdwerkwoord van het Mod-cluster een invloed heeft op de voorkeur voor *hebben* of *zijn*. In de enquêtezinnen van Hofmans (1981:73) komen tien verschillende hoofdwerkwoorden voor die *zijn* selecteren (*gaan, komen, stijgen, landen, blijven, ontsnappen, vertrekken, verschijnen, sterven en groeien*). Bepaalde hoofdwerkwoorden worden vaker met *zijn* gecombineerd, andere vaker met *hebben*. Zo wordt in clusters met *gaan* als hoofdwerkwoord, zoals in (25a), in Vlaanderen in 91.7% van de gevallen voor *zijn* gekozen, en in Nederland in 30.8% van de gevallen.⁶ Hoewel het percentage in Nederland kleiner is dan in Vlaanderen, ligt het in het licht van de nationale variatie relatief hoog (zie §2.3.2.1). In Mod-clusters met *stijgen* als hoofdwerkwoord, zoals in (25b), wordt er daarentegen meer voor *hebben* gekozen: in Nederland in 96.6% van de gevallen, in Vlaanderen in 80.2% (Hofmans 1981:81-84).

- (25) a. Hoewel zijn vrienden erop aandrongen, **heeft/is** hij niet naar dat feestje **willen gaan**.
(Hofmans 1981:112, zin 7)
- b. De berggids zei dat ze niet verder **had/was mogen stijgen**, omdat ze te zwak was.
(Hofmans 1981:112, zin 8)

Daarnaast zijn er ook hoofdwerkwoorden die in de Mod-clusters niet duidelijk meer met *hebben* of met *zijn* gecombineerd worden, zoals *verschijnen* in (26). Uit de zinnen met die hoofdwerkwoorden blijkt geen duidelijke voorkeur van de respondenten voor een bepaald hulpwerkwoord (Hofmans 1981:85).

⁶ Bij de illustratie van de enquêtezinnen voeg ik de vormen van *hebben* en *zijn* toe voor het leesgemak, maar in de werkelijke zinnen van Hofmans (1981) staat een lege plaats die door de respondent moet worden ingevuld (zie §2.3.2.1).

(26) Waarom **heeft/is** ze niet voor de rechtbank **willen verschijnen**? (Hofmans 1981:113, zin 29)

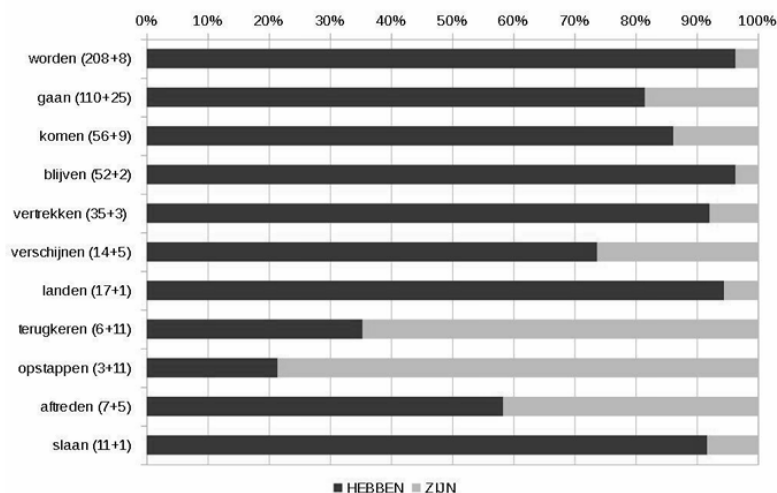
Hofmans (1981:86) beargumenteert zo dat Z-werkwoorden toch een voorkeur kunnen hebben voor het hulpwerkwoord van voltooid aspect van het Mod-cluster waarin ze als hoofdwkwoord fungeren. Werkwoorden als *gaan* worden door Hofmans (1981:82) bestempeld als “pro-*zijn*”: ze worden het vaakst gecombineerd met *zijn* in Mod-clusters en hebben dus een positieve invloed op de selectie van *zijn*. Werkwoorden als *stijgen* daarentegen zijn “pro-*hebben*”: ze komen in Mod-clusters het vaakst voor met *hebben* en hebben dus een positieve invloed op de selectie van *hebben*. Werkwoorden als *verschijnen* bestempelt Hofmans (1981:84) als “neutrale” werkwoorden, die evenveel met *hebben* als met *zijn* voorkomen. Aangezien bij die hoofdwkwoorden de invloed van de andere taalinterne factoren het sterkst naar voren komt, meer dan bij *gaan* of *stijgen* bijvoorbeeld, hebben ze volgens Hofmans (1981:84) de minst dominante invloed op de selectie van het hulpwerkwoord in Mod-clusters. De voorkeur voor *hebben* of *zijn* in Mod-clusters wordt voor de tien gebruikte hoofdwkwoorden gevisualiseerd in Figuur 2.2 (overgenomen uit Hofmans 1981:86). Wanneer een werkwoord op de *zijn/hebben*-as dicht bij *zijn* staat, is het pro-*zijn*, wanneer het dicht bij *hebben* staat is het pro-*hebben* (Hofmans 1981:86).



Figuur 2.2: Selectievoorkeur van hoofdwkwoorden bij Mod-clusters op *zijn/hebben*-as (Hofmans 1981:86)

Figuur 2.2 laat zien dat de selectievoorkeur niet voor elk hoofdwkwoord even sterk is. *Gaan*, *vertrekken* en *ontsnappen* zijn alle drie pro-*zijn*, maar *gaan* is nog meer pro-*zijn* dan *ontsnappen*. De positieve invloed van *gaan* op de selectie van *zijn* is dus sterker dan die van *ontsnappen*. Hetzelfde geldt voor de pro-*hebben*-werkwoorden: de positieve invloed van *stijgen* op de selectie van *hebben* is sterker dan die van *sterven*. Bij de neutrale werkwoorden *komen*, *verschijnen* en *blijven* is *komen* meer pro-*zijn*, en *blijven* meer pro-*hebben*. Hoewel de drie werkwoorden weinig invloed hebben op de hulpwerkwoordselectie, is die van *verschijnen* dus het kleinst (Hofmans 1981:86).

Ook uit het corpusonderzoek van Van Eynde et al. (2016) blijkt dat bepaalde Z-hoofdwkwoorden in Mod-clusters met *moeten* en *kunnen* (zie §2.3.1) vaker gecombineerd worden met *hebben* of *zijn*. Figuur 2.3 illustreert het voorkomen van *hebben* en *zijn* in Mod-clusters met *moeten* voor de hoofdwkwoorden die in het corpusonderzoek meer dan tien keer voorkomen (Van Eynde et al. 2016:23).



Figuur 2.3: Voorkomen van *hebben* en *zijn* bij hoofdwerkwoorden in Mod-clusters met *moeten* (Van Eynde et al. 2016:23)

Uit Figuur 2.3 blijkt bijvoorbeeld dat in clusters met het hoofdwerkwoord *opstappen* het hulpwerkwoord meestal *zijn* is (in 78.9% van de gevallen), terwijl in clusters met *blijven* het hulpwerkwoord *hebben* vooral voorkomt (96.3%). Bij het hoofdwerkwoord *af treden* is het voorkomen van *hebben* en *zijn* vergelijkbaar (respectievelijk 58.3% en 41.7%). Ik ga hier niet verder in op de voorkeur van alle hoofdwerkwoorden in het onderzoek van Van Eynde et al. (2016) of op de verschillen en gelijkenissen met de resultaten van Hofmans (1981). Beide studies maken echter duidelijk dat de hoofdwerkwoorden in Mod-clusters, hoewel ze allemaal Z-werkwoorden zijn (zie §2.3.1), een verschillende voorkeur vertonen voor het hulpwerkwoord van voltooid aspect van het cluster. Sommige werkwoorden bevorderen de selectie van *zijn* (bv. *gaan*, *opstappen*), andere stimuleren de selectie van *hebben* (bv. *stijgen*, *blijven*), terwijl andere hoofdwerkwoorden geen grote invloed op de hulpwerkwoordselectie hebben (bv. *verschijnen*, *af treden*).

2.3.3.3 Afstand

De invloed van de afstand tussen het hulpwerkwoord en het modale werkwoord wordt onderzocht door Hofmans (1981) en Dhont (2024). Hofmans (1981:79-82) test die invloed aan de hand van het verschil tussen hoofd- en bijzinnen. In bijzinnen zijn het hulpwerkwoord en het modale werkwoord steeds adjacent, zoals in (27a), waarbij geen woorden tussen hulpwerkwoord en het modale werkwoord *kunnen* staan. In de hoofdzinnen van de enquête plaatste Hofmans (1981:82) daarentegen steeds een of meerdere woorden tussen het hulpwerkwoord en het modale werkwoord, zoals in (27b).

- (27) a. Ik zeg je toch dat ik niet langer **heb/ben kunnen blijven**. (Hofmans 1981:113, zin 20)
 b. Hij **heeft/is** toch naar Zaïre **moeten vertrekken**. (Hofmans 1981:113, zin 31)

Uit de resultaten blijkt dat *hebben* meer gekozen wordt in bijzinnen als (27a), terwijl *zijn* de voorkeur krijgt in hoofdzinnen als (27b). Op basis daarvan concludeert Hofmans (1981:82) dat een afstand tussen het hulp- en het modale werkwoord een positieve invloed heeft op de selectie van *zijn*. Hij veronderstelt dat het modale werkwoord een directe invloed op de hulpwerkwoordselectie kan uitoefenen wanneer het naast het hulpwerkwoord staat, terwijl die invloed kleiner wordt bij een afstand. In (27a) komt *hebben* volgens hem vaker voor omdat het H-werkwoord *kunnen* naast het hulpwerkwoord staat, terwijl in (27b) vaker *zijn* wordt geselecteerd omdat het H-werkwoord *moeten* verder van het hulpwerkwoord af staat en daardoor minder invloed heeft op de selectie ervan (Hofmans 1981:82).

(28) a. Ja, hij **heeft/is kunnen gaan**.
b. Nee, hij **heeft/is** gisteren niet naar de tentoonstelling **kunnen gaan**. (Hofmans 1981:103)

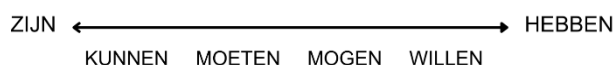
Concluderend stellen Hofmans (1981) en Dhont (2024) vast dat niet het zinstype, maar de afstand tussen het hulpwerkwoord en het modale werkwoord de hulpwerkwoordselectie beïnvloedt. Uit de resultaten van Hofmans (1981) blijkt dat afstand een positieve invloed heeft op de selectie van *zijn*. De invloed van de grootte van de afstand - het aantal woorden - wordt verondersteld, maar is niet onderzocht.

De invloed van de aard van het modale werkwoord op hulpwerkwoordselectie blijkt zowel uit het enquêteonderzoek van Hofmans (1981) als uit het corpusonderzoek van Dhont (2024) en Van Eynde et al. (2016). Hoewel de modale werkwoorden H-werkwoorden zijn, worden bepaalde modale werkwoorden in Mod-clusters meer gecombineerd met *hebben*, en andere meer met *zijn*.

(29)

- a. Ik zeg je toch dat ik niet lager **had/was kunnen blijven**. (Hofmans 1980:113, zin 20)
- b. Waarom **heb/ben ik moeten komen** en hij niet? (Hofmans 1980:112, zin 15)
- c. Ze zei dat ze niet naar het feestje **had/was mogen gaan**. (Hofmans 1980:114, zin 40)
- d. Waarom **heeft/is** ze niet voor de rechtbank **willen verschijnen**?
(Hofmans 1980:113, zin 29)

Die selectievoorkeur van de modale werkwoorden in Mod-clusters wordt gevisualiseerd op de as tussen *zijn* en *hebben* in Figuur 2.4 (overgenomen uit Hofmans 1980:88).



Figuur 2.4: Selectievoorkeur van modale werkwoorden bij Mod-clusters op *zijn/hebben*-as (Hofmans 1981:88)

Dhont (2024:191) bespreekt enkel Mod-clusters met modale werkwoorden die meer dan tien keer voorkomen in zijn dataset, met name *kunnen*, *moeten*, *mogen* en *willen*. De verdeling tussen *zijn* en *hebben* is in zijn onderzoek significant verschillend naargelang het modale werkwoord. Bij Mod-clusters met *kunnen* en *moeten* is er een significante voorkeur voor *zijn*, bij die met *mogen* en *willen* voor *hebben*. Die positieve invloed van *kunnen* en *moeten* op de selectie van *zijn* is echter enkel zichtbaar in Vlaanderen, niet in Nederland (Dhont 2024:192). Van Eynde et al. (2016) onderzoeken de verhouding tussen *hebben* en *zijn* in Mod-clusters met *moeten* en *kunnen* (zie §2.3.1). De resultaten tonen dat in Mod-clusters met *moeten* het hulpwerkwoord in 18% van de gevallen *zijn*, en in 82% *hebben* is. In Mod-clusters met *kunnen* komt *zijn* vaker voor (23.5%) dan bij *moeten* (Van Eynde et al. 2016:22).

Concluderend zien we dat de verschillende modale werkwoorden, net als de hoofdwkwoorden (zie §2.3.3.2), niet dezelfde voorkeur vertonen voor het hulpwerkwoord van het Mod-cluster. Zonder erg in detail te gaan, suggereert het onderzoek van Hofmans (1981), Dhont (2024) en Van Eynde et al. (2016) dat het modale werkwoord *kunnen* de meest positieve invloed heeft op de selectie van *zijn*, gevolgd door *moeten*, terwijl *mogen* en *willen* een positieve invloed hebben op de selectie van *hebben*.

2.3.3.5 *Irrealis*

Hoewel Hofmans (1981:106) in zijn enquêteonderzoek geen expliciete aandacht besteedt aan modaliteit, vermoedt hij dat een irrealislezing een positieve invloed heeft op de selectie van *hebben* in Mod-clusters. Dat illustreert hij met de zin in (30), die twee Mod-clusters bevat waarbij volgens de besproken invloed van het modale werkwoord en het hoofdwkwoord een voorkeur voor *zijn* zou bestaan. Het modale werkwoord *kunnen* komt immers vaker voor met *zijn* dan met *hebben* (zie §2.3.3.4), en de hoofdwkwoorden *komen* en *(weg)gaan* zijn volgens Hofmans pro-*zijn* (zie §2.3.3.2). Toch voorspelt Hofmans (1981:106) in zin (30) een voorkeur voor *hebben*, omwille van de irrealislezing.

- (30) Indien je vroeger **had/³was kunnen komen**, **had/³was** je ook vroeger **kunnen weggaan**!
(Hofmans 1981:106)

De hypothese van Hofmans (1981:106) over de positieve invloed van irrealis op de selectie van *hebben* in Mod-clusters is aannemelijk, aangezien blijkt dat werkwoorden die *zijn* selecteren soms met *hebben* gecombineerd worden wanneer de zin een irrealislezing uitdrukt (De Vooys 1967:141, Paardekooper 1978:89, Haeseryn et al. 1997:75, Ackema & Sorace 2017:9). Dat wordt geïllustreerd in (31).

- (31) a. Marie **is**/*heeft gisteren een uur bij Klaas **gebleven**. AUX_Z - HWW_Z
b. Marie **was**/*had gisteren een uur bij Klaas **gebleven**. AUX_Z - HWW_Z
c. **Was/had** Marie bij Klaas **gebleven**, dan was het niet gebeurd. AUX_{Z/H} - HWW_Z

Het hoofdwerkwoord *blijven* in (31) selecteert *zijn* als hulpwerkwoord van voltooid aspect. De zinnen in (31a) en (31b) krijgen een realislezing, en dan is enkel het hulpwerkwoord *zijn* grammaticaal, zowel in de onvoltooid tegenwoordige (31a) als in de onvoltooid verleden tijd (31b). De zin in (31c) staat echter in irrealis: de onvoltooid verleden tijd drukt in dat geval een niet-werkelijkheidsmodaliteit uit (Haeseryn et al. 1997:75). Dan kan zowel *zijn* als *hebben* het hulpwerkwoord van voltooid aspect zijn (Paardekooper 1978:89). Het gebruik van *hebben* bij Z-werkwoorden in irrealis, zoals in (31c), komt vrijwel enkel voor in gesproken taal en wordt als informeel taalgebruik beschouwd. Bovendien is het niet in alle delen van het taalgebied gangbaar (Haeseryn et al. 1997:75, Onze Taal 2024).

In verder onderzoek naar hulpwerkwoordselectie in Mod-clusters (Dhont 2024, Van Eynde et al. 2016) is de invloed van irrealis niet onderzocht. Dhont (2024:181) vermeldt expliciet dat hij het irrealiseffect niet onderzoekt omdat hij zich toespitst op de voltooid tegenwoordige tijd; de irrealislezing wordt meestal uitgedrukt aan de hand van de voltooid verleden tijd, zoals in (31c) (Haeseryn et al. 1997:75). Het enquêteonderzoek van Van Leeuwen (2009) naar hulpwerkwoordselectie in vierledige clusters suggereert echter dat irrealis bij drieledige Mod-clusters een belangrijke rol speelt. Van Leeuwen (2009) onderzoekt hulpwerkwoordselectie in zinnen als (32), die een vierledige cluster bevatten die bestaat uit een hulpwerkwoord van voltooid aspect, het modale werkwoord *moeten*, een aspectueel werkwoord (*gaan* of *blijven*) en een hoofdwerkwoord dat ofwel *zijn* ofwel *hebben* selecteert. Elke cluster komt steeds voor in drie condities: de voltooid tegenwoordige tijd (32a), de voltooid verleden tijd met realislezing (32b) en de voltooid verleden tijd met irrealislezing (32c) (Van Leeuwen 2009:63).

- (32) a. Het jongetje **is/heeft moeten gaan groeien**.
 b. Door de medicijnen **was/had** het jongetje **moeten gaan groeien** en nu was hij dan ook een flinke kerel.
 c. Het jongetje **was/had moeten gaan groeien**, maar dat gebeurde niet.
 (Van Leeuwen 2009:63)

In het enquêteonderzoek werden 208 respondenten 75 zinnen zoals in (32) voorgelegd, telkens in de twee hulpwerkwoordvarianten, zoals geïllustreerd in (33). De respondenten gaven aan welke van de twee zinnen voor hen het natuurlijkst klonk (Van Leeuwen 2009:28).

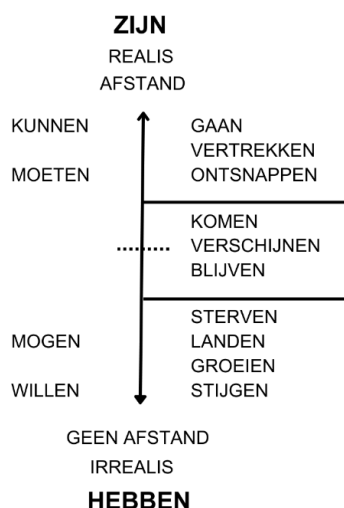
- (33) a. Jan **had moeten blijven werken**, maar hij stopte voortijdig.
 b. Jan **was moeten blijven werken**, maar hij stopte voortijdig. (Van Leeuwen 2009:63)

De resultaten tonen aan dat de irrealislezing in vierledige clusters met een modaal en aspectueel werkwoord een sterke positieve invloed heeft op de selectie van *hebben*. In zin (32a) en (32b) wordt *hebben* bijvoorbeeld verkozen in respectievelijk 55% en 50% van de gevallen, terwijl *hebben* in de irrealiszin (32c) in 93% van de gevallen de voorkeur krijgt (Van Leeuwen 2009:32). Dat positieve effect van irrealis is bij alle testzinnen aanwezig: de irrealiszin scoort steeds tussen de 90% en 100% voor *hebben* en die score verschilt telkens significant van de *hebben*-score van de andere zinnen (Van Leeuwen 2009:48).

De positieve invloed van irrealis op de voorkeur voor *hebben* in Mod-clusters, zoals voorspeld door Hofmans (1981), is dus niet onderzocht. Zijn hypothese wordt wel ondersteund door het algemene effect van irrealis op hulpwerkwoordselectie bij Z-werkwoorden, en door het onderzoek van Van Leeuwen (2009), dat het irrealiseffect bevestigt bij vierledige clusters met een modaal en aspectueel werkwoord.

2.3.3.6 Conclusie: taalinterne factoren

Op basis van het onderzoek van Hofmans (1981), Van Eynde et al. (2016) en Dhont (2024) werden vier taalinterne factoren geïdentificeerd die een invloed hebben op hulpwerkwoordselectie in Mod-clusters. De invloed van die factoren wordt gevisualiseerd op een *zijn/hebben*-as in Figuur 2.5.



Figuur 2.5: Selectievoorkeur in Mod-clusters bij vier taalinterne factoren op *zijn/hebben*-as

In de eerste plaats blijkt dat de hoofdwerkwoorden in Mod-clusters niet evenveel met *hebben* of *zijn* gecombineerd worden. Bepaalde hoofdwerkwoorden (bv. *vertrekken*) hebben een positieve invloed op de selectie van *zijn* als hulpwerkwoord van het Mod-cluster, terwijl andere hoofdwerkwoorden (bv. *stijgen*) een positieve invloed hebben op de selectie van *hebben*. Daarnaast zijn er neutralere hoofdwerkwoorden (bv. *verschijnen*) die minder invloed uitoefenen op hulpwerkwoordselectie. Ten tweede heeft afstand tussen het hulpwerkwoord en het modale werkwoord een positieve invloed op de selectie van *zijn*. Die invloed is vermoedelijk sterker bij een grotere afstand. Ten derde komen de modale werkwoorden *kunnen* en *moeten* vaker voor met het hulpwerkwoord *zijn*, terwijl *mogen* en *willen* een positieve invloed hebben op de selectie van *hebben*. Tot slot is het op basis van de bestudeerde invloed van irrealis bij Z-werkwoorden (o.a. Paardekooper 1978) en het resultaat van Van Leeuwen (2009) aannemelijk dat een irrealisreading een positieve invloed heeft op de selectie van *hebben* in Mod-clusters.

2.3.4 Conclusie: hulpwerkwoordvariatie in Mod-clusters

In driedelige clusters met een modaal werkwoord kan hulpwerkwoordvariatie voorkomen wanneer het hoofdwerkwoord *zijn* selecteert. Hoewel het hulpwerkwoord *hebben* meer voorkomt dan *zijn*, kent de hulpwerkwoordselectie in Mod-clusters een grote geografische variatie. In Nederland krijgt *hebben* de voorkeur, in Vlaanderen *zijn*. In Nederland komt *zijn* meer voor in het zuiden dan in het noorden; in Vlaanderen onderscheidt West-Vlaanderen zich vanwege de sterkere voorkeur voor *hebben*. Daarnaast wordt hulpwerkwoordselectie in Mod-clusters ook bepaald door taalinterne factoren. Zo komen de verschillende hoofdwerkwoorden en modale werkwoorden in Mod-clusters niet evenveel voor met *hebben* of *zijn*, en vertonen ze dus een verschillende voorkeur voor het hulpwerkwoord. Daarnaast heeft een afstand tussen het hulpwerkwoord en het modale werkwoord een positieve invloed op de selectie van *zijn*, en irrealis een positieve invloed op de selectie van *hebben*. In de volgende sectie breng ik in kaart wat geweten is over de minder bestudeerde hulpwerkwoordvariatie in Asp-clusters.

2.4 Hulpwerkwoordvariatie in clusters met een aspectueel werkwoord (Asp-clusters)

In deze sectie wordt de hulpwerkwoordvariatie in drieledige werkwoordsclusters met een aspectueel werkwoord (Asp-clusters) beschreven. Ik volg daarbij dezelfde structuur als bij de Mod-clusters (§2.3). Eerst bespreek ik de algemene verhouding tussen *zijn* en *hebben* in Asp-clusters (§2.4.1), en vervolgens de geografische variatie (§2.4.2). De invloed van taalinterne factoren komt in §2.4.3 aan bod. In §2.4.4 vat ik de belangrijkste bevindingen voor hulpwerkwoordvariatie in Asp-clusters samen.

2.4.1 Verhouding tussen *zijn* en *hebben*

De algemene verhouding tussen *zijn* en *hebben* in Asp-clusters kan enkel worden onderzocht wanneer hulpwerkwoordvariatie mogelijk is. De aspectuele werkwoorden *gaan*, *blijven*, *komen* en *beginnen* selecteren *zijn*, dus variatie is mogelijk als het hoofdwkwoord van het cluster *hebben* selecteert (zie §2.2.2). In (34a) is het hoofdwkwoord *bellen* van het Asp-cluster met *blijven* een H-werkwoord, dus variatie is mogelijk. In (34b) selecteert het hoofdwkwoord *ontsnappen* net als *blijven* het hulpwerkwoord *zijn*. Dan is het hulpwerkwoord van voltooid aspect altijd *zijn* (zie §2.2.2).

- (34) a. Ze **zijn/hebben** hem maar **blijven bellen**. $AUX_{Z/H} - ASP_Z - HWW_H$
 b. Ze **zijn/*hebben** maar uit de kooi **blijven ontsnappen**. $AUX_Z - ASP_Z - HWW_Z$

Van Eynde et al. (2016) voerden naar hulpwerkwoordvariatie in Asp-clusters een corpusonderzoek uit in het Corpus Gesproken Nederlands (Oostdijk 2002), Lassy Small (Van Noord et al. 2013) en het SoNaR-500-corpus (Oostdijk et al. 2013). De Asp-clusters komen in de dataset 228 keer voor met *zijn*, en slechts 11 keer met *hebben* (Van Eynde et al. 2016:21). Die resultaten werden echter niet gefilterd op basis van het hulpwerkwoord dat het hoofdwkwoord selecteert, dus er zijn ook Asp-clusters met een Z-hoofdwkwoord opgenomen die geen variatie toestaan, zoals in (34b). Een exacte verhouding kunnen ze uit de resultaten dus niet afleiden, maar de absolute aantallen maken volgens Van Eynde et al. (2016:21) duidelijk dat het hulpwerkwoord *zijn* in Asp-clusters de ‘normale’ variant is, en dat de ‘alternatieve’ variant *hebben* slechts sporadisch voorkomt.

2.4.2 Geografische variatie

De hulpwerkwoordselectie in Asp-clusters vertoont geografische variatie, zowel tussen Vlaanderen en Nederland (§2.4.2.1) als binnen de provincies van Vlaanderen (§2.4.2.2).

2.4.2.1 Nationale variatie

De nationale tendens voor Asp-clusters is dat in Nederland enkel *zijn* voorkomt, terwijl in Vlaanderen variatie tussen *hebben* en *zijn* mogelijk is, zoals geïllustreerd in (35) (Draye & Van der Horst 2006:238).

- (35) a. Die **ben/*heb** ik in Gent **gaan kopiëren**. NEDERLAND $AUX_Z - ASP_Z - HWW_H$
 b. Die **ben/heb** ik in Gent **gaan kopiëren**. VLAANDEREN $AUX_{Z/H} - ASP_Z - HWW_H$

De geografische variatie in Asp-clusters werd voor het eerst onderzocht in een tweede enquêteonderzoek van Hofmans (1982b, 1985), waarin de hulpwerkwoordvariatie voor verschillende types drie- en vierledige werkwoordsclusters in kaart wordt gebracht.⁷ Voor Asp-clusters zijn er vier zinnen opgenomen in de enquête. De clusters bevatten de aspectuele werkwoorden *gaan* en *komen* en de H-hoofdwkwoorden *halen* en *spelen*, zoals de zinnen in (36) (Hofmans 1982b:90).

⁷ Over het aantal respondenten of de specifieke taak van de tweede enquête verschaft Hofmans (1982b, 1985) geen informatie.

- (36) a. Ze **is/heeft** hem gisteren niet **komen halen**.
b. Hij **is/heeft** in de kelder **gaan spelen**.

(Hofmans 1982b:84)

De resultaten tonen dat de Vlaamse respondenten in 6% van de gevallen *hebben* verkiezen, terwijl de Nederlandse respondenten in slechts 1.9% van de clusters voor *hebben* kiezen (Hofmans 1982b:90). Het hulpwerkwoord *hebben* krijgt dus weinig voorkeur, maar wel meer in Vlaanderen dan in Nederland. Die nationale tendens wordt bevestigd in corpusonderzoek. In het onderzoek van Van Eynde et al. (2016:18) zijn alle voorbeelden van Asp-clusters met *hebben* uit het Corpus Gesproken Nederlands afkomstig uit het Vlaamse subcorpus (zie §2.4.1). Dhont (2024) onderzoekt de nationale variatie in hulpwerkwoordselectie bij Asp-clusters aan de hand van de krantencomponent van het SoNaR-500-corpus. De resultaten tonen aan dat de verhouding tussen *zijn* en *hebben* significant verschilt in Vlaanderen en Nederland. In Vlaanderen wordt in 8 van de 251 Asp-clusters (3.2%) *hebben* gebruikt, in Nederland daarentegen komt uitsluitend *zijn* voor (Dhont 2024:195).

Hoewel *hebben* in Asp-clusters meer voorkomt in Vlaanderen dan in Nederland, is het gebruik van *hebben* in Vlaanderen niet algemeen. Het maakt dus geen deel uit van de standaardvariëteit (Dhont 2024:197). Op basis van persoonlijke observaties in informeel gesproken taalgebruik vermoedt Hofmans (1985:75) echter dat *hebben* in het gesproken Vlaams meer voorkomt dan aangegeven in zijn enquêteonderzoek. Ook Van der Horst (2005:102) gaat ervan uit dat de *hebben*-variant meer gebruikt wordt in gesproken dan in geschreven taal.

2.4.2.2 Regionale variatie

Aangezien *hebben* in Nederland niet voorkomt, wordt er geen regionale variatie in Nederland aangenomen. Binnen Vlaanderen blijkt West-Vlaanderen zich te onderscheiden van de andere Vlaamse provincies. Hofmans (1982b, 1985) vergelijkt de voorkeur van de respondenten voor *hebben* of *zijn* per provincie (West-Vlaanderen, Oost-Vlaanderen, Antwerpen, Vlaams-Brabant en Limburg). Het resultaat toont aan dat *hebben* in West-Vlaanderen vaker gekozen wordt dan in de andere provincies, namelijk in 12% van de gevallen tegenover respectievelijk 7%, 6.5%, 2.3% en 2.2% van de gevallen (Hofmans 1982b:90). Draye en Van der Horst (2006:239) vermoeden ook dat *hebben* in West-Vlaanderen meer voorkomt dan in de andere Vlaamse provincies. In verschillende West-Vlaamse dialecten worden immers enkele werkwoorden die in het Standaardnederlands met *zijn* worden vervoegd, zoals *gaan* en *blijven*, met *hebben* gecombineerd, zoals geïllustreerd in (37) (Draye & Van der Horst 2006:238-239).

- (37) a. Hij **heeft** naar huis **gegaan**.
b. Ze **hebben** lang **gebleven**.

(Draye & Van der Horst 2006:238)

2.4.3 Taalinterne factoren

Aan de taalinterne factoren die de hulpwerkwoordselectie in Asp-clusters mogelijk beïnvloeden is niet veel aandacht besteed. Hofmans (1982b:90) stelt enkel een invloed van de aard van het aspectuele werkwoord vast. Hij beargumenteert dat *gaan* een sterkere positieve invloed heeft op de selectie van *hebben* dan *komen*, omdat in Asp-clusters met *gaan hebben* meer wordt verkozen (in 4.25% van de gevallen) dan in clusters met *komen* (3.77%) (zie §2.4.2.1). Andere taalinterne factoren onderzoekt Hofmans (1982b) niet. In het corpusonderzoek van Van Eynde et al. (2016) en Dhont (2024) wordt de invloed van taalinterne factoren niet expliciet onderzocht. De resultaten tonen enkel aan dat hulpwerkwoordvariatie mogelijk is bij de vier aspectuele werkwoorden *gaan*, *blijven*, *komen* en *beginnen* (Van Eynde et al. 2016:17-19, Dhont 2024:195).

2.4.4 Conclusie: hulpwerkwoordvariatie in Asp-clusters

In driedelige werkwoordsclusters met een aspectueel werkwoord kan hulpwerkwoordvariatie voorkomen wanneer het hoofdwerkwoord *hebben* selecteert. Uit het beperkte onderzoek blijkt dat *hebben* in Asp-clusters slechts sporadisch en veel minder dan *zijn* voorkomt. Hofmans (1985) vermoedt echter dat het werkelijke voorkomen van *hebben* in het onderzoek nog niet is vastgelegd aangezien de *hebben*-variant vooral voorkomt in informele spreektaal. De hulpwerkwoordselectie vertoont daarnaast geografische variatie: als *hebben* voorkomt, is dat in Vlaanderen; in Nederland blijkt de variatie niet mogelijk. Binnen Vlaanderen vertoont de provincie West-Vlaanderen de grootste voorkeur voor *hebben*. Tot slot is er, afgezien van een verschil tussen *gaan* en *komen*, niets bekend over de invloed van taalinterne factoren zoals het hoofdwerkwoord, afstand, het aspectuele werkwoord of modaliteit op hulpwerkwoordselectie in Asp-clusters.

In §2.3 en §2.4 bracht ik de variatie tussen *hebben* en *zijn* voor Mod- en Asp-clusters in kaart. Daarbij ging de aandacht uit naar de algemene verhouding, de geografische variatie en de taalinterne factoren die de hulpwerkwoordselectie beïnvloeden. In de volgende sectie (§2.5) verplaats ik de focus van de beschrijving van de hulpwerkwoordvariatie naar mogelijke analyses van de variatie.

2.5 Analyse van hulpwerkwoordvariatie in driedelige werkwoordsclusters

In deze sectie worden drie voorgestelde analyses voor hulpwerkwoordvariatie in Mod- en/of Asp-clusters besproken. Eerst behandel ik de grammaticalisatie van het groepsvormende werkwoord (§2.5.1), gevolgd door de analyse van Coussé en Van de Velde (2014), die de eerste benadering kritisch herzien. Tot slot beschrijf ik de formele analyse waarbij de mogelijkheid tot hulpwerkwoordvariatie verklaard wordt door de functionele, semi-lexicale of lexicale status van het groepsvormende werkwoord (§2.5.3).

2.5.1 Grammaticalisatie van het groepsvormende werkwoord

Hulpwerkwoordvariatie in Mod- en Asp-clusters wordt door Hofmans (1982a, 1985), Draye en Van der Horst (2006) en Van der Horst (2008) beschouwd als een indicatie van de grammaticalisatie van de groepsvormende werkwoorden.⁸ De hulpwerkwoordvarianten weerspiegelen in de analyse verschillende stadia in een grammaticalisatieproces. De evidentie voor de analyse is het meest uitgebreid voor Mod-clusters (§2.5.1.1), maar wordt naar analogie aangenomen voor Asp-clusters (§2.5.1.2).

2.5.1.1 Mod-clusters

Volgens Hofmans (1982a), Draye en Van der Horst (2006) en Van der Horst (2008) vindt er in het Nederlands een evolutie plaats waarbij het modale werkwoord het vermogen verliest om de keuze van het hulpwerkwoord van voltooid aspect te bepalen wanneer het groepsvormend gebruikt is. Van der Horst staat dat idee met historische voorbeelden van Mod-clusters. Volgens Van der Horst (2008:422) was het hulpwerkwoord in Mod-clusters in het Middelnederlands altijd *hebben*, zoals in voorbeeld (38).

- (38) **Hadde** dese ledich **moeten** **bliven**.
 Had deze leeg moeten blijven.
 ‘Had deze leeg moeten blijven.’

(Van der Horst 2008:422, 1350-1500)

⁸ Hofmans (1982a, 1985) gebruikt de term “deverbalisering”, maar verwijst daarmee naar wat Draye en Van der Horst (2006) en Van der Horst (2008) grammaticalisatie noemen.

Mod-clusters met *zijn* worden volgens Van der Horst (2008:893) daarentegen pas vanaf de 16^{de} eeuw aangetroffen. In de 17^{de} en 18^{de} eeuw komen ze steeds vaker voor, zoals in (39); in de 19^{de} eeuw worden ze veelvuldig gebruikt. Van der Horst benadrukt daarbij dat de vroegste voorbeelden van Mod-clusters met *zijn* afkomstig zijn uit het zuidelijke deel van het taalgebied, en dat zulke clusters in latere periodes ook voornamelijk aangetroffen worden in zuidelijke teksten (Van der Horst 2008:893).

- (39) En **syn** daer **moeten** **blyven**.
 En *zijn* daer moeten blijven.
 ‘En *zijn* daer moeten blijven.’

(Van der Horst 2008:1173, 17^{de} eeuw)

Op die manier wordt een overgang van *hebben* naar *zijn* in Mod-clusters geschetst, waarbij de *hebben*-variant de oudere vorm, en de *zijn*-variant de nieuwere vorm is (Draye & Van der Horst 2006:230-236). Die evolutie van *hebben* naar *zijn* toont volgens Van der Horst (2008) aan dat de keuze van het hulpwerkwoord in Mod-clusters steeds meer bepaald wordt door het hoofdwkwoord van het Mod-cluster. In de *hebben*-variant in (38) wordt het hulpwerkwoord van voltooid aspect volgens Van der Horst bepaald door het modale werkwoord *moeten*, en niet door het hoofdwkwoord *blijven*, dat *zijn* selecteert. In (39) wordt *zijn* daarentegen geselecteerd door het hoofdwkwoord *blijven*, en niet door het modale *moeten*, dat *hebben* selecteert. Samengevat: als het hulpwerkwoord van een Mod-cluster *hebben* is, wordt het door het modale werkwoord geselecteerd; als het *zijn* is, wordt het door het hoofdwkwoord geselecteerd.⁹ Aangezien er een overgang plaatsvindt van *hebben* naar *zijn*, vindt er dus ook een overgang plaats van het modale werkwoord naar het hoofdwkwoord als selecteerder van het hulpwerkwoord van voltooid aspect (Draye & Van der Horst 2006:230-236, Van der Horst 2008:893-894).

Het modale werkwoord verliest in die analyse dus het vermogen om de keuze van het hulpwerkwoord te bepalen wanneer het groepsvormend gebruikt is, en draagt dat vermogen over aan het hoofdwkwoord. De twee hulpwerkwoordvarianten in Mod-clusters worden zo opgevat als twee stadia in een grammaticalisatieproces (Hofmans 1982a, Draye & Van der Horst 2006:237, Van der Horst 2008:893-894). Wanneer *hebben* het hulpwerkwoord in een Mod-cluster is, is het geselecteerd door het modale werkwoord, zoals door *moeten* in (40a). Het modale werkwoord heeft het vermogen dus nog om de keuze van het hulpwerkwoord te bepalen: het bevindt zich in een vroeg stadium van het grammaticalisatieproces. Wanneer *zijn* het hulpwerkwoord is, is het geselecteerd door het hoofdwkwoord, zoals *gaan* in (40b). Het modale werkwoord heeft dus het vermogen om de keuze van het hulpwerkwoord te bepalen afgestaan aan het hoofdwkwoord, en bevindt zich dus in een later stadium in het grammaticalisatieproces (Draye & Van der Horst 2006:237).

- | | | |
|------|---|--|
| (40) | a. Alex heeft naar de kerk moeten gaan. | AUX _H – MOD _H – HWW _Z |
| | b. Alex is naar de kerk moeten gaan . | AUX _Z – MOD _H – HWW _Z |

Twee implicaties van die analyse zijn van belang. In de eerste plaats veronderstellen Hofmans (1982a:83-84) en Draye en Van der Horst (2006:237) dat de verschillende modale werkwoorden zich in verschillende stadia van het grammaticalisatieproces kunnen bevinden. De verschillende selectievoordeur van de modale werkwoorden in Mod-clusters (zie §2.3.3.4) wordt dan verklaard doordat het ene werkwoord zich verder dan het andere in het grammaticalisatieproces bevindt.

⁹ Van der Horst (2008) houdt enkel rekening met Mod-clusters waarbij het hoofdwkwoord *zijn* selecteert en waarbij dus variatie mogelijk is (zie §2.2.2 en §2.3.1).

In de tweede plaats stellen Draye en Van der Horst (2006:236) dat het modale werkwoord het grammaticalisatieproces in Vlaanderen verder doorlopen heeft dan in Nederland. De oudere vorm, de *hebben*-variant, komt immers het meest voor in Nederland, terwijl de nieuwere vorm, de *zijn*-variant, in Vlaanderen de voorkeur krijgt (zie §2.3.2). Concreet wordt volgens Draye en Van der Horst (2006:236) de keuze van het hulpwerkwoord van voltooid aspect in Nederland dus vaker bepaald door het modale werkwoord, en in Vlaanderen vaker door het hoofdwerkwoord. Het feit dat de eerste historische voorbeelden van Mod-clusters met *zijn* afkomstig zijn uit zuidelijke teksten, biedt ondersteuning aan die analyse van de nationale variatie (Draye & Van der Horst 2006:236).

2.5.1.2 Asp-clusters

Naar analogie met Mod-clusters gaan Hofmans (1985:77-79) en Draye en Van der Horst (2006:239) ervan uit dat de twee hulpwerkwoordvarianten in Asp-clusters ook een weerspiegeling zijn van een grammaticalisatieproces van het aspectuele werkwoord. Net als bij de modale werkwoorden staan de aspectuele werkwoorden dan steeds meer het vermogen om de keuze van het hulpwerkwoord te bepalen bij groepsvorming af aan het hoofdwerkwoord. In (41a) wordt het hulpwerkwoord *zijn* geselecteerd door *gaan*, en niet door het H-hoofdwerkwoord *halen*. Het aspectuele werkwoord heeft het vermogen om het hulpwerkwoord te selecteren niet afgestaan en bevindt zich dus in een vroeg stadium van grammaticalisatie. In (41b) wordt *hebben* geselecteerd door het hoofdwerkwoord *halen*. Het aspectuele werkwoord heeft het vermogen dus afgestaan en bevindt zich in een verder stadium.

- | | | |
|------|--|--|
| (41) | a. Toen ben ik mijn rapport gaan halen. | AUX _Z – ASP _Z – HWW _H |
| | b. Toen heb ik mijn rapport gaan halen . | AUX _H – ASP _Z – HWW _H |

De analyse van een grammaticalisatieproces van de aspectuele werkwoorden wordt niet, zoals voor de modale werkwoorden (§2.5.1.1), ondersteund aan de hand van een diachrone evolutie. Asp-clusters komen volgens Draye en Van der Horst (2006:233-234) immers al vanaf het Middelnederlands zowel met *zijn*, zoals in (42), als met *hebben*, zoals in (43), voor.

- | | | |
|------|---|---|
| (42) | Daer naer zyn wy gaen slaepen.
Daar na zijn wij gaan slapen.
'Daarna zijn wij gaan slapen.' | (Draye & Van der Horst 2006:234, 16 ^{de} eeuw) |
| (43) | Dien auent hebben wy gaen slaepen.
Die avond hebben wij gaan slapen.
'Die avond hebben wij gaan slapen.' | (Draye & Van der Horst 2006:234, 16 ^{de} eeuw) |

Toch toont de *hebben*-variant uit de 16^{de} eeuw in (43) volgens Draye en Van der Horst (2006:233) niet aan dat de keuze van het hulpwerkwoord van voltooid aspect al sinds het Middelnederlands door het hoofdwerkwoord van het cluster bepaald kan worden. De werkwoorden *gaan*, *blijven*, *komen* en *beginnen* werden in de vroegere periodes van het Nederlands namelijk zowel met *zijn* als met *hebben* vervoegd (Van der Horst 2008:422-423), zoals geïllustreerd voor *gaan* in (44).

- | | | |
|------|--|---|
| (44) | Hoe dattet Christoffel Wagenaer gegaen heeft na zijns heren doot.
Hoe dat het Christoffel Wagenaer gegaan heeft na van zijn heer dood.
'Hoe het Christoffel Wagenaer vergaan is na de dood van zijn heer.' | (Van der Horst 2008:856; 16 ^{de} eeuw) |
|------|--|---|

Op die manier bestaat er volgens Van der Horst (2008:893) geen zekerheid over welk werkwoord de keuze van het hulpwerkwoord van voltooid aspect bepaalt in clusters als (43). Zowel het aspectuele werkwoord *gaan* als het hoofdwerkwoord *slapen* konden immers *hebben* selecteren. Hiermee wordt het gebrek aan historische evidentie dat in Asp-clusters *zijn* de oudere en *hebben* de nieuwere variant is, genuanceerd (Draye & Van der Horst 2006:236-239).

Ter conclusie, Hofmans (1982a, 1985), Draye en Van der Horst (2006) en Van der Horst (2008) analyseren hulpwerkwoordvariatie in Mod- en Asp-clusters als twee verschillende stadia in een grammaticalisatieproces van de modale en aspectuele werkwoorden. De richting van de ontwikkeling is verschillend (van *hebben* naar *zijn* in Mod-clusters en van *zijn* naar *hebben* bij Asp-clusters), maar het gaat om hetzelfde proces waarbij het groepsvormende werkwoord steeds minder, en het hoofdwerkwoord steeds meer de keuze van het hulpwerkwoord van voltooid aspect bepaalt.

2.5.2 Coussé en Van de Velde (2014)

Coussé en Van de Velde (2014) herzien de grammaticalisatie-analyse (§2.5.1) en stellen een alternatieve analyse voor hulpwerkwoordvariatie in Mod-clusters voor. Aan de hand van drie observaties benadrukken ze het belang van meer aandacht voor modaliteit bij de analyse van hulpwerkwoordvariatie. Ten eerste merken Coussé en Van de Velde (2014:353) op dat de voorbeelden die Van der Horst (2008) gebruikt om aan te tonen dat in het Middelnederlands het hulpwerkwoord geselecteerd werd door het modale werkwoord, allemaal voorkomen in irrealis.¹⁰ Ik herhaal het voorbeeld van Van der Horst (2008:422) in (38) hieronder als (45) (zie §2.5.1.1). In een voorbeeld als (45) is het volgens Coussé en Van de Velde (2014:353) niet zeker of de keuze van het hulpwerkwoord bepaald is door het modale werkwoord, of door de irrealis zelf. Al sinds het Middelnederlands kunnen werkwoorden die *zijn* selecteren, zoals *blijven* in (45), namelijk gecombineerd worden met *hebben* bij een irrealislezing (Kern 1912:272, De Rooij 1988:24, Coussé en Van de Velde 2014:353, zie ook §2.3.3.5).

(45) **Hadde** dese ledich **moeten** bliven.

Had deze leeg moeten bliven.

‘Had deze leeg moeten bliven.’

(Van der Horst 2008:422, 1350-1500)

Mod-clusters met *hebben* zonder irrealis duiken volgens Coussé en Van de Velde (2014:355) pas vanaf de 19^{de} eeuw op in het materiaal van Van der Horst (2008). Ze vinden daarvan slechts twee voorbeelden, waaronder (46). Mod-clusters met *hebben* in realis blijken dus zeldzaam in het historische materiaal.

(46) tot welker behandeling men toch eindelijk **heeft moeten komen**

(Van der Horst 2008:1760, 19^{de}-20^{ste} eeuw)

Ten derde komen Mod-clusters met *zijn* uitsluitend in de realis voor in het materiaal. Van der Horst (2008) gebruikt voorbeelden met *zijn*, zoals (39), hieronder herhaald als (47), om aan te tonen dat het hulpwerkwoord geselecteerd is door het hoofdwerkwoord, maar stelt zelf niet aan de orde dat zulke voorbeelden nooit in irrealis voorkomen (Coussé en Van de Velde 2014:354).

¹⁰ Ik richt me enkel op de observaties en analyse van Coussé en Van de Velde (2014) over Mod-clusters met een hoofdwerkwoord dat *zijn* selecteert en waarin dus variatie mogelijk is (zie §2.3.1).

- (47) En **syn** daer **moeten blyven**.
En zijn daar moeten blijven.
'En zijn daar moeten blijven.'

(Van der Horst 2008:1173, 17^{de} eeuw)

Op basis van hun observaties stellen Coussé en Van de Velde (2014) dat in de geschetste overgang van *hebben* naar *zijn* (§2.5.1) niet voldoende rekening werd gehouden met het onderscheid tussen irrealis en realis. Ze voeren aanvullend corpusonderzoek uit in het Compilatiecorpus Historisch Nederlands, en stellen op basis daarvan een alternatieve, constructiegebaseerde analyse voor. Zonder in detail op de analyse in te gaan, suggereren ze dat Mod-clusters met *hebben* in irrealis en in realis op verschillende manieren geanalyseerd moeten worden (Coussé en Van de Velde 2014:361-362).

Voor Mod-clusters met *hebben* in irrealis veronderstellen ze dat het bredere gebruik van *hebben* in irrealis bij werkwoorden die *zijn* selecteren doorgetrokken is tot het gebruik van *hebben* in Mod-clusters met een Z-hoofdwkwoord in irrealis. In dergelijke clusters zou de keuze voor *hebben* volgens hen dan aan de irrealislezing toegeschreven kunnen worden (Coussé en Van de Velde 2014:360). Mod-clusters met *hebben* in realis beschouwen ze daarentegen als een zelfstandige constructie die bestaat uit het vaste hulpwerkwoord *hebben*, een modaal werkwoord en een open plaats voor een hoofdwkwoord (*hebben kunnen* infinitief). In die analyse kan *hebben* met elk hoofdwkwoord gecombineerd worden, ongeacht het type van dat hoofdwkwoord (Coussé en Van de Velde 2014:362).

2.5.3 De status van het groepsvormende werkwoord

In deze sectie licht ik de analyse van hulpwerkwoordvariatie in Mod- en Asp-clusters binnen de formele benaderingen van Cardinaletti en Shlonsky (2004) en Pots (2020) toe. Daarbij blijkt de functionele, lexicale of semi-lexicale status van het groepsvormende werkwoord de mogelijkheid tot hulpwerkwoordvariatie te verklaren. Ik bespreek eerst de status van de modale werkwoorden (§2.5.3.1), daarna die van de aspectuele werkwoorden (§2.5.3.2).

2.5.3.1 Modale werkwoorden

Hulpwerkwoordvariatie in Mod-clusters wordt door Cardinaletti en Shlonsky (2004) en Pots (2020) in verband gebracht met het verschil tussen functionele en lexicale groepsvormende werkwoorden. Om te bepalen of een groepsvormend werkwoord functioneel of lexicaal is, baseert Pots zich op de tests van Wurmbrand (2001), die de morfosyntactische verschillen tussen de twee types werkwoorden voor het Duits beschrijft. Ten eerste kunnen functionele groepsvormende werkwoorden gecombineerd worden met een loos onderwerp, zoals geïllustreerd in (48a), maar lexicale niet, zoals in (48b). Ten tweede is extrapositie van de ingebedde infinitief enkel mogelijk bij lexicale groepsvormende werkwoorden (49b), niet bij functionele (49a) (Wurmbrand 2001:156-171, Pots 2020:22-24).¹¹

- (48) a. **Es muß** schneien.
Het moet sneeuwen.
b. ***Es versuchte** zu schneien.
Het probeert te sneeuwen.

(Wurmbrand 2001:167-168)

¹¹ Wurmbrand (2001) beschrijft ook dat functionele groepsvormende werkwoorden het IPP-effect ondergaan en lexicale niet. Pots (2020:27) laat echter zien dat dat morfosyntactische verschil minder relevant is voor het Nederlands dan voor het Duits, aangezien IPP veel meer voorkomt in het Nederlands. Ik laat het in de rest van de bespreking daarom buiten beschouwing.

- (49) a. *dass Hans **muß** [den Kuchen **essen**].
 dat Hans moet [de koek eten].
 b. dass Hans **versuchte** [den Kuchen zu **essen**].
 dat Hans probeert [de koek te eten]. (Wurmbrand 2001:156-157)

Volgens Pots (2020:86) staan de functionele groepsvormende werkwoorden in het Nederlands hulpwerkwoordvariatie toe, maar de lexicale niet.¹² In clusters met een modaal werkwoord zoals *moeten* in (50) is zowel de vervoeging met *hebben* (50a) als met *zijn* (50b) mogelijk. In een cluster met *beloven* zoals in (51) is echter enkel *hebben* grammaticaal (51a), *zijn* niet (51b). *Moeten* is een functioneel groepsvormend werkwoord, *beloven* een lexicaal (Pots 2020:22, 86).

- (50) a. dat ik naar huis **heb** moeten gaan.
 b. dat ik naar huis **ben** moeten gaan. (Pots 2020:86)
- (51) a. dat ik **heb** beloofd naar huis te gaan.
 b. *dat ik **ben** beloofd naar huis te gaan. (Pots 2020:86)

De kenmerken van de functionele en lexicale groepsvormende werkwoorden worden samengevat in Tabel 2.1 (gebaseerd op Pots (2020:24)).

Tabel 2.1: Kenmerken van functionele en lexicale groepsvormende werkwoorden

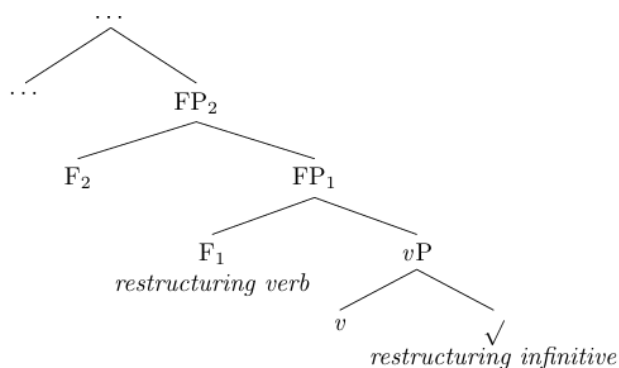
Groepsvormend werkwoord	Loos onderwerp	Extrapositie	Hulpwerkwoordvariatie
Functioneel	JA	NEE	JA
Lexicaal	NEE	JA	NEE

Groepsvorming met een functioneel werkwoord heeft een andere syntactische structuur dan groepsvorming met een lexicaal werkwoord. Die structuur verklaart of er hulpwerkwoordvariatie mogelijk is of niet. De structurele analyse die Pots (2020:21) aanneemt voor functionele groepsvorming wordt weergegeven in (52). In (52) bevindt het groepsvormende werkwoord (*restructuring verb*) zich in een functioneel hoofd (F) in het functionele domein van het hoofdwkwoord (*restructuring infinitive*).¹³ Het functionele groepsvormende werkwoord is dus de uitspelling van syntactische kenmerken (Pots 2020:21). Volgens Cardinaletti en Shlonsky (2004:521-522) is hulpwerkwoordvariatie bij functionele groepsvorming mogelijk omdat functionele groepsvormende werkwoorden “transparant” kunnen zijn voor hulpwerkwoordselectie. Het hulpwerkwoord wordt dan geselecteerd door het lagere, lexicale hoofdwkwoord en niet door het hogere, functionele werkwoord.

¹² Hulpwerkwoordvariatie in Mod-clusters komt niet voor in het Duits en werd dus niet beschreven door Wurmbrand (2001) (Pots 2020:22-24).

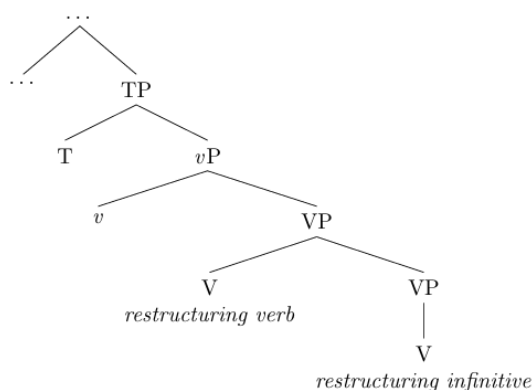
¹³ Ik bespreek enkel het verschil in de plaats van het groepsvormende werkwoord in de structuren omwille van het belang voor de analyse van hulpwerkwoordvariatie; voor de andere verschillen verwijs ik naar het werk van Pots (2020).

(52)



De structuur die Pots (2020:21) aanneemt voor lexicale groepsvorming is weergegeven in (53). Het groepsvormende werkwoord is in (53) een lexicaal werkwoord dat een eigen lexicale semantiek uitdrukt. Het wordt dus niet gegenereerd in een functioneel hoofd, maar is het hoofd van zijn eigen verbale projectie (VP), net als het hoofdwerkwoord. Bij lexicale groepsvorming is hulpwerkwoordvariatie ongrammaticaal: lexicale groepsvormende werkwoorden kunnen niet transparant zijn voor hulpwerkwoordselectie. Het hulpwerkwoord van voltooid aspect wordt dan steeds geselecteerd door het hoogste lexicale werkwoord (Cardinaletti & Shlonsky 2004:521-522, Pots 2020:86).

(53)



De mogelijkheid tot hulpwerkwoordvariatie in de voorbeelden in (50) en (51) wordt zo verklaard door de functionele of lexicale status van het groepsvormende werkwoord en de mate waarin het daardoor transparant is voor hulpwerkwoordselectie. In (50) is *moeten* het hogere, functionele groepsvormende werkwoord en *gaan* het lagere, lexicale hoofdwerkwoord. In (50a) wordt *hebben* geselecteerd door het H-werkwoord *moeten*; in (50b) is het functionele *moeten* transparant voor hulpwerkwoordselectie en wordt *zijn* geselecteerd door het Z-werkwoord *gaan*. In (51) daarentegen is *beloven* het hogere, lexicale groepsvormende werkwoord en *gaan* het lagere, lexicale hoofdwerkwoord. Aangezien het om lexicale groepsvorming gaat, moet het hulpwerkwoord van voltooid aspect geselecteerd worden door het hoogste werkwoord: het H-werkwoord *beloven* selecteert *hebben* (51a); *zijn* is ongrammaticaal (51b).

2.5.3.2 Aspectuele werkwoorden

Een formele analyse voor hulpwerkwoordvariatie in Asp-clusters in het Nederlands zoals die voor Mod-clusters (§2.5.3.1) ontbreekt in de literatuur. Cardinaletti en Shlonsky (2004) en Wurmbrand (2001) bespreken wel de groepsvormende werkwoorden *andare* ‘gaan’ voor het Italiaans en *gehen* ‘gaan’ en *kommen* ‘komen’ voor het Duits. Voor die werkwoorden blijkt de binaire opdeling tussen functionele en lexicale groepsvorming niet voldoende: ze behoren tot een derde klasse, tussen de functionele en

Cardinaletti en Shlonsky (2004:537) nemen een derde klasse aan die ze “quasi-functioneel” noemen, waartoe het Italiaanse werkwoord *andare* ‘gaan’ behoort. Ze tonen ten eerste aan dat *andare* in (54) een functioneel groepsvormend werkwoord is. Dat doen ze aan de hand van het fenomeen *clitic climbing*; daar ga ik hier verder niet op in, zie Cardinaletti en Shlonsky (2004:537). Hoewel functionele werkwoorden in het Italiaans hulpwerkwoordvariatie toestaan, staat *andare* die niet toe. In (54) is *andare* een Z-werkwoord: het selecteert *essere* ‘zijn’. Het hoofdwkwoord *trovare* ‘bezoeken’ selecteert *avere* ‘hebben’. Het enige grammaticale hulpwerkwoord in (54) is *zijn*; *hebben* is ongrammaticaal.

- Quasi-functionele werkwoorden gedragen zich volgens Cardinaletti en Shlonsky (2004:537) dus als functionele werkwoorden, maar net als lexicale werkwoorden zijn ze niet transparant voor hulpwerkwoordselectie. Ze bepalen zelf het hulpwerkwoord van voltooid aspect in een Asp-cluster: in (54) wordt *zijn* geselecteerd door het groepsvormende werkwoord *andare*, niet door het lagere, lexicale werkwoord *trovare*. Samenvattend stellen Cardinaletti en Shlonsky (2004:537-539) dat in het Italiaans het hulpwerkwoord van voltooid aspect steeds bepaald wordt door het hoogste werkwoord dat zelf een hulpwerkwoord kan selecteren. De functionele groepsvormende werkwoorden, zoals de modale werkwoorden, kunnen dat niet; de lexicale en quasi-functionele, zoals *andare*, wel.

(56) a. *dass Hans **geht** [den Kuchen **essen**].
 dat Hans gaat [de koek eten].
 b. *dass Hans **komt** [den Kuchen **essen**].
 dat Hans komt [de koek eten]. (Wurmbrand 2001:158)

35

Tabel 2.2: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende werkwoorden in het Duits

Groepsvormend werkwoord	Loos onderwerp	Extrapositie
Functioneel	JA	NEE
Semi-lexicaal	JA	NEE
Lexicaal	NEE	JA

Het onderzoek van Pots (2020) gaat niet specifiek in op aspectuele werkwoorden, maar is relevant omdat het inzicht brengt in de klasse van de semi-lexicale groepsvormende werkwoorden in het Nederlands. Pots (2020:84-87) brengt de morfosyntactische kenmerken van de twee semi-lexicale groepsvormende werkwoorden *hoeven* en *zitten* in kaart aan de hand van de tests van Wurmbrand (2001). Het semi-lexicale *hoeven* staat een loos onderwerp toe (57a), maar *zitten* niet (57b). *Hoeven* sluit daarmee aan bij de functionele werkwoorden, *zitten* bij de lexicale. Geen van beide werkwoorden laat extrapositie van de ingebedde infinitief toe, zoals geïllustreerd in (58), wat overeenkomt met de functionele groepsvormende werkwoorden. Tot slot staat *hoeven* hulpwerkwoordvariatie toe (59a), *zitten* niet (59b).¹⁴

- (57) a. **Het hoeft** niet te sneeuwen.
b. ***Het zit** niet te sneeuwen. (Pots 2020:84-85)
- (58) a. *dat Frans niet **hoeft** [de koek te **eten**].
b. *dat Frans **zit** [de koek te **eten**]. (Pots 2020:84)
- (59) a. dat ik niet naar huis **heb/ben hoeven** gaan.
b. dat ik naar huis ??**heb/*ben zitten** gaan. (Pots 2020:86-87)

De kenmerken van functionele, semi-lexicale en lexicale groepsvormende werkwoorden voor het Nederlands worden weergegeven in Tabel 2.3. Die toont dat van de twee onderzochte semi-lexicale groepsvormende werkwoorden, *hoeven* functioneler is dan *zitten*. *Hoeven* komt in de drie besproken kenmerken zelfs overeen met de functionele groepsvormende werkwoorden. Aangezien *hoeven* een infinitief met *te* als aanvulling selecteert, is het volgens Pots (2020:87) echter nog steeds semi-lexicaal.

Tabel 2.3: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende werkwoorden in het Nederlands

Groepsvormend werkwoord	Loos onderwerp	Extrapositie	Hulpwerkwoordvariatie
Functioneel	JA	NEE	JA
Semi-lexicaal	Hoeven	JA	NEE
	Zitten	NEE	NEE
Lexicaal	NEE	JA	NEE

¹⁴ Pots (2020:86) vermeldt dat het voor *zitten* moeilijk te testen is of hulpwerkwoordvariatie mogelijk is, aangezien *zitten* meestal geen Z-werkwoord als *gaan* selecteert bij groepsvorming. Ze kan niet definitief vaststellen of de onaanvaardbaarheid van de *zijn*-variant te wijten is aan de hulpwerkwoordvariatie of aan de aard van het complement.

(60)

$$\begin{array}{c}
 vP \\
 \swarrow \quad \searrow \\
 \checkmark \quad \quad vP \\
 \text{semi-lexical} \quad \swarrow \quad \searrow \\
 \quad \quad v \quad \quad \checkmark \\
 \quad \quad \quad \text{lexical}
 \end{array}$$

```

graph TD
    FP --> F1[F]
    FP --> vP[vP]
    F1 --> F2[F]
    F1 --> c1[✓]
    F2 --- s[semi-lexicaal]
    vP --> v[v]
    vP --> c2[✓]
    v --- l[lexicaal]
  
```

2.5.3.3 Conclusie: de status van het groepsvormende werkwoord

Dergelijke analyse voor de aspectuele werkwoorden in het Nederlands ontbreekt. De analyses voor *gaan* en *komen* in het Duits (Wurmbbrand 2001) en *gaan* in het Italiaans (Cardinaletti & Shlonsky 2004) suggereren echter dat de aspectuele werkwoorden een semi-lexicale status hebben, tussen de functionele en lexicale groepsvormende werkwoorden in. Pots (2020) beargumenteert dat semi-lexicale werkwoorden in het Nederlands zich in verschillende fases van een grammaticalisatieproces kunnen bevinden. *Zitten* zit in de eerste fase en staat geen hulpwerkwoordvariatie toe, *hoeven* is in transitie van de eerste naar de tweede fase en staat wel variatie tussen *hebben* en *zijn* toe.

37

2.6 Conclusie

In dit hoofdstuk kwam de onderzoeksliteratuur over hulpwerkwoordselectie in drieledige werkwoordsclusters met modale (Mod-clusters) en aspectuele werkwoorden (Asp-clusters) aan bod. Daarbij werd aandacht besteed aan zowel de beschrijving als de analyse van hulpwerkwoordvariatie.

In drieledige werkwoordsclusters wordt de keuze van het hulpwerkwoord van voltooid aspect meestal bepaald door het groepsvormende werkwoord. Het modale werkwoord van een Mod-cluster selecteert *hebben* als hulpwerkwoord; het aspectuele werkwoord van een Asp-cluster selecteert *zijn* (Haeseryn et al. 1997:80). Er kan echter hulpwerkwoordvariatie optreden wanneer het groepsvormende werkwoord en het hoofdwkwoord een verschillende selectievoorkeur hebben. Als het hoofdwkwoord van een Mod-cluster *zijn* selecteert, kan *zijn* als hulpwerkwoord voorkomen; in een Asp-cluster kan *hebben* voorkomen als het hoofdwkwoord *hebben* selecteert. Hulpwerkwoordvariatie in Asp- en Mod-clusters heeft echter niet dezelfde omvang. In Asp-clusters komt de alternatieve variant *hebben* veel minder voor dan de alternatieve variant *zijn* in Mod-clusters (Hofmans 1981, Van Eynde et al. 2016, Dhont 2024).

De variatie tussen *hebben* en *zijn* vertoont zowel bij Mod- als bij Asp-clusters geografische variatie. Uitgebreid enquête- en corpusonderzoek voor Mod-clusters (Hofmans 1981, Van Eynde et al. 2016, Dhont 2024) toont aan dat de alternatieve variant *zijn* in Vlaanderen meer voorkomt dan in Nederland. In Nederland bestaat een voorkeur voor *hebben*, die het sterkst is in het noorden van het land. Het beperkte onderzoek voor Asp-clusters (Hofmans 1985, Van Eynde et al. 2016, Dhont 2024) suggereert dat in Asp-clusters ook de alternatieve variant, *hebben*, meer voorkomt in Vlaanderen; in Nederland wordt geen hulpwerkwoordvariatie geattesteerd. Tot slot onderscheidt West-Vlaanderen zich zowel in Mod- als in Asp-clusters van de andere Vlaamse provincies door een sterkere voorkeur voor *hebben*.

Naast de geografische regio blijken in Mod-clusters ook enkele taalinterne factoren een invloed te hebben op de hulpwerkwoordselectie. Zo oefenen sommige hoofdwkwoorden een sterkere invloed uit op de selectie van *hebben* of *zijn* in het Mod-cluster dan andere hoofdwkwoorden. Hetzelfde geldt voor de modale werkwoorden: *moeten* en *kunnen* komen vaker voor met *zijn*, maar *mogen* en *willen* vaker met *hebben*. Daarnaast komt het hulpwerkwoord *zijn* vaker voor wanneer er een afstand is tussen het hulpwerkwoord en het modale werkwoord, en heeft een irrealislezing een positieve invloed op de selectie van *hebben* (Hofmans 1981, Van Eynde et al. 2016, Dhont 2024). Of die taalinterne factoren ook de hulpwerkwoordselectie in Asp-clusters beïnvloeden, werd niet onderzocht.

De hulpwerkwoordvariatie in Mod- en Asp-clusters wordt op verschillende manieren verklaard. Volgens Hofmans (1982a, 1985), Draye en Van der Horst (2006) en Van der Horst (2008) bevinden zowel de modale als de aspectuele werkwoorden zich in een grammaticalisatieproces waarbij ze het vermogen verliezen om de keuze van het hulpwerkwoord van voltooid aspect te bepalen. Dat vermogen wordt dan steeds meer overgedragen aan het hoofdwkwoord. Coussé en Van de Velde (2014) vestigen de aandacht op de rol van irrealis in de analyse van hulpwerkwoordselectie. Binnen formele analyses (Cardinaletti & Shlonsky 2004, Pots 2020) wordt hulpwerkwoordvariatie verklaard door de verschillende status van de groepsvormende werkwoorden: functionele groepsvormende werkwoorden, zoals modale werkwoorden, laten variatie toe, lexicale niet, semi-lexicale werkwoorden bevinden zich daartussenin.

Het literatuuroverzicht toont dat de factoren die hulpwerkwoordselectie in Asp-clusters beïnvloeden nog niet uitgebreid in kaart zijn gebracht. Deze masterproef richt zich dan ook op dat vraagstuk, zoals in het volgende hoofdstuk wordt beschreven.

3 Onderzoeksvragen en hypothesen

In dit hoofdstuk licht ik de focus van het onderzoek in deze masterproef toe. Eerst worden de onderzoeksvragen besproken (§3.1), vervolgens de hypothesen die bij de onderzoeksvragen horen (§3.2).

3.1 Onderzoeksvragen

Uit het literatuuronderzoek (§2) blijkt dat er in clusters met een aspectueel werkwoord (Asp-clusters) hulpwerkwoordvariatie bestaat, net als in clusters met een modaal werkwoord (Mod-clusters). In de onderzoeksliteratuur bestaat er echter nog geen uitgebreide evidentie voor die variatie (Hofmans 1982b, Draye & Van der Horst 2006, Van Eynde et al. 2016, Dhont 2024). Ik beoog met dit onderzoek de hulpwerkwoordvariatie in Asp-clusters in kaart te brengen en zo de factoren te identificeren die een invloed uitoefenen op hulpwerkwoordselectie in Asp-clusters. Op basis van die factoren stel ik een analyse voor van hulpwerkwoordvariatie in Asp-clusters. De globale onderzoeksvraag (OV) waarop ik in deze masterproef een antwoord zoek is op die manier:

OV Wat bepaalt de selectie van het hulpwerkwoord van voltooid aspect in Asp-clusters?

Om een antwoord op de globale onderzoeksvraag te formuleren wordt de invloed van vijf taalinterne en -externe factoren op hulpwerkwoordselectie in Asp-clusters onderzocht. De selectie van de taalinterne factoren is gebaseerd op de factoren die in het onderzoek naar Mod-clusters belangrijk bleken voor hulpwerkwoordvariatie (zie §2.3.3). Ik focus me op het hoofdwkwoord, afstand en modaliteit (irrealis vs. realis). De invloed van de aard van het aspectuele werkwoord wordt in deze masterproef niet specifiek onderzocht. De deelvragen voor de onderzochte taalinterne factoren zijn:

OV1 Welke invloed heeft het hoofdwkwoord op hulpwerkwoordselectie in Asp-clusters?

OV2 Welke invloed heeft de afstand tussen het hulpwerkwoord van voltooid aspect en het aspectuele werkwoord op hulpwerkwoordselectie in Asp-clusters?

OV3 Welke invloed heeft irrealis op hulpwerkwoordselectie in Asp-clusters?

De taalexterne factoren die in deze masterproef worden onderzocht zijn geografische regio en register. De geografische regio blijkt uit het onderzoek van Hofmans (1982b) en Dhont (2024) erg van belang, zowel wat nationale als regionale variatie betreft. De invloed van register werd nog niet onderzocht, maar de factor wordt opgenomen in het onderzoek op basis van de veronderstelling van Hofmans (1985:75) en Van der Horst (2005:102) over registervariatie. De deelvragen voor de taalexterne factoren zijn:

OV4 Vertoont hulpwerkwoordselectie in Asp-clusters geografische variatie?

OV5 Vertoont hulpwerkwoordselectie in Asp-clusters registervariatie?

3.2 Hypotheses

Op basis van de onderzoeksvragen formuleer ik zes hypothesen. De basishypothese (§3.2.1) gaat over de verhouding tussen *zijn* en *hebben* in Asp-clusters. De andere vijf hypothesen horen specifiek bij een van de deelonderzoeksvragen: de hoofdwerkwoordhypothese bij OV1 (§3.2.2), de afstandshypothese bij OV2 (§3.2.3), de irrealishypothese bij OV3 (§3.2.4), de geografische hypothese bij OV4 (§3.2.5) en de registerhypothese bij OV5 (§3.2.6). Ik licht hieronder de hypothesen toe, die telkens gebaseerd zijn op het onderzoek naar hulpwerkwoordvariatie in zowel Asp-als Mod-clusters zoals besproken in §2.

3.2.1 Basishypothese

Ik neem op basis van §2.4.1 aan dat hulpwerkwoordvariatie in Asp-clusters enkel optreedt wanneer het hoofdwerkwoord van het cluster een H-werkwoord is en dus *hebben* selecteert. De variatie wordt dan ook enkel onderzocht in clusters zoals in (61), die bestaan uit een hulpwerkwoord van voltooid aspect (*hebben* of *zijn*), een aspectueel werkwoord zoals *beginnen*, en een H-hoofdwerkwoord zoals *zeggen*.

- | | | |
|------|--|-------------------------|
| (61) | a. Ze is dat al beginnen zeggen tegen haar man. | $AUX_Z - ASP_Z - HWW_H$ |
| | b. Ze heeft dat al beginnen zeggen tegen haar man. | $AUX_H - ASP_Z - HWW_H$ |

Ik neem aan dat het hulpwerkwoord van voltooid aspect in Asp-clusters als (61) meestal geselecteerd wordt door het groepsvormende werkwoord (zie §2.2.2). Aangezien de aspectuele werkwoorden *gaan*, *blijven*, *komen* en *beginnen* Z-werkwoorden zijn, is het hulpwerkwoord van voltooid aspect dus in de meerderheid van de gevallen *zijn*, zoals in (61a). In §2.4 wordt aangetoond dat het hulpwerkwoord van voltooid aspect ook *hebben* kan zijn en dus geselecteerd kan worden door het hoofdwerkwoord, zoals in (61b). De *hebben*-variant komt echter slechts sporadisch voor in het corpusonderzoek van Van Eynde et al. (2016) en Dhont (2024) en krijgt erg weinig voorkeur in het enquêteonderzoek van Hofmans (1982b). Op basis daarvan formuleer ik de basishypothese voor hulpwerkwoordselectie in Asp-clusters:

Basishypothese

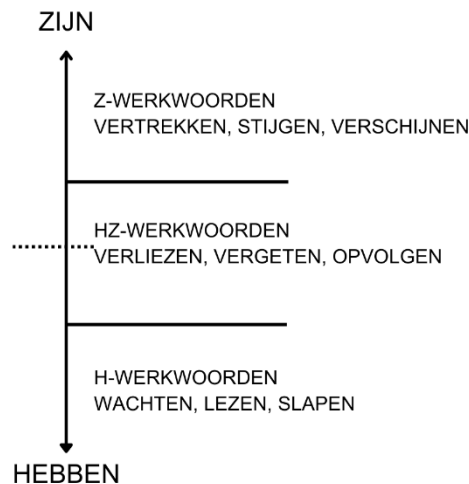
Zijn komt frequenter dan *hebben* voor als hulpwerkwoord van voltooid aspect in Asp-clusters.

3.2.2 Hoofdwerkwoordhypothese

Ik verwacht naar analogie met de resultaten van Hofmans (1981) en Van Eynde et al. (2016) voor Mod-clusters dat de aard van het hoofdwerkwoord een invloed heeft op de selectie van *hebben* of *zijn* als hulpwerkwoord van voltooid aspect in Asp-clusters (OV1). Net als bij de Z-hoofdwerkwoorden in Mod-clusters (zie §2.3.3.2) verwacht ik dat de H-hoofdwerkwoorden in verschillende types kunnen worden ondergebracht: H-werkwoorden met een voorkeur voor *hebben* in Asp-clusters, (“pro-*hebben*” in de terminologie van Hofmans (1981)), H-werkwoorden met een voorkeur voor *zijn* in Asp-clusters, en neutrale H-werkwoorden die geen sterke voorkeur voor *hebben* of *zijn* vertonen in Asp-clusters. Over welke H-werkwoorden tot welk type behoren, is echter geen informatie beschikbaar.

Om de invloed van het type hoofdwerkwoord te meten in Asp-clusters, vergelijk ik de invloed van H-hoofdwerkwoorden met die van HZ-hoofdwerkwoorden. H-werkwoorden zijn werkwoorden die steeds *hebben* selecteren, zoals *wachten*, *lezen* en *slapen*; de kleine groep van de HZ-werkwoorden bestaat uit werkwoorden die zowel *hebben* als *zijn* kunnen combineren, zoals *verliezen*, *vergeten* en *opvolgen* (zie §2.2.1). Ik neem daarbij aan dat HZ-werkwoorden geen sterke voorkeur voor een bepaald hulpwerkwoord hebben, omdat ze beide hulpwerkwoorden kunnen selecteren. De H-werkwoorden kunnen enkel

hebben selecteren en hebben dus volgens mij een sterkere voorkeur voor *hebben* dan de HZ-werkwoorden. Dat idee wordt gevisualiseerd in Figuur 3.1, naar analogie met de visualisatie van Hofmans (1981:86), zie §2.3.3.2. Figuur 3.1 toont aan dat ik verwacht dat in Asp-clusters de Z-hoofdwerkwoorden de sterkste voorkeur hebben voor *zijn*: hulpwerkwoordvariatie in Asp-clusters met Z-werkwoorden is niet mogelijk (zie §2.2.2). Ik verwacht een sterkere voorkeur voor *hebben* bij de H-werkwoorden dan bij de HZ-werkwoorden, omdat die enkel *hebben* kunnen selecteren.



Figuur 3.1: Verwachte selectievoorkeur van hoofdwerkwoorden in Asp-clusters

Ik formuleer de hypothese voor de invloed van het type hoofdwerkwoord op de hulpwerkwoordselectie in Asp-clusters als volgt:

Hoofdwerkwoordhypothese

De aanwezigheid van een H-hoofdwerkwoord heeft een sterkere positieve invloed op de selectie van *hebben* in Asp-clusters dan de aanwezigheid van een HZ-hoofdwerkwoord.

De hypothese wordt geïllustreerd aan de hand van (62). In (62a) is het hoofdwerkwoord van het Asp-cluster het H-werkwoord *slapen*, in (62b) is het hoofdwerkwoord het HZ-werkwoord *opvolgen* (zie §2.2.1). Volgens de hypothese komt het hulpwerkwoord *hebben* dan gemakkelijker voor in (62a) dan in (62b). Merk op dat dat niet betekent dat de aanwezigheid van een H-hoofdwerkwoord de selectie van *hebben* afdwingt: *zijn* kan nog steeds voorkomen als hulpwerkwoord. De hypothese is enkel dat de aanwezigheid van een H-werkwoord een grotere positieve invloed heeft op de selectie van *hebben* dan de aanwezigheid van een HZ-werkwoord, die minder invloed op de hulpwerkwoordselectie uitoefent.

- | | | |
|------|--|--------------------------------|
| (62) | a. Klaas is/heeft op tijd gaan slapen . | $AUX_{Z/H} - ASP_Z - HWW_H$ |
| | b. Klaas is/heeft zijn vader gaan opvolgen . | $AUX_{Z/H} - ASP_Z - HWW_{HZ}$ |

3.2.3 Afstandshypothese

Wat de invloed van afstand (OV2) betreft verwacht ik op basis van Hofmans (1981) dat *hebben* meer voorkomt in Asp-clusters wanneer er afstand is tussen het hulpwerkwoord en het aspectuele werkwoord (zie §2.3.3.3). Het aspectuele werkwoord kan dan immers minder direct een invloed uitoefenen op de selectie van het hulpwerkwoord. Ik maak wat de invloed van afstand betreft geen onderscheid tussen

hoofd- en bijzinnen (zie §2.3.3.3), maar neem de hypothese van Hofmans (1981:103) aan dat de voorkeur voor *hebben* groter wordt bij een grotere afstand. Ik formuleer de afstandshypothese als volgt:

Afstandshypothese

Een grotere afstand tussen het hulpwerkwoord van voltooid aspect en het aspectuele werkwoord heeft een positieve invloed op de selectie van *hebben* in Asp-clusters.

Ik illustreer de hypothese aan de hand van het voorbeeld in (63). In (63a) zijn het hulpwerkwoord en het aspectuele werkwoord adjacent, dus dan komt *hebben* moeilijker voor. In (63b) staat er een woord tussen de werkwoorden en komt *hebben* dus gemakkelijker voor. In (63c) is de afstand nog groter, want er staan drie woorden tussen het hulpwerkwoord en het aspectuele werkwoord. Dan komt *hebben* nog gemakkelijker voor, want dan heeft de selectievoordeur van het aspectuele *blijven* nog minder invloed.

- | | | |
|------|---|--|
| (63) | a. Isaak is/heeft blijven slapen. | AUX _{Z/H} – ASP _Z – HWW _H |
| | b. Isaak is/heeft gisteren blijven slapen. | AUX _{Z/H} – ASP _Z – HWW _H |
| | c. Isaak is/heeft gisteren bij ons blijven slapen. | AUX _{Z/H} – ASP _Z – HWW _H |

3.2.4 Irrealishypothese

De basisaanname voor de invloed van irrealis (OV3) op hulpwerkwoordselectie in Asp-clusters is dat een hoofdwkwoord dat *zijn* selecteert bij een irrealislezing ook *hebben* kan selecteren (zie §2.3.3.5). De irrealislezing heeft dan een positieve invloed op de selectie van *hebben* bij Z-werkwoorden in voltooid aspect. In (64) fungeert het werkwoord *komen* als hoofdwkwoord. *Komen* selecteert *zijn* in realislezing, *hebben* is ongrammaticaal (64a). In (64b) komt het Z-werkwoord *komen* echter voor in irrealislezing. Dan zijn zowel *hebben* als *zijn* mogelijk als hulpwerkwoord van voltooid aspect.

- | | | | |
|------|---|----------|---------------------------------------|
| (64) | a. Axelle is/*heeft niet naar huis gekomen. | REALIS | AUX _Z – HWW _Z |
| | b. Was/had dan toch naar huis gekomen! | IRREALIS | AUX _{Z/H} – HWW _Z |

Als die basisaanname klopt, verwacht ik dat ook in Asp-clusters *hebben* vaker wordt geselecteerd bij irrealis dan bij realis, net zoals verwacht wordt voor Mod-clusters (zie §2.3.3.5). In (65) fungeert het Z-werkwoord *komen* als groepsvormend werkwoord in plaats van als hoofdwkwoord zoals in (64). In (65a) staat het Asp-cluster in realis, in (65b) in irrealis. De selectie van *hebben* krijgt vermoedelijk meer voorkeur in (65b) dan in (65a), aangezien *komen* in irrealislezing wel *hebben* kan selecteren, en in realislezing niet (zie (64)).

- | | | | |
|------|---|----------|--|
| (65) | a. Axelle is/heeft niets meer komen zeggen. | REALIS | AUX _{Z/H} – ASP _Z – HWW _H |
| | b. Was/had dan toch iets komen zeggen! | IRREALIS | AUX _{Z/H} – ASP _Z – HWW _H |

Op basis daarvan formuleer ik de basishypothese als volgt:

Irrealishypothese

Een irrealislezing heeft een positieve invloed op de selectie van *hebben* als hulpwerkwoord van voltooid aspect in Asp-clusters.

3.2.5 Geografische hypothese

Wat de geografische variatie (OV4) betreft verwacht ik op basis van het onderzoek van Hofmans (1982b) en Dhont (2024) dat hulpwerkwoordvariatie in Asp-clusters enkel voorkomt in Vlaanderen (zie §2.4.2). In Nederland is dan enkel *zijn* mogelijk als hulpwerkwoord van voltooid aspect in Asp-clusters, terwijl in Vlaanderen zowel *zijn* als *hebben* mogelijk is. Daarnaast verwacht ik dat er binnen Vlaanderen ook regionale variatie optreedt, met name dat *hebben* in West-Vlaanderen meer voorkomt dan in de andere Vlaamse provincies (zie §2.4.2). Ik formuleer de geografische hypothese als volgt:

Geografische hypothese

In Asp-clusters komt *hebben* als hulpwerkwoord van voltooid aspect frequenter voor in Vlaanderen dan in Nederland. Binnen Vlaanderen komt *hebben* frequenter voor in West-Vlaanderen dan in de andere Vlaamse provincies.

3.2.6 Registerhypothese

Naar registervariatie (OV5) bij hulpwerkwoordselectie in Asp-of Mod-clusters is nog geen onderzoek gevoerd. Volgens zowel Hofmans (1985:75) als Van der Horst (2005:102) wordt *hebben* in Asp-clusters echter meer uitgesproken dan neergeschreven, en komt het vooral voor in informeel taalgebruik (zie §2.4.2). Op basis daarvan formuleer ik de registerhypothese:

Registerhypothese

In Asp-clusters komt *hebben* als hulpwerkwoord van voltooid aspect frequenter voor in gesproken en informeel taalgebruik dan in geschreven en formeel taalgebruik.

3.3 Conclusie

In deze masterproef onderzoek ik welke factoren de hulpwerkwoordselectie in Asp-clusters bepalen, en zo beoog ik in kaart te brengen wat ervoor zorgt dat de alternatieve *hebben*-variant in Asp-clusters geselecteerd wordt. Daarvoor worden taalinterne factoren bekeken die een mogelijke positieve invloed hebben op de selectie van *hebben*, met name de aanwezigheid van een H-hoofdwerkwoord, een grote afstand tussen het hulpwerkwoord en het aspectuele werkwoord, en een irrealisreading. Daarnaast onderzoek ik of *hebben* meer voorkomt in Vlaanderen en in informeel, gesproken taalgebruik.

De hypothesen worden in wat volgt getoetst aan de hand van twee onderzoeksmethodes, namelijk een experiment en een corpusonderzoek. In het experimentele onderzoek (§4) komen de basis-, hoofdwerkwoord-, afstands-, irrealis- en geografische hypothese aan bod. Aangezien in het experiment hulpwerkwoordselectie in geëliciteerd, geschreven taalgebruik wordt onderzocht, wordt de registerhypothese daar niet getoetst. In het corpusonderzoek (§5) ga ik wel in op registervariatie. Daarnaast worden daarin ook - zij het in minder detail dan in het experiment - de andere factoren behandeld. Ik breng tot slot de resultaten van het experiment en het corpusonderzoek samen in §6, waar ik terugkeer naar de onderzoeksvragen en op basis daarvan een analyse voorstel.

4 Experimenteel onderzoek

In dit hoofdstuk beschrijf ik het experimentele onderzoek naar hulpwerkwoordselectie in Asp-clusters. Op basis van de acceptabiliteitsoordelen van 1105 Vlaamse en Nederlandse respondenten onderzocht ik met lineaire regressieanalyses de invloed van het type hoofdwerkwoord, afstand, modaliteit en geografische regio op hulpwerkwoordselectie in Asp-clusters. Ik licht eerst de methodologie van het experiment toe (§4.1), en vervolgens de resultaten (§4.2). Ik sluit af met een conclusie in §4.3.

4.1 Methodologie

Deze methodologische sectie vangt aan met een beschrijving van de experimentele stimuli (§4.1.1). Vervolgens bespreek ik de respondenten die aan het experiment deelnamen (§4.1.2), en de taak die ze uitvoerden (§4.1.3). In §4.1.4 licht ik de procedure van het experiment toe, gevolgd door analysemethode (§4.1.5). In §4.1.6 vat ik de methodologie van het experiment samen.

4.1.1 Stimuli

De stimuli van het experiment omvatten zowel kritische stimuli als fillers. De kritische stimuli zijn in het experimentele design verdeeld over vijf deelexperimenten op basis van de onderzochte factoren, maar ze hebben een gelijkaardige structuur. In §4.1.1.1 behandel ik de algemene opbouw van de kritische stimuli, en vervolgens ga ik specifiek in op de stimuli per deelexperiment (§4.1.1.2). De fillers komen aan bod in §4.1.1.3, en §4.1.1.4 bevat tot slot een schematisch overzicht van de stimuli.

4.1.1.1 Opbouw van de kritische stimuli

Hoewel de deelexperimenten de invloed van verschillende factoren onderzoeken, hebben de stimuli uit de deelexperimenten een gelijkaardige opbouw en bevatten ze dezelfde selectie aan werkwoorden. Ik beschrijf eerst de algemene structuur van de stimuli (§4.1.1.1.1). Daarna licht ik toe welke onderzoeksvragen en hypothesen in ieder deelexperiment behandeld worden (§4.1.1.1.2). Vervolgens bespreek ik de selectie van de hoofdwerkwoorden (§4.1.1.1.3) en de selectie van de groepsvormende werkwoorden (§4.1.1.1.4).

4.1.1.1.1 Algemene structuur

Elk deelexperiment bevat zinnen die telkens bestaan uit een gemanipuleerde kritische zin (κ) en een contextzin (c), zoals de voorbeelden in (66). Het hulpwerkwoord van voltooid aspect, dat de focus vormt van het onderzoek, staat steeds in de kritische zin. De contextzin wordt toegevoegd zodat de respondenten een betekenisvolle en natuurlijke zin beoordelen waaruit niet meteen af te leiden is waar het onderzoek over gaat. In de contextzin vermijd ik systematisch de werkwoorden *hebben* en *zijn* om *priming* te voorkomen: de aanwezigheid van *hebben* of *zijn* in de contextzin zou het oordeel over het hulpwerkwoord in de kritische zin kunnen beïnvloeden.

- (66)
- | | |
|--|----------------|
| a. [κ De Argentijnen zijn blijven winnen] en [c kwalificeerden zich zo voor het WK]. | E1_1_A0_Z |
| b. [κ Michelle heeft snel moeten trouwen], want [c ze bleek zwanger]. | E2B_2_tro_Z_H |
| c. [κ Had de spion haar blijven volgen], [c kende hij nu haar geheim]. | E4_1_irr_inv_H |

Elk deelexperiment omvat een verschillend aantal zinnen, waarbij elke zin in meerdere condities voorkomt en dus meerdere varianten kent. Het aantal varianten is afhankelijk van de manipulaties die in het deelexperiment op de zin worden toegepast. Omdat elke zin in elk experiment zowel met *hebben* als met *zijn* voorkomt, heeft elke zin minimaal twee varianten. Bij de bespreking van de deelexperimenten

(§4.1.1.2) licht ik steeds het aantal zinnen en het aantal varianten toe. Appendix A bevat een overzicht van alle varianten uit het experiment.

Elke variant heeft een unieke code die de kenmerken ervan beschrijft. Die bevat steeds het nummer van het experiment (E1-E4), het nummer van de zin (1-6), de codering voor de specifieke manipulatie(s), en het gebruikte hulpwerkwoord van voltooid aspect (H vs. Z). De code 'E1_1_A0_Z' verwijst bijvoorbeeld naar de variant van de eerste zin ('1') van Experiment 1 ('E1'), waarin de afstandsmanipulatie 'A0' is en *zijn* ('Z') het hulpwerkwoord van voltooid aspect. Die code verwijst dan eenduidig naar de zin in (66a).

4.1.1.1.2 Onderzoeksvragen en hypothesen per deelexperiment

De stimuli van de experimenten zijn onderverdeeld in vijf deelexperimenten: Experiment 1, Experiment 2A, Experiment 2B, Experiment 3 en Experiment 4. In elk deelexperiment wordt de invloed van een van de taalinterne factoren (zie §3.1) onderzocht en wordt dus een specifieke hypothese getoetst (zie §3.2). In Experiment 1 onderzoek ik de invloed van afstand (OV2) in Asp-clusters en toets ik dus de afstandshypothese. Experiment 2A en 2B zijn gericht op de invloed van het type hoofdwerkwoord (OV1). In Experiment 2A wordt de hoofdwerkwoordhypothese getoetst in Asp-clusters; in Experiment 2B onderzoek ik de invloed van het type hoofdwerkwoord in Mod-clusters. Op die manier wordt het mogelijk de hulpwerkwoordvariatie in Mod- en Asp-clusters met elkaar te vergelijken, en ontstaat er een referentiepunt voor de resultaten van Experiment 2A. In Experiment 3 en 4 wordt ten slotte de invloed van modaliteit (OV3) onderzocht. Experiment 3 toetst de basisaanname dat een irrealislezing voor de selectie van *hebben* kan zorgen bij Z-werkwoorden in voltooid aspect (zie §3.2.4). De resultaten van die basisaanname zijn belangrijk voor de interpretatie van Experiment 4, waarin de invloed van irrealis in Asp-clusters wordt onderzocht en dus de irrealishypothese wordt getoetst.

Daarnaast wordt de basishypothese over de verhouding tussen *zijn* en *hebben* in Asp-clusters getoetst in Experiment 1, 2A en 4, aangezien die Asp-clusters bevatten. Experiment 2B bevat Mod-clusters, en Experiment 3 geen driedelige clusters. Aangezien voor elk deelexperiment oordelen van zowel Vlaamse als Nederlandse respondenten voorhanden zijn, wordt daarnaast in elk deelexperiment met Asp-clusters (Experiment 1, 2A en 4) de geografische variatie onderzocht (OV4). De registervariatie (OV5) komt, zoals toegelicht in §3.3, in het experiment niet aan bod. Tabel 4.1 bevat een overzicht van de onderzoeksvragen en hypothesen die worden behandeld per deelexperiment, en specificeert welk type cluster in het experiment voorkomt.

Tabel 4.1: Overzicht van onderzoeksvragen en hypothesen per deelexperiment

Experiment	Cluster	Onderzoeksvraag	Hypothesen
E1	Asp	OV2	Basishypothese
		OV4	Afstandshypothese
			Geografische hypothese
E2A	Asp	OV1	Basishypothese
		OV4	Hoofdwerkwoordhypothese
			Geografische hypothese
E2B	Mod	OV1	/
E3	/	OV3	/
E4	Asp	OV3	Basishypothese
		OV4	Irrealishypothese
			Geografische hypothese

4.1.1.1.3 Selectie van de hoofdwerkwoorden

Ik selecteerde zes werkwoorden die in de experimenten fungeren als hoofdwerkwoorden van Asp- en Mod-clusters: *slapen*, *lezen*, *stoppen*, *trouwen*, *winnen* en *volgen*. Die hoofdwerkwoorden behoren tot drie types: werkwoorden die *hebben* selecteren, werkwoorden die *zijn* selecteren en werkwoorden die zowel *hebben* als *zijn* kunnen selecteren (zie §2.2.1). Tabel 4.2 geeft de drie types hoofdwerkwoorden en de bijbehorende coderingen weer. In wat volgt licht ik eerst toe waarom de drie types hoofdwerkwoorden in het experiment zijn opgenomen. Vervolgens bespreek ik de selectiecriteria van de specifieke hoofdwerkwoorden en beschrijf ik de zes geselecteerde werkwoorden in detail. Tot slot behandel ik de verdeling van de hoofdwerkwoorden over de dealexperimenten.

Tabel 4.2: Types hoofdwerkwoorden die in de experimenten gebruikt worden

Code	Toelichting
H	Het werkwoord selecteert enkel <i>hebben</i> .
Z	Het werkwoord selecteert enkel <i>zijn</i> .
HZ	Het werkwoord kan zowel <i>hebben</i> als <i>zijn</i> selecteren.

De H- en HZ-hoofdwerkwoorden zijn in de eerste plaats nodig in het experiment om de hoofdwerkwoordhypothese te testen (zie §3.2.2). De hypothese is dat de aanwezigheid van een H-werkwoord een sterkere positieve invloed heeft op de selectie van *hebben* dan de aanwezigheid van een HZ-werkwoord. De Z-werkwoorden daarentegen worden opgenomen in het experiment om als hoofdwerkwoord in Mod-clusters te fungeren. In Mod-clusters is hulpwerkwoordvariatie mogelijk wanneer het hoofdwerkwoord *zijn* kan selecteren (zie §2.3.1), dus Z-werkwoorden zijn nodig om die variatie in Mod-clusters toe te laten. Naar analogie met de hoofdwerkwoordhypothese voor Asp-clusters (zie §3.2.2), neem ik aan dat de aanwezigheid van een Z-hoofdwerkwoord een sterkere positieve invloed heeft op de selectie van *zijn* dan de aanwezigheid van een HZ-hoofdwerkwoord.

Naast het feit dat ik hoofdwerkwoorden van de drie verschillende types zocht, wilde ik hoofdwerkwoorden selecteren die (i) ongeveer even frequent voorkomen in het Nederlands en (ii) uit evenveel lettergrepen bestaan. Op die manier probeer ik de invloed van de keuze van de specifieke hoofdwerkwoorden zo veel mogelijk te beperken. De frequentie heb ik bepaald aan de hand van SUBTLEX-NL (Keuleers et al. 2010), een databank met Nederlandse woordfrequenties gebaseerd op 44 miljoen woorden uit televisie- en filmondertitels. Concreet bekeek ik de kolom ‘FREQlemma’, die de frequentie bevat van het lemma waartoe het woord behoort (Keuleers et al. 2010:648). Daarenboven was het ook belangrijk voor de selectie dat de hoofdwerkwoorden betekenisvol konden voorkomen in een Asp- of Mod-cluster. Mits het kleine aantal mogelijkheden voor HZ-werkwoorden, was dat zeker voor dat type een uitdaging, zoals blijkt uit (67): de HZ-werkwoorden *opvolgen* en *verliezen* (zie §2.2.1) zijn niet gemakkelijk te combineren tot een betekenisvolle of natuurlijke Asp-cluster.

- (67) a. ?Hij **is** zijn vader **gaan opvolgen** toen hij stierf. AUX_Z – ASP_Z - HWW_{HZ}
b. ?Ze **is** haar stem **gaan verliezen** tijdens de drukke festivalzomer. AUX_Z – ASP_Z - HWW_{HZ}

De zes hoofdwerkwoorden die ik op basis van bovenstaande criteria selecteerde zijn *slapen*, *lezen*, *stoppen*, *trouwen*, *volgen* en *winnen*. Ze zijn allemaal tweelettergrepig en vallen binnen het middensegment van frequentie: ze behoren niet tot de meest noch tot de minst frequent voorkomende werkwoorden in het Nederlands. Tabel 4.3 toont de geselecteerde hoofdwerkwoorden (‘HWW’), samen

met het type waartoe ze behoren ('TypeHWW'), het aantal lettergrepen waaruit ze bestaan ('Syllaben') en het aantal keer dat het lemma voorkomt in SUBTLEX-NL ('Frequentie').

Tabel 4.3: Type, aantal syllaben en frequentie van geselecteerde hoofdwerkwoorden

HWW	TypeHWW	Syllaben	Frequentie
slapen	H	2	19 761
lezen	H	2	12 747
stoppen	Z	2	26 875
trouwen	Z	2	15 804
volgen	HZ	2	25 591
winnen	HZ	2	12 385

De H-hoofdwerkwoorden *slapen* en *lezen* selecteren systematisch *hebben* als hulpwerkwoord van voltooid aspect, zoals geïllustreerd in (68) (Haeseryn et al. 1997:75). De Z-werkwoorden *stoppen* en *trouwen* selecteren *zijn* als ze intransitief worden gebruikt, zoals in (69a) en (69c), maar in de (minder vaak voorkomende) gevallen dat ze transitief worden gebruikt, zoals in (69b) en (69d), selecteren ze *hebben* (Haeseryn et al. 1997:76). Aangezien ik *stoppen* en *trouwen* in de stimuli enkel gebruik in intransitieve vorm, kunnen ze in het experiment als Z-werkwoorden worden beschouwd.

- (68) a. Ze **heeft**/*is de hele nacht **geslapen**.
b. Hij **heeft**/*is het artikel al **gelezen**.
- (69) a. De trein **is**/*heeft net **gestopt** aan het perron.
b. De politieagent **heeft**/*is de verdachte op de hoek van de straat **gestopt**.
c. We **zijn**/*hebben in augustus **getrouwd**.
d. De pastoor **heeft**/*is ons vorig jaar **getrouwd**.

Het transitieve HZ-werkwoord *volgen* kan zowel met *hebben* als met *zijn* vervoegd worden wanneer het als werkwoord van beweging fungeert en dus de letterlijke betekenis 'achter iemand aan gaan' uitdrukt, zoals in (70a). In de figuurlijke betekenissen 'luisteren naar' (70b) of 'nadoen' (70c) kan *volgen* echter enkel *hebben* selecteren (Haeseryn et al. 1997:79-80). Bijgevolg komt *volgen* in het experiment enkel voor in de letterlijke betekenis van (70a), zodat het beschouwd kan worden als HZ-werkwoord.

- (70) a. Die hond **is/heeft** ons tot aan het hek **gevolgd**.
b. Ik **heb**/*ben het nieuws de laatste tijd aandachtig **gevolgd**.
c. Ik **heb**/*ben het verkeerde voorbeeld **gevolgd**.

Het HZ-werkwoord *winnen* kan zowel *hebben* als *zijn* selecteren wanneer het intransitief wordt gebruikt, zoals in (71a). De vervoeging met *zijn* is dan geen standaardtaal, maar komt vooral voor in de gesproken taal in België. Wanneer *winnen* transitief wordt gebruikt, zoals in (71b), kan de voltooide tijd enkel met *hebben* worden gevormd, ook in België (Taaladvies 2024). In de stimuli gebruik ik bijgevolg enkel het intransitieve *winnen*, dat ik als HZ-werkwoord beschouw.

- (71) a. Wie **heeft/is** er gisteren **gewonnen**?
b. Ik **heb**/*ben die wedstrijd **gewonnen**.

De selectie van *winnen* als HZ-hoofdwerkwoord is omwille van de niet-standaardtalige status en de nationale variatie niet de meest voor de hand liggende keuze. Zoals echter uit het voorgaande blijkt, was de selectie van de hoofdwerkwoorden een complex proces waarbij de selectiecriteria niet gemakkelijk met elkaar te combineren waren, waardoor ik mogelijk niet de meest representatieve Z- en HZ-hoofdwerkwoorden selecteerde. Die overwegingen neem ik echter mee in de analyse van de resultaten.

Tot slot bespreek ik de verdeling van de geselecteerde hoofdwerkwoorden over de deelexperimenten. In Experiment 2A wordt de hoofdwerkwoordhypothese getoetst, dus daarin zijn zowel de H-werkwoorden *slapen* en *lezen* als de HZ-werkwoorden *winnen* en *volgen* opgenomen. Z-werkwoorden komen niet voor in Experiment 2A, aangezien ik ervan uitga dat *hebben* niet geselecteerd kan worden wanneer zowel het aspectuele werkwoord als het hoofdwerkwoord *zijn* selecteren (zie §2.2.2). Naar analogie komen in Experiment 2B de Z-werkwoorden *stoppen* en *trouwen* en de HZ-werkwoorden *winnen* en *volgen* voor. H-werkwoorden komen in Experiment 2B niet voor aangezien ik ervan uitga dat *zijn* niet wordt geselecteerd wanneer zowel het modale werkwoord als het hoofdwerkwoord *hebben* selecteren (zie §2.2.2). In Experiment 1 en 4 gebruik ik enkel de HZ-werkwoorden. Ik zet zo de aangenomen neutraliteit van de HZ-werkwoorden in om de invloed van afstand (E1) en irrealis (E4) te meten zonder dat het hoofdwerkwoord een uitgesproken invloed uitoefent op de selectie van *hebben* of *zijn* (zie §2.3.3.2). De keuze tussen de twee geselecteerde HZ-hoofdwerkwoorden maakte ik willekeurig: *winnen* is het hoofdwerkwoord in Experiment 1, *volgen* in Experiment 4. De hoofdwerkwoorden in Experiment 3 zijn tot slot andere werkwoorden dan de zes die ik hier bespreek, aangezien ze niet in een werkwoordcluster voorkomen. Die worden besproken in §4.1.1.1.4.

4.1.1.1.4 Selectie van de groepsvormende werkwoorden

In de stimuli komen drie werkwoorden voor die fungeren als groepsvormend werkwoord in Asp- en Mod-clusters: *gaan*, *blijven* en *moeten*. Ik koos ervoor om per deelexperiment steeds slechts één groepsvormend werkwoord te gebruiken. Ik neem namelijk aan dat het geselecteerde groepsvormende werkwoord ook een invloed kan uitoefenen op de aanvaardbaarheid van het hulpwerkwoord (zie §2.3.3.4). Als ik per experiment steeds hetzelfde groepsvormende werkwoord gebruik, kan de variatie in de acceptabiliteitsscore niet toe te schrijven zijn aan het aspectuele werkwoord, maar wel aan de specifieke manipulatie die gebeurde in het experiment.

In de experimenten met Asp-clusters (Experiment 1, 2A en 4) komt niet elk aspectueel werkwoord (*gaan*, *komen*, *blijven*, *beginnen*) voor. Ik selecteerde het specifieke aspectuele werkwoord voor de Asp-clusters op basis van hoe gemakkelijk ze te combineren waren met de geselecteerde hoofdwerkwoorden van het betreffende experiment tot meerdere betekenisvolle zinnen. In Experimenten 1 en 4 wordt het aspectuele werkwoord gecombineerd met de HZ-werkwoorden *volgen* en *winnen* (zie §4.1.1.1.3). Die twee werkwoorden bleken in voltooid aspect betekenisvoller en natuurlijker gecombineerd te worden met *blijven* dan met *gaan*, *komen* of *beginnen*, zoals blijkt uit (72) voor *winnen* en (73) voor *volgen*.

- (72)
- a. Jan is **blijven** winnen, hij was heel blij!
 - b. ?Jan is meer **gaan** winnen toen hij beter zijn best deed.
 - c. ??Jan is toen **komen** winnen en daarna weer weggegaan.
 - d. ??Jan is **beginnen** winnen toen hij beter zijn best deed.

- (73)
- a. De man is haar **blijven** volgen, ze werd er bang van!
 - b. ?De man is haar **gaan** volgen, maar hij kreeg haar niet te pakken.
 - c. *De man is haar **komen** volgen.
 - d. ?De man is haar **beginnen** volgen toen hij haar voorbij zag wandelen.

Bijgevolg selecteerde ik het aspectuele werkwoord *blijven* voor Experimenten 1 en 4. Aangezien ik niet in alle experimenten met Asp-clusters hetzelfde aspectuele werkwoord wilde gebruiken, maakte ik in Experiment 2A gebruik van een ander aspectueel werkwoord. In Experiment 1 wordt het aspectuele werkwoord gecombineerd met vier verschillende hoofdwerkwoorden (*slapen*, *lezen*, *volgen* en *winnen*), dus ik selecteerde het aspectuele werkwoord dat naast *blijven* het gemakkelijkst combineerde met die vier hoofdwerkwoorden in voltooid aspect, en dat was voor mij *gaan*, zoals blijkt uit (74).

- (74)
- a. Marie is (vroeg/het boek/hem/de wedstrijden) **gaan** slapen/lezen/volgen/winnen.
 - b. Marie is (vroeg/het boek/hem/de wedstrijden) **komen** slapen/?lezen/*volgen/?winnen.
 - c. Marie is (vroeg/het boek/hem/de wedstrijden) **beginnen**?slapen/lezen/volgen/?winnen.

De aspectuele werkwoorden in de experimenten met Asp-clusters zijn dus *gaan* (E2A) en *blijven* (E1 en E4). De andere twee aspectuele werkwoorden, *komen* en *beginnen*, komen dus niet voor in het experiment als groepsvormend werkwoord van een Asp-cluster. Ik includeer ze echter wel in Experiment 3, waarbij de vier werkwoorden *gaan*, *blijven*, *komen* en *beginnen* als hoofdwerkwoord fungeren, en dus niet als groepsvormend werkwoord worden gebruikt (zie §4.1.1.2.4).

Wat de groepsvormende werkwoorden voor de Mod-clusters betreft, selecteerde ik slechts één modaal groepsvormend werkwoord, aangezien Mod-clusters enkel in Experiment 2B voorkomen. Ik koos voor het werkwoord *moeten* omdat het in voltooid aspect gemakkelijk combineerbaar was met de geselecteerde Z- en HZ-hoofdwerkwoorden die voorkomen in Experiment 2B, zoals geïllustreerd in (75).

- (75)
- Marie is/heeft (daar/toen/hem/de wedstrijd) **moeten** stoppen/trouwen/volgen/winnen.

4.1.1.2 Stimuli per deexperiment

Na de beschrijving van de algemene opbouw van de kritische stimuli, licht ik in deze sectie de stimuli per deexperiment toe (Experiment 1 – Experiment 4).

4.1.1.2.1 Experiment 1

Elke kritische zin van Experiment 1 bevat een Asp-cluster met een hulpwerkwoord van voltooid aspect (*hebben* of *zijn*), het aspectuele werkwoord *blijven* en het HZ-hoofdwkwoord *winnen*, zoals in (76).

(76) [_kDe pianist **is blijven winnen**] en [_czit nu in de finale van de muziekwedstrijd]. E1_3_A0_Z

In Experiment 1 wordt het effect van afstand op hulpwerkwoordselectie in Asp-clusters onderzocht. Ik toets met name of een grotere afstand tussen het hulpwerkwoord en het aspectuele werkwoord een positief effect heeft op de aanvaardbaarheid van *hebben*. Om het effect van afstand te meten, paste ik drie afstandsmanipulaties toe op de kritische zinnen. De drie manipulaties, evenals de codering die er verder voor gebruikt wordt, zijn weergegeven in Tabel 4.4.

Tabel 4.4: Afstandsmanipulaties gebruikt in Experiment 1

Code	Toelichting
A0	Het hulpwerkwoord en het aspectuele werkwoord zijn adjacent.
A1	Er staat 1 bepaling tussen het hulpwerkwoord en het aspectuele werkwoord.
A2	Er staan 2 bepalingen tussen het hulpwerkwoord en het aspectuele werkwoord.

De bepalingen die gebruikt worden om de afstand te vergroten zijn *toen*, *daar*, *maar* en *lang*. Ik koos voor die bepalingen omdat ze allemaal bestaan uit één lettergreep en omdat het mogelijk is om ze met elkaar te combineren, zoals bijvoorbeeld *toen maar*, *daar lang* en *daar toen*. In (77) wordt van elke afstandsmanipulatie een voorbeeld gegeven. In (77a) zijn het hulpwerkwoord en het aspectuele werkwoord *blijven* adjacent (A0), in (77b) staat er één bepaling, *toen*, tussen het hulpwerkwoord en *blijven* (A1), en in (77c) staan er twee bepalingen, *toen* en *maar*, tussen (A2).

- (77) a. De pianist **heeft blijven winnen** en zit nu in de finale [...]. E1_3_A0_H
b. De pianist **heeft toen blijven winnen** en zit nu in de finale [...]. E1_3_A1_H
c. De pianist **heeft toen maar blijven winnen** en zit nu in de finale [...]. E1_3_A2_H

Experiment 1 bestaat uit zes zinnen die elk in zes condities voorkomen. Die zes condities zijn het resultaat van twee hulpwerkwoordvarianten (H vs. Z) en drie manipulaties op basis van afstand (A0 vs. A1 vs. A2). Tabel 4.5 toont ter illustratie de zes varianten van de eerste zin van Experiment 1.

Tabel 4.5: Alle varianten van de eerste zin van Experiment 1 (E1_1)

Hulpww	Afstand	Variant
Z	A0	De Argentijnen zijn blijven winnen en kwalificeerden zich zo voor het WK.
H	A0	De Argentijnen hebben blijven winnen en kwalificeerden zich zo voor het WK.
Z	A1	De Argentijnen zijn toen blijven winnen en kwalificeerden zich zo voor het WK.
H	A1	De Argentijnen hebben toen blijven winnen en kwalificeerden zich zo voor het WK.
Z	A2	De Argentijnen zijn toen maar blijven winnen en kwalificeerden zich zo voor het WK.
H	A2	De Argentijnen hebben toen maar blijven winnen en kwalificeerden zich zo voor het WK.

4.1.1.2.2 Experiment 2A

De kritische zinnen van Experiment 2A bevatten een Asp-cluster die bestaat uit een hulpwerkwoord van voltooid aspect (*hebben* of *zijn*), het aspectuele werkwoord *gaan* en een van de hoofdwerkwoorden *slapen*, *lezen*, *volgen* of *winnen*, zoals geïllustreerd in (78).

(78) [_kDe politie **is** hem **gaan volgen**], maar [_cze kregen hem niet te pakken]. E2A_4_vol_HZ_Z

In Experiment 2A wordt de invloed van het type hoofdwerkwoord onderzocht en dus de hoofdwerkwoordhypothese getoetst. Ik onderzoek met name of *hebben* meer aanvaard wordt in Asp-clusters met een H-hoofdwerkwoord dan met een HZ-hoofdwerkwoord. Ik gebruik vier verschillende hoofdwerkwoorden om die hypothese te toetsen: twee van het H-type en twee van het HZ-type (zie §4.1.1.1.3). Tabel 4.6 toont de hoofdwerkwoorden die voorkomen in Experiment 2A, samen met de code voor het specifieke hoofdwerkwoord ('HWW') en het type waartoe het behoort ('TypeHWW').

Tabel 4.6: Hoofdwerkwoorden die gebruikt worden in Experiment 2A

Code	HWW	TypeHWW
sla	slapen	H
lez	lezen	H
vol	volgen	HZ
win	winnen	HZ

In alle zinnen van Experiment 2A is de afstand tussen het hulpwerkwoord en *gaan* steeds beperkt tot één lettergreep, om ervoor te zorgen dat de afstand geen invloed kan uitoefenen op de acceptabiliteitscores. Bij *slapen*, *lezen* en *winnen* plaatste ik er een eenlettergrepig bijwoord (*vroeg*, *veel* of *meer*) tussen, bij het transitieve werkwoord *volgen* gaat het om het lijdend voorwerp *hem*. Doordat er in Experiment 2A vier hoofdwerkwoorden gebruikt worden, bestaat Experiment 2A uit vier verschillende zinnen. Elke zin komt in twee condities voor, namelijk in de twee hulpwerkwoordvarianten. Tabel 4.7 presenteert ter illustratie de twee varianten van de eerste zin van Experiment 2A.

Tabel 4.7: Alle varianten van de eerste zin van Experiment 2A (E2A_1)

Hulpww	HWW	TypeHWW	Variant
Z	sla	H	Alex is vroeg gaan slapen, want hij moet morgen om zes uur op.
H	sla	H	Alex heeft vroeg gaan slapen, want hij moet morgen om zes uur op.

4.1.1.2.3 Experiment 2B

Elke kritische zin van Experiment 2B bevat een Mod-cluster bestaande uit een hulpwerkwoord van voltooid aspect (*hebben* of *zijn*), het modale werkwoord *moeten* en een van de hoofdwerkwoorden *stoppen*, *trouwen*, *volgen* of *winnen*, zoals geïllustreerd in (79).

(79) [_kMichelle **heeft** snel **moeten trouwen**], want [_cze bleek zwanger]. E2B_2_tro_Z_H

In Experiment 2B wordt net als in Experiment 2A de invloed van het type hoofdwerkwoord op hulpwerkwoordselectie onderzocht, maar dan in Mod-clusters. Ik toets met name of in Mod-clusters het hulpwerkwoord *zijn* meer aanvaard wordt bij een Z-hoofdwerkwoord dan bij een HZ-hoofdwerkwoord (zie §4.1.1.1.3). Om die hypothese te toetsen worden vier verschillende hoofdwerkwoorden gebruikt: twee Z-werkwoorden en twee HZ-werkwoorden.

Tabel 4.8 toont de hoofdwkwoorden die voorkomen in Experiment 2B, samen met de codering voor het specifieke hoofdwkwoord en het type waartoe het behoort.

Tabel 4.8: Hoofdwkwoorden die gebruikt worden in Experiment 2B

Code	HWW	TypeHWW
tro	trouwen	Z
sto	stoppen	Z
vol	volgen	HZ
win	winnen	HZ

In alle zinnen van Experiment 2B is de afstand tussen het hulpwerkwoord en *gaan* steeds beperkt tot één lettergreep, om te verhinderen dat de afstand een invloed uitoefent op de acceptabiliteitsscores. Bij *trouwen*, *stoppen* en *winnen* plaatste ik er een eenlettergrepig bijwoord tussen (*toen* of *snel*), bij het transitieve *volgen* het lijdend voorwerp *hem*. Aangezien er in Experiment 2B vier hoofdwkwoorden gebruikt worden, bestaat het experiment uit vier verschillende zinnen. Elke zin komt voor in twee condities, met name in de twee hulpwerkwoordvarianten (H vs. Z). Tabel 4.9 toont ter illustratie de twee varianten van de eerste zin uit Experiment 2B.

Tabel 4.9: Alle varianten van de eerste zin van Experiment 2B (E2B_1)

Hulpww	HWW	TypeHWW	Variant
Z	sto	Z	Lotte is toen moeten stoppen met studeren omdat haar vader erg ziek werd.
H	sto	Z	Lotte heeft toen moeten stoppen met studeren omdat haar vader erg ziek werd.

4.1.1.2.4 Experiment 3

Elke kritische zin van Experiment 3 bevat een voltooidetijdsvorm van de werkwoorden *gaan*, *komen*, *blijven* en *beginnen*, zoals in (80). Die werkwoorden worden in Experiment 3 dus niet als groepsvormend werkwoord, maar als hoofdwkwoord gebruikt, zonder een Asp-cluster te vormen.

(80) [cMoest ik alles nog klaarzetten], [kik **was** wat vroeger **gekomen**]! E3_2_irr_Z

In Experiment 3 wordt het effect van irrealis op hulpwerkwoordselectie onderzocht. Ik test namelijk de basisaanname voor de irrealishypothese, met name dat een irrealislezing een positief effect heeft op de aanvaardbaarheid van *hebben* bij Z-werkwoorden die als hoofdwkwoord fungeren (zie §3.2.4). Om het irrealiseffect te onderzoeken, paste ik twee modaliteitsmanipulaties toe op de zinnen, die samen met de gebruikte codering worden weergegeven in Tabel 4.10. Wanneer een zin een realisbetekenis uitdrukt, is de inhoud ervan in overeenstemming met de werkelijkheid, terwijl een irrealisbetekenis van een zin wil zeggen dat de inhoud ervan niet overeenstemt met de werkelijkheid (Haeseryn et al. 1997:1610).

Tabel 4.10: Modaliteitsmanipulaties gebruikt in Experiment 3

Code	Toelichting
rea	De zin drukt een realisbetekenis uit.
irr	De zin drukt een irrealisbetekenis uit.

Concreet wordt in de zinnen van Experiment 3 de realisbetekenis uitgedrukt door een werkwoord in de voltooid tegenwoordige tijd, zoals in (81a), en de irrealisbetekenis door de voltooid verleden tijd (81b). Om ervoor te zorgen dat de woordvolgorde in de kritische zinnen van de realisvariant en de irrealisvariant gelijk bleef, liet ik in de irrealisvariant (81b) geen inversie van onderwerp en persoonsvorm plaatsvinden.

- (81) a. Ik wist dat van tevoren, dus **ik heb** pas later begonnen. E3_4_rea_H
 b. Wist ik dat van tevoren, **ik had** pas later begonnen! E3_4_irr_H

Ik stelde in alle zinnen van Experiment 3 de afstand tussen het hulpwerkwoord en het aspectuele werkwoord gelijk aan drie lettergrepen (*wat vroeger, snel naar huis, wat langer, pas later*), om het effect van irrealis te kunnen onderzoeken zonder afstand als bijkomende beïnvloedende factor. Doordat er in Experiment 3 vier verschillende hoofdwerkwoorden voorkomen, bevat het experiment vier verschillende zinnen. Elke zin komt in vier condities voor, die het resultaat zijn van de twee hulpwerkwoordvarianten (*hebben* en *zijn*) en de twee modaliteitsmanipulaties (Realis). Tabel 4.11 presenteert ter illustratie de vier varianten van de eerste zin van Experiment 3.

Tabel 4.11: Varianten van de eerste zin van Experiment 3 (E3_1)

Hulpww	Realis	Variant
Z	rea	Ik maakte me veel zorgen, dus ik ben snel naar huis gegaan.
H	rea	Ik maakte me veel zorgen, dus ik heb snel naar huis gegaan.
Z	irr	Maakte ik me zo veel zorgen, ik was snel naar huis gegaan!
H	irr	Maakte ik me zo veel zorgen, ik had snel naar huis gegaan!

4.1.1.2.5 Experiment 4

Elke kritische zin van Experiment 4 bevat een Asp-cluster met een hulpwerkwoord van voltooid aspect, het aspectuele werkwoord *blijven* en het HZ-hoofdwkwoord *volgen*, zoals de zin in (82).

- (82) [_k**Had** die man mij **blijven volgen**], [_cvoelde ik mij onveilig]. E4_3_irr_inv_H

In Experiment 4 onderzoek ik het effect van irrealis op hulpwerkwoordselectie in Asp-clusters. Op die manier wordt de irrealishypothese getoetst en wordt er met name bekeken of een irrealislezing een positief effect heeft op de aanvaardbaarheid van *hebben* in Asp-clusters. Net als in Experiment 3 pas ik daarom de modaliteitsmanipulaties (zie Tabel 4.10) toe op de stimuli. Om van een realiszin als (83a) een irrealiszin te maken zijn er in het Nederlands twee veelgebruikte manieren (Haeseryn et al. 1997:1619). De meest voorkomende manier is een conditionele bijzin met *als*, zoals in (83b). Die alterneert met de voorwaardelijke zin met voor-pv, zoals in (83c) (Broekhuis & Corver 2016:164). Ik koos ervoor om in Experiment 4 de irrealisvarianten te formuleren als hoofdzinsconstructies met voor-pv, zoals (83c), en niet als conditionele bijzin, zoals (83b). Als de irrealis steeds in bijzinsconstructie (83b) zou worden uitgedrukt, en de realis door een hoofdzinsconstructie (83a), zou het verschil tussen realis en irrealis samenvallen met het verschil tussen hoofd- en bijzinnen. Om te vermijden dat het effect van irrealis niet te onderscheiden is van het effect van de bijzinsconstructie, wordt de irrealis in Experiment 4 dus steeds in hoofdzinsconstructie met voor-pv geformuleerd.

- (83) a. Dirk **was** haar blijven volgen, hij wist nu waar ze woonde.
 b. **Als** Dirk haar was blijven volgen, dan wist hij nu waar ze woonde.
 c. **Was** Dirk haar blijven volgen, **dan** wist hij nu waar ze woonde.

Als de zin in (83a) gebruiken als realisvariant en de zin in (83c) als irrealisvariant, verschillen de zinnen in zinsstructuur door de aanwezigheid van *dan* in contextzin van de irrealisvariant. Aangezien ik ernaar streefde de zinnen zo gelijkaardig mogelijk te maken afgezien van de bewuste manipulaties, kies ik ervoor om *dan* in de irrealisconstructie weg te laten vallen. Dat kan op twee manieren, zoals geïllustreerd in

(84). In (84a) is *dan* weggelaten en vindt er inversie van het onderwerp en de persoonsvorm plaats in de contextzin; in (84b) is *dan* ook weggelaten, maar vindt er geen inversie plaats. Aangezien de aanwezigheid van *dan* in zinnen als (84) in de standaardtaal verplicht is, zijn zowel (84a) als (84b) geen neutraal taalgebruik. Zinnen met inversie (84a) behoren tot het informele taalgebruik, terwijl zinnen zonder inversie (84b) tot het formele taalgebruik behoren (Haeseryn et al. 1997:560-561).

- (84) a. Was Dirk haar blijven volgen, **wist hij** nu waar ze woonde. E4_4_irr_inv_Z
 b. Was Dirk haar blijven volgen, **hij wist** nu waar ze woonde. E4_4_irr_ninv_Z

Aangezien beide zinnen in (84) wel grammaticaal maar gemarkeerd zijn, kies ik ervoor om beide opties in Experiment 4 op te nemen. De twee inversiem manipulaties, evenals de gebruikte codering, wordt weergegeven in Tabel 4.12. Op die manier vermijd ik dat de voorkeur van de respondenten voor de aan- of afwezigheid van inversie een invloed uitoefent op de acceptabiliteitscores voor *hebben* of *zijn*. Ik kan op die manier namelijk controleren of de scores voor *hebben* in de irrealis van elkaar verschillen afhankelijk van of er inversie plaatsvindt of niet. Merk op dat het bij deze manipulaties gaat om de contextzin, en niet om de kritische zin, zoals bij alle voorgaande manipulaties.

Tabel 4.12: Inversiem manipulaties gebruikt in Experiment 4

Code	Toelichting
inv	Er vindt inversie van het onderwerp en de persoonsvorm plaats in de contextzin.
ninv	Er vindt geen inversie van het onderwerp en de persoonsvorm plaats in de contextzin.

Experiment 4 bestaat uit zes zinnen die in zes condities voorkomen. Die zes condities zijn het resultaat van twee hulpwerkwoordvarianten (H vs. Z), twee modaliteitsmanipulaties (Realis, rea vs. irr) en twee inversiem manipulaties (inv vs. ninv). Tabel 4.13 presenteert ter illustratie de zes varianten van de eerste zin van Experiment 4.

Tabel 4.13: Alle varianten van de eerste zin van Experiment 4 (E4_1)

Hulpww	Realis	Inversie	Variant
Z	rea	ninv	De spion was haar blijven volgen, hij kende nu haar geheim.
H	rea	ninv	De spion had haar blijven volgen, hij kende nu haar geheim.
Z	irr	ninv	Was de spion haar blijven volgen, hij kende nu haar geheim.
H	irr	ninv	Had de spion haar blijven volgen, hij kende nu haar geheim.
Z	irr	inv	Was de spion haar blijven volgen, kende hij nu haar geheim.
H	irr	inv	Had de spion haar blijven volgen, kende hij nu haar geheim.

4.1.1.3 Fillers

De kritische stimuli uit de deelexperimenten worden aangevuld met 24 fillers. Die hebben dezelfde structuur als de zinnen van het experiment: ze bestaan ook steeds uit twee zinnen die verbonden zijn door een komma en/of een voegwoord, zoals in (85).

- (85) [Oma kan hier rustig zitten te lezen], maar [ik moet studeren]. F11

In de fillers komen verschillende factoren en constructies aan bod die aansluiten bij het onderwerp van deze masterproef. Zo gaat het bijvoorbeeld om clusters met een modaal en aspectueel werkwoord (F1-F4), zoals in (86a), om hulpwerkwoordselectie zonder clusters (F4-F7, zoals bij het HZ-werkwoord *verliezen* in (86b) en om irrealisconstructies (F14-F18), zoals in (86c). Daarnaast komen er enkele nationale syntactische verschillen aan bod (F19-F24), zoals de variabele plaatsing van de negatielementen *niet (...) meer* in zinnen als (86d). Die laatste fillers baseerde ik op het onderzoek naar nationale grammaticale variatie van De Troij (2023). Appendix A bevat een overzicht van de 24 fillers.

- (86)
- | | |
|--|-----|
| a. Kaat is lang moeten blijven wachten omdat de trein vertraging had. | F2 |
| b. Ik had mijn handschoenen verloren , maar ik heb ze inmiddels teruggevonden. | F5 |
| c. Ik had moeten meegaan hebben , dan leefde Eva nu nog! | F14 |
| d. Ik vind zwemmen niet leuk meer , dus ik ga niet meer. | F23 |

4.1.1.4 Conclusie: stimuli

Het experiment bestaat uit een combinatie van kritische stimuli en fillers. De kritische stimuli zijn opgedeeld in vijf deelexperimenten, die elk de invloed van een andere taalinterne factor onderzoeken. In Tabel 4.14 geef ik een overzicht van de kenmerken van de stimuli per deelexperiment. De tabel bevat informatie over de manipulatie(s), het type cluster, het groepsvormende werkwoord ('Groepsww'), het hoofdwkwoord ('HWW'), het aantal zinnen ('# zin') en het aantal varianten ('# var') per deelexperiment. Daarnaast toont de tabel ook op welke onderzoeksvragen ('OV') en hypothesen aan de hand van het deelexperiment een antwoord wordt gezocht.

Tabel 4.14: Overzicht stimuli per deelexperiment

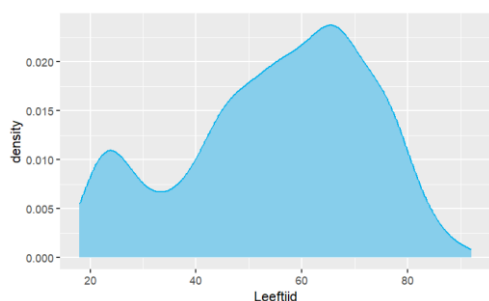
Exp	Manipulatie	Cluster	Groepsww	HWW	# zin	# var	OV	Hypothesen
E1	Afstand	Asp	blijven	winnen	6	36	OV2 OV4	Basis Afstand Geografisch
E2A	Type HWW	Asp	gaan	slapen, lezen volgen, winnen	4	8	OV1 OV4	Basis Hoofdwkwoord Geografisch
E2B	Type HWW	Mod	moeten	stoppen, trouwen volgen, winnen	4	8	OV1	/
E3	Realis	/	/	gaan, komen, blijven, beginnen	4	16	OV3	/
E4	Realis Inversie	Asp	blijven	volgen	6	36	OV3 OV4	Basis Irrealis Geografisch

4.1.2 Respondenten

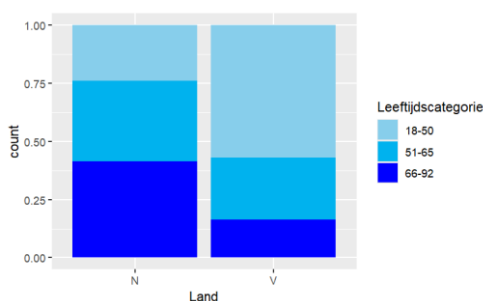
De respondenten werden uitgenodigd om deel te nemen aan het experiment via een Qualtrics-link,¹⁶ die ik verspreidde binnen mijn kennissenkring, op sociale media en via het online tijdschrift *Neerlandistiek*.¹⁷ In de inleidende tekst bij het experiment (Appendix D) maakte ik duidelijk dat de respondenten minstens 18 jaar oud moeten zijn, dat het Nederlands hun moedertaal moet zijn en dat hun geïnformeerde toestemming vereist is om te kunnen deelnemen. In deze sectie bespreek ik de samenstelling van de steekproef en de demografische kenmerken van de respondenten, waaronder hun geboorteland, geslacht, opleidingsniveau, leeftijd, taalprofiel, regio van herkomst en huidige woonplaats.

Aan het experiment namen in totaal 1562 Belgische en Nederlandse respondenten deel, in een convenience sample dat via *snowballing* het huidige formaat bereikte. De respondenten die het experiment niet volledig voltooiden ($N = 402$) en/of die niet akkoord gingen met de geïnformeerde toestemming ($N = 13$) werden uit de steekproef verwijderd. Ook respondenten met een andere moedertaal dan het Nederlands ($N = 7$) liet ik buiten de analyse. Ten slotte werden een aantal groepen respondenten van de analyse uitgesloten omdat ze niet in voldoende aantal voorkwamen om op een zinvolle manier statistische analyses uit te voeren. Het gaat om de groep respondenten waarvan we het geslacht niet kennen ($N = 26$), respondenten met ‘lager onderwijs’ als hoogste opleidingsniveau ($N = 2$) en respondenten die tot hun achttiende ergens anders dan in Vlaanderen of Nederland woonden ($N = 7$). Op die manier bleven 1105 respondenten over voor de analyse.

Van de 1105 respondenten zijn er 713 (64.5%) afkomstig uit Nederland en 392 (35.5%) uit Vlaanderen. Er zijn 616 vrouwen (55.7%) en 489 mannen (44.3%). Voor 983 respondenten (89%) is het hoogst behaalde diploma dat van het hoger/universitair onderwijs, voor 122 (11%) dat van het middelbaar onderwijs. De leeftijden variëren van 18 tot 92 jaar, met een gemiddelde van 55 jaar. Figuur 4.1 toont de dichtheidsverdeling van de leeftijd van de respondenten. Die wijst erop dat de leeftijd niet normaal verdeeld is, maar eerder een asymmetrische verdeling kent met twee pieken rond de leeftijden van 25 en 65 jaar. Voor de analyse werden de respondenten gelijkmatig verdeeld over drie leeftijdscategorieën: 394 respondenten (35.6%) behoren tot de leeftijdscategorie ‘18-50’, 351 (31.8%) tot ‘51-65’ en 360 (32.6%) tot ‘66-92’. Het is echter belangrijk te benadrukken dat de leeftijdscategorieën niet gelijk verdeeld zijn in Nederland en Vlaanderen. Uit Figuur 4.2 blijkt dat het aandeel jongere respondenten (18-50) groter is in Vlaanderen (67.8%) dan in Nederland (24%), terwijl de oudste leeftijdsgroep (66-92) sterker vertegenwoordigd is in Nederland (41.5%) dan in Vlaanderen (19.5%). Die associatie tussen leeftijdscategorie en land is statistisch significant ($\chi^2(2) = 130.8, p < 0.001$, Cramer’s $V = 0.34$).



Figuur 4.1: Dichtheidsverdeling leeftijd van respondenten



Figuur 4.2: Relatieve verdeling leeftijdsgroep per land

¹⁶ https://kuleuven.eu.qualtrics.com/jfe/form/SV_1z4XLU2N9RnmJg

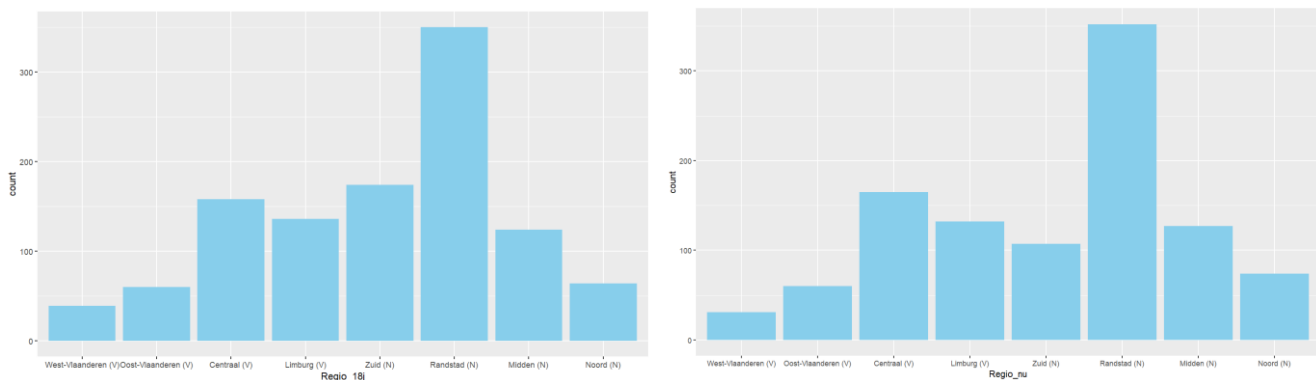
¹⁷ <https://neerlandistiek.nl/2024/04/doe-mee-onderzoek-over-aanvaardbaarheid-zinnen/>

De meerderheid van de respondenten ($N = 780$, 70.6%) blijkt geen dialect te spreken, terwijl 29.4% ($N = 325$) wel aangeeft dialectspreker te zijn. De gesproken dialecten variëren van het Houthalens, Waaslands en Oostends in Vlaanderen tot het Maastrichts, Twents en Gronings in Nederland. De provincies waarvan de respondenten afkomstig zijn of waarin ze nu wonen werden voor de analyse geclusterd in acht regio's: vier voor Vlaanderen (V) en vier voor Nederland (N). Tabel 4.15 geeft een overzicht van de regio's en welke provincies ertoe behoren.

Tabel 4.15: Overzicht van verdeling provincies in regio's

Regio	Provincie(s)
West-Vlaanderen (V)	West-Vlaanderen
Oost-Vlaanderen (V)	Oost-Vlaanderen
Centraal (V)	Vlaams-Brabant, Antwerpen, Brussel
Limburg (V)	Limburg
Zuid (N)	Limburg, Zeeland, Noord-Brabant
Midden (N)	Gelderland, Overijssel, Flevoland
Randstad (N)	Utrecht, Noord-Holland, Zuid-Holland
Noord (N)	Drenthe, Groningen, Friesland

Figuur 4.3 toont de verdeling van de respondenten op basis van de regio waar ze vóór hun achttiende het langst hebben gewoond (Regio_18j) en waar ze nu wonen (Regio_nu). Omdat de verdeling grotendeels overeenkomt, richt ik me hier op 'Regio_18j'. De regio Randstad (N) is met 350 respondenten (31.7%) het sterkst vertegenwoordigd. De minst vertegenwoordigde regio is West-Vlaanderen (V), met slechts 39 respondenten (3.5%), gevolgd door Oost-Vlaanderen (V) en Noord (N), met respectievelijk 60 (5.4%) en 64 respondenten (5.8%). Het aantal respondenten uit de overige regio's is vergelijkbaar: 124 (11.2%) voor Midden (N), 136 (12.3%) voor Limburg (V), 158 (14.3%) voor Centraal (V) en 174 (15.7%) voor Zuid (N).



Figuur 4.3: Absolute verdeling Regio_18j en Regio_nu

4.1.3 Taak

Tijdens het experiment beoordeelden de respondenten de stimuli op hun acceptabiliteit, aan de hand van de vraag in (87). In de inleidende tekst van het experiment (Appendix D) werd benadrukt dat ze hun oordeel moeten baseren op taalgevoel, en dat er dus geen juiste of foute antwoorden zijn.

- (87) Hoe aanvaardbaar vind je deze zin? Kan je je voorstellen dat jij of iemand in je omgeving deze zin zou gebruiken?

De beoordeling gebeurde aan de hand van een schuiver op een schaal met als polen '1' ('helemaal niet aanvaardbaar') en '7' ('helemaal aanvaardbaar'). Figuur 4.4 illustreert de opmaak van de sliderschaal en de presentatie van de stimuli aan de respondenten.



Figuur 4.4: Illustratie sliderschaal experiment

4.1.4 Procedure

De respondenten voerden het volledige experiment uit via de online enquêtetool Qualtrics.¹⁸ Het experiment bestond uit vier delen. In het eerste deel voerden de respondenten de acceptabiliteitsbeoordelingstaak (zie §4.1.3) uit op twee testzinnen, met name op de eerste twee zinnen die ze te zien kregen. Die testzinnen (T1-T2), weergegeven in (88), hebben niets te maken met hulpwerkwoordselectie, maar hebben een gelijkaardige zinsstructuur als de stimuluszinnen. Op die manier maken ze de respondenten vertrouwd met de aard van de stimuli en met het experimentele opzet.

- (88) a. Als ik had geweten dat hij er ook zou zijn, dan was ik niet gekomen. T1
b. Toen hij haar na lange tijd ontmoette, hij was heel blij. T2

In het tweede deel van het experiment voerden de respondenten de acceptabiliteitsbeoordelingstaak uit op de eigenlijke stimuli. Aangezien het voor de zuiverheid van de intuïties niet wenselijk en logistiek niet haalbaar was om iedere respondent elke variant te laten beoordelen, werd er gebruikgemaakt van een *between-subjects* design. Ik verdeelde alle varianten uit de experimenten, in verschillende combinaties, onder in zes condities. Elke conditie bevat steeds één variant van iedere zin van elk experiment, en alle fillers. Aan de hand van de *randomizer*-functie van Qualtrics werden de respondenten willekeurig toegewezen aan een van de zes condities. Op die manier beoordeelde elke respondent 48 stimuli, bestaande uit 24 gemanipuleerde zinnen en 24 fillers. Appendix B bevat een overzicht van de condities.

¹⁸ <https://www.qualtrics.com/nl/>

In het derde deel verstrekten de respondenten demografische gegevens over hun leeftijd, geslacht, opleidingsniveau, geboorteland, provincie van opgroeien en huidige woonplaats. Daarnaast gaven ze informatie over hun taalkundig profiel, met name over hun moedertaal en het dialect dat ze (eventueel) spreken. De concrete demografische vragen (D1 – D10) worden opgelijst in Appendix C. Het vierde en laatste deel bestond uit een *debriefing*, waarbij de respondenten hun inschatting over het onderwerp van het onderzoek deelden aan de hand van de vraag in (89).

(89) Waarover denk je dat dit onderzoek ging?

Tabel 4.16 geeft een schematisch overzicht van de opbouw van het experiment, zoals het eruitziet voor een respondent die werd toegewezen aan de eerste conditie.

Tabel 4.16: Schematische opbouw experiment voor conditie 1

Deel	Inhoud	Code
Deel 1	Testzinnen	T1 – T2
Deel 2	Gemanipuleerde zinnen	E1_1_A0_Z
		E1_2_A2_H
		E1_3_A2_Z
		E1_4_A1_H
		E1_5_A1_Z
		E1_6_A0_H
		E2A_1_sla_H_Z
		E2A_2_lez_H_H
		E2A_3_win_HZ_Z
		E2A_4_vol_HZ_H
		E2B_1_sto_Z_H
		E2B_2_tro_Z_Z
		E2B_3_win_HZ_H
		E2B_4_vol_HZ_Z
		E3_1_rea_Z
		E3_2_rea_H
		E3_3_irr_Z
		E3_4_irr_H
		E4_1_rea_ninv_Z
		E4_2_irr_inv_H
		E4_3_irr_inv_Z
		E4_4_irr_ninv_H
		E4_5_irr_ninv_Z
		E4_6_rea_ninv_H
	Fillers	F1 – F24
Deel 3	Demografische gegevens	D1 – D10
Deel 4	Debriefing	

4.1.5 Analyse

Ik voerde de analyse van het experiment uit via het statistische programma R (R Core Team 2024) en de gebruiksvriendelijke werkomgeving RStudio (Posit Team 2024).¹⁹ In de eerste plaats maakte ik voor elk Experiment een afzonderlijk resultatenbestand aan, met verschillende kolommen die de onderzochte variabelen in het betreffende dealexperiment bevatten. De taalinterne variabelen voegde ik aan de bestanden toe aan de hand van de variantencodes (zie §4.1.1): Hulpww (H vs. Z), Afstand (A0 vs. A1 vs. A2), TypeHWW (H vs. Z vs. HZ), HWW (sla vs. lez vs. sto vs. tro vs. vol vs. win), Realis (rea vs. irr) en Inversie (inv vs. ninv). De taalexterne variabelen die zijn toegevoegd zijn Leeftijdscategorie (18-50 vs. 51-60 vs. 65-92), Geslacht (Man vs. Vrouw), Opleiding (Middelbaar onderwijs vs. Hoger/universitair onderwijs), Land (V vs. N)²⁰ en Regio. Voor die laatste variabele gebruikte ik 'Regio_18j' als referentie voor de regio van de respondent, aangezien de waarden in 'Regio_18j' en 'Regio_nu' nauwelijks van elkaar verschillen (zie §4.1.2). De mogelijke waarden van Regio zijn de geclusterde regio's zoals toegelicht in Tabel 4.15. De variabele Dialect, die aangeeft of de respondent een dialectspreker is, wordt omwille van tijd- en plaatsgebrek niet meegenomen in de analyse, net als de kolom Debriefing, die het antwoord van de respondent op vraag (89) bevat. Tot slot bevat elk resultatenbestand een kolom met de variabele Score (1-7), met de score die de respondent gaf aan de betreffende variant, en een kolom met RespoID, waarin ik aan elke specifieke respondent een nummer koppelde.

Na het voorbereidende werk in de resultatenbestanden voerde ik in de eerste plaats een verkennende analyse uit per Experiment. Daarin berekende ik de gemiddelde scores en standaarddeviaties voor alle variabelen die een rol spelen in het betreffende dealexperiment, om een algemeen inzicht in de data te verkrijgen. Voor de variabelen die buiten de regressieanalyse werden gehouden (Regio, HWW en Inversie; zie verder), heb ik de data nader verkend door middel van grafieken en, waar mogelijk, *t*-toetsen om de gemiddelden tussen groepen te vergelijken.

Na de verkennende analyse analyseerde ik de acceptabiliteitsoordelen via lineaire *mixed-effects* regressies. De analyse via een lineaire regressie was mogelijk doordat de oordelen via een sliderschaal, en niet via een ordinale 7-punts Likertschaal geëliciteerd werden (zie §4.1.3). Experiment 2A, 2B en 3 zijn minder geschikt voor een regressieanalyse dan Experiment 1 en 4 aangezien er minder meetpunten beschikbaar zijn vanwege het kleinere aantal varianten (zie Tabel 4.14). Desondanks maakte het grote aantal respondenten een regressieanalyse alsnog mogelijk. Voor elk Experiment stelde ik verschillende regressiemodellen op. Daarbij was Score telkens de afhankelijke variabele. In de eerste plaats werd er steeds een basismodel opgesteld, waarin twee gemanipuleerde *fixed effects* werden opgenomen: Afstand, TypeHWW of Realis en Hulpww, afhankelijk van het dealexperiment. In de tweede plaats voerde ik voor elk Experiment enkele driewegsanalyses uit waarbij ik aan het basismodel een covariaat (Land, Opleiding, Geslacht, Leeftijdscategorie) toevoegde. In elk regressiemodel includeerde ik daarnaast RespoID als *random effect*, zodat het model kon controleren op variatie tussen de verschillende respondenten. Drie variabelen integreerde ik niet in de regressieanalyse. Het specifieke hoofdwerkwoord (HWW) werd niet opgenomen vanwege multicollineariteit met TypeHWW, en Inversie

¹⁹ Volgende R-pakketten werden opgeladen: car (Fox & Weisberg 2019), dplyr (Wickham et al. 2023), effects (Fox 2003), ggplot2 (Wickham 2016), lmerTest (Kuznetsova et al. 2017), multcomp (Hothorn et al. 2008), MuMIn (Bartón 2024), readxl (Wickham & Bryan 2023), sjPlot (Lüdtke 2024), sjstats (Lüdtke 2024), tidyverse (Wickham et al. 2019) en vcd (Meyer et al. 2023).

²⁰ Hoewel Vlaanderen strikt genomen geen 'land' is, gebruik ik de variabele 'Land' om het onderscheid tussen Nederland en Vlaanderen aan te duiden. Die term is relevant aangezien ik onderzoek of er een verschil is aan weerszijden van de staatsgrens tussen Nederland en België.

integreerde ik niet wegens de overlap met Realis: de inversievariantie bestaat enkel voor irrealis (zie §4.1.1.2.5). De variabele Regio is ten slotte uitgesloten omdat die acht niveaus omvat met een ongelijke verdeling van datapunten (zie §4.1.2). Tabel 4.17 biedt een overzicht van welke variabelen worden opgenomen in de verkennende analyse en in welke in de regressieanalyse als *fixed effects*, covariaten en *random effect*. Zoals hierboven vermeld, is Score steeds de afhankelijke variabele.

Tabel 4.17: Overzicht opgenomen variabelen in verkennende analyse en regressieanalyse per deexperiment

Experiment	Verkennende analyse	Regressieanalyse		
		Fixed effects	Covariaten	Random effect
E1	Hulpww Afstand Land	Hulpww Afstand	Land Opleiding Geslacht Leeftijdscategorie	RespID
E2A	Hulpww TypeHWW HWW Land Regio	Hulpww TypeHWW		
E2B	Hulpww TypeHWW HWW Land Regio	Hulpww TypeHWW		
E3	Hulpww Realis Land	Hulpww Realis		
E4	Hulpww Realis Inversie Land	Hulpww Realis		

Om de lineaire modellen te evalueren bekeek ik de AIC-waarde (*Akaike Information Criterion*), aangezien een lagere AIC-waarde duidt op een beter model (Levshina 2015:149). Daarnaast rapporteer ik de marginale R^2 , een maat voor de variantie verklaard door de *fixed effects*, en de conditionele R^2 , die aangeeft hoeveel variantie verklaard wordt door het gehele model (Gries 2021:432). Tot slot voerde ik post-hoc analyses uit (Tukey Honest Significant Differences) om paarsgewijze vergelijkingen te kunnen maken bij interacties (Levshina 2015:149).

4.1.6 Conclusie

In het experimentele onderzoek voerden meer dan duizend Nederlandse en Vlaamse respondenten een acceptabiliteitsbeoordelingstaak uit. Ze beoordeelden elk 48 stimuli, voor de helft bestaande uit gemanipuleerde zinnen, voor de andere helft uit fillers. De gemanipuleerde zinnen komen uit vijf verschillende deexperimenten die elk de invloed van een specifieke taalinterne variabele op de aanvaardbaarheid van de hulpwerkwoorden *hebben* of *zijn* onderzoeken: Afstand in Experiment 1, TypeHWW in Experimenten 2A en 2B en Realis in Experimenten 3 en 4. Aangezien de respondenten in het experiment ook demografische gegevens verschaften, kon ook de invloed van Land en Regio op de acceptabiliteitsoordelen worden onderzocht, net als effect van Leeftijd, Opleidingsniveau en Geslacht. De oordelen van de respondenten worden geanalyseerd aan de hand van een combinatie van een verkennende analyse en een lineaire regressieanalyse. De resultaten van de analyses worden in de volgende sectie (§4.2) uiteengezet.

4.2 Resultaten

De resultaten van het experiment worden beschreven per deexperiment: Experiment 1 (§4.2.1), Experiment 2A (§4.2.2), Experiment 2B (§4.2.3), Experiment 3 (§4.2.4) en Experiment 4 (§4.2.5). Voor elk Experiment bespreek ik eerst de verkennende analyse, vervolgens de lineaire regressieanalyse, en tot slot een conclusie. In §4.2.6 vat ik de resultaten van alle deexperimenten samen.

4.2.1 Experiment 1

In Experiment 1 wordt specifiek de afstandshypothese getoetst. Ik onderzoek met name of een grotere afstand tussen het hulpwerkwoord en het aspectuele werkwoord in Asp-clusters een positief effect heeft op de aanvaardbaarheid van *hebben*. Zowel in de verkennende analyse (§4.2.1.1) als in de regressieanalyse (§4.2.1.2) worden de variabelen Hulpww, Afstand en Land opgenomen. In §4.2.1.3 volgt een conclusie in het licht van de hypothesen.

4.2.1.1 Verkennende analyse

Tabel 4.18 bevat de gemiddelde scores (*M*) en standaarddeviaties (*SD*) voor Experiment 1, per hulpwerkwoord (H vs. Z), afstand (A0 vs. A1 vs. A2), land (Nederland vs. Vlaanderen) en in totaal.

Tabel 4.18: Exp1 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Afstand en Land

Hulpww	Afstand	Nederland		Vlaanderen		Totaal	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
H	A0	1.42	0.87	2.10	1.37	1.66	1.12
H	A1	1.42	0.79	2.02	1.23	1.63	1.01
H	A2	1.36	0.71	2.01	1.28	1.59	0.99
Z	A0	4.76	2.21	5.24	1.91	4.93	2.12
Z	A1	4.21	2.27	4.53	2.08	4.32	2.21
Z	A2	3.68	2.22	4.18	2.10	3.86	2.19

De gemiddelde score voor het hulpwerkwoord *hebben* ($M = 1.63$) ligt beduidend lager dan die voor *zijn* ($M = 4.37$). De standaardafwijking voor *hebben* ($SD = 1.04$) is kleiner dan die voor *zijn* ($SD = 2.22$), wat wijst op weinig spreiding in de negatieve oordelen over *hebben*. Die sterke voorkeur voor *zijn* zien we zowel in Nederland als in Vlaanderen, maar er zijn ook nationale verschillen. *Hebben* wordt meer aanvaard in Vlaanderen ($M = 2.04$) dan in Nederland ($M = 1.40$), terwijl ook *zijn* in Vlaanderen hogere scores krijgt ($M = 4.65$) dan in Nederland ($M = 4.22$). In Nederland is de standaardafwijking voor *hebben* ($SD = 0.79$) lager dan in Vlaanderen ($SD = 1.29$), wat duidt op een nagenoeg categorische afwijzing van *hebben* in Nederland, terwijl de scores van Vlamingen voor *hebben* een grotere spreiding vertonen.

Wat de afstand betreft blijkt dat de gemiddelde score voor *hebben* afneemt naarmate de afstand toeneemt (A0 ($M = 1.66$) > A1 ($M = 1.63$) > A2 ($M = 1.59$)), met een gering verschil tussen de gemiddelden. Een vergelijkbare trend is zichtbaar bij het hulpwerkwoord *zijn*, waarbij de gemiddelde score eveneens afneemt bij een grotere afstand (A0 ($M = 4.93$) > A1 ($M = 4.32$) > A2 ($M = 3.86$)). De verschillen tussen de gemiddelden zijn bij *zijn* groter dan bij *hebben*, wat suggereert dat de invloed van Afstand op de aanvaardbaarheid sterker is voor *zijn* dan voor *hebben*. Tot slot toont Tabel 4.18 dat de effecten van Afstand op de aanvaardbaarheid erg vergelijkbaar zijn in Vlaanderen en Nederland.

4.2.1.2 Lineaire regressieanalyse

In het basismodel van Experiment 1 (§4.2.1.2.1) wordt het effect van de variabelen Afstand en Hulpww, evenals hun interactie, onderzocht. In de driewegsanalyses (§4.2.1.2.2) worden daarna de covariaten aan het model toegevoegd. Eerst voeg ik Land toe om na te gaan of de effecten anders verdeeld zijn in Nederland en Vlaanderen, en daarna stel ik modellen op met de covariaten Opleiding, Geslacht en Leeftijdscategorie. Zoals toegelicht in §4.1.5 is Score in elk model de afhankelijke variabele, en werd RespoID in elk model opgenomen als *random effect*.

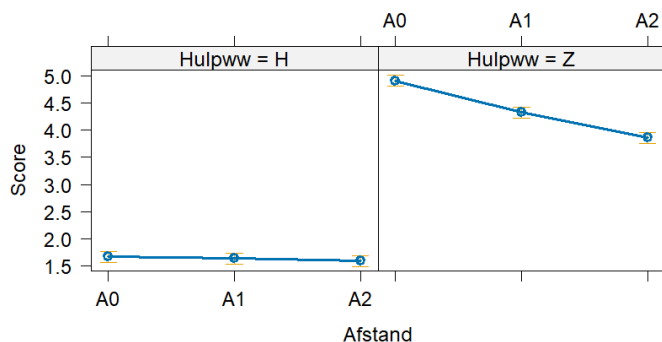
4.2.1.2.1 Basismodel

In het basismodel (AIC = 23921.38; marginale $R^2 = 0.406$; conditionele $R^2 = 0.564$) zijn Afstand, Hulpww en hun eventuele interactie opgenomen. Tabel 4.19 bevat de regressieoutput van het basismodel, met de geschatte regressiecoëfficiënten (*estimate*), de standaardafwijking (*Std. Error*), de vrijheidsgraden (*df*), de *t*-waarden (*t value*), de *p*-waarden (*Pr(>|t|)*) en de significantiecodes.²¹

Tabel 4.19: Exp1 – Regressieoutput basismodel (Afstand*Hulpww)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.67	0.05145	4775.81	32.36	< 0.001	***
AfstandA1	-0.03	0.06254	5248.72	-0.49	0.623	
AfstandA2	-0.08	0.06245	5244.30	-1.23	0.221	
HulpwwZ	3.24	0.06323	5261.63	51.24	< 0.001	***
AfstandA1:HulpwwZ	-0.55	0.08947	5260.72	-6.20	< 0.001	***
AfstandA2:HulpwwZ	-0.97	0.08950	5261.26	-10.86	< 0.001	***

Tabel 4.19 laat zien dat Afstand geen hoofdeffect is ($p = 0.623$ en $p = 0.221$). De varianten worden dus niet meer of minder aanvaard naarmate de afstand tussen het hulpwerkwoord en het aspectuele werkwoord toeneemt, ongeacht of het om *hebben* of *zijn* gaat. Hulpww is daarentegen wel een hoofdeffect ($p < 0.001$): onafhankelijk van de afstand wordt *zijn* significant geprefereerd boven *hebben*. De significante interacties tussen Afstand en Hulpww ($p < 0.001$) bewijzen dat Afstand een significant effect heeft op de scores voor *zijn*. Uit de post-hoc analyse blijken de scores voor *zijn* significant lager bij A1 in vergelijking met A0 (*estimate* = -0.59, $p < 0.001$) en eveneens lager bij A2 in vergelijking met A1 (*estimate* = -0.46, $p < 0.001$). De aanvaardbaarheid van *zijn* neemt dus significant af naarmate de afstand toeneemt. Op de aanvaardbaarheid van *hebben* heeft Afstand echter geen significant effect. De post-hoc analyse toont aan dat de scores van A1 en A0 niet significant van elkaar verschillen ($p = 0.996$), net als die van A2 en A1 ($p = 0.978$). Figuur 4.5 illustreert de interactie tussen Afstand en Hulpww.



Figuur 4.5: Exp1 - Plot interactie Afstand*Hulpww

²¹ Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.2.1.2.2 Driewegsanalyses

In de eerste plaats voeg ik aan een tweede regressiemodel ook het covariaat Land toe om te onderzoeken of de interactie tussen Hulpww en Afstand afhankelijk is van het land. Het model bevat dan de twee *fixed effects* (Hulpww en Afstand), het covariaat Land, en de twee- en driewegsinteractie tussen die predictoren. Dit model (AIC = 23871.06; marginale $R^2 = 0.421$; conditionele $R^2 = 0.565$) is beter dan het vorige (AIC 23871.06 < AIC 23921.38). Tabel 4.20 toont de regressieoutput van het model.

Tabel 4.20: Exp1 – Regressieoutput driewegsanalyse (Afstand*Hulpww*Land)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.43	0.06319	4896	22.56	< 0.001	***
LandV	0.69	0.1064	4939	6.49	< 0.001	***
AfstandA1	-0.00	0.07751	5236	-0.04	0.970	
AfstandA2	-0.07	0.07748	5235	-0.85	0.397	
HulpwwZ	3.32	0.07842	5252	42.33	< 0.001	***
LandV:AfstandA1	-0.08	0.1310	5255	-0.60	0.548	
LandV:AfstandA2	-0.03	0.1307	5247	-0.24	0.811	
LandV:HulpwwZ	-0.22	0.1324	5266	-1.69	0.091	.
AfstandA1:HulpwwZ	-0.53	0.1109	5251	-4.80	< 0.001	***
AfstandA2:HulpwwZ	-0.99	0.1110	5256	-8.93	< 0.001	***
LandV:AfstandA1:HulpwwZ	-0.07	0.1873	5266	-0.35	0.724	
LandV:AfstandA2:HulpwwZ	0.05	0.1873	5262	0.28	0.778	

Het tweede model bevestigt de effecten uit het basismodel, met name het hoofdeffect van Hulpww ($p < 0.001$) en de interacties tussen Afstand en Hulpww ($p < 0.001$), maar verdiept het met een aantal nationale tendensen. Het hoofdeffect van Land ($p < 0.001$) toont ten eerste dat Vlamingen, ongeacht de manipulaties, hogere scores toekennen in alle condities. Daarnaast suggereert de bijna-significante interactie tussen Land en Hulpww ($p = 0.091$) dat *hebben* meer aanvaard wordt in Vlaanderen dan in Nederland. Het feit dat de driewegsinteractie niet significant is ($p = 0.724$ en $p = 0.778$), toont ten slotte aan dat de interactie tussen Afstand en Hulpww op dezelfde manier werkt in Vlaanderen en Nederland. Het afstandseffect voor *zijn* is dus vergelijkbaar in beide landen, terwijl een dergelijk effect voor *hebben* in beide landen ontbreekt.

In de tweede plaats voer ik bijkomende driewegsanalyses uit waarbij de covariaten Geslacht, Opleiding en Leeftijdscategorie aan regressiemodellen worden toegevoegd. De modellen bevatten steeds de twee *fixed effects* Hulpww en Afstand, een van de covariaten, en alle mogelijke twee- en driewegsinteracties ertussen. Ik beperk me hier tot de bespreking van de driewegsinteractie. In het model met Opleiding (AIC = 23928.46; marginale $R^2 = 0.407$; conditionele $R^2 = 0.565$) is de driewegsinteractie tussen Opleiding, Afstand en Hulpww niet significant ($p = 0.544$). Ook in het model met Geslacht (AIC = 23834.36; marginale $R^2 = 0.420$; conditionele $R^2 = 0.570$) en dat met Leeftijdscategorie (AIC = 23888.56; marginale $R^2 = 0.413$; conditionele $R^2 = 0.570$) zijn de driewegsinteracties niet significant (respectievelijk $p = 0.766$ en $p = 0.066$). Die resultaten wijzen erop dat de interactie tussen Afstand en Hulpww niet afhankelijk is van het opleidingsniveau, het geslacht of de leeftijdscategorie van de respondenten.

4.2.1.3 Conclusie: Experiment 1

De analyse van Experiment 1 laat zien dat het hulpwerkwoord *hebben* in Asp-clusters significant minder aanvaard wordt dan het hulpwerkwoord *zijn*. Dat bevestigt de basishypothese. Die sterke voorkeur voor *zijn* zien we zowel in Vlaanderen als in Nederland. Toch wordt het hulpwerkwoord *hebben* meer aanvaard door de Vlaamse dan door de Nederlandse respondenten, wat de geografische hypothese ondersteunt. De afstandshypothese vindt geen ondersteuning in de resultaten van Experiment 1: de afstand tussen het hulpwerkwoord van voltooid aspect en het aspectuele werkwoord heeft in beide landen geen significante invloed op de aanvaardbaarheid van *hebben*. Het is dus niet zo dat *hebben* meer aanvaard wordt bij een grotere afstand (A2) dan bij een kleinere afstand (A1) of adjacentie (A0). Op de aanvaardbaarheid van *zijn* heeft de afstand echter wel een significante invloed in beide landen: hoe meer de afstand toeneemt, hoe minder *zijn* aanvaard wordt.

4.2.2 Experiment 2A

In Experiment 2A test ik specifiek de hoofdwkwoordhypothese. Ik onderzoek dus of in Asp-clusters *hebben* meer aanvaard wordt bij hoofdwkwoorden die *hebben* selecteren (H-hoofdwkwoorden) dan bij hoofdwkwoorden die zowel *hebben* als *zijn* kunnen selecteren (HZ-hoofdwkwoorden). In de verkennende analyse (§4.2.2.1) bespreek ik eerst Hulpww, TypeHWW en Land (§4.2.2.1.1), en daarna in meer detail de variabelen HWW (§4.2.2.1.2) en Regio (§4.2.2.1.3), die niet worden opgenomen in de regressieanalyse (zie §4.1.5). Vervolgens licht ik de resultaten van de lineaire regressie toe (§4.2.2.2), en in §4.2.2.3 vat ik de resultaten van Experiment 2A samen met betrekking tot de hypotheses.

4.2.2.1 Verkennende analyse

4.2.2.1.1 Hulpwerkwoord, Type hoofdwkwoord en Land

Tabel 4.21 geeft een overzicht van de gemiddelde scores en standaarddeviaties van Experiment 2A per hulpwerkwoord, type hoofdwkwoord (HZ vs. H), land (Nederland vs. Vlaanderen) en in totaal.

Tabel 4.21: Exp2A - Gemiddelde scores (met standaarddeviatie) als functie van Hulpww, TypeHWW en Land

Hulpww	TypeHWW	Nederland		Vlaanderen		Totaal	
		M	SD	M	SD	M	SD
H	HZ	1.26	0.57	1.54	0.89	1.36	0.71
H	H	1.22	0.58	1.41	0.69	1.28	0.63
Z	HZ	5.69	1.76	4.83	1.93	5.39	1.86
Z	H	6.58	0.98	6.04	1.41	6.39	1.18

De gemiddelde score voor het hulpwerkwoord *hebben* ($M = 1.32$) ligt lager dan die voor *zijn* ($M = 5.90$). Daarnaast is de standaarddeviatie voor *hebben* ($SD = 0.67$) kleiner dan die voor *zijn* ($SD = 1.63$), wat wijst op weinig variatie in de negatieve oordelen over *hebben*. De sterke voorkeur voor *zijn* bestaat zowel in Vlaanderen als in Nederland, maar er zijn ook nationale verschillen. Het hulpwerkwoord *hebben* wordt meer aanvaard in Vlaanderen ($M = 1.48$) dan in Nederland ($M = 1.24$), terwijl *zijn* hogere scores krijgt in Nederland ($M = 6.14$) dan in Vlaanderen ($M = 5.44$). De lage standaardafwijking voor *hebben* is in Nederland ($SD = 0.58$) nog lager dan in Vlaanderen ($SD = 0.80$), wat suggereert dat er in Nederland een hogere mate van consensus bestaat over de onaanvaardbaarheid van *hebben*.

Wat het type hoofdwerkwoord betreft, ligt de gemiddelde score voor *hebben* hoger bij de HZ-hoofdwerkwoorden *volgen* en *winnen* ($M = 1.36$) dan bij de H-hoofdwerkwoorden *slapen* en *lezen* ($M = 1.28$). Voor *zijn* is de gemiddelde score daarentegen hoger bij de H-hoofdwerkwoorden ($M = 6.39$) dan bij de HZ-hoofdwerkwoorden ($M = 5.39$). Die trend is zowel in Nederland als in Vlaanderen zichtbaar.

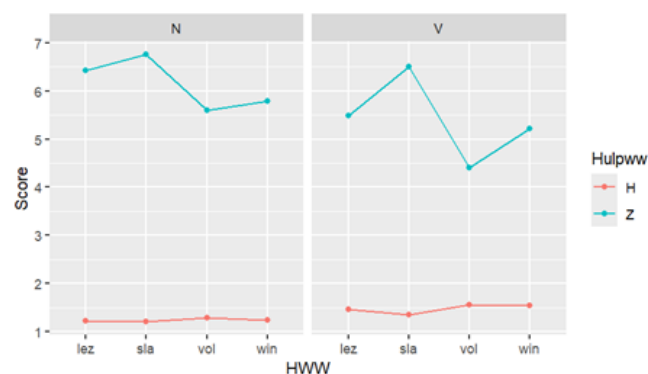
4.2.2.1.2 Hoofdwerkwoord

Tabel 4.22 toont de gemiddelde scores en standaardafwijkingen per hulpwerkwoord en per afzonderlijk hoofdwerkwoord (HWW).

Tabel 4.22: Exp2A - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww en HWW

Hulpww	HWW	M	SD
H	lezen	1.31	0.60
H	slapen	1.26	0.66
H	volgen	1.39	0.72
H	winnen	1.33	0.71
Z	lezen	6.12	1.36
Z	slapen	6.66	0.90
Z	volgen	5.20	1.97
Z	winnen	5.58	1.73

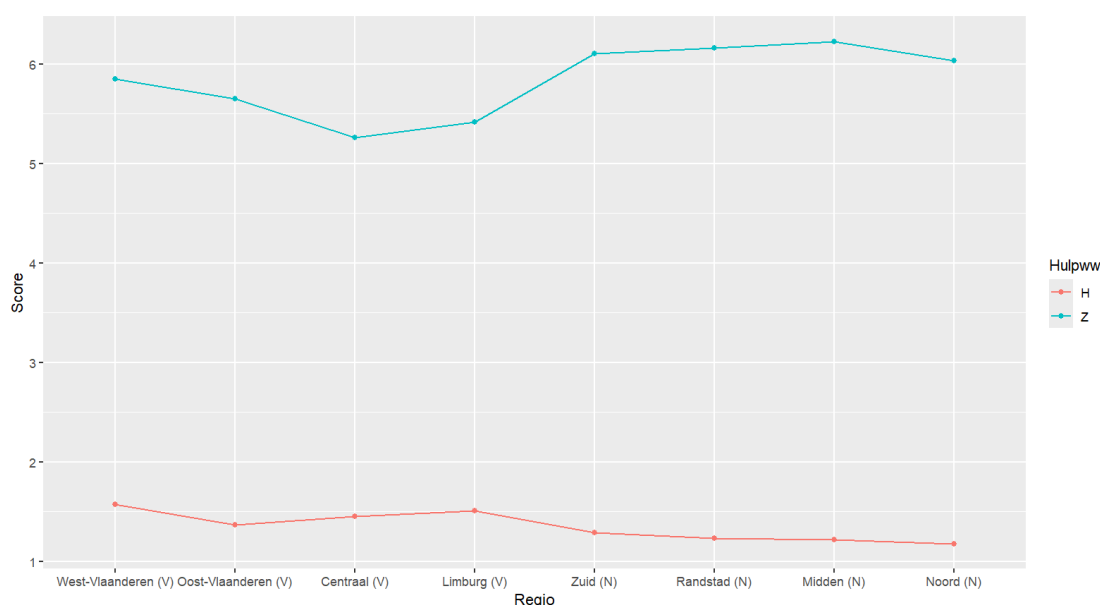
De gemiddelde score voor *hebben* verschilt weinig afhankelijk van het hoofdwerkwoord: de scores voor de HZ-hoofdwerkwoorden *volgen* ($M = 1.39$) en *winnen* ($M = 1.33$) liggen dicht bij die voor de H-hoofdwerkwoorden *lezen* ($M = 1.31$) en *slapen* ($M = 1.26$). Het verschil tussen de gemiddelden bij *zijn* is groter: *slapen* ($M = 6.66$) krijgt de hoogste score, gevolgd door *lezen* ($M = 6.12$), *winnen* ($M = 5.58$) en *volgen* ($M = 5.20$). Figuur 4.6 visualiseert de gemiddelde scores voor de afzonderlijke hoofdwerkwoorden per land en laat zien dat het effect van de hoofdwerkwoorden gelijkaardig is in Nederland en Vlaanderen.



Figuur 4.6: Exp2A - Gemiddelde acceptabiliteitsscore als functie van Hulpww, HWW en Land

4.2.2.1.3 Regio

Figuur 4.7 toont de gemiddelde score per hulpwerkwoord en per regio. De gemiddelde score voor *hebben* is het hoogst in West-Vlaanderen ($M = 1.57$). Daarna volgen Limburg ($M = 1.51$), Centraal-Vlaanderen ($M = 1.45$) en Oost-Vlaanderen ($M = 1.36$). In de Nederlandse regio's liggen de gemiddelden voor *hebben* consistent lager, met Zuid-Nederland bovenaan ($M = 1.29$), gevolgd door Randstad ($M = 1.23$) en Midden-Nederland ($M = 1.22$). In Noord-Nederland krijgt *hebben* de laagste scores ($M = 1.18$). Voor het hulpwerkwoord *zijn* ligt de gemiddelde score daarentegen het hoogst in Midden-Nederland ($M = 6.23$) en het laagst in Centraal-Vlaanderen ($M = 5.26$).



Figuur 4.7: Exp2A - Gemiddelde acceptabiliteitsscore als functie van Hulpww en Regio

4.2.2.2 Lineaire regressieanalyse

In de eerste plaats wordt voor Experiment 2A een basismodel opgesteld (§2.2.2.1), waarin het effect van TypeHWW en Hulpww, evenals hun interactie, wordt geanalyseerd. Vervolgens worden in de driewegsanalyses (§2.2.2.2) de covariaten toegevoegd. Eerst voeg ik Land toe aan een tweede model, en daarna stel ik modellen op met Opleiding, Geslacht en Leeftijdscategorie.

4.2.2.2.1 Basismodel

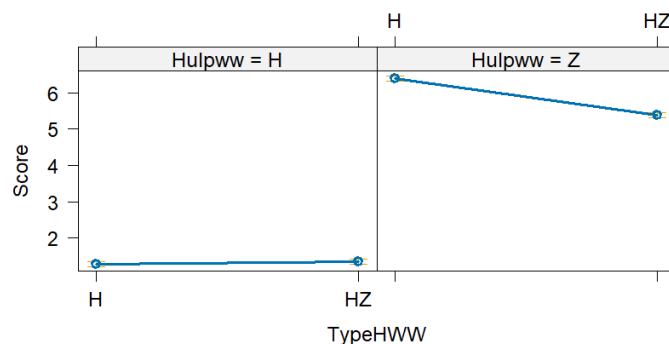
Het basismodel van Experiment 2A (AIC = 13504; marginale $R^2 = 0.791$; conditionele $R^2 = 0.813$) bevat Hulpww, TypeHWW en hun eventuele interactie. Tabel 4.23 toont de regressieoutput van het model.

Tabel 4.23: Exp2A – Regressieoutput basismodel (TypeHWW*Hulpww)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.28	0.03614	4124.13	35.477	< 0.001	***
TypeHWWHZ	0.08	0.04840	3164.94	1.583	0.114	
HulpwwZ	5.11	0.04873	3181.42	104.814	< 0.001	***
TypeHWWHZ:HulpwwZ	-1.08	0.06917	3183.79	-15.597	< 0.001	***

De regressiegegevens laten zien dat TypeHWW geen hoofdeffect is ($p = 0.114$). De varianten worden dus niet meer of minder aanvaard afhankelijk van het type hoofdwerkwoord van het Asp-cluster, ongeacht of het hulpwerkwoord *hebben* of *zijn* is. Het hulpwerkwoord is wel een hoofdeffect: *zijn* krijgt significant hogere scores dan *hebben* ($p < 0.001$). De significante interactie tussen TypeHWW en Hulpww toont aan

dat het type hoofdwerkwoord een significant effect heeft op de aanvaardbaarheid van *zijn* ($p < 0.001$). Uit de post-hoc analyse blijkt dat de score voor *zijn* significant lager is bij HZ-hoofdwerkwoorden (*volgen* en *winnen*) dan bij H-hoofdwerkwoorden (*slapen* en *lezen*), met een geschat verschil van -1.001 ($p < 0.001$). Op de aanvaardbaarheid van *hebben* heeft het type hoofdwerkwoord geen significant effect. De post-hoc analyse toont dat de score hoger is bij HZ-werkwoorden dan bij H-werkwoorden, met een geschat verschil van slechts 0.08 ($p = 0.114$). Figuur 4.8 visualiseert de interactie tussen TypeHWW en Hulpww.



Figuur 4.8: Exp2A - Plot interactie TypeHWW*Hulpww

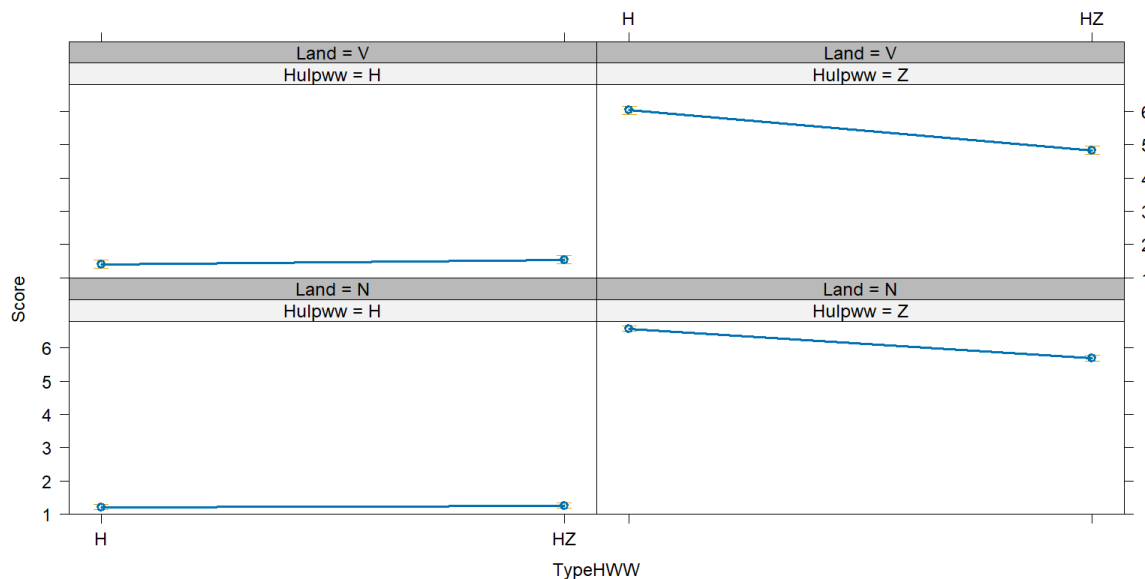
4.2.2.2.2 Driewegsanalyses

In de eerste plaats voeg ik aan het basismodel het covariaat Land toe, zodat in het model TypeHWW en Hulpww, Land, en alle mogelijke twee- en driewegsinteracties zijn opgenomen. Dit model (AIC = 13317.81; marginale $R^2 = 0.801$; conditionele $R^2 = 0.823$) is beter dan het vorige (AIC 13317.81 < AIC 13504). De regressieoutput van het model wordt weergegeven in Tabel 4.24.

Tabel 4.24: Exp2A – Regressieoutput driewegsinteractie (TypeHWW*Hulpww*Land)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.22	0.04390	4094.43	27.699	< 0.001	***
TypeHWWHZ	0.04	0.05852	3152.73	0.732	0.464	
HulpwwZ	5.37	0.05876	3166.77	91.307	< 0.001	***
LandV	0.19	0.07396	4101.55	2.563	0.010	*
TypeHWWHZ:HulpwwZ	-0.94	0.08344	3170.50	-11.233	< 0.001	***
TypeHWWHZ:LandV	0.09	0.09849	3170.36	0.944	0.345	
HulpwwZ:LandV	-0.74	0.09941	3188.72	-7.436	< 0.001	***
TypeHWWHZ:HulpwwZ:LandV	-0.40	0.14107	3189.68	-2.848	0.004	**

Het tweede model bevestigt de tendensen van het basismodel, met name het hoofdeffect van Hulpww ($p < 0.001$) en de interactie tussen TypeHWW en Hulpww ($p < 0.001$). Daarnaast voegt het model nationale diepgang toe. Het hoofdeffect van Land ($p = 0.010$) suggereert dat Vlamingen hogere scores toekennen aan alle varianten, onafhankelijk van de manipulaties. Daarnaast toont de significante interactie tussen Land en Hulpww ($p < 0.001$) dat *hebben* meer aanvaard wordt in Vlaanderen dan in Nederland. De driewegsinteractie tussen Land, TypeHWW en Hulpww is significant ($p = 0.004$), wat betekent dat de interactie tussen TypeHWW en Hulpww verschillend is in Vlaanderen en Nederland. De post-hoc analyse toont aan dat het geschatte verschil in de scores voor *zijn* tussen H-hoofdwerkwoorden en HZ-hoofdwerkwoorden groter is in Vlaanderen ($estimate = -1.20$, $p < 0.001$) dan in Nederland ($estimate = -0.89$, $p < 0.001$), wat suggereert dat het negatieve effect van HZ-hoofdwerkwoorden voor *zijn* sterker is in Vlaanderen dan in Nederland. Een hoofdwerkwoordeffect voor *hebben* blijft zowel in Vlaanderen als in Nederland afwezig. Figuur 4.9 visualiseert de driewegsinteractie.



Figuur 4.9: Exp2A – Plot driewegsinteractie TypeHWW*Hulpww*Land

In de tweede plaats voer ik bijkomende driewegsanalyses uit waarbij Opleiding, Geslacht en Leeftijdscategorie aan regressiemodellen worden toegevoegd. In de modellen neem ik steeds de twee *fixed effects* TypeHWW en Hulpww op, een van de covariaten, en alle twee- en driewegsinteracties. Ik beperk me hier tot het bespreken van de driewegsinteracties. Zowel in het model met Opleiding (AIC = 13507.23; marginale $R^2 = 0.792$; conditionele $R^2 = 0.814$), als in het model met Geslacht (AIC = 13507.03; marginale $R^2 = 0.792$; conditionele $R^2 = 0.813$) en het model met Leeftijdscategorie (AIC = 13502.23; marginale $R^2 = 0.793$; conditionele $R^2 = 0.815$) is de driewegsinteractie niet significant (respectievelijk $p = 0.754$, $p = 0.076$ en $p = 0.092$). Dat betekent dat de interactie tussen TypeHWW en Hulpww niet significant verschillend is afhankelijk van opleidingsniveau, geslacht of leeftijdscategorie.

4.2.2.3 Conclusie: Experiment 2A

In Experiment 2A wordt aangetoond dat het hulpwerkwoord *hebben* in Asp-clusters significant minder aanvaard wordt dan *zijn*. Dat bevestigt de basishypothese. Zowel in Vlaanderen als in Nederland bestaat een voorkeur voor *zijn*, maar in Vlaanderen wordt *hebben* significant meer aanvaard dan in Nederland. Wat regionale variatie betreft, krijgt *hebben* gemiddeld de hoogste score in West-Vlaanderen en de laagste in Noord-Nederland. De geografische hypothese vindt dus ondersteuning in de resultaten. De hoofdworkeffecthypothese daarentegen wordt niet bevestigd in Experiment 2A: het hulpwerkwoord *hebben* wordt in beide landen niet meer aanvaard bij H-hoofdworkeffecten dan bij HZ-hoofdworkeffecten. Op de aanvaardbaarheid van *zijn* zien we wel een significant effect van het type hoofdworkeffect: de score voor *zijn* is significant lager bij HZ-hoofdworkeffecten dan bij H-hoofdworkeffecten. Dat hoofdworkeffect voor *zijn* is in Vlaanderen sterker dan in Nederland.

4.2.3 Experiment 2B

In Experiment 2B wordt net als in Experiment 2A de invloed van het type hoofdwerkwoord onderzocht, maar dan in Mod-clusters. Ik test met name of het hulpwerkwoord *zijn* meer aanvaard wordt bij hoofdwerkwoorden die *zijn* selecteren (Z-hoofdwerkwoorden) dan bij hoofdwerkwoorden die zowel *hebben* als *zijn* kunnen selecteren (HZ-hoofdwerkwoorden). Doordat in Experiment 2B ook de nationale en regionale variatie voor Mod-clusters in kaart wordt gebracht, kunnen de resultaten dienen als referentiepunt voor die van Experiment 2A. In de verkennende analyse (§4.2.3.1) komen eerst de variabelen Hulpww, TypeHWW en Land aan bod (§4.2.3.1.1) en daarna behandel ik de variabelen HWW (§4.2.3.1.2) en Regio (§4.2.3.1.3), die niet zijn opgenomen in de regressieanalyse (zie §4.1.5). Vervolgens beschrijf ik de resultaten van de lineaire regressie (§4.2.3.2), en in §4.2.3.3 vat ik de resultaten samen.

4.2.3.1 Verkennende analyse

4.2.3.1.1 Hulpwerkwoord, Type hoofdwerkwoord en Land

In Tabel 4.25 worden de gemiddelde scores en standaarddeviaties voor Experiment 2B getoond per hulpwerkwoord, type hoofdwerkwoord (HZ vs. Z), land (Nederland vs. Vlaanderen) en in totaal.

Tabel 4.25: Exp2B - Gemiddelde acceptabiliteitscore (met standaarddeviatie) als functie van Hulpww, TypeHWW en Land

Hulpww	TypeHWW	Nederland		Vlaanderen		Totaal	
		M	SD	M	SD	M	SD
H	HZ	4.34	2.25	4.02	2.15	4.22	2.22
H	Z	4.56	2.16	2.82	1.85	3.93	2.22
Z	HZ	1.78	1.44	3.55	2.20	2.41	1.94
Z	Z	2.54	1.98	5.30	1.81	3.52	2.33

De gemiddelde scores voor de hulpwerkwoorden liggen in Experiment 2B relatief dicht bij elkaar, hoewel de score voor *hebben* ($M = 4.08$) gemiddeld hoger is dan die voor *zijn* ($M = 2.96$). De standaarddeviaties voor *hebben* ($SD = 2.22$) en *zijn* ($SD = 2.21$) zijn vrijwel identiek, wat wijst op een vergelijkbare variabiliteit in de oordelen over beide hulpwerkwoorden. Uit Tabel 4.25 blijkt een diametrale omkeer van voorkeur in Vlaanderen en Nederland: de gemiddelde score voor *zijn* ligt beduidend hoger in Vlaanderen ($M = 4.42$) dan in Nederland ($M = 2.16$), terwijl *hebben* gemiddeld een hogere score krijgt in Nederland ($M = 4.45$) dan in Vlaanderen ($M = 3.31$). De standaardafwijking voor *zijn* in Nederland ($SD = 1.77$) ligt lager in vergelijking met die voor *zijn* in Vlaanderen ($SD = 2.19$) en die voor *hebben* in beide landen ($SD = 2.21$, $SD = 2.09$), wat wijst op een kleinere variabiliteit in de Nederlandse oordelen over *zijn* in Mod-clusters.

Wat het type hoofdwerkwoord betreft, ligt de gemiddelde score voor *zijn* hoger bij de Z-hoofdwerkwoorden *stoppen* en *trouwen* ($M = 3.93$) dan bij de HZ-hoofdwerkwoorden *volgen* en *winnen* ($M = 2.41$). Dat positieve effect van Z-hoofdwerkwoorden op de aanvaardbaarheid van *zijn* is zowel in Nederland (HZ ($M = 1.78$) < Z ($M = 2.54$)) als in Vlaanderen (HZ ($M = 3.55$) < Z ($M = 5.30$)) aanwezig. Voor het hulpwerkwoord *hebben* is de score gemiddeld hoger bij de HZ-hoofdwerkwoorden ($M = 4.22$) dan bij de Z-hoofdwerkwoorden ($M = 3.93$). Dat positieve effect van HZ-hoofdwerkwoorden op de scores van *hebben* zien we echter enkel in Vlaanderen (HZ ($M = 4.02$) > Z ($M = 2.82$)). In Nederland liggen de scores voor *hebben* dichter bij elkaar (HZ ($M = 4.34$) < Z ($M = 4.56$)).

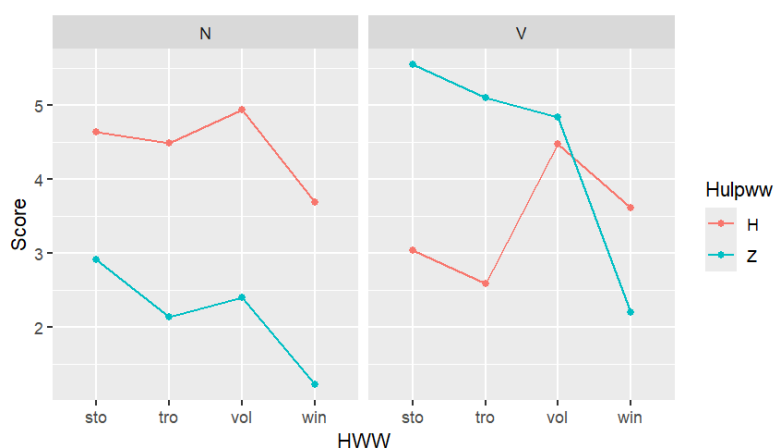
4.2.3.1.2 Hoofdwerkwoord

Tabel 4.26 toont de gemiddelde scores en standaardafwijkingen per hulpwerkwoord en per afzonderlijk hoofdwerkwoord (HWW). Voor *hebben* is de gemiddelde score het hoogst bij *volgen* ($M = 4.79$), gevolgd door *stoppen* ($M = 4.03$), *trouwen* ($M = 3.84$), en het laagst bij *winnen* ($M = 3.66$). Voor *zijn* is de gemiddelde score opvallend lager bij *winnen* ($M = 1.55$) dan bij *stoppen* ($M = 3.79$), *volgen* ($M = 3.31$) en *trouwen* ($M = 3.25$). Ook de standaardafwijking van *winnen* bij *zijn* ($SD = 1.02$) is opmerkelijk lager dan de rest, wat wijst op een hoge mate van uniformiteit in de afwijzing van *zijn* in combinatie met het hoofdwerkwoord *winnen*.

Tabel 4.26: Exp2B - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww en HWW

Hulpww	HWW	M	SD
H	stoppen	4.03	2.19
H	trouwen	3.84	2.24
H	volgen	4.79	2.10
H	winnen	3.66	2.19
Z	stoppen	3.79	2.33
Z	trouwen	3.25	2.30
Z	volgen	3.31	2.25
Z	winnen	1.55	1.02

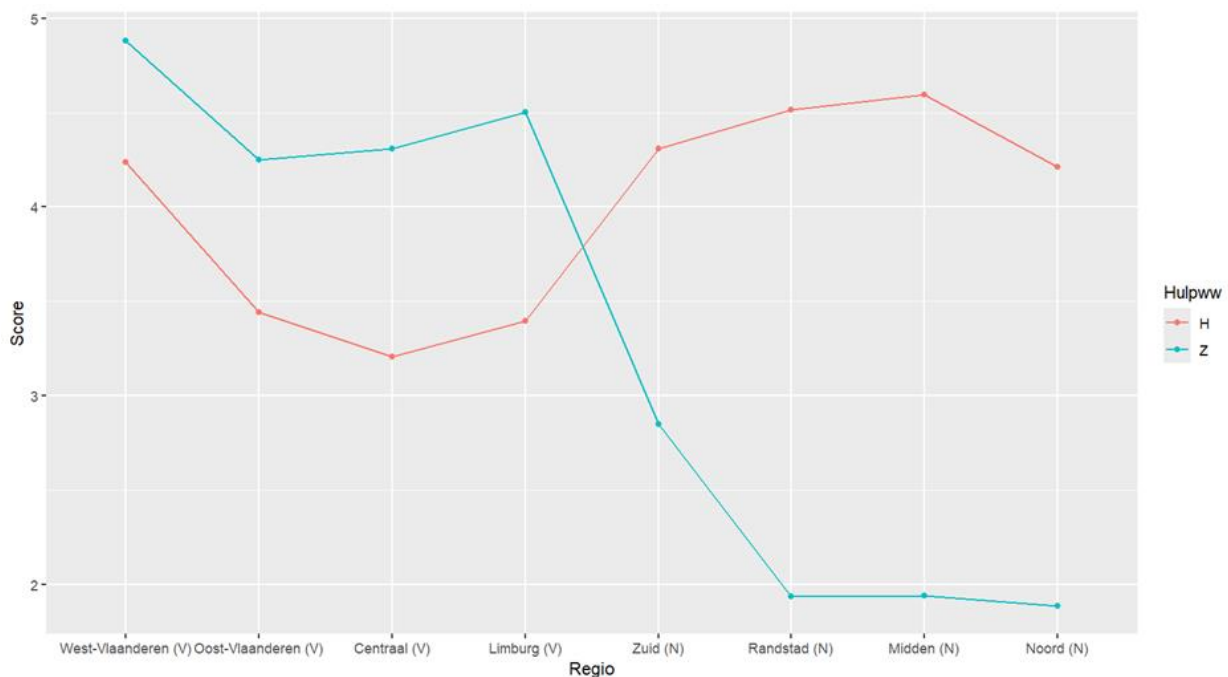
Figuur 4.10 visualiseert de gemiddelde scores per hulpwerkwoord en hoofdwerkwoord, en per land. De figuur bevestigt ten eerste de opmerkelijke status van het HZ-hoofdwkwoord *winnen* in combinatie met *zijn* in Mod-clusters. Hoewel *winnen* over het algemeen minder geaccepteerd wordt dan de andere hoofdwkwoorden, aanvaarden de Vlaamse respondenten *winnen* met *zijn* meer ($M = 2.20$) dan Nederlandse respondenten ($M = 1.23$). Toch gaat de voorkeur in Vlaanderen uit naar het hulpwerkwoord *hebben* bij Mod-clusters met *winnen*, terwijl bij de andere hoofdwkwoorden de voorkeur gegeven wordt aan *zijn*. De lage score voor *zijn* in Vlaanderen zorgt op die manier voor een doorbreking van de diametrale omkeer in de voorkeur voor het hulpwerkwoord in Nederland en Vlaanderen. Daarnaast valt ook het nationale verschil in de gemiddelde scores voor het HZ-hoofdwkwoord *volgen* op. In Vlaanderen worden *hebben* ($M = 4.48$) en *zijn* ($M = 4.83$) evenveel aanvaard, terwijl er in Nederland bij Mod-clusters met *volgen* een duidelijke voorkeur bestaat voor *hebben* ($M = 4.94$) ten opzichte van *zijn* ($M = 2.40$).



Figuur 4.10: Exp2B - Gemiddelde acceptabiliteitsscore als functie van Hulpww, HWW en Land

4.2.3.1.3 Regio

Figuur 4.11 toont de gemiddelde score per hulpwerkwoord en per regio. De gemiddelde score voor *zijn* is het hoogst in West-Vlaanderen ($M = 4.88$). Daarna volgen Limburg ($M = 4.50$), Centraal-Vlaanderen ($M = 4.31$) en Oost-Vlaanderen ($M = 4.25$). In de Nederlandse regio's liggen de scores voor *zijn* consistent lager dan in de Vlaamse. In Noord-Nederland krijgt *zijn* gemiddeld de laagste score ($M = 1.89$), in Midden-Nederland ($M = 1.94$) en Randstad ($M = 1.94$) ligt die iets hoger. In Zuid-Nederland ligt het gemiddelde beduidend hoger ($M = 2.85$) dan in de andere Nederlandse regio's. Wat het hulpwerkwoord *hebben* betreft liggen de laagste scores in Centraal-Vlaanderen ($M = 3.20$), Limburg ($M = 3.40$) en Oost-Vlaanderen ($M = 3.44$). De gemiddelde score in West-Vlaanderen ($M = 4.24$) daarentegen ligt hoger dan die in Noord-Nederland ($M = 4.21$) en is gelijkaardig aan die in Zuid-Nederland ($M = 4.31$). De hoogste scores voor *hebben* vinden we in Randstad ($M = 4.52$) en Midden-Nederland ($M = 4.59$).



Figuur 4.11: Exp2B - Gemiddelde acceptabiliteitscore als functie van Hulpww en Regio

Figuur 4.11 laat op die manier duidelijk het diametrale nationale verschil in de resultaten van Experiment 2B zien. De *zijn*-lijn en de *hebben*-lijn kruisen elkaar immers net tussen Limburg (V) en Zuid (N), waardoor de omkeer van de selectievoorkoor tussen Vlaanderen en Nederland wordt gevisualiseerd. Daarnaast liggen de gemiddelde scores in de Vlaamse regio's steeds dichterbij elkaar dan in Nederland, wat illustreert dat er in Nederland een categorischer onderscheid wordt gemaakt tussen *hebben* en *zijn*. Tot slot maakt Figuur 4.11 ook de opmerkelijke status van West-Vlaanderen duidelijk, waar *hebben* meer dan in de andere Vlaamse provincies en evenveel als in Nederland aanvaard wordt.

4.2.3.2 Lineaire regressieanalyse

Net zoals bij Experiment 2A wordt in het basismodel van Experiment 2B (§4.2.3.2.1) het effect van de variabelen TypeHWW en Hulpww en hun interactie geanalyseerd. In §4.2.3.2.2 worden de covariaten aan regressiemodellen toegevoegd.

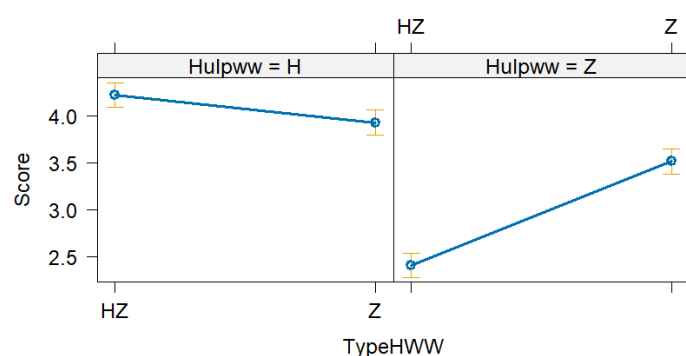
4.2.3.2.1 Basismodel

Het basismodel van Experiment 2B (AIC = 18430.87; marginale $R^2 = 0.092$; conditionele $R^2 = 0.186$) bevat TypeHWW, Hulpww en hun eventuele interactie. Tabel 4.27 toont de regressieoutput van het model.

Tabel 4.27: Exp2B – Regressieoutput basismodel (TypeHWW*Hulpww)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	4.22	0.06750	4092.13	62.564	< 0.001	***
TypeHWWZ	-0.29	0.09053	3176.75	-3.223	0.001	**
HulpwwZ	-1.82	0.08996	3166.72	-20.199	< 0.001	***
TypeHWWZ:HulpwwZ	1.40	0.12756	3170.21	10.984	< 0.001	***

De regressiegegevens tonen aan dat TypeHWW een hoofdeffect is ($p = 0.001$). De hulpwerkwoorden worden dus, ongeacht of het om *hebben* of *zijn* gaat, meer of minder aanvaard afhankelijk van het type hoofdwerkwoord. Ook Hulpww is een hoofdeffect ($p < 0.001$): het hulpwerkwoord *hebben* wordt significant meer aanvaard dan hulpwerkwoord *zijn*. Daarnaast is de interactie tussen TypeHWW en Hulpww ook statistisch significant ($p < 0.001$). Uit de post-hoc analyse blijkt dat de scores voor *zijn* significant hoger liggen bij de Z-hoofdwerkwoorden (*stoppen* en *trouwen*) dan bij de HZ-hoofdwerkwoorden (*volgen* en *winnen*) ($estimate = 1.40$, $p < 0.001$). Voor *hebben* blijkt uit de post-hoc analyse dat de scores significant lager liggen bij de Z-hoofdwerkwoorden dan bij de HZ-hoofdwerkwoorden ($estimate = -0.29$, $p = 0.001$). Het verschil in schatters toont dat het negatieve effect van HZ-werkwoorden voor *zijn* groter is dan het negatieve effect van Z-werkwoorden voor *hebben*. Figuur 4.12 visualiseert de significante interactie tussen TypeHWW en Hulpww.



Figuur 4.12: Exp2B - Plot interactie TypeHWW*Hulpww

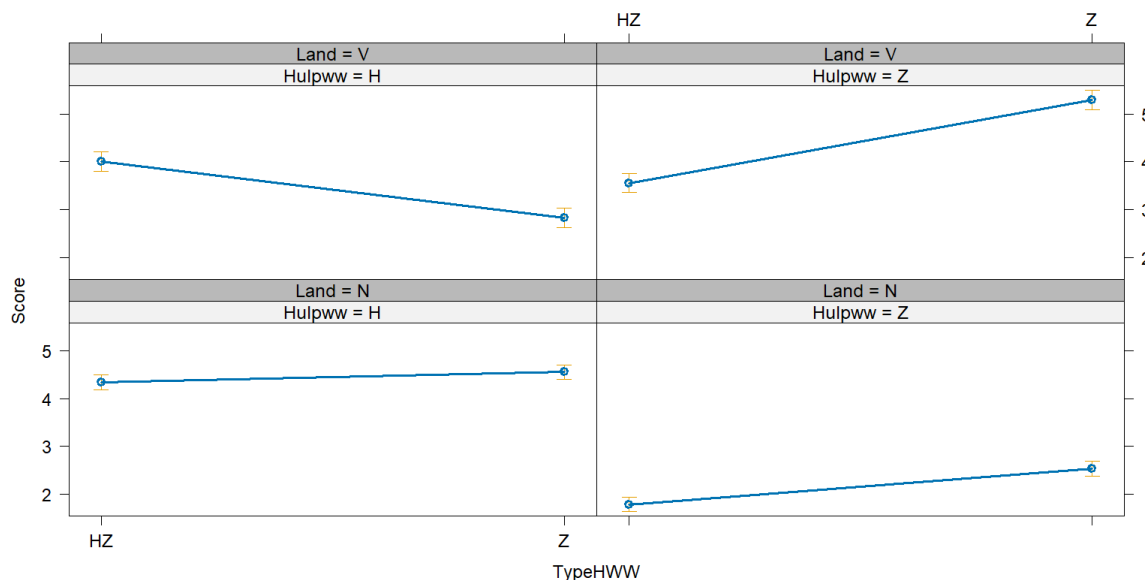
4.2.3.2.2 Driewegsanalyses

In de eerste plaats neem ik in een tweede model de *fixed effects* TypeHWW en Hulpww op, samen met het covariaat Land en alle mogelijke twee- en driewegsinteracties. Dit model (AIC = 17603.17; marginale $R^2 = 0.244$; conditionele $R^2 = 0.364$) is beter dan het basismodel (AIC 17603.17 < AIC 18430.87). Tabel 4.28 bevat de regressieoutput van het model.

Tabel 4.28: Exp2B – Regressieoutput driewegsinteractie (TypeHWW*Hulpww*Land)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	4.34	0.07673	3951.99	56.577	< 0.001	***
TypeHWWZ	0.21	0.09998	3168.97	2.147	0.032	*
HulpwwZ	-2.56	0.09901	3150.16	-25.849	< 0.001	***
LandV	-0.33	0.12852	3953.13	-2.498	0.009	**
TypeHWWZ:HulpwwZ	0.54	0.14060	3158.54	3.828	< 0.001	***
TypeHWWZ:LandV	-1.40	0.16703	3163.56	-8.370	< 0.001	***
HulpwwZ:LandV	2.11	0.16643	3163.76	12.655	< 0.001	***
TypeHWWZ:HulpwwZ:LandV	2.38	0.23571	3161.87	10.102	< 0.001	***

Het tweede model bevestigt de effecten uit het basismodel, met name de hoofdeffecten van TypeHWW ($p = 0.032$) en Hulpww ($p = < 0.001$), en de interactie tussen TypeHWW en Hulpww ($p < 0.001$), maar verdiept het ook met nationale tendensen. Het hoofdeffect van Land ($p = 0.009$) laat ten eerste zien dat Vlamingen, onafhankelijk van de manipulaties, lagere scores toekennen in alle condities. Daarnaast laat de significante interactie tussen Land en Hulpww ($p < 0.001$) zien dat de aanvaardbaarheid van het hulpwerkwoord afhankelijk is van het land. De post-hoc analyse toont aan dat *hebben* significant minder aanvaard wordt in Vlaanderen ($estimate = -1.04, p < 0.001$), terwijl *zijn* significant minder aanvaard wordt in Nederland ($estimate = 2.26, p < 0.001$). De driewegsinteractie tussen Land, TypeHWW en Hulpww is ook significant ($p < 0.001$), wat betekent dat de interactie tussen TypeHWW en Hulpww verschillend werkt in Vlaanderen en Nederland. De post-hoc analyse laat zien dat *hebben* in Vlaanderen significant lagere scores krijgt bij Z-hoofdwerkwoorden dan bij HZ-hoofdwerkwoorden ($estimate = -1.18, p < 0.001$), terwijl het type hoofdwerkwoord in Nederland geen significant effect op de scores voor *hebben* ($p = 0.129$). Voor *zijn* daarentegen blijkt uit de post-hoc analyse dat het significant meer wordt aanvaard bij Z-hoofdwerkwoorden dan bij HZ-hoofdwerkwoorden, zowel in Nederland ($estimate = 0.75, p < 0.001$) als in Vlaanderen ($estimate = 1.74, p < 0.001$). In Figuur 4.13 wordt de driewegsinteractie gevisualiseerd.



Figuur 4.13: Exp2B - Plot driewegsinteractie TypeHWW*Hulpww*Land

In de tweede plaats voer ik driewegsanalyses uit waarbij Opleiding, Geslacht en Leeftijdscategorie in regressiemodellen worden opgenomen. De modellen bevatten steeds de twee *fixed effects* TypeHWW en Hulpww, een van de covariaten, en alle twee- en driewegsinteracties ertussen. Ik beperk me hier tot het bespreken van de driewegsinteracties. In het model met Opleiding (AIC = 18438.49; marginale R^2 = 0.093; conditionele R^2 = 0.187) is de driewegsinteractie niet significant (p = 0.496), net als in de modellen met Geslacht (AIC = 18424.59; marginale R^2 = 0.098; conditionele R^2 = 0.188) en Leeftijdscategorie (AIC = 18382.85; marginale R^2 = 0.11; conditionele R^2 = 0.199) (respectievelijk p = 0.604 en p = 0.390). De interactie tussen het hulpwerkwoord en het type hoofdwerkwoord is in Experiment 2B dus onafhankelijk van het opleidingsniveau, het geslacht of de leeftijdscategorie van de respondenten.

4.2.3.3 Conclusie: Experiment 2B

De resultaten van Experiment 2B tonen aan dat de aanvaardbaarheid van *hebben* en *zijn* in Mod-clusters sterk afhankelijk is van nationale variatie. In Nederland bestaat een significante voorkeur voor *hebben*, in Vlaanderen een significante voorkeur voor *zijn*. Wat regionale variatie betreft zien we dat in Zuid-Nederland *zijn* meer aanvaard wordt dan in de andere Nederlandse regio's, en dat in West-Vlaanderen *hebben* meer aanvaard wordt dan in de andere Vlaamse regio's. Wat het effect van het type hoofdwerkwoord betreft toont Experiment 2B aan dat het type hoofdwerkwoord een significante invloed heeft op hulpwerkwoordselectie in Mod-clusters. Het hulpwerkwoord *zijn* wordt meer aanvaard in Mod-clusters met een Z-hoofdwerkwoord dan met een HZ-hoofdwerkwoord, zowel in Vlaanderen als in Nederland. Het hulpwerkwoord *hebben* wordt daarnaast significant meer aanvaard bij HZ- dan bij Z-hoofdwerkwoorden. Dat effect zien we enkel in Vlaanderen, niet in Nederland. Uit de analyse van de afzonderlijke hoofdwerkwoorden blijkt daarbij dat *zijn* bij Mod-clusters met het HZ-hoofdwerkwoord *winnen* in Vlaanderen veel minder aanvaard wordt dan bij de andere hoofdwerkwoorden.

4.2.4 Experiment 3

In Experiment 3 wordt de invloed van modaliteit (irrealis vs. realis) op hulpwerkwoordselectie onderzocht. Ik test de basisaanname voor de irrealishypothese dat een irrealislezing een positieve invloed heeft op de aanvaardbaarheid van *hebben* bij hoofdwerkwoorden die *zijn* selecteren in voltooid aspect. De hulpwerkwoordselectie werd onderzocht bij de werkwoorden *gaan*, *komen*, *blijven* en *beginnen* wanneer ze niet als groepsvormend werkwoord, maar als hoofdwerkwoord fungeren. Zowel in de verkennende analyse (§4.2.4.1) als in de regressieanalyse (§4.2.4.2) komen de variabelen Hulpww, Realis en Land aan bod. In §4.2.4.3 vat ik de resultaten van Experiment 3 samen.

4.2.4.1 Verkennende analyse

Tabel 4.29 toont de gemiddelde score en standaardafwijking voor Experiment 3 per hulpwerkwoord, per modaliteit (Realis, irr vs. rea), per land (Nederland vs. Vlaanderen) en in totaal.

Tabel 4.29: Exp3 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Land

Hulpww	Realis	Nederland		Vlaanderen		Totaal	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
H	rea	1.28	0.69	1.50	0.98	1.36	0.81
H	irr	1.55	1.04	2.15	1.53	1.76	1.26
Z	rea	6.38	1.20	5.96	1.54	6.23	1.34
Z	irr	1.96	1.42	2.78	1.86	2.25	1.64

De gemiddelde score voor *hebben* ($M = 1.56$) ligt beduidend lager dan die voor *zijn* ($M = 4.26$). Daarnaast is de standaardafwijking voor *hebben* ($SD = 1.08$) kleiner dan die voor *zijn* ($SD = 2.49$), wat impliceert dat de acceptabiliteitsoordelen over *hebben* minder variatie vertonen dan die voor *zijn*. De voorkeur voor *zijn* is zowel in Nederland als in Vlaanderen aanwezig, maar in Nederland ligt de gemiddelde score voor zowel *hebben* ($M = 1.42$) als *zijn* ($M = 4.19$) lager dan die voor *hebben* ($M = 1.82$) en *zijn* ($M = 4.39$) in Vlaanderen. Daarnaast zien we dat de variatie in de oordelen over *hebben* groter is in Vlaanderen ($SD = 1.32$) dan in Nederland ($SD = 0.89$), waar er meer consensus bestaat over de onaanvaardbaarheid van *hebben*.

Wat de modaliteit betreft ligt de gemiddelde score voor *hebben* hoger bij de irrealis ($M = 1.76$) dan bij de realis ($M = 1.36$). Het hulpwerkwoord *zijn* scoort daarentegen beduidend hoger in de realis ($M = 6.23$) dan in de irrealis ($M = 2.25$). Uit Tabel 4.29 blijkt dat het positieve effect van irrealis op *hebben* en het negatieve effect van irrealis op *zijn* zowel in Vlaanderen als in Nederland voorkomen.

4.2.4.2 Lineaire regressieanalyse

In het basismodel van Experiment 3 (§4.2.4.2.1) onderzoek ik de invloed van de variabelen Realis en Hulpww, evenals hun interactie, op de afhankelijke variabele Score. In de driewegsanalyses (§4.2.4.2.2) worden de covariaten Land, Opleiding, Geslacht en Leeftijdscategorie aan modellen toegevoegd.

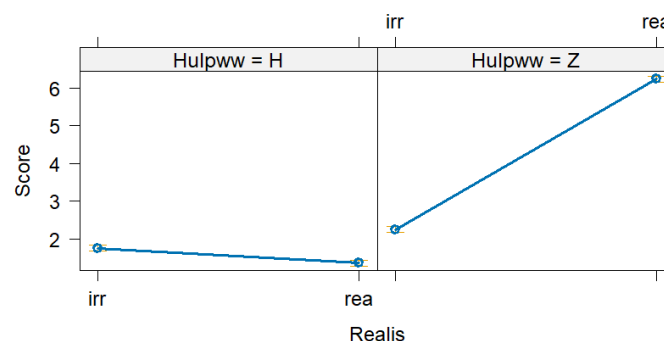
4.2.4.2.1 Basismodel

In het basismodel ($AIC = 14200.28$; marginale $R^2 = 0.695$; conditionele $R^2 = 0.729$) worden Realis, Hulpww en hun eventuele interactie opgenomen. Tabel 4.30 geeft de regressieoutput van het basismodel weer.

Tabel 4.30: Exp3 – Regressieoutput basismodel (Realis*Hulpww)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.76	0.03944	4099.48	44.59	< 0.001	***
Realisrea	-0.40	0.05248	3172.33	-7.61	< 0.001	***
HulpwwZ	0.49	0.05305	3195.45	9.23	< 0.001	***
Realisrea:HulpwwZ	4.38	0.07477	3188.50	58.62	< 0.001	***

In Tabel 4.30 zien we dat Realis een hoofdeffect is ($p < 0.001$), wat suggereert dat de varianten meer worden aanvaard in realis dan in irrealis, ongeacht of het om *hebben* of *zijn* gaat. Daarnaast is Hulpww ook een hoofdeffect ($p < 0.001$): het hulpwerkwoord *zijn* wordt, onafhankelijk van de modaliteit, geprefereerd boven *hebben*. Tot slot is ook de interactie tussen Realis en Hulpww statistisch significant ($p < 0.001$). De post-hoc analyse toont aan dat de score voor *hebben* significant lager is bij realis dan bij irrealis ($estimate = -0.40$, $p < 0.001$), en voor *zijn* significant hoger bij realis dan bij irrealis ($estimate = 3.89$, $p < 0.001$). Figuur 4.14 visualiseert de interactie tussen Hulpww en Realis.



Figuur 4.14: Exp3 – Plot interactie Realis*Hulpww

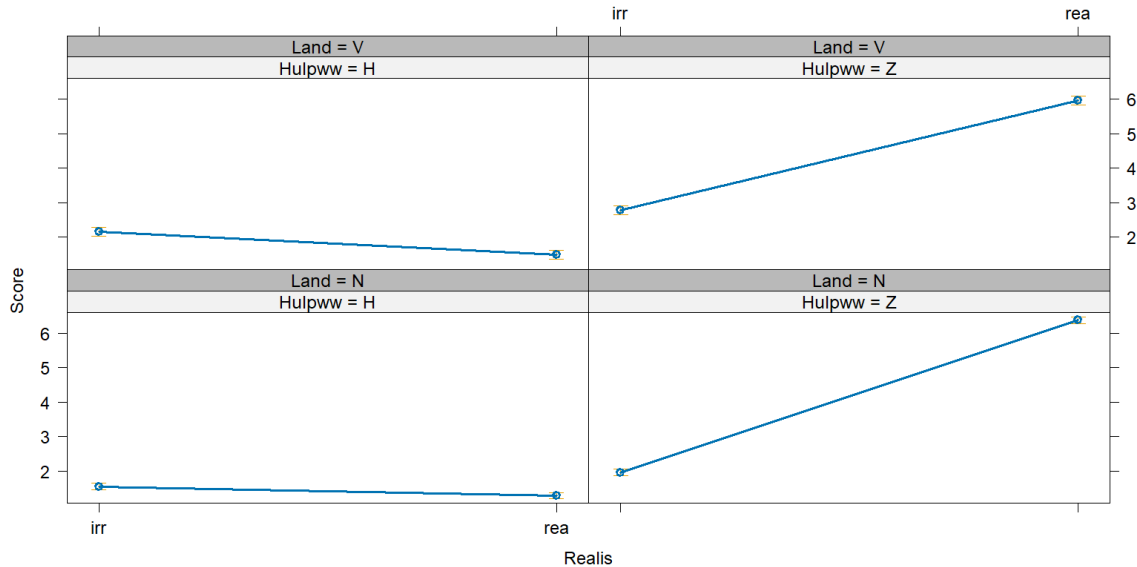
4.2.4.2.2 Driewegsanalyse

In de eerste plaats voer ik een driewegsanalyse uit waarbij ik het covariaat Land toevoeg aan het regressiemodel, zodat het Hulpww, Realis, Land en de twee- en driewegsinteracties ertussen bevat. Dit model (AIC = 14031.45; marginale $R^2 = 0.707$; conditionele $R^2 = 0.742$) is beter dan het basismodel (AIC 14031.45 < AIC 14200.28). De regressieoutput van model wordt weergegeven in Tabel 4.31.

Tabel 4.31: Exp3 – Regressieoutput driewegsinteractie (Realis*Hulpww*Land)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.55	0.04783	4082.22	32.355	< 0.001	***
Realisrea	-0.26	0.06358	3158.61	-4.142	< 0.001	***
HulpwwZ	0.41	0.06420	3180.93	6.422	< 0.001	***
LandV	0.61	0.08103	4093.68	7.482	< 0.001	***
Realisrea:HulpwwZ	4.68	0.09055	3174.21	51.716	< 0.001	***
Realisrea:LandV	-0.39	0.10755	3179.95	-3.655	< 0.001	***
HulpwwZ:LandV	0.22	0.10884	3203.59	2.043	0.041	*
Realisrea:HulpwwZ:LandV	-0.86	0.15331	3196.44	-5.597	< 0.001	***

Het tweede model bevestigt de effecten van het basismodel, namelijk de hoofdeffecten van Hulpww ($p < 0.001$) en Realis ($p < 0.001$) en de interactie tussen Realis en Hulpww ($p < 0.001$), maar verdiept het ook met nationale tendensen. Ten eerste toont het hoofdeffect van Land ($p < 0.001$) dat Vlamingen, los van alle manipulaties, in alle condities hogere scores toekennen. Daarnaast toont de significante interactie ($p = 0.041$) tussen Hulpww en Land dat *hebben* over het algemeen meer aanvaard wordt in Vlaanderen dan in Nederland. De significante interactie tussen Realis en Land ($p < 0.001$) laat vervolgens zien dat de modaliteit verschillend aanvaard wordt in beide landen. Uit de post-hoc analyse blijkt dat de irrealislezing, onafhankelijk van het hulpwoord, significant meer aanvaard wordt in Vlaanderen dan in Nederland ($estimate = 0.714$, $p < 0.001$), terwijl er geen significant verschil is in de aanvaardbaarheid van de realislezing tussen de landen ($p = 0.629$). Tot slot laat de significante driewegsinteractie ($p < 0.001$) zien dat de interactie tussen Realis en Hulpww verbonden is aan nationale effecten. Uit de post-hoc analyse blijkt dat het geschatte verschil tussen irrealis en realis bij *hebben* groter is in Vlaanderen ($estimate = -0.66$, $p < 0.001$) dan in Nederland ($estimate = -0.26$, $p < 0.001$). Dat laat zien dat het irrealiseffect voor *hebben* sterker is in Vlaanderen dan in Nederland. Wat *zijn* betreft toont de post-hoc analyse aan dat het geschatte verschil tussen irrealis en realis groter is in Nederland ($estimate = 4.42$, $p < 0.001$) dan in Vlaanderen ($estimate = 3.17$, $p < 0.001$), wat erop wijst dat het positieve effect van realis op de aanvaardbaarheid van *zijn* sterker is in Nederland dan in Vlaanderen. In Figuur 4.15 wordt de driewegsinteractie gevisualiseerd.



Figuur 4.15: Exp3 - Plot driewegsinteractie Realis*Hulpww*Land

In de tweede plaats voer ik driewegsanalyses uit waarbij ik Opleiding, Geslacht en Leeftijdscategorie toevoeg aan regressiemodellen. In die modellen neem ik steeds Hulpww en Realis, een van de covariaten, en alle mogelijke twee- en driewegsinteracties op. Ik beperk me tot de bespreking van de driewegsinteracties. Zowel in het model met Opleiding (AIC = 14191; marginale $R^2 = 0.696$; conditionele $R^2 = 0.731$), als in het model met Geslacht (AIC = 14211.65; marginale $R^2 = 0.695$; conditionele $R^2 = 0.730$) en Leeftijdscategorie (AIC = 14207.14; marginale $R^2 = 0.697$; conditionele $R^2 = 0.730$) is er geen significante driewegsinteractie (respectievelijk $p = 0.189$, $p = 0.438$ en $p = 0.652$). Dat toont aan dat de interactie tussen de modaliteit en het hulpwerkwoord in Experiment 3 onafhankelijk is van het opleidingsniveau, het geslacht en de leeftijdscategorie van de respondenten.

4.2.4.3 Conclusie: Experiment 3

Op basis van Experiment 3 concludeer ik dat een irrealislezing een significant positief effect heeft op de aanvaardbaarheid van *hebben* bij de hoofdwerkwoorden *gaan*, *blijven*, *komen* en *beginnen*. Dat irrealiseffect is sterker in Vlaanderen dan in Nederland. Daarnaast heeft de irrealislezing in Experiment 3 een significant negatief effect op de aanvaardbaarheid van *zijn*. Dat effect is sterker in Nederland.

4.2.5 Experiment 4

In Experiment 4 test ik specifiek de irrealishypothese. Ik onderzoek met name of een irrealislezing een positieve invloed heeft op de acceptabiliteitscores van *hebben* in Asp-clusters. In de verkennende analyse (§4.2.5.1) bespreek ik eerst de variabelen Hulpww, Realis en Land (§4.2.5.1.1), en daarna in meer detail de variabele Inversie (§4.2.5.1.2), die niet wordt opgenomen in de regressieanalyse wegens de overlap met Realis (zie §4.1.5). In §4.2.5.2 licht ik de resultaten van de lineaire regressieanalyse toe, en in §4.2.5.3 vat ik de resultaten samen in het licht van de hypothesen.

4.2.5.1 Verkennende analyse

4.2.5.1.1 Hulpwerkwoord, Realis en Land

Tabel 4.32 geeft de gemiddelde scores en standaardafwijkingen van Experiment 4 weer per hulpwerkwoord, modaliteit (Realis, irr vs. rea), land (Nederland vs. Vlaanderen) en in totaal.

Tabel 4.32: Exp4 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Land

Hulpww	Realis	Nederland		Vlaanderen		Totaal	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
H	rea	1.34	0.73	1.91	1.33	1.54	1.01
H	irr	1.60	1.12	2.48	1.73	1.91	1.12
Z	rea	4.98	2.07	4.74	2.02	4.89	2.21
Z	irr	2.06	1.56	2.96	1.97	2.37	2.12

De gemiddelde score voor *hebben* ($M = 1.78$) ligt lager dan die voor *zijn* ($M = 3.20$). De standaarddeviatie voor *zijn* ($SD = 2.21$) is groter dan die voor *hebben* ($SD = 1.32$), wat erop wijst dat de negatieve oordelen over *hebben* minder variatie vertonen dan die over *zijn*. De voorkeur voor *zijn* bestaat zowel in Vlaanderen als in Nederland, maar in Vlaanderen krijgen zowel *hebben* ($M = 2.29$) als *zijn* ($M = 3.55$) hogere scores dan *hebben* ($M = 1.51$) en *zijn* ($M = 3.01$) in Nederland. Daarnaast bestaat er in Nederland meer consensus over de onaanvaardbaarheid van *hebben* ($SD = 1.01$), terwijl de oordelen van de Vlaamse respondenten over *hebben* meer variatie vertonen ($SD = 1.63$).

Wat modaliteit betreft, ligt de gemiddelde score voor *hebben* hoger bij irrealis ($M = 1.91$) dan bij realis ($M = 1.45$), terwijl die voor *zijn* hoger is bij een realislezing ($M = 4.89$) dan bij een irrealislezing ($M = 2.37$). Het verschil tussen die gemiddelden is groter bij *zijn* dan bij *hebben*, wat suggereert dat het effect van de modaliteit sterker is bij *zijn* dan bij *hebben*. Tot slot is de invloed van modaliteit gelijkaardig in beide landen, al suggereert het verschil tussen de gemiddelden dat het irrealiseffect voor *hebben* sterker is in Vlaanderen (rea ($M = 1.91$) < irr ($M = 2.48$)) dan in Nederland (rea ($M = 1.34$) < irr ($M = 1.60$)).

4.2.5.1.2 Inversie

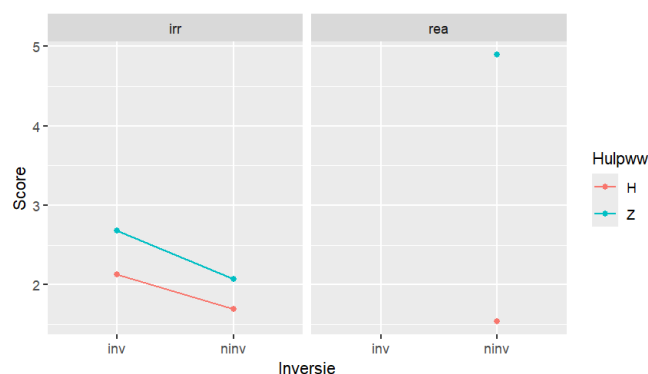
Tabel 4.33 toont de gemiddelde scores en standaarddeviaties per hulpwerkwoord, per modaliteit en per woordvolgorde (Inversie). Het effect van inversie is enkel zichtbaar bij zinnen in irrealis, aangezien enkel de irrealislezing zowel met inversie (inv) als zonder inversie (ninv) voorkomt (zie §4.1.1.2.5).

Tabel 4.33: Exp4 - Gemiddelde acceptabiliteitsscore (met standaarddeviatie) als functie van Hulpww, Realis en Inversie

Hulpww	Realis	Inversie	<i>M</i>	<i>SD</i>
H	irr	inv	2.13	1.61
H	irr	ninv	1.69	1.19
H	rea	ninv	1.54	1.02
Z	irr	inv	2.68	1.91
Z	irr	ninv	2.07	1.56
Z	rea	ninv	4.89	2.06

De gemiddelde score voor *hebben* in irrealis is hoger met inversie ($M = 2.13$) dan zonder inversie ($M = 1.69$). Dat verschil is statistisch significant ($t(1964.6) = 7.189$, $p < 0.001$). Voor *zijn* geldt hetzelfde: de score is gemiddeld hoger met inversie ($M = 2.68$) dan zonder inversie ($M = 2.07$). Ook dat verschil is significant ($t(2011.9) = 7.952$, $p < 0.001$). Binnen de irrealiscontext van Experiment 4 heeft de afwezigheid

van inversie dus een significant negatief effect op de aanvaardbaarheid van beide hulpwerkwoorden. Figuur 4.16 illustreert dat effect van Inversie op de gemiddelde scores voor *hebben* en *zijn* in irrealis.



Figuur 4.16: Exp4 – Gemiddelde acceptabiliteitscore als functie van Hulpww, Realis en Inversie

Concreet wordt in Experiment 4 een zin als (90a) dus meer aanvaard dan een zin als (90b), aangezien er in (90a) inversie plaatsvindt van het onderwerp en de persoonsvorm in de contextzin, en in (90b) niet.

- (90) a. Had die man mij blijven volgen, **voelde ik** mij onveilig. E4_3_irr_inv_H
b. Had die man mij blijven volgen, **ik voelde** mij onveilig. E4_3_irr_ninv_H

4.2.5.2 Lineaire regressieanalyse

In het basismodel van Experiment 4 (§4.2.5.2.1) wordt het effect van Realis en Hulpww, evenals hun interactie, op de acceptabiliteitscore onderzocht. In de driewegsanalyses (§4.2.5.2.2) voeg ik de covariaten Land, Opleiding, Geslacht en Leeftijdscategorie aan regressiemodellen toe.

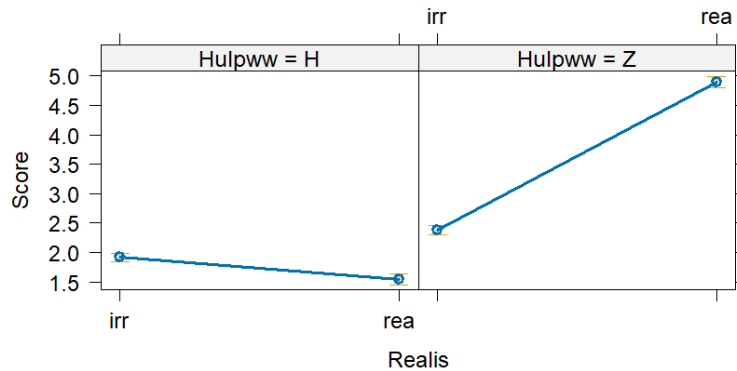
4.2.5.2.1 Basismodel

Het basismodel (AIC = 23699.17; marginale $R^2 = 0.318$; conditionele $R^2 = 0.447$) bevat Realis, Hulpww en hun eventuele interactie. Tabel 4.34 toont de regressieoutput van het model.

Tabel 4.34: Exp4 – Regressieoutput basismodel (Realis*Hulpww)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.91	0.03781	3188.31	50.642	< 0.001	***
Realisrea	-0.37	0.05423	5266.34	-6.825	< 0.001	***
HulpwwZ	0.47	0.04453	5277.15	10.477	< 0.001	***
Realisrea:HulpwwZ	2.88	0.07733	5275.43	37.293	< 0.001	***

De regressiegegevens tonen aan dat Realis een hoofdeffect is ($p < 0.001$). Dat suggereert dat de varianten meer worden aanvaard bij realis dan bij irrealis, onafhankelijk van het hulpwerkwoord. Daarnaast is ook Hulpww een hoofdeffect ($p < 0.001$): het hulpwerkwoord *hebben* wordt, los van de manipulaties, minder aanvaard dan *zijn*. De interactie tussen Realis en Hulpww is ten slotte ook statistisch significant ($p < 0.001$). Uit de post-hoc analyse blijkt dat de aanvaardbaarheid van *hebben* toeneemt bij irrealis ($estimate = -0.37$, $p < 0.001$), terwijl die van *zijn* toeneemt bij realis ($estimate = 2.51$, $p < 0.001$). Het feit dat het geschatte verschil bij *zijn* groter is dan bij *hebben*, toont aan dat het positieve effect van realis op *zijn* sterker is dan het positieve effect van irrealis op *hebben*. Figuur 4.17 visualiseert de significante interactie tussen het hulpwerkwoord en de modaliteit.



Figuur 4.17: Exp4 - Plot interactie Realis*Hulpww

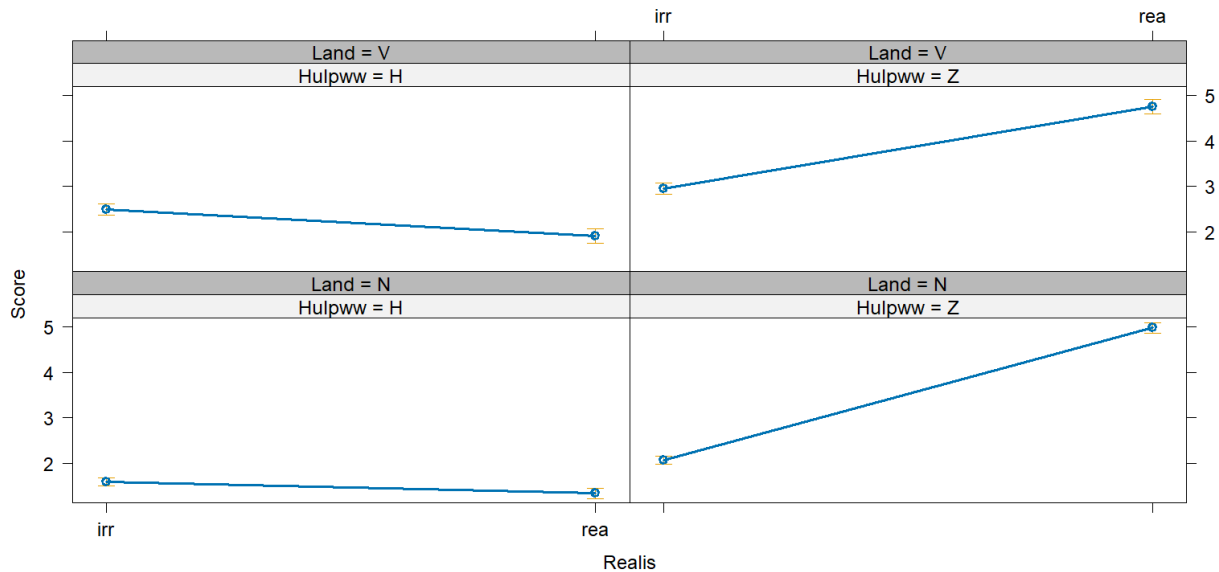
4.2.5.2.2 Driewegsanalyse

In de eerste plaats voer ik een driewegsanalyse uit waarbij ik het covariaat Land aan het basismodel toevoeg, zodat het model Realis en Hulpww, Land en alle mogelijke twee- en driewegsinteracties bevat. Dit model (AIC = 23470.08; marginale $R^2 = 0.354$; conditionele $R^2 = 0.459$) is beter dan het basismodel (AIC 23470.08 < AIC 23699.17). Tabel 4.35 bevat de regressieoutput van het model.

Tabel 4.35: Exp4 – Regressieoutput driewegsinteractie (Realis*Hulpww*Land)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.60	0.0453	3391	35.285	< 0.001	***
Realisrea	-0.26	0.06658	5261	-3.892	< 0.001	***
HulpwwZ	0.47	0.05468	5271	8.559	< 0.001	***
LandV	0.89	0.07617	3425	11.705	< 0.001	***
Realisrea:HulpwwZ	3.17	0.09511	5280	33.314	< 0.001	***
Realisrea:LandV	-0.31	0.01124	5276	-2.790	0.005	**
HulpwwZ:LandV	0.00	0.09231	5289	0.006	0.995	
Realisrea:HulpwwZ:LandV	-0.80	0.01601	5278	-5.020	< 0.001	***

Het tweede model bevestigt de effecten van het basismodel, met name de hoofdeffecten van Hulpww ($p < 0.001$) en Realis ($p < 0.001$) en de interactie tussen Realis en Hulpww ($p < 0.001$), maar biedt ook inzicht in nationale variatie. Ten eerste toont het hoofdeffect van Land ($p < 0.001$) dat Vlamingen, los van de manipulaties, in alle condities hogere scores toekennen. De significante interactie tussen Realis en Land ($p = 0.005$) suggereert dat het effect van de modaliteitsmanipulatie op de aanvaardbaarheid verschillend is per land. Uit de post-hoc analyse blijkt dat de irrealislezing, onafhankelijk van het hulpwerkwoord, significant meer aanvaard wordt in Vlaanderen dan in Nederland (*estimate* = 0.888, $p < 0.001$), terwijl er geen significant verschil is in de aanvaardbaarheid van de realislezing tussen beide landen ($p = 0.090$). Het feit dat de driewegsinteractie statistisch significant is ($p < 0.001$) toont ten slotte aan dat de interactie tussen Realis en Hulpww verbonden is aan nationale effecten. Uit de post-hoc analyse blijkt dat het geschatte verschil tussen irrealis en realis bij *hebben* groter is in Vlaanderen (*estimate* = -0.57, $p < 0.001$) dan in Nederland (*estimate* = -0.26, $p < 0.001$), terwijl dat voor *zijn* groter is in Nederland (*estimate* = 2.91, $p < 0.001$) dan in Vlaanderen (*estimate* = 1.79, $p < 0.001$). Dat laat zien dat het positieve effect van irrealis voor *hebben* sterker is in Vlaanderen, en het negatieve effect van irrealis voor *zijn* sterker is in Nederland. In Figuur 4.18 wordt de driewegsinteractie gevisualiseerd.



Figuur 4.18: Exp4 - Plot driewegsinteractie Realis*Hulpww*Land

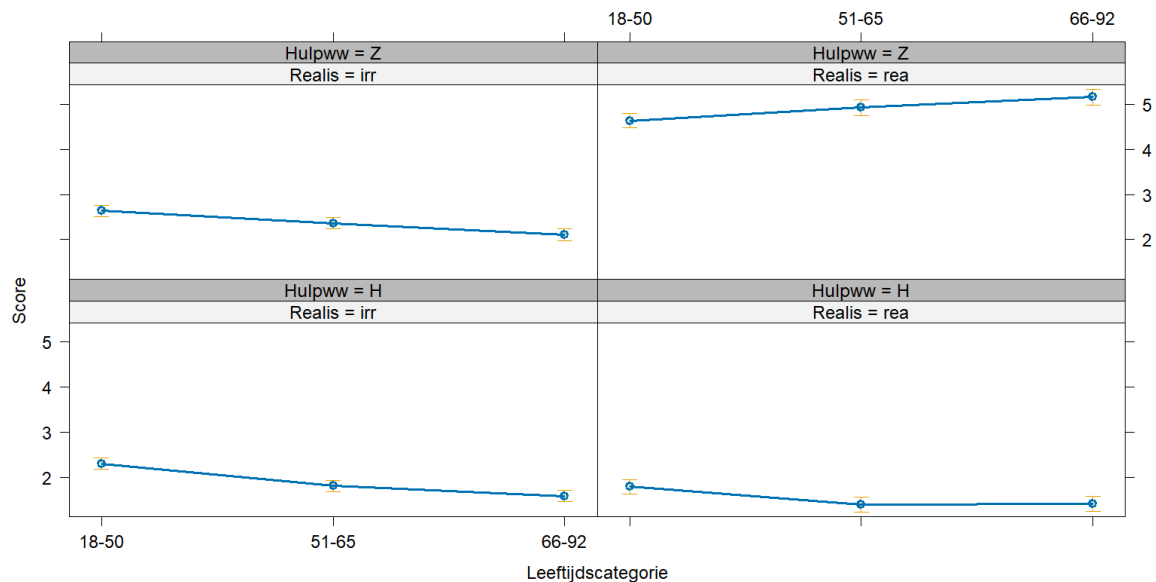
In de tweede plaats bekijk ik de driewegsinteracties voor Experiment 4 met de covariaten Geslacht, Opleiding en Leeftijdscategorie. De modellen bevatten steeds Hulpww en Realis, een van de covariaten en alle mogelijke twee- en driewegsinteracties ertussen. In de modellen met Opleiding (AIC = 23694.4; marginale $R^2 = 0.322$; conditionele $R^2 = 0.447$) en met Geslacht (AIC = 23714.89; marginale $R^2 = 0.319$; conditionele $R^2 = 0.447$) is de driewegsinteractie niet significant (respectievelijk $p = 0.976$ en $p = 0.850$). In het model met Leeftijdscategorie (AIC = 23602.37; marginale $R^2 = 0.334$; conditionele $R^2 = 0.457$) is de driewegsinteractie tussen Leeftijdscategorie, Realis en Hulpww echter wel significant. Tabel 4.36 bevat de regressieoutput van het model.

Tabel 4.36: Exp4 - Regressieoutput driewegsinteractie (Realis*Hulpww*Leeftijdscategorie)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	2.30	0.06240	3234.39	36.939	< 0.001	***
Leeftijdscategorie51-65	-0.50	0.09081	3226.09	-5.470	< 0.001	***
Leeftijdscategorie66-92	-0.72	0.09059	3219.56	-7.923	< 0.001	***
Realisrea	-0.51	0.09009	5266.30	-5.651	< 0.001	***
HulpwwZ	0.33	0.07371	5277.23	4.543	< 0.001	***
Leeftijdscategorie51-65:Realisrea	0.09	0.13092	5263.82	0.716	0.474	
Leeftijdscategorie66-92:Realisrea	0.34	0.13059	5261.14	2.587	0.010	**
Leeftijdscategorie51-65:HulpwwZ	0.22	0.10752	5276.09	2.080	0.038	*
Leeftijdscategorie66-92:HulpwwZ	0.18	0.10708	5270.63	1.718	0.086	.
Realisrea:HulpwwZ	2.50	0.12801	5267.60	19.561	< 0.001	***
Leeftijdscategorie51-65:Realisrea:HulpwwZ	0.47	0.18654	5269.65	2.525	0.012	*
Leeftijdscategorie66-92:Realisrea:HulpwwZ	0.72	0.18612	5270.94	3.869	< 0.001	***

Uit Tabel 4.36 blijkt dat Leeftijdscategorie een hoofdeffect ($p < 0.001$) is, en dat de interacties tussen Leeftijdscategorie en Hulpww (bijna-)significant zijn ($p = 0.038$ en $p = 0.086$). De post-hoc analyse laat zien dat de jongere respondenten (18-50), onafhankelijk van de modaliteit, *hebben* significant meer aanvaarden dan de respondenten tussen de 51 en 65 jaar (*estimate* = -0.46, $p < 0.001$) en dan de respondenten tussen de 66 en 92 jaar (*estimate* = -0.60, $p < 0.001$). De significante driewegsinteracties ($p = 0.012$ en $p < 0.001$) tonen daarnaast aan dat de interactie tussen Realis en Hulpww verbonden is aan leeftijdseffecten. Uit de post-hoc analyse blijkt dat het geschatte verschil tussen irrealis en realis bij

hebben groter is bij de jongste leeftijdsgroep (18-50, $estimate = -0.51$, $p < 0.001$) dan bij de middengroep (51-65, $estimate = -0.42$, $p < 0.001$). Bij de oudste groep (66-92) is er geen significant effect van irrealis op de aanvaardbaarheid van *hebben* ($estimate = -0.17$, $p = 0.201$). De oudste respondenten vinden *hebben* dus even onaanvaardbaar in realis als in irrealis, en naarmate de respondenten jonger worden aanvaarden ze *hebben* meer in irrealis. De driewegsinteractie wordt gevisualiseerd in Figuur 4.19.



Figuur 4.19: Exp4 - Plot driewegsinteractie Realis*Hulpww*Leeftijdscategorie

4.2.5.3 Conclusie: Experiment 4

De resultaten van Experiment 4 tonen aan dat het hulpwerkwoord *hebben* in Asp-clusters significant minder wordt aanvaard dan het hulpwerkwoord *zijn*, waardoor de basishypothese bevestigd wordt. De voorkeur voor *zijn* bestaat in beide landen. Hoewel de gemiddelde score voor *hebben* in Vlaanderen hoger ligt, wordt *hebben* niet significant meer aanvaard in Vlaanderen dan in Nederland. De geografische hypothese vindt in Experiment 4 dus geen duidelijke ondersteuning. De irrealishypothese kan daarentegen wel duidelijk bevestigd worden: de irrealislezing heeft een significant positief effect op de aanvaardbaarheid van *hebben* in Asp-clusters. Dat irrealiseffect blijkt sterker in Vlaanderen dan in Nederland, en sterker bij jongere dan bij oudere respondenten. Op de aanvaardbaarheid van *zijn* heeft de irrealislezing een significant negatief effect, dat sterker is in Nederland dan in Vlaanderen.

4.2.6 Conclusie: resultaten

De resultaten van het experimentele onderzoek tonen in de eerste plaats aan dat het hulpwerkwoord *hebben* in clusters met een aspectueel werkwoord (Asp-clusters) minder aanvaard wordt dan het hulpwerkwoord *zijn*. In de drie deelexperimenten waarin Asp-clusters voorkomen (Experiment 1, 2A en 4) krijgt *zijn* immers significant hogere acceptabiliteitsscores dan *hebben*. Op die manier bevestigt het experimentele onderzoek de basishypothese (§3.2.1).

Het type hoofdwkwoord en afstand blijken weinig tot geen invloed uit te oefenen op de aanvaardbaarheid van *hebben* in Asp-clusters. In Experiment 2A wordt geen significant effect van het type hoofdwkwoord op de scores van *hebben* gevonden. De hoofdwkwoordhypothese (§3.2.2) vindt dus geen ondersteuning in het experiment: *hebben* wordt niet meer aanvaard in Asp-clusters met een H-hoofdwkwoord dan in Asp-clusters met een HZ-hoofdwkwoord. In clusters met een modaal werkwoord (Mod-clusters) wordt er daarentegen wel een hoofdwkwoordeffect aangetoond. Uit de resultaten van Experiment 2B blijkt dat de aanwezigheid van een Z-hoofdwkwoord een significante positieve invloed heeft op de selectie van de alternatieve variant *zijn* in Mod-clusters, zowel in Vlaanderen als in Nederland. Wat afstand betreft blijkt uit Experiment 1 geen significant effect op de scores voor *hebben*: een grotere afstand tussen het hulpwerkwoord en het aspectuele werkwoord leidt niet tot een grotere acceptatie van *hebben*. De afstandshypothese (§3.2.3) wordt dus niet bevestigd.

De irrealis daarentegen oefent wel een significante invloed uit op de acceptabiliteitsscores van *hebben*. Ten eerste toont Experiment 3 dat de irrealis een significant positief effect heeft op de aanvaardbaarheid van *hebben* bij *gaan*, *blijven*, *komen* en *beginnen* in voltooid aspect. De basisaannname dat irrealis *hebben* bevordert bij werkwoorden die *zijn* selecteren wordt op die manier aangetoond. Die significante positieve invloed van irrealis op de acceptabiliteitsscores van *hebben* zien we ook bij de Asp-clusters van Experiment 4. Dat irrealiseffect blijkt sterker in Vlaanderen en bij jongere respondenten. Het experimentele onderzoek bevestigt dus de irrealishypothese (§2.3.4): Asp-clusters met *hebben* worden meer aanvaard bij een irrealis- dan bij een realislezing.

Tot slot laten de resultaten van het experiment een invloed van geografische regio zien. Het hulpwerkwoord *hebben* wordt in Asp-clusters meer aanvaard door de Vlaamse dan door de Nederlandse respondenten. In Experiment 1 en 2A krijgt *hebben* namelijk significant hogere acceptabiliteitsscores in Vlaanderen dan in Nederland. De hulpwerkwoordselectie in Asp-clusters vertoont echter niet zo'n sterke nationale variatie als die in Mod-clusters. Uit Experiment 2B blijkt immers een diametrale omkering van de selectievoorkeur in Mod-clusters: in Nederland is er een significante voorkeur voor *hebben*, in Vlaanderen een significante voorkeur voor *zijn*. Bij Asp-clusters daarentegen krijgt *zijn* in beide landen de voorkeur, maar is de voorkeur voor *hebben* sterker in Vlaanderen. Wat de regionale variatie betreft is de gemiddelde score voor *hebben* in Asp-clusters het grootst in West-Vlaanderen. Die variatie is echter opnieuw niet zo sterk als bij Mod-clusters, waar West-Vlaanderen zich duidelijk onderscheidt van de andere Vlaamse provincies door de voorkeur voor *hebben*. De geografische hypothese (§3.2.4) vindt dus ook ondersteuning in de resultaten van het experimentele onderzoek.

4.3 Conclusie

Aan de hand van acceptabiliteitsoordelen van 1105 Vlaamse en Nederlandse respondenten onderzocht ik in het experimentele onderzoek de invloed van het type hoofdwerkwoord, afstand, irrealis en geografische regio op hulpwerkwoordselectie in Asp-clusters. De stimuli van het experiment werden onderverdeeld in vijf deelexperimenten die elk een van de taalinterne factoren onderzochten. In Experiment 1, 2A en 4 bevatten de stimuli Asp-clusters met de aspectuele werkwoorden *gaan* en *blijven* en de hoofdwerkwoorden *slapen*, *lezen*, *winnen* en *volgen*. De stimuli van Experiment 2B bevatten Mod-clusters, die van Experiment 3 de hoofdwerkwoorden *gaan*, *blijven*, *komen* en *beginnen*. De oordelen werden geanalyseerd aan de hand van verkennende analyses met gemiddelde scores en lineaire regressieanalyses waarin ook werd gecontroleerd voor leeftijd, geslacht en opleidingsniveau.

Op basis van de analyses concludeer ik dat het hulpwerkwoord *hebben* in Asp-clusters minder aanvaard wordt dan *zijn*. Dat betekent echter niet dat het volledig wordt afgewezen. *Hebben* wordt meer geaccepteerd in Vlaanderen dan in Nederland, en meer in irrealis- dan in realislezing. Het type hoofdwerkwoord en de afstand tussen het hulpwerkwoord en het aspectuele werkwoorden blijken daarentegen de aanvaardbaarheid van *hebben* niet te beïnvloeden. Op die manier worden in het experimentele onderzoek zowel de basishypothese als de irrealishypothese en de geografische hypothese bevestigd, terwijl de hoofdwerkwoordhypothese en de afstandshypothese niet ondersteund worden door de resultaten van het experiment.

In dit hoofdstuk heb ik de voorkeur voor *hebben* of *zijn* in Asp-clusters experimenteel getest in een acceptabiliteitsdesign. Er zijn echter vragen te stellen bij de ecologische validiteit van die methode: in welke mate sluiten de acceptabiliteitsoordelen aan bij het werkelijke taalgebruik? Ik voer daarom een aanvullend corpusonderzoek uit waarbij ik de invloed van het type hoofdwerkwoord, afstand, irrealis en geografische regio onderzoek in Asp-clusters met *hebben* in werkelijk, niet-geëliciteerd taalgebruik. Daarnaast biedt het corpusonderzoek de mogelijkheid om de registerhypothese (§3.2.6) te toetsen, die omwille van het experimentele design met geschreven stimuli nog niet aan bod kwam. In het volgende hoofdstuk (§5) bespreek ik de methodologie en de resultaten van het corpusonderzoek. Daarna breng ik in het discussiehoofdstuk (§6) de resultaten van het corpusonderzoek en het experiment samen om een antwoord te formuleren op de onderzoeksvragen en een analyse voor te stellen.

5 Corpusonderzoek

5.1 Inleiding

In dit hoofdstuk beschrijf ik het corpusonderzoek dat ik uitvoerde naar hulpwerkwoordselectie in driedelige clusters met een aspectueel werkwoord (Asp-clusters). Aan de hand van het SoNaR-corpus (Oostdijk et al. 2013) en het Corpus Gesproken Nederlands (Oostdijk 2002) stelde ik een dataset samen van zinnen met Asp-clusters waarin *hebben* het hulpwerkwoord van voltooid aspect is, zoals in (91).

(91) Ik **heb** dat altijd wel **blijven doen**.

AUX_H – ASP_Z – HWW_H

Aan de hand van frequentieanalyses onderzocht ik of Asp-clusters met *hebben* vaker voorkomen afhankelijk van het type hoofdwkwoord (OV1), de afstand tussen het hulpwerkwoord en het aspectuele werkwoord (OV2), modaliteit (OV3), geografische regio (OV4) en register (OV5). Het corpusonderzoek vult de resultaten van het experiment aan door na te gaan of het irrealiseffect en de geografische variatie die geattesteerd werden in het experiment (zie §4.2.6) ook in niet-geëliciteerd taalgebruik zichtbaar zijn, en of het hoofdwkwoord en/of afstand in werkelijk taalgebruik wel een rol spelen. Daarnaast wordt in het corpusonderzoek de registerhypothese (§3.2.6) getoetst, die in het experiment niet aan bod kwam (zie §3.3). Ik bespreek eerst de methodologie van het corpusonderzoek (§5.2), daarna de resultaten (§5.3).

5.2 Methodologie

Deze methodologische sectie begint met de beschrijving van de geraadpleegde corpora voor het corpusonderzoek (§5.2.1). Daarna komen de dataverzameling (§5.2.2) en de data-annotatie (§5.2.3) aan bod. Tot slot licht ik in §5.2.4 toe op welke manier de dataset geanalyseerd werd.

5.2.1 Corpora

Ik gebruikte de webapplicatie OpenSoNaR voor het corpusonderzoek.²² Via die online interface is het mogelijk om zowel het SoNaR-corpus (Oostdijk et al. 2013) als het Corpus Gesproken Nederlands (CGN, Oostdijk 2002) te bevragen. Die corpora bieden de mogelijkheid om zowel de geografische variatie als de registervariatie te onderzoeken, aangezien ze geschreven en gesproken taalgebruik van zowel Vlamingen als Nederlanders bevatten en een onderscheid maken tussen verschillende registers. Ik bespreek hieronder de belangrijkste kenmerken van het SoNaR-corpus, gevolgd door die van het CGN.

Het SoNaR-500-corpus (Oostdijk et al. 2013) is een tekstcorpus dat meer dan 500 miljoen woorden geschreven taal bevat, afkomstig uit België en Nederland. In het corpus zijn enkel Nederlandstalige teksten opgenomen die verschenen na 1954. Alle teksten zijn getokeniseerd, ge-POS-tagd en gelemmatiseerd. Het omvat zowel teksten die geschreven zijn om te lezen als teksten die bedoeld zijn om uit te spreken, en zowel gepubliceerde als ongepubliceerde teksten. Daarnaast bestaat het corpus uit een mix van verschillende teksttypes, zoals onder andere kranten, beleidsdocumenten, discussiefora, chatberichten, juridische teksten en tweets (Oostdijk et al. 2013:222-233). Appendix E bevat een schematische weergave van het SoNaR-corpusontwerp, inclusief een overzicht van de teksttypes die in het corpus zijn opgenomen.

²² <https://opensonar.ivdnt.org/>

Het Corpus Gesproken Nederlands (Oostdijk 2002) bevat bijna dertienduizend spraakfragmenten van Nederlandse en Vlaamse sprekers, verzameld tijdens een vijfjarig project dat in 2004 werd afgerond. De omvang van het corpus is negen miljoen woorden, waarvan een derde afkomstig is uit Vlaanderen, en twee derde uit Nederland (Oostdijk 2004). De fragmenten zijn allemaal orthografisch getranscribeerd en verrijkt met lemmata en POS-tags. Het CGN bevat zowel monologische als dialogische spraak en zowel spontane als min of meer voorbereide spraak. Daarnaast omvat het corpus spraakfragmenten uit diverse taalgebruikssituaties die zijn onderverdeeld in verschillende componenten, zoals spontane conversaties, debatten, interviews, radio- en televisiefragmenten en colleges (Van Eerten 2007:2-3).

5.2.2 Dataverzameling

In deze sectie beschrijf ik op welke manier ik de dataset samenstelde. Eerst bespreek ik de zoekcriteria (§5.2.2.1) en vervolgens licht ik de zoekinstructies toe die gebruikt werden om zinnen te vinden die aan de zoekcriteria voldoen (§5.2.2.2). In §5.2.2.3 bespreek ik ten slotte op welke manier ik de resultaten van de zoekinstructies handmatig corrigeerde om tot de finale dataset van 346 zinnen te komen.

5.2.2.1 Zoekcriteria

In het corpusonderzoek ga ik op zoek naar zinnen met een Asp-cluster waarvan het hulpwerkwoord van voltooid aspect *hebben* is, zoals de zinnen in (92). De zinnen die het resultaat zijn van de zoekinstructies in OpenSoNaR moeten aan vijf criteria voldoen, die hieronder worden beschreven.

- (92) a. Ik **heb** nog nooit naar die website **gaan kijken**. (SoNaR, WR-P-E-A-0000627724)
b. Het is een film die me **heeft blijven boeien** tot het einde. (SoNaR, WR-P-E-A-0000355156)

In de eerste plaats ga ik op zoek naar Asp-clusters zoals die gedefinieerd werden in §2.1. Het moet dus gaan om driedelige, en niet om twee- of vierledige clusters, die bestaan uit een hulpwerkwoord van voltooid aspect, een aspectueel werkwoord en een hoofdwerkwoord. Het hulpwerkwoord van voltooid aspect is het finiete werkwoord van het cluster. Het finiete werkwoord kan zowel in de eerste pool staan, zoals in (92a), als in de tweede pool, zoals in (92b). Het is dus niet van belang of het om een hoofd- of bijzin gaat. Ik formuleer het eerste criterium als volgt:

Criterion 1 De zin bevat een drieledige cluster met een aspectueel werkwoord.

Ten tweede bepaal ik dat er in het corpusonderzoek enkel Asp-clusters met *hebben* als hulpwerkwoord van voltooid aspect worden opgenomen. Hoewel het includeren van Asp-clusters met *zijn* een interessante alternantiestudie zou kunnen opleveren, kies ik er omwille van haalbaarheid voor om enkel de *hebben*-variant in de dataset op te nemen. Daardoor gaat het niet om een typische variationistische analyse. Het tweede criterium is dan:

Criterion 2 Het hulpwerkwoord van voltooid aspect is *hebben*.

Vervolgens kunnen de groepsvormende werkwoorden van de Asp-clusters in de dataset zowel *gaan* en *blijven* als *komen* en *beginnen* zijn. In de stimuli van het experimentele onderzoek zijn enkel *gaan* en *blijven* als groepsvormende werkwoorden gebruikt (zie §4.1.1.1.4), waardoor in het corpusonderzoek meer verschillende Asp-clusters aan bod komen. Het derde criterium luidt als volgt:

Criterion 3 Het aspectuele werkwoord is *gaan*, *blijven*, *komen* of *beginnen*.

Ten vierde moet het hoofdwkwoord van het Asp-cluster het hulpwerkwoord *hebben* kunnen selecteren, zodat er hulpwerkwoordvariatie mogelijk is (zie §2.4.1). Het moet dus ofwel om een H-werkwoord *gaan*, dat *hebben* selecteert, zoals *zoeken* in (92a) en *boeien* in (92b), ofwel om een HZ-werkwoord dat zowel *hebben* als *zijn* kan selecteren, zoals *volgen* in de betekenis ‘achter iemand aan gaan’ (zie §4.1.1.1.3). Zolang het hoofdwkwoord geen Z-werkwoord is, zijn er geen restricties op de selectie van de hoofdwkwoorden. Ik formuleer het vierde criterium als volgt:

Criterion 4 Het hoofdwkwoord is een H- of HZ-werkwoord.

Tot slot bepaal ik aan de hand van het vijfde criterium een maximale afstand tussen *hebben* en het aspectuele werkwoord. Ofwel zijn de werkwoorden adjacent, zoals in (92b), ofwel worden ze van elkaar gescheiden door maximaal vijf woorden, zoals in (92a). Ik koos voor een afstand van vijf woorden aangezien er op die manier zinnen in de dataset terechtkomen met een grotere afstand tussen het hulpwerkwoord en het aspectuele werkwoord dan in het experimentele onderzoek, waar de afstandsmanipulatie slechts tot twee eenlettergrepige bepalingen ging (zie §4.1.1.2.1). Door de maximale afstand te beperken tot vijf woorden, beperk ik een toename aan valse treffers. Het vijfde criterium luidt als volgt:

Criterion 5 Er staan maximaal 5 woorden tussen *hebben* en het aspectuele werkwoord.

Samengevat ga ik in het corpusonderzoek op zoek naar driedelige werkwoordsclusters die bestaan uit het hulpwerkwoord van voltooid aspect *hebben*, een van de aspectuele werkwoorden *gaan*, *blijven*, *komen* of *beginnen*, en een H- of HZ-hoofdwkwoord, waarbij er maximaal vijf woorden tussen *hebben* en het aspectuele werkwoord staan.

5.2.2.2 Zoekinstructies

Om zo veel mogelijk zinnen op te sporen die voldoen aan de zoekcriteria, voerde ik verschillende zoekinstructies uit met behulp van de optie ‘advanced search’ in de OpenSoNaR-interface. In (93) worden de zoekinstructies weergegeven voor wanneer *hebben* en *gaan* (93a), *blijven* (93b), *komen* (93c) of *beginnen* (93d) adjacent zijn.

- (93)
- a. [lemma = “hebben”] [word = “gaan”] [pos_wvorm = “inf”]
 - b. [lemma = “hebben”] [word = “blijven”] [pos_wvorm = “inf”]
 - c. [lemma = “hebben”] [word = “komen”] [word = “te”?] [pos_wvorm = “inf”]
 - d. [lemma = “hebben”] [word = “beginnen”] [word = “te”?] [pos_wvorm = “inf”]

Merk op dat de zoekinstructies voor *komen* en *beginnen* enigszins verschillen van die voor *blijven* en *gaan*. Aangezien de met *blijven* of *gaan* verbonden infinitief nooit gecombineerd wordt met de infinitivale marker *te* (Haeseryn et al. 1997:975), verschijnt *te* niet in zoekinstructies (93a) en (93b). De infinitief bij *komen* en *beginnen* kan daarentegen wel voorafgegaan worden door *te*. Bij *komen* gebeurt dat enkel bij een specifieke betekenis: het gebruik van *te* geeft dan aan dat het onderwerp geen invloed uitoefent op de situatie, zoals geïllustreerd in (94a) (Haeseryn et al. 1997:982). Om ervoor te zorgen dat Asp-clusters met *komen* in die betekenis ook in de dataset terechtkomen, voegde ik aan zoekinstructie (93c)

een optionele *te* toe. Het werkwoord *beginnen* wordt in de standaardtaal altijd verbonden met een infinitief met *te*, zoals in (94b). In informele gesproken taal in België wordt *beginnen* echter ook gecombineerd met een infinitief zonder *te*, zoals in (94c) (Haeseryn et al. 1997:1043, Taaladvies 2024). Aangezien ik ernaar streef om alle Asp-clusters met *beginnen* op te sporen, voegde ik ook in zoekinstructie (93d) *te* als optioneel woord toe.

- (94) a. Ik **kwam** naast een brutaal ogend jongetje **te zitten**. (Haeseryn et al. 1997:982)
 b. Ik hoor dat Jan zijn viool al is **beginnen te stemmen**. (Haeseryn et al. 1997:1043)
 c. Ik ben vandaag **beginnen werken**. (Haeseryn et al. 1997:1043)

In de zoekinstructies in (93) zijn *hebben* en het aspectuele werkwoord adjacent. Om de afstand tussen de werkwoorden te vergroten, werden aan de zoekinstructies systematisch woorden toegevoegd aan de hand van de instructie in (95), van één tot vijf keer. Met zoekinstructie (96) wordt dan bijvoorbeeld gezocht naar een zin waarbij er twee woorden tussen *hebben* en *blijven* staan.

- (95) [pos_head = "adj|bw|lid|n|spec|tsw|tw|vg|vnw|vz"]
 (96) [lemma = "hebben"] [pos_head = "adj|bw|lid|n|spec|tsw|tw|vg|vnw|vz"] [pos_head = "adj|bw|lid|n|spec|tsw|tw|vg|vnw|vz"] [word = "blijven"] [pos_wvorm = "inf"]

Het woord dat wordt toegevoegd aan de hand van (95) kan de vorm aannemen van alle POS-tags die in SoNaR geannoteerd zijn ('adjective', 'adverb', 'article', 'noun', 'special', 'interjection', 'numeral', 'conjunction', 'pronoun', 'adposition'), behalve een leesteken ('punctuation') of een werkwoord ('verb'). Als leestekens tussen *hebben* en het aspectuele werkwoorden zouden worden toegestaan, kwamen zinnen als (97a), waarbij er geen sprake is van een Asp-cluster omdat *hebben* geen hulpwerkwoord van aspect is in de dataset terecht. Als werkwoorden niet werden uitgesloten, kwamen zinnen met vierledige clusters, zoals in (97b), in de dataset terecht. Door de instructie in (95) worden die zinnen, die niet voldoen aan Criterium 1, uitgesloten.

- (97) a. We kunnen, als we daar zin in **hebben, gaan liggen**. (SoNaR, WR-P-P-H-0000091743)
 b. Maar ze **had mogen komen helpen**. (SoNaR, WR-P-E-G-0000003792)

Bovenstaande zoekinstructies werden steeds toegepast op beide corpora ('SoNaR' en 'CGN') en afzonderlijk gefilterd op land ('Belgium' of 'Netherlands').

5.2.2.3 Handmatige correctie

Ten slotte corrigeerde ik de resultaten van het corpusonderzoek nog handmatig, aangezien de zoekinstructies nog valse treffers opleverden die niet voldeden aan de zoekcriteria. In de eerste plaats gaat het om de twee zinnen in (98), waarbij de hoofdwkwoorden *vervellen* en *flippen* Z-werkwoorden zijn en daardoor niet voldoen aan Criterium 4. In de zoekinstructies is het immers niet mogelijk om te filteren op basis van de selectievoorkeur van het hoofdwkwoord.²³

²³ Het voorkomen van Z-hoofdwkwoorden in Asp-clusters met *hebben* zoals in (98) is onverwacht. *Hebben* lijkt hier namelijk noch door het aspectuele werkwoord, noch door het hoofdwkwoord geselecteerd te worden, waardoor het niet overeenkomt met de generalisatie van hulpwerkwoordselectie in clusters (§2.2.2). Wegens de erg lage frequentie ga ik hier niet verder op in.

- (98) a. En dat **heeft beginnen vervellen**. (CGN, fv400370)
 b. Die **heeft beginnen flippen** en heeft gemoord. (SoNaR, WR-P-E-A-0000628875)

In de tweede plaats nam ik zinnen zoals in (99) niet op in de dataset, aangezien *hebben* daar niet fungeert als hulpwerkwoord van voltooid aspect maar als hoofdwkwoord van de zin op bovenliggend niveau. De zinnen bevatten dus geen Asp-clusters en voldoen daarom niet aan Criterium 1.

- (99) a. [Ik vind [dat ik het recht **heb** [om te **gaan zitten**]]] (SoNaR, WR-P-E-A-0000034140)
 b. [Opvallend is [dat de twee partijen toch kiezers **hebben** [die **blijven zweren** bij het Belgische niveau]]] (SoNaR, WS-U-E-A-0000109685)

Na de handmatige correctie resulteerde het corpusonderzoek in een dataset van 346 zinnen die een Asp-cluster met *hebben* als hulpwerkwoord van voltooid aspect bevatten.

5.2.3 Annotatie

De dataset werd geannoteerd voor een aantal taalinterne en -externe factoren. Wat de taalinterne factoren betreft werden in de eerste plaats de werkwoorden van het cluster geïnventariseerd. Het aspectuele werkwoord (gaan vs. blijven vs. komen vs. beginnen) komt in het veld Aspww terecht en het hoofdwkwoord is in het veld HWW geannoteerd. In het veld TypeHWW deelde ik de hoofdwkwoorden in per type: 'H' voor hoofdwkwoorden die *hebben* selecteren en 'HZ' voor hoofdwkwoorden die zowel *hebben* als *zijn* kunnen selecteren. Daarnaast annoteerde ik de data ook om de invloed van afstand en irrealis te onderzoeken. Het veld Afstand bevat het aantal woorden dat tussen *hebben* en het aspectuele werkwoord staat: 'A0' bij adjacentie en 'A1' bij één woord tot 'A5' bij vijf woorden. De modaliteit van de zin is geannoteerd in het veld Realis (irr vs. rea). Aangezien de niet-werkelijkheidsmodaliteit vaak wordt uitgedrukt door de onvoltooid verleden tijd, annoteerde ik in het veld Tijd de werkwoordstijd van het finiete werkwoord *hebben*: 'OTT' voor onvoltooid tegenwoordige tijd en 'OVT' voor onvoltooid tegenwoordige tijd (Haeseryn et al. 1997:75, zie §2.3.3.5). Als andere indicatoren voor een irrealislezing gebruikte ik bepalingen van modaliteit als *beter* zoals in (100a) (Haeseryn et al. 1997:75) of de conditionele bijzin met *als*, zoals in (100b) (Haeseryn et al. 1997:1619). Bij twijfel over de modaliteit van de zin bekeek ik ook de context van de zin naderbij in OpenSoNaR.

- (100) a. Gij **had** daar **beter** gaan staan hè. (SoNaR, WR-U-E-A-0000104039)
 b. **Als** ik een auto moest hebben **had** ik ze al lang gaan halen.
 (SoNaR, WR-U-E-A-0000104063)

Wat de taalexterne factoren betreft annoteerde ik in het veld Corpus (SoNaR vs. CGN) uit welk corpus de zin afkomstig is, en in het veld Land (V vs. N) of de zin uit het Vlaamse of het Nederlandse subcorpus komt. Tot slot voegde ik het veld Register toe. Bij de zinnen die afkomstig zijn uit het SoNaR-corpus voegde ik daar het teksttype aan toe dat in OpenSoNaR onder het tabblad 'metadata' bij 'text type' staat (o.a. chats, discussion lists, newspapers; zie Appendix E). Voor de zinnen die uit het CGN komen voeg ik aan dat veld de taalgebruikssituatie toe die onder het tabblad 'metadata' bij 'collection' staat (o.a. telefoondialogen, discussies, spontane conversaties).

5.2.4 Analyse

Aangezien de dataset enkel Asp-clusters met *hebben* als hulpwerkwoord van voltooid aspect bevat, is het niet mogelijk om een alternantieonderzoek uit te voeren naar de varianten *hebben* en *zijn*.²⁴ In het corpusonderzoek voer ik dus geen variationistische analyse uit, zoals ik wel deed in het experimentele onderzoek (§4). De dataset leent zich wel uitstekend om de frequentie van de factoren te onderzoeken die de voorkeur voor het minder geaccepteerde *hebben* in Asp-clusters kunnen verklaren. In wat volgt bespreek ik eerst de methode van de frequentieanalyses die ik uitvoerde, en daarna op welke manier die analyses een aanvulling vormen op de resultaten van het experiment.

Ik voerde frequentieanalyses uit met het statistische programma R (R Core Team 2024) in de werkomgeving RStudio (Posit Team 2024).²⁵ In de eerste plaats bracht ik de absolute en relatieve frequentie van de taalinterne (Aspww, HWW, TypeHWW, Afstand, Realis) en taalexterne factoren (Land, Corpus, Register) in de volledige dataset in kaart. In de tweede plaats vergeleek ik de frequentie van de factoren tussen Vlaanderen en Nederland om na te gaan of er nationale verschillen zijn in de verhouding. Om te bepalen of de nationale verschillen significant zijn, maak ik gebruik van de Fisher-exacttoets. De meer gebruikelijke chikwadraattoets is niet mogelijk vanwege de te lage verwachte frequenties voor het Nederlandse subcorpus (Levshina 2015:200-201).²⁶

Aan de hand van de frequentieanalyses is het mogelijk na te gaan of Asp-clusters met *hebben* meer voorkomen afhankelijk van de taalinterne en -externe factoren uit de onderzoeksvragen (§3.1), waarvan de invloed al dan niet werd aangetoond in het experiment (§4.2.6). Door het in kaart brengen van het aantal zinnen uit het Vlaamse en Nederlandse subcorpus test ik of de geografische hypothese (§3.2.5) ook in niet-geëliciteerd taalgebruik bevestigd wordt en of Asp-clusters met *hebben* dus meer voorkomen in Vlaanderen dan in Nederland. Ook voor de irrealishypothese (§3.2.4) ga ik na of er ook in het werkelijk taalgebruik een positieve invloed van irrealis op *hebben* wordt gevonden door te onderzoeken of Asp-clusters met *hebben* meer voorkomen in irrealis- dan in realislezing. De frequentieanalyse van de verschillende afstanden (A0-A5) maakt het daarnaast mogelijk te onderzoeken of Asp-clusters met *hebben* meer voorkomen bij een grotere dan bij een kleinere afstand (§3.2.3) en of er in niet-geëliciteerd taalgebruik dus wel een invloed van afstand wordt gevonden. Een analyse van de frequentie van de voorkomende hoofdwerkwoorden en het type waartoe ze behoren biedt tot slot inzicht in welke hoofdwerkwoorden vaker voorkomen met *hebben* in Asp-clusters, wat ondersteuning kan bieden voor de hoofdwerkwoordhypothese (§3.2.2), die in het experiment ook niet bevestigd werd. Daarnaast analyseer ik ook of de hoofdwerkwoorden die gebruikt zijn in de stimuli van het experiment (§4.1.1) in niet-geëliciteerd taalgebruik wel degelijk als hoofdwerkwoord van een Asp-cluster met *hebben* kunnen fungeren. Ook de frequentie van de aspectuele werkwoorden wordt geanalyseerd om de stimuli van het experiment te evalueren. In het experimentele onderzoek worden immers enkel Asp-clusters met *gaan* en *blijven* gebruikt, en de frequentieanalyse biedt inzicht in of Asp-clusters wel degelijk met *gaan* en *blijven* voorkomen en of er een verschil is in frequentie met de andere aspectuele werkwoorden *komen* en *beginnen*.

²⁴ Een voorbeeld van zo'n alternantieonderzoek is het onderzoek naar *zijn* en *hun* van Grondelaers et al. (2023).

²⁵ Volgende R-pakketten werden opgeladen: dplyr (Wickham et al. 2023), ggplot2 (Wickham 2016), mclm (Speelman & Montes 2022), readxl (Wickham & Bryan 2023) en tidyverse (Wickham et al. 2019).

²⁶ Om de chikwadraattoets te kunnen gebruiken moet 80% van de verwachte frequenties groter zijn dan 5 (Gries 2021:178-181).

Ten slotte biedt de frequentieanalyse van de corpora en de registers de mogelijkheid om de registerhypothese (§3.2.6) te toetsen, die in het experiment niet is behandeld. De verwachting is dat Asp-clusters met *hebben* meer voorkomen in gesproken en informeel taalgebruik dan in geschreven en formeel taalgebruik. Wat het verschil tussen gesproken en geschreven taal betreft analyseer ik of Asp-clusters met *hebben* zowel voorkomen in het SoNaR-corpus met geschreven taal als in het CGN met gesproken taal. Vanwege het verschil in omvang van de corpora – het SoNaR-corpus bevat veel meer woorden dan het CGN (zie §5.2.1) – is een uitspraak over de verhouding tussen Asp-clusters met *hebben* in gesproken en geschreven taal niet valide. De analyse maakt het wel mogelijk in kaart te brengen of Asp-clusters met *hebben* meer voorkomen afhankelijk van de formaliteit van het register. Voor de teksttypes van het SoNaR-corpus volg ik de formaliteitsclassificatie van Dhont (2024:143) die aangeeft dat de SoNaR-componenten *newspapers, periodicals & magazines, legal texts* en *policy documents* tot het formele taalgebruik behoren, en de componenten *chats, tweets, SMS* en *discussion lists* tot het informele taalgebruik. Voor de taalgebruikssituaties van het CGN maak ik gebruik van de indeling in Van Eerten (2007:2), die wordt weergegeven in Tabel 5.1. De taalgebruikssituaties worden in Tabel 5.1 onderverdeeld op basis van het aantal deelnemers en de mate van spontaniteit. CGN-componenten met een dialoog zijn dan informeler dan die met een monoloog, componenten met spontaan taalgebruik zijn informeler dan componenten met min of meer voorbereide spraak.

Tabel 5.1: Indeling CGN-componenten op basis van aantal deelnemers en mate van spontaniteit (Van Eerten 2007:2)

# deelnemers	mate van spontaniteit	CGN-component
dialoog/multiloog	spontaan	a: spontane conversaties (<i>face-to-face</i>) b: interviews met leraren Nederlands c/d: telefoondialogen e: zakelijke onderhandelingen g: discussies, debatten, vergaderingen h: lessen (middelbare school)
	min of meer voorbereid	f: interviews en discussies
monoloog	spontaan	i: spontane commentaren (radio en tv)
	min of meer voorbereid	j: actualiteitenrubrieken, reportages (radio en tv) k: nieuwsbulletins (radio en tv) l: beschouwingen, commentaren (radio en tv) m: missen, lezingen, plechtige toespraken n: colleges, voordrachten, lezingen o: voorgelezen teksten

In de volgende sectie (§5.2) worden de resultaten van het corpusonderzoek gepresenteerd. In het discussiehoofdstuk (§6) worden die in verband gebracht met de resultaten van het experiment.

5.3 Resultaten

In deze sectie bespreek ik de resultaten van de analyse van de dataset. Ik begin met de aspectuele werkwoorden (§5.3.1) en de hoofdwerkwoorden (§5.3.2), gevolgd door de afstand (§5.3.3) en de modaliteit (§5.3.4). Vervolgens worden de nationale variatie (§5.3.5) en het register (§5.3.6) behandeld. In §5.3.7 vat ik de resultaten samen in het licht van de hypothesen.

5.3.1 Aspectuele werkwoorden

Tabel 5.2 toont de absolute (abs) en relatieve (rel) frequentie van de aspectuele werkwoorden (Aspww) in de volledige dataset (Totaal), evenals uitgesplitst per land (Vlaanderen vs. Nederland).

Tabel 5.2: Absolute en relatieve frequentie van aspectuele werkwoorden als functie van land

Aspww	Vlaanderen		Nederland		Totaal	
	abs	rel	abs	rel	abs	rel
gaan	177	53.8	5	29.4	182	52.6
blijven	66	20.1	7	41.2	73	21.1
komen	57	17.3	5	29.4	62	17.9
beginnen	29	8.8	0	0	29	8.4
totaal	329	100	17	100	346	100

Het hulpwerkwoord *hebben* komt voor in combinatie met alle vier de aspectuele werkwoorden. Clusters met *gaan* zijn het meest frequent, met een totaal van 182 zinnen (52.6% van de gevallen). Hierop volgen *blijven* en *komen*, die respectievelijk 73 (21.1%) en 62 (17.9%) keer voorkomen. Clusters met *beginnen* zijn het minst frequent, met slechts 29 zinnen (8.4%). De verdeling van de aspectuele werkwoorden verschilt significant in Vlaanderen en Nederland (Fisher exact: $p = 0.047$). In Vlaanderen is *gaan* veruit het meest frequent (53.8%), terwijl het in Nederland even vaak voorkomt als *komen* (29.4%). In Nederland is *blijven* het meest frequent (41.2%), terwijl het in Vlaanderen slechts in 20.1% van de gevallen voorkomt. Het aspectuele werkwoord *beginnen* komt tot slot in het Vlaamse subcorpus in 8.8% van de zinnen voor, maar in het Nederlandse in geen enkele zin.²⁷

5.3.2 Hoofdwerkwoorden

In de dataset komen 161 verschillende hoofdwerkwoorden voor. Tabel 5.3 toont de absolute en relatieve frequentie van de tien meest frequente hoofdwerkwoorden, in de volledige dataset en per land. Een volledig overzicht van de frequenties van alle hoofdwerkwoorden is opgenomen in Appendix F.

²⁷ De verdeling van de vier aspectuele werkwoorden in Asp-clusters met *hebben* weerspiegelt mogelijk algemene verschillen in het voorkomen van de werkwoorden. Zo is *gaan* in SUBTLEX-NL (zie §4.1.1.1.3) van de vier werkwoorden het meest frequent (375 370 voorkomens) en *beginnen* het minst (27 038 voorkomens). Ook het nationale verschil in de verdeling kan bredere verschillen tussen Vlaanderen en Nederland weerspiegelen. In Vlaanderen wordt *gaan* bijvoorbeeld algemener gebruikt dan in Nederland (Taaladvies 2024). Aangezien de invloed van het specifieke aspectuele werkwoord geen deel uitmaakt van de onderzoeksvragen in deze masterproef, ga ik hier verder niet op in.

Tabel 5.3: Absolute en relatieve frequentie van tien meest frequente hoofdwerkwoorden als functie van land

HWW	Vlaanderen		Nederland		Totaal	
	abs	rel	abs	rel	abs	rel
halen	32	9.2	0	0.0	32	9.2
kijken	21	6.0	2	0.6	23	6.6
zien	17	4.9	0	0.0	17	4.9
doen	9	2.6	1	0.3	10	2.9
bekijken	8	2.3	0	0.0	8	2.3
eten	8	2.3	0	0.0	8	2.3
staan	8	2.3	0	0.0	8	2.3
vragen	8	2.3	0	0.0	8	2.3
zoeken	7	2.0	0	0.0	7	2.0
kopen	6	1.7	0	0.0	6	1.7

In (101) geef ik een voorbeeld uit de dataset van de drie werkwoorden die het vaakst voorkomen in Asp-clusters met *hebben*. Het hoofdwerkwoord *halen* (101a) komt in 32 zinnen (9.2%) voor, *kijken* (101b) in 23 zinnen (6.6%) en *zien* (101c) in 17 zinnen (4.9%).

- (101) a. Als het aan hem lag **had** die mij gisteren **komen halen**. (SoNaR, WR-U-E-A-0000104146)
 b. Ik **heb** dan **blijven kijken** om meer details te weten. (SoNaR, WR-P-E-A-0000336819)
 c. Ik **heb** die indertijd in de cinema **gaan zien**. (SoNaR, WR-P-E-A-0000358407)

Die algemene frequenties komen overeen met de verdeling in Vlaanderen. In Nederland is *kijken* het meest frequente hoofdwerkwoord (2 zinnen, 11.8% van het Nederlandse subcorpus). De andere hoofdwerkwoorden die voorkomen in het Nederlandse subcorpus (o.a. *doen*, *staan* en *twijfelen*) komen steeds slechts één keer voor. De hoofdwerkwoorden die in Experiment 2A (§4.1.1.2.2) gebruikt worden om de invloed van het hoofdwerkwoord in Asp-clusters te onderzoeken (*slapen*, *volgen*, *lezen* en *winnen*) behoren niet tot de tien meest frequente hoofdwerkwoorden, maar ze komen wel alle vier voor in de dataset. Tabel 5.4 toont de absolute en relatieve frequentie van die werkwoorden, verdeeld per land.

Tabel 5.4: Absolute en relatieve frequentie van *slapen*, *lezen*, *volgen* en *winnen* als functie van land

HWW	Vlaanderen		Nederland		Totaal	
	abs	rel	abs	rel	abs	rel
slapen	2	0.6	1	0.3	3	0.9
volgen	2	0.6	0	0	2	0.6
lezen	1	0.3	0	0	1	0.3
winnen	1	0.3	0	0	1	0.3

Slapen is het meest frequent (3 zinnen, 0.9%) en komt in zowel Vlaanderen als Nederland voor. *Volgen* komt in twee zinnen (0.6%) voor, *lezen* en *winnen* in één zin (0.3%), steeds enkel in Vlaanderen. In (102) geef ik een voorbeeld van *slapen* (102a), *volgen* (102b), *lezen* (102c) en *winnen* (102d) uit de dataset.

- (102) a. En toen **hebben** we daar **blijven slapen**. (CGN, fn000591)
 b. Ik **heb** de ganse discussie **blijven volgen**. (SoNaR, WR-P-E-A-0000176683)
 c. Ge **hebt** nog niet **gaan lezen**? (CGN, fv400056)
 d. Waarmee ik natuurlijk niet wil zeggen dat hij alles **had blijven winnen**. (SoNaR, WR-P-P-H-0000135912)

Van de 161 hoofdwerkwoorden in de dataset is er één werkwoord van het HZ-type: *optreden* in zin (103). Wanneer *optreden* in de betekenis ‘actie ondernemen’ wordt gebruikt, kan het namelijk zowel *hebben* als *zijn* selecteren (Onze Taal 2024, zie §2.2.1).

(103) Men **heeft** te laat hard **beginnen** *optreden*. (SoNaR, WR-P-E-A-0000805723)

De overige 160 hoofdwerkwoorden behoren allemaal tot het H-type. Ook de werkwoorden *volgen* en *winnen*, die in het experimentele onderzoek het HZ-type vertegenwoordigen (§4.1.1.1.3), worden in het corpusonderzoek tot de H-hoofdwerkwoorden gerekend. In de zinnen waarin *volgen* voorkomt, zoals in (102b), gaat het namelijk niet om de letterlijke betekenis ‘achter iemand aan gaan’ en in de zin met *winnen* (102d) wordt het werkwoord transitief gebruikt. Dan is enkel *hebben* mogelijk (zie §4.1.1.1.3).

5.3.3 Afstand

Tabel 5.5 geeft de absolute en relatieve frequentie van de verschillende afstanden (A0-A5) weer voor de volledige dataset en afzonderlijk voor het Vlaamse en Nederlandse subcorpus.

Tabel 5.5: Absolute en relatieve frequentie van afstand als functie van land

Afstand	Vlaanderen		Nederland		Totaal	
	abs	rel	abs	rel	abs	rel
A0	39	11.9	2	11.8	41	11.8
A1	50	15.2	1	5.9	51	14.7
A2	55	16.7	4	23.5	59	17.1
A3	69	21.0	4	23.5	73	21.1
A4	60	18.2	3	17.6	63	18.2
A5	56	17.0	3	17.6	59	17.1
totaal	329	100	17	100	346	100

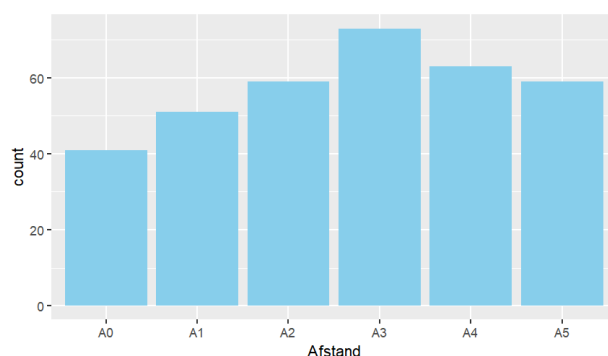
Zinnen waarbij *hebben* en het aspectuele werkwoord adjacent zijn (A0), zoals in (104a), zijn het minst frequent (41 zinnen, 11.8%). Zinnen waarbij er één woord tussen de werkwoorden staat (A1), zoals (104b), komen meer voor (51 zinnen, 14.7%) en zinnen met twee woorden tussen de werkwoorden (A2), zoals (104c) nog meer (59 zinnen, 17.1%). Zinnen met een afstand van drie woorden tussen *hebben* en het aspectuele werkwoord (A3), zoals (104d), zijn het meest frequent in de dataset (73 zinnen, 21.1%).

- (104) a. Het is een film die me **heeft blijven boeien** tot het einde. (SoNaR, WR-P-E-A-0000355156)
 b. Die **heeft ook komen helpen** met de verhuis enzo hè. (CGN, fv400235)
 c. **Heb je niet gaan kijken?** (SoNaR, WR-P-E-A-0000263633)
 d. Ik **heb al die tijd blijven werken** op training. (SoNaR, WR-P-P-G-0000531414)

Naarmate de afstand groter wordt, komen Asp-clusters met *hebben* dus meer voor. Dat stijgend effect heeft echter een bovengrens bij de afstand van drie woorden (A3). Zinnen met een afstand van vier woorden (A4), zoals (105a), komen namelijk minder voor dan A3-zinnen (63 zinnen, 18.2%) en zinnen waarbij er vijf woorden tussen de werkwoorden staan, zoals (105b), nog minder (59 zinnen, 17.1%).

- (105) a. Groenten **heb ik pas later graag beginnen eten**. (SoNaR, WR-P-E-A-0000427033)
 b. Het bord zelf **hebben beide heren op mijn grond komen gooien**. (SoNaR, WR-P-P-G-0000683894)

Figuur 5.1 geeft de absolute frequentie van de afstanden weer en illustreert zo het ordinaal effect. Asp-clusters met *hebben* komen vaker voor naarmate de afstand toeneemt, met een bovengrens bij de afstand van drie woorden. Vanaf een afstand van vier woorden neemt de frequentie van de Asp-clusters met *hebben* weer af. Dat effect is zowel in Vlaanderen als in Nederland zichtbaar. Er zijn dan ook geen significante nationale verschillen in de verdeling van de afstanden (Fisher exact: $p = 0.92$).



Figuur 5.1: Absolute frequentie van afstand

5.3.4 Modaliteit

Tabel 5.6 toont de absolute en relatieve frequentie van de werkwoordstijden (OTT vs. OVT) en de modaliteit (realis vs. irrealis) in de volledige dataset, evenals uitgesplitst per land.

Tabel 5.6: Absolute en relatieve frequentie van tijd en modaliteit als functie van land

Tijd	Modaliteit	Vlaanderen		Nederland		Totaal	
		abs	rel	abs	rel	abs	rel
OTT	realis	256	77.8	5	29.4	261	75.4
	irrealis	0	0	0	0	0	0
OVT	realis	23	7.0	1	5.9	24	6.9
	irrealis	50	15.2	11	64.7	61	17.7
totaal		329	100	17	100	346	100

In de eerste plaats laat Tabel 5.6 zien dat de realislezing zowel in de onvoltooid tegenwoordige tijd (OTT), zoals in (106a), als in de onvoltooid verleden tijd (OVT), zoals in (106b), voorkomt, terwijl de irrealislezing enkel voorkomt in de onvoltooid verleden tijd (OVT), zoals in (106c).

- (106) a. En dan **hebben** ze dat systeem beginnen versoepelen. (CGN, fv701044)
b. Ja 'k **had** gaan volleyballen hè dus 'k was helemaal bezweet. (CGN, fv901013)
c. We **hadden** liever zelf alles blijven financieren. (SoNaR, WR-P-P-H-0000073249)

Daarnaast zien we dat het finiete werkwoord *hebben* in Vlaanderen in de meerderheid van de gevallen in de OTT voorkomt (256 zinnen, 77.8%), terwijl *hebben* in Nederland in de meerderheid van de gevallen in de OVT staat (12 zinnen, 70.6%). Dat nationaal verschil in de verhouding van de werkwoordstijden is significant (Fisher exact: $p < 0.01$). Daaruit volgt logischerwijze dat in Vlaanderen de realislezing het meest frequent is (279 zinnen, 84.8%); in Nederland staat de meerderheid van de zinnen in irrealis (11 zinnen, 64.7%), terwijl de minderheid een realislezing krijgt (6 zinnen, 35.3%). Ook het nationaal verschil in de verhouding tussen irrealis en realis is significant (Fisher exact: $p < 0.01$). Als we echter enkel de zinnen in rekening nemen waarbij een irrealislezing mogelijk is, met name de OVT-zinnen, zien we geen significant verschil tussen het Vlaamse en Nederlandse subcorpus (Fisher exact: $p = 0.17$). Binnen de 85

zinnen in de OVT is de irrealislezing in beide landen het meest frequent (61 zinnen, 71.8%) en de realislezing het minst frequent (24 zinnen, 28.2%).

Concluderend tonen de resultaten aan dat Asp-clusters met *hebben* in Vlaanderen meer voorkomen in de onvoltooid tegenwoordige tijd en dus ook in realis, terwijl ze in Nederland vooral in de onvoltooid verleden tijd en in irrealis voorkomen. Wanneer de irrealislezing mogelijk is, met name in OVT, komt *hebben* meer voor in irrealis- dan in realislezing, ongeacht de geografische regio.

5.3.5 Nationale variatie

In Tabel 5.7 wordt het absolute en relatieve aantal zinnen weergegeven voor de gehele dataset, evenals afzonderlijk voor Vlaanderen en Nederland. Van de 346 zinnen die een Asp-cluster met *hebben* bevatten komt de overgrote meerderheid (329 zinnen, 95.1%) uit het Vlaamse subcorpus, terwijl slechts 17 zinnen (4.9%) uit het Nederlandse subcorpus afkomstig zijn. Asp-clusters waarin het hulpwerkwoord van voltooid aspect *hebben* is komen dus beduidend meer voor in Vlaanderen dan in Nederland.

Tabel 5.7: Absolute en relatieve frequentie van Asp-clusters met *hebben* als functie van land

Vlaanderen		Nederland		Totaal	
abs	rel	abs	rel	abs	rel
329	95.1	17	4.9	346	100

5.3.6 Register

De meeste zinnen van de dataset komen uit de geschreven taal van het SoNaR-corpus (276 zinnen, 79.8%) en 70 zinnen (20.2%) komen uit de spraakfragmenten van het Corpus Gesproken Nederlands. In deze sectie beschrijf ik het voorkomen van de verschillende teksttypes in het SoNaR-corpus (§5.3.6.1) en de verschillende taalgebruikssituaties in het Corpus Gesproken Nederlands (§5.3.6.2).

5.3.6.1 Teksttypes in het SoNaR-corpus

Tabel 5.8 bevat de absolute en relatieve frequentie van de teksttypes van het SoNaR-corpus die voorkomen in de dataset, verdeeld over Vlaanderen en Nederland.

Tabel 5.8: Absolute en relatieve frequentie van voorkomende teksttypes van het SoNaR-corpus als functie van land

Teksttype	Vlaanderen		Nederland		Totaal	
	abs	rel	abs	rel	abs	rel
discussion lists	136	51.7	5	38.5	141	51.1
newspapers	55	20.9	0	0	55	19.9
chats	36	13.7	1	7.7	37	13.4
subtitles	17	6.5	0	0	17	6.2
periodicals, magazines	11	4.2	0	0	11	4.0
tweets	2	0.8	6	46.1	8	2.9
e-magazines	2	0.8	0	0	2	0.7
sms	2	0.8	0	0	2	0.7
autocues	1	0.4	0	0	1	0.4
books	0	0	1	7.7	1	0.4
teletext pages	1	0.4	0	0	1	0.4
totaal	263	100	13	100	276	100

In Vlaanderen is het meest voorkomende teksttype in de dataset de discussiefora (141 zinnen, 51.1%). Daarnaast komen in het Vlaamse subcorpus ook veel kranten (55 zinnen, 19.9%) en chatberichten (37

zinnen, 13.4%) voor. In Nederland komen de meeste zinnen uit tweets (6 zinnen, 46.1%) en discussiefora (5 zinnen, 38.5%). Opvallend daarbij is dat die veelvoorkomende teksttypes behoren tot het informeel geschreven taalgebruik, terwijl de teksttypes die helemaal niet voorkomen in de dataset (bv. handleidingen, juridische teksten, beleidsdocumenten en scripties) behoren tot het formeel geschreven taalgebruik. Zowel in Vlaanderen als in Nederland blijken Asp-clusters met *hebben* dus vooral voor te komen in het informele taalgebruik.

Een uitzondering daarop vormt het frequente voorkomen van de krantentaal in het Vlaamse subcorpus. Krantentaal wordt immers doorgaans gezien als een representatie van formeel geschreven taalgebruik (Dhont 2024:107). Dhont (2024) benadrukt echter dat niet alle passages in krantentaal (even) formeel zijn, aangezien bijvoorbeeld in interviewartikelen of reportages regelmatig informele, spreektaalige items bewaard worden, zoals in citaten (Dhont 2024:109-110). Om na te gaan of de zinnen van de dataset uit krantentaal als formeel bestempeld kunnen worden, bekeek ik een selectie van 15 van de 55 zinnen uit krantenartikels van naderbij, in de context. Daaruit blijkt dat 12 van de 15 zinnen (80%) afkomstig zijn uit citaten uit interviews, zoals het voorbeeld in (107). Op basis daarvan vermoed ik dat veel zinnen uit krantentaal in de dataset eerder een transcriptie zijn van gesproken, informeel taalgebruik dan het formele taalgebruik van de auteur van het krantenartikel.

- (107) “Ik zat veel op de bank, speelde zelfs een viertal wedstrijden met de reserves en was vorige week, tegen Gistel, zelfs niet geselecteerd”, legt De Beer uit. “Ik **heb** al die tijd **blijven werken** op training en ik bleef geloven in mijn speelmansen.”
(SoNaR, WR-P-P-G-0000531414)

Concluderend blijken Asp-clusters met *hebben* in geschreven taalgebruik vooral voor te komen in teksttypes met informeel taalgebruik, zoals discussiefora, chatberichten en tweets, maar ook als informeel item binnen formeel taalgebruik, zoals krantentaal.

5.3.6.2 Taalgebruikssituaties in het CGN

Tabel 5.9 bevat de absolute en relatieve frequentie van de taalgebruikssituaties van het CGN die voorkomen in de dataset. De tweede kolom bevat de bijbehorende letter van de CGN-component.

Tabel 5.9: Absolute en relatieve frequentie van voorkomende taalgebruikssituaties van het CGN als functie van land

Taalgebruikssituatie	Component	Vlaanderen		Nederland		Totaal	
		abs	rel	abs	rel	abs	rel
spontane conversaties (<i>face-to-face</i>)	a	33	50	1	25	34	48.6
telefoondialogen	c/d	26	39.4	1	25	27	38.6
interviews met leraren Nederlands	b	3	4.6	2	50	5	7.1
discussies, debatten, vergaderingen	g	2	3	0	0	2	2.9
colleges, voordrachten, lezingen	n	1	1.5	0	0	1	1.4
lessen	h	1	1.5	0	0	1	1.4
totaal		66	100	4	100	70	100

De twee taalgebruikssituaties die in de dataset het meest frequent voorkomen, en zowel in Vlaanderen als in Nederland, zijn de spontane conversaties (34 zinnen, 48.6%) en de telefoondialogen (27 zinnen, 38.6%). Die twee componenten behoren typisch tot het spontane, informele taalgebruik. Meer nog, 98.6% van de zinnen in de dataset is afkomstig van componenten die dialogisch en spontaan taalgebruik bevatten. Er komt immers slechts één zin (1.4%) uit component n (colleges, voordrachten, lezingen), die

tot de monologische en voorbereide spraak behoort. De acht componenten die niet voorkomen in de dataset, en dus geen Asp-cluster met *hebben* bevatten, zijn zakelijke onderhandelingen (comp. e), interviews en discussies (comp. f.), alle soorten radio- en televisiefragmenten (comp. i – l), missen, lezingen en plechtige toespraken (comp. m) en voorgelezen teksten (comp. o). Bij zes van die acht ontbrekende componenten gaat het om monologisch taalgebruik en om min of meer voorbereide spraak.

Op die manier wordt duidelijk dat Asp-clusters met *hebben* in het gesproken taalgebruik vooral voorkomen in situaties waarbij taalgebruikers spontaan met elkaar in interactie gaan, zoals in *face-to-face* conversaties. In situaties waarbij taalgebruikers hun spraak min of meer hebben voorbereid en niet met elkaar in interactie gaan, komen Asp-clusters met *hebben* veel minder voor.

5.3.7 Conclusie

Op basis van de dataset van 346 zinnen die een Asp-cluster met *hebben* bevatten, blijkt ten eerste dat *hebben* vaker voorkomt in Asp-clusters met *gaan*, *blijven* en *komen* dan met *beginnen*. In Vlaanderen is de combinatie van *hebben* en *gaan* duidelijk het meest frequent, in Nederland zijn de verschillen tussen de werkwoorden minder groot. Daarnaast worden Asp-clusters met *hebben* in de dataset het vaakst gecombineerd worden met de hoofdwkwoorden *halen*, *kijken* en *zien*, maar de combinatie met veel verschillende hoofdwkwoorden is mogelijk.

De hoofdwkwoorden zijn bijna altijd van het H-type: er komt in de dataset slechts één HZ-hoofdwkwoord voor. Dat lijkt de hoofdwkwoordhypothese (§3.2.2) te ondersteunen. Wat de afstand betreft zien we dat Asp-clusters met *hebben* in beide landen meer voorkomen bij een grotere afstand tussen het hulpwerkwoord en het aspectuele werkwoord dan bij adjacentie of een kleinere afstand. Dat ondersteunt de afstandshypothese (§3.2.3). Er bestaat echter een bovengrens van drie woorden: wanneer de afstand groter is, komen Asp-clusters met *hebben* weer minder voor. Daarnaast tonen de resultaten dat Asp-clusters met *hebben* zowel in realis als in irrealis voorkomen. In Vlaanderen komen ze vaker voor in realis, in Nederland vaker in irrealis. In de onvoltooid verleden tijd komen Asp-clusters met *hebben* echter in beide landen het meest voor in irrealis. Dat ondersteunt de irrealishypothese (§3.2.4). Vervolgens toont de dataset dat Asp-clusters met *hebben* aanzienlijk vaker voorkomen in Vlaanderen dan in Nederland. Dat ondersteunt de geografische hypothese (§3.2.5). Tot slot komen Asp-clusters met *hebben* zowel voor in geschreven als in gesproken taal. Ze worden eerder gebruikt in informele registers met spontaan taalgebruik en interactie dan in formelere registers met voorbereid taalgebruik en zonder interactie. Dat bevestigt de registerhypothese (§3.2.6).

In dit hoofdstuk onderzocht ik het voorkomen van *hebben* als hulpwerkwoord van Asp-clusters in het werkelijk, niet-geëliciteerd taalgebruik aan de hand van het SoNaR-corpus en het Corpus Gesproken Nederlands. In het volgende hoofdstuk (§6) worden de resultaten van het experimentele onderzoek en het corpusonderzoek bij elkaar gebracht, om zo een antwoord te formuleren op de onderzoeksvragen en op basis daarvan een analyse voor te stellen.

6 Discussie

In dit hoofdstuk breng ik de resultaten van het experimentele onderzoek (§4) en het corpusonderzoek (§5) met elkaar in verband in het licht van de onderzoeksvragen (§3.1). In §6.1 schets ik het algemene beeld van hulpwerkwoordvariatie in Asp-clusters dat in het onderzoek werd vastgesteld. Daarna ga ik in op de invloed van de taalinterne (§6.2) en taalexterne factoren (§6.3). In §6.4 stel ik op basis van de resultaten een formele analyse voor hulpwerkwoordvariatie in Asp- en Mod-clusters voor. Tot slot bespreek ik in §6.5 de beperkingen van het onderzoek en suggesties voor vervolgonderzoek.

6.1 Hulpwerkwoordvariatie in Asp-clusters

Het onderzoek in deze masterproef toont aan dat het hulpwerkwoord *hebben* in driedelige werkwoordsclusters met een aspectueel werkwoord significant minder aanvaard wordt dan *zijn*. De voorkeur voor *zijn* bestaat zowel in Vlaanderen als in Nederland (zie §4.2.6). Hoewel in het onderzoek de precieze verhouding tussen *hebben* en *zijn* niet is onderzocht vanwege de aard van het corpusonderzoek, laten de lage gemiddelde acceptabiliteitsscores en de kleine spreiding voor *hebben* in het experiment zien dat *hebben* over het algemeen erg weinig aanvaard wordt en dat *zijn* dus frequenter voorkomt, zoals in de basishypothese werd gesteld.

Desalniettemin toont de omvang van de dataset van het corpusonderzoek dat Asp-clusters met *hebben* wel degelijk voorkomen in het Nederlands. Gezien het lage aantal voorbeelden met *hebben* in het corpusonderzoek van Van Eynde et al. (2016) en Dhont (2024) was de dataset van 346 voorbeelden eerder onverwacht (zie §2.4). De corpusanalyse toont aan dat *hebben* het meest frequent voorkomt in clusters met het aspectuele werkwoord *gaan*, gevolgd door clusters met *blijven* (§5.3.1). Het feit dat *gaan* en *blijven* gebruikt werden als aspectuele werkwoorden in de stimuli van het experiment, suggereert dat de stimuli overeenkomen met de Asp-clusters die in werkelijk taalgebruik het meest voorkomen.

6.2 Taalinterne factoren

In wat volgt bespreek ik de resultaten wat betreft de invloed van de taalinterne factoren: het hoofdwkwoord (§6.2.1), afstand (§6.2.2) en irrealis (§6.2.3). Uit de resultaten blijkt enkel een duidelijke invloed van irrealis; de invloed van het type hoofdwkwoord of de afstand wordt niet aangetoond.

6.2.1 De invloed van het hoofdwkwoord

In het onderzoek wordt geen duidelijke invloed van het type hoofdwkwoord op de aanvaardbaarheid van *hebben* vastgesteld (OV1). In Experiment 2A heeft het type hoofdwkwoord geen significante invloed op de acceptabiliteitsscores voor *hebben*. Hoewel het niet significant is, toont het resultaat van Experiment 2A een omgekeerd effect van het type hoofdwkwoord dan verwacht op basis van de hoofdwkwoordhypothese: *hebben* wordt niet meer aanvaard in clusters met een H-hoofdwkwoord (*slapen* en *lezen*), maar net meer in clusters met een HZ-hoofdwkwoord (*volgen*, *winnen*). Uit de corpusanalyse blijkt dat de hoofdwkwoorden die voorkomen in Asp-clusters met *hebben* bijna uitsluitend H-werkwoorden zijn. Aangezien er in het Nederlands slechts enkele HZ-werkwoorden bestaan (zie §2.2.1), is dat echter geen opmerkelijk resultaat en neem ik aan dat dat geen inzicht biedt in de invloed van het type hoofdwkwoord.

Ik vermoed dat het resultaat wat betreft de invloed van het type hoofdwerkwoord sterk beïnvloed is door de selectie van de hoofdwerkwoorden in de stimuli. Ik stel namelijk in vraag of de vier geselecteerde hoofdwerkwoorden wel degelijk tot twee afzonderlijke types behoren.

Ik neem ten eerste aan dat de categorisatie van *winnen* als HZ-hoofdwerkwoord niet valide was. Zoals vermeld is het hulpwerkwoord *hebben* bij *winnen* veel algemener aanvaard dan *zijn* en is de hulpwerkwoordselectie bij *winnen* sterk verbonden aan nationale variatie: enkel in Vlaanderen is *zijn* mogelijk (§4.1.1.1.3). Uit de resultaten van Experiment 2B blijkt dat er in Vlaanderen bij Mod-clusters steeds een voorkeur voor *zijn* bestaat, behalve in clusters met het hoofdwerkwoord *winnen* (§4.2.3.1.2). Dat suggereert dat er ook in Vlaanderen een voorkeur bestaat voor *hebben*, net als in Nederland. Ik veronderstel op basis daarvan dat het hulpwerkwoord *zijn* bij *winnen* over het algemeen weinig aanvaard wordt. Vermoedelijk is *winnen* dan niet zo neutraal als verwacht, en kan het eerder bestempeld worden als een H- dan als een HZ-werkwoord. In dat opzicht is het niet verbazend dat er geen sterk verschil optreedt tussen de H- en de HZ-werkwoorden. Ten tweede suggereert het feit dat de H-werkwoorden *slapen* en *lezen* in de dataset bijna niet voorkomen dat die werkwoorden geen grote voorkeur voor *hebben* vertonen. Naar het idee van Hofmans (1981) zijn die werkwoorden vermoedelijk niet echt ‘pro-*hebben*’ (zie §2.3.3.3). In dat opzicht had het opnemen van de werkwoorden *halen*, *zien* of *kijken*, die het vaakst voorkomen in de dataset, vermoedelijk andere resultaten opgeleverd.

6.2.2 De invloed van afstand

Wat de invloed van de afstand tussen het hulpwerkwoord van voltooid aspect en het aspectuele werkwoord betreft (OV2), divergeren de resultaten van het experiment en het corpusonderzoek. De resultaten van Experiment 1 laten zien dat afstand geen significant effect heeft op de aanvaardbaarheid van *hebben*. De gemiddelde score voor *hebben* daalt zelfs wanneer de afstand toeneemt. De corpusanalyse suggereert echter dat een grotere afstand wel een factor is die *hebben* bevordert. De discrepantie tussen de resultaten heeft volgens mij verschillende verklaringen.

Het ontbrekende afstandseffect kan in de eerste plaats het resultaat zijn van de categorische voorkeur voor *zijn* in Experiment 1 (zie §4.2.1). Vermoedelijk wordt *hebben* in zo’n mate afgekeurd dat de intuïties over de aanvaardbaarheid op hun grenzen stuiten en afstand geen rol kan spelen. Daarnaast is het verschil in resultaat volgens mij te wijten aan de voorzichtige afstandsmanipulaties in het experiment. In de stimuli van Experiment 1 staan er maximaal twee eenlettergrepige bepalingen tussen de werkwoorden, zoals in (108a). Uit het corpusonderzoek blijkt echter dat de meeste Asp-clusters met *hebben* voorkomen bij een afstand van drie woorden, zoals in (108b).²⁸ Grotere afstandsmanipulaties in de experimentele stimuli hadden op die manier misschien wel een afstandseffect aangetoond.

(108) a. De turnster **heeft** daar toen **blijven winnen**, dus nu mag ze naar de Olympische Spelen.

E1_2_A2_H

b. Ik **had** graag een stukje **gaan fietsen** ofzo.

(SoNaR, WR-P-E-A-0006586788)

Het corpusonderzoek laat echter ook zien dat wanneer de afstand erg groot is, meer dan drie woorden, Asp-clusters met *hebben* weer minder voorkomen. Ik vermoed echter dat drieledige werkwoordsclusters

²⁸ Merk op dat de afstand in het experimentele onderzoek werd gemeten aan de hand van lettergrepen, en in het corpusonderzoek aan de hand van woorden. In OpenSoNaR is het immers niet mogelijk woorden te selecteren op basis van het aantal lettergrepen. Een afstand van drie woorden kan bijgevolg bestaan uit meer dan drie lettergrepen, zoals in (109b) waar de afstand drie woorden, maar vier lettergrepen betreft.

waarbij het hulpwerkwoord en het groepsvormende werkwoord met meer dan drie woorden van elkaar gescheiden zijn, zoals in (109), over het algemeen minder frequent voorkomen en daardoor ook minder in de dataset aanwezig zijn.

- (109) Ze **hebben** dat meisje na jaren terug **gaan zoeken**. (SoNaR, WR-P-E-A-0000610602)

Tot slot suggereert het significante negatieve effect van afstand op de aanvaardbaarheid van *zijn* in Experiment 1 dat de respondenten zinnen met afstand over het algemeen minder aanvaarden, door de toenemende complexiteit van de stimuluszin. Ongeacht de sterke voorkeur voor *zijn* in Experiment 1 worden zinnen als (110b), met een afstand van twee woorden, significant minder aanvaard dan zinnen als (110a), waarbij de werkwoorden adjacent zijn. Naar analogie kan de toenemende complexiteit ook een verklaring zijn voor de dalende scores voor *hebben* bij toenemende afstand.

- (110) a. De pianist **is blijven winnen** en zit nu in de finale van de muziekwedstrijd. E1_3_A0_Z
b. De pianist **is** toen maar **blijven winnen** en zit nu in de finale van de muziekwedstrijd. E1_3_A2_Z

Samengevat tonen de resultaten van het onderzoek geen eenduidige invloed van afstand op de aanvaardbaarheid van *hebben* aan. Vervolgonderzoek met grotere afstandsmanipulaties dat het negatieve effect van toenemende complexiteit omzeilt, kan een antwoord bieden op de divergerende resultaten van het experiment en het corpusonderzoek.

6.2.3 De invloed van irrealis

Zowel het experiment als het corpusonderzoek tonen duidelijk een positieve invloed van irrealis op de selectie van *hebben* aan (OV3). In Experiment 4 wordt immers een significant positief effect van irrealis op de scores voor *hebben* vastgesteld, en uit de corpusanalyse blijkt dat irrealis meer voorkomt dan realis in de onvoltooid verleden tijd. In tegenstelling tot het type hoofdwerkwoord en afstand, wordt de invloed van irrealis in deze masterproef zo wel eenduidig aangetoond.

Ik vermoed echter dat de bandbreedte van het irrealiseffect in het experimentele onderzoek nog niet volledig in kaart is gebracht. Ik veronderstel namelijk dat de irrealisconstructies die voorkomen in de stimuli de acceptabiliteitscores voor de irrealiszinnen sterk hebben beïnvloed. De aan-of afwezigheid van inversie heeft een sterke invloed op de aanvaardbaarheid van de irrealiszinnen. De analyse van de inversievariabele in Experiment 4 (§4.2.5.1.2) toont namelijk aan dat irrealiszinnen waarbij geen inversie van het onderwerp en de persoonsvorm optreedt in de contextzin, zoals in (111a), significant minder aanvaard worden dan irrealiszinnen met inversie, zoals in (111b).

- (111) a. Had die man mij blijven volgen, **ik voelde** mij onveilig. E4_3_irr_ninv_H
b. Had die man mij blijven volgen, **voelde ik** mij onveilig. E4_3_irr_inv_H

Daarnaast blijkt uit de resultaten van Experiment 4 een significant negatief effect van irrealis op de aanvaardbaarheid van *zijn*. Een realiszin met *zijn* zoals (112a) krijgt significant hogere scores dan de irrealiszinnen (112a) en (112b). Dat suggereert volgens mij dat de manier waarop de irrealislezing is uitgedrukt in de stimuli van Experiment 4 de score over het algemeen, zowel bij *hebben* als bij *zijn*, heeft doen dalen.

- | | | |
|-------|--|-----------------|
| (112) | a. De spion was haar blijven volgen, hij kende nu haar geheim. | E4_1_rea_ninv_Z |
| | b. Was de spion haar blijven volgen, kende hij nu haar geheim. | E4_1_irr_inv_Z |
| | c. Was de spion haar blijven volgen, hij kende nu haar geheim. | E4_1_rea_ninv_Z |

Ik gaf in de opbouw van de stimuli de voorkeur aan een gelijkenis in zinsstructuur en gebruikte daarom geen conditionele bijzin met *als* en geen voegwoord *dan*, die meer gebruikelijk zijn om een irrealislezing uit te drukken (zie §4.1.1.2.4 en §4.1.1.2.5). De gemarkeerde status van de hoofdzinsconstructie met voor-pv zonder *dan* heeft vermoedelijk de acceptabiliteitscores voor de irrealiszinnen doen dalen.²⁹ Ik verwacht dan ook dat het positieve effect van irrealis op de aanvaardbaarheid van *hebben* nog sterker naar voren was gekomen in het experiment als de stimuli waren opgebouwd zoals in (115), aangezien de invloed van inversie op die manier wordt vermeden.

- (113) Was de spion haar blijven volgen, **dan** kende hij nu haar geheim.

Tot slot is het feit dat het irrealiseffect gestratificeerd is voor leeftijd (zie §4.2.5.2.2) een onverwacht resultaat. Aangezien de aanwezigheid van *hebben* bij werkwoorden die *zijn* selecteren in de irrealis al sinds het Middelnederlands geattesteerd is (zie §2.3.4), verwacht ik immers geen voortgaande verandering wat de invloed van irrealis betreft. Het leeftijdseffect dient echter met voorzichtigheid geïnterpreteerd te worden aangezien er een significante associatie bestaat tussen de leeftijd en het land van de respondenten (zie §4.1.2). Binnen het bestek van deze masterproef ga ik hier verder niet op in.

6.3 Taalexterne factoren

6.3.1 De invloed van geografische regio

De resultaten van experiment en het corpusonderzoek tonen aan dat hulpwerkwoordselectie in Asp-clusters geografische variatie vertoont (OV4). Ik bespreek eerst de geattesteerde nationale variatie (§6.3.1.1), daarna de regionale variatie (§6.3.1.2).

6.3.1.1 Nationale variatie

Het hulpwerkwoord *hebben* wordt in Asp-clusters meer aanvaard in Vlaanderen dan in Nederland. In de experimenten met Asp-clusters (Experiment 1, 2A en 4) liggen de gemiddelde scores voor *hebben* telkens hoger in Vlaanderen dan in Nederland; in Experiment 1 en 2A is die interactie tussen het land en het hulpwerkwoord ook significant (zie §4.2.6). Daarnaast tonen de lage standaardafwijkingen dat Nederlandse respondenten *hebben* telkens resoluter afwijzen, terwijl de oordelen van de Vlamingen over *hebben* steeds meer variabiliteit vertonen. Daarenboven toont de corpusanalyse dat Asp-clusters met *hebben* beduidend meer voorkomen in Vlaanderen dan in Nederland: slechts 4.9% van de voorbeelden uit de dataset zijn afkomstig uit Nederland. De resultaten van het onderzoek komen dus overeen met de nationale variatie die geschetst wordt in het onderzoek van Hofmans (1982b) en Dhont (2024) (zie §2.4.2): hulpwerkwoordvariatie in Asp-clusters komt meer voor in Vlaanderen dan in Nederland, waar *hebben* niet aanvaard wordt en slechts sporadisch voorkomt.

Daarnaast wordt in het experimentele onderzoek aangetoond dat de invloed van geografische regio bij Asp-clusters niet zo sterk is als bij Mod-clusters. In Mod-clusters wordt de alternatieve variant *zijn* meer

²⁹ Dit idee wordt ook ondersteund doordat verschillende respondenten in de *debriefing* (zie §4.1.4) aangeven dat ze vermoeden dat het onderzoek gaat over “woordvolgorde in bijzinnen”, over “wanneer signaalwoorden (dus, maar, dan) mogen weggelaten worden” of over “inversie in de hoofdzin”.

aanvaard in Vlaanderen dan in Nederland, net zoals de alternatieve variant *hebben* in Asp-clusters door Vlamingen meer aanvaard wordt. De omvang van de verhouding is echter verschillend. Waar bij Mod-clusters een diametraal verschil bestaat – in Vlaanderen *zijn*, in Nederland *hebben* – gaat het bij Asp-clusters om een kleiner verschil. In Asp-clusters wordt *hebben* meer aanvaard in Vlaanderen dan in Nederland, maar zowel Vlamingen als Nederlanders hebben een voorkeur voor *zijn* (zie §4.2.6).

6.3.1.2 Regionale variatie

De resultaten van Experiment 2A tonen aan dat hulpwerkwoordselectie in Asp-clusters regionale variatie vertoont: binnen Vlaanderen krijgt *hebben* gemiddeld de hoogste score in West-Vlaanderen; binnen Nederland wordt *hebben* het minst aanvaard in Noord-Nederland. Hoewel *hebben* in West-Vlaanderen het meest aanvaard wordt, wordt in het experiment niet aangetoond dat die regio zich met een veel sterkere voorkeur voor *hebben* duidelijk onderscheidt van de andere Vlaamse regio's. Het verschil met de andere regio's is niet zo groot als in het onderzoek van Hofmans (1982b), waar *hebben* in West-Vlaanderen bijna dubbel zo veel verkozen wordt als in andere provincies (zie §2.4.2). Het onderzoek naar regionale variatie bleef in deze masterproef echter beperkt tot het berekenen van gemiddelde scores, en biedt vermoedelijk geen volledig inzicht in de regionale variatie. West-Vlaanderen is bovendien de regio met het kleinste aantal vertegenwoordigde respondenten in de steekproef (zie §4.1.2), waardoor een conclusie over de positie van West-Vlaanderen uit het onderzoek niet volgt.

6.3.2 De invloed van register

Op basis van het onderzoek naar de invloed van register (OV5) in het corpusonderzoek stel ik vast dat Asp-clusters met *hebben* zowel voorkomen in gesproken als in geschreven taal. Daarnaast komt *hebben* in Asp-clusters vooral voor in informeel, spontaan taalgebruik, zoals in discussiefora en spontane conversaties, en minder in formeel, voorbereid taalgebruik. Op die manier wordt de veronderstelling van Hofmans (1985:75) en Van der Horst (2005:102) (zie §2.4.2) aangetoond in het corpusonderzoek.

De registervariatie biedt een mogelijke verklaring voor de verrassende omvang van de dataset (§6.1). Dat in dit onderzoek meer voorbeelden van Asp-clusters met *hebben* worden gevonden dan in het corpusonderzoek van Dhont (2024) is dan vermoedelijk te verklaren door het feit dat Dhont uitsluitend de krantencomponent van het SoNaR-corpus onderzocht, waarin vooral formele taal vervat zit (zie §5.3.6). Ten tweede kan de discrepantie tussen de lage acceptabiliteitsscores voor *hebben* in het experiment en de omvang van de dataset verklaard worden doordat het experiment een eerder formele context vertegenwoordigt. Ondanks dat ik aangaf dat de respondenten zich op hun eigen taalgevoel en niet op standaardtaal moesten baseren (zie §4.1.3), gaat het in het experiment om formeel, geschreven taalgebruik dat geen interactie vertoont en niet spontaan is. Net zoals Hofmans (1985:75) vermoedt dat de *hebben*-variant in zijn enquêteonderzoek minder verkozen wordt dan dat die daadwerkelijk uitgesproken wordt in informele situaties (zie §2.4.2), verwacht ik ook dat de lage scores voor *hebben* in het experiment gedeeltelijk door de formaliteit van de context verklaard kunnen worden.

6.4 Analyse

In deze sectie stel ik op basis van de resultaten van het onderzoek een formele analyse van hulpwerkwoordselectie in Asp-clusters voor. Ik baseer me daarvoor op de analyses die in §2.5 werden toegelicht. In §6.4.1 stel ik een analyse voor de hulpwerkwoordvariatie in Asp-clusters voor. Vervolgens breid ik de analyse uit naar Mod-clusters, en toon ik aan dat beide types clusters op dezelfde manier geanalyseerd kunnen worden (§6.4.2). Tot slot beargumenteer ik dat de invloed van irrealis op een andere manier geanalyseerd kan worden (§6.4.3).

6.4.1 Hulpwerkwoordvariatie bij Asp-clusters

Het vertrekpunt voor de analyse voor hulpwerkwoordselectie in Asp-clusters is het bepalen van de status van de groepsvormende werkwoorden *gaan* en *blijven*.³⁰ Uit de formele analyse van hulpwerkwoordvariatie (zie §2.5.2) blijkt immers dat de status van het groepsvormende werkwoord samenhangt met de mogelijkheid tot hulpwerkwoordvariatie. Ik test de status van de aspectuele werkwoorden aan de hand van de drie morfosyntactische kenmerken die Wurmbrand (2001) en Pots (2020) naar voren brengen. Die kenmerken staan in Tabel 2.5, hieronder herhaald als Tabel 6.1.

Tabel 6.1: Kenmerken van functionele, semi-lexicale en lexicale groepsvormende werkwoorden

Groepsvormend werkwoord		Loos onderwerp	Extrapositie	Hulpwerkwoordvariatie
Functioneel		JA	NEE	JA
Semi-lexicaal	Hoeven	JA	NEE	JA
	Zitten	NEE	NEE	NEE
Lexicaal		NEE	JA	NEE

Gaan en *blijven* staan een loos onderwerp toe, zoals geïllustreerd in (114). Dat wijst op een functionele status. Daarnaast illustreren de zinnen in (115) dat de aspectuele werkwoorden *gaan* en *blijven* geen extrapositie van de ingebedde infinitief toelaten, wat ook op een functionele status wijst.

- (114) a. **Het gaat** sneeuwen.
b. **Het blijft** sneeuwen.

- (115) a. *dat ik **ga** [een koekje **eten**].
b. *dat ik **blijf** [een nachtje **slapen**].

Ten derde laten de resultaten van deze masterproef zien dat *gaan* en *blijven* hulpwerkwoordvariatie toestaan. Het corpusonderzoek toont dat de aanwezigheid van *hebben* bij Asp-clusters met *gaan* en *blijven* mogelijk is (zie §6.1). De geattesteerde hulpwerkwoordvariatie wordt geïllustreerd in (116).

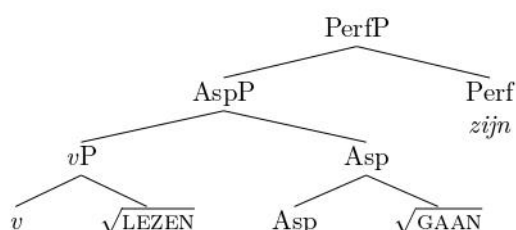
- (116) a. Emma **is/heeft** vroeg **gaan slapen**.
b. Emma **is/heeft** toen **blijven slapen**.

³⁰ Ik beperk me in de analyse tot de twee werkwoorden *gaan* en *blijven* aangezien enkel die in het experimentele onderzoek als groepsvormend werkwoord in een Asp-cluster fungeerden. Later in de analyse beperk ik me tot het werkwoord *moeten* aangezien dat als groepsvormend werkwoord in Experiment 2B fungeerde.

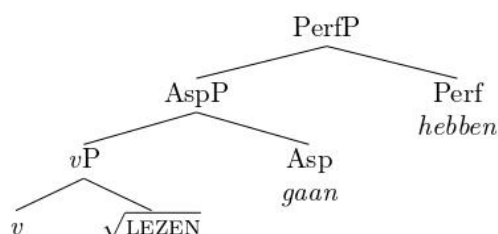
De combinatie van die drie kenmerken van *gaan* en *blijven* komt in Tabel 6.1 overeen met de kenmerken van zowel de functionele groepsvormende werkwoorden als het semi-lexicale groepsvormende werkwoord *hoeven*. Dat werkwoord bevindt zich volgens Pots (2020) in een grammaticalisatieproces van de eerste naar de tweede semi-lexicale fase (zie §2.5.2).

In het licht hiervan stel ik een analyse voor waarbij de aspectuele werkwoorden *gaan* en *blijven* zich in een grammaticalisatieproces bevinden van de tweede semi-lexicale fase van Pots (2020) naar functionele groepsvorming. *Gaan* en *blijven* bevinden zich zo tussen de status van *hoeven* en die van de functionele groepsvormende werkwoorden in. De structuur in (117a) representeert de tweede fase van semi-lexicaliteit, die in (117b) de functionele groepsvorming. Asp-clusters zijn binnen mijn analyse dus in transitie van de structuur in (117a) naar (117b).³¹

(117) a.



b.



In (117a) is het aspectuele werkwoord de uitspelling van een complex hoofd, waarin een wortel en een functioneel hoofd worden gecombineerd.³² Dat complexe hoofd projecteert vervolgens in het functionele domein van het lexicale hoofdwerkwoord. In (117b) is het aspectuele werkwoord echter de uitspelling van een functioneel hoofd in het functionele domein van het lexicale hoofdwerkwoord. Het functionele hoofd in (117b) is niet complex en bevat dus geen wortel.

Wat ik voorstel is dat variatie in hulpwerkwoordselectie bij Asp-clusters verklaard kan worden door de aan- of afwezigheid van een wortel in het aspectuele hoofd (Asp). Ik volg daarbij het uitgangspunt van het Minimalistisch Programma dat kenmerken op functionele hoofden weggestreept moeten worden tegen lexicale kenmerken (Sybesma 2002:146). In mijn analyse heeft het functionele hoofd van de PerfP een lexicaal element nodig dat voor de hulpwerkwoordselectie kan zorgen: het perfectieve hoofd heeft een wortel nodig om op het moment van Spell Out de vorm van *hebben* of *zijn* aan te nemen. Het aspectuele hoofd bevindt zich het dichtst bij het perfectieve hoofd. Wanneer dat hoofd complex is en dus een wortel bevat, zoals in (117a), wordt het lexicale kenmerk op het perfectieve hoofd weggestreept door de wortel van het aspectuele hoofd. In dat geval wordt het hulpwerkwoord van voltooid aspect geselecteerd door het aspectuele werkwoord en wordt het hulpwerkwoord van het cluster dus *zijn*, zoals in (117a).

³¹ Ik volg hier Cinque (1999) in de aanname dat het clausale functionele domein is onderverdeeld in verschillende functionele projecties, met name hier PerfP, ModP en AspP. De hiërarchische relatie tussen PerfP, ModP en AspP baseer ik op Van Craenenbroeck & Van Koppen (2016). De structuren zijn hoofdfinaal gemaakt, maar daar neem ik verder geen structurele theoretische implicaties mee aan.

³² Ik gebruik in de analyse het idee van *Distributed Morphology* (Halle & Marantz 1993) dat lexicale woorden de uitspelling zijn van wortels, en functionele woorden de uitspelling van syntactische kenmerken op functionele hoofden. Met de term 'wortel' verwijs ik dus naar de lexicale kern van individuele woorden.

Bij de functionele groepsvorming in (117b) daarentegen bevat het aspectuele hoofd geen wortel. Het lexicale kenmerk van het perfectieve hoofd kan niet worden weggestreept, dus het gaat verder in de structuur op zoek naar een wortel. Het lexicale hoofdwerkwoord bevat wel een wortel die kan zorgen voor het wegstrepen van het lexicale kenmerk van het perfectieve hoofd. In dat geval wordt de keuze van het hulpwerkwoord van voltooid aspect bepaald door het hoofdwerkwoord. In (117b) is het hulpwerkwoord van het cluster *hebben* omdat *lezen hebben* selecteert. Aangezien Asp-clusters in de analyse in transitie zijn van de structuur in (107a) naar (107b), neem ik aan dat er sprake is van structurele ambiguïteit wanneer hulpwerkwoordvariatie voorkomt. Afhankelijk van de fase van het grammaticalisatieproces waarin het aspectuele werkwoord zich bevindt, bevat het aspectuele hoofd wel of geen wortel.

De voorgestelde analyse maakt de ‘transparantie’ van de functionele groepsvormende werkwoorden in de analyse van Cardinaletti en Shlonsky (2004) concreet door de aandacht te vestigen op de aan- of afwezigheid van de wortel in het hoofd van het groepsvormende werkwoord. Bij functionele groepsvorming is het groepsvormende werkwoord transparant voor hulpwerkwoordselectie omdat het geen wortel bevat die de keuze van het hulpwerkwoord kan bepalen. Het feit dat hulpwerkwoordvariatie ongrammaticaal is wanneer het groepsvormende werkwoord een lexicaal werkwoord is (Cardinaletti & Shlonsky 2004, Pots 2020), wordt aan de hand van de voorgestelde analyse ook verklaard. Een lexicaal werkwoord is steeds de uitspelling van een wortel, en streept de kenmerken van het functionele hoofd dus al weg; er wordt niet verder naar een wortel gezocht bij het hoofdwerkwoord.

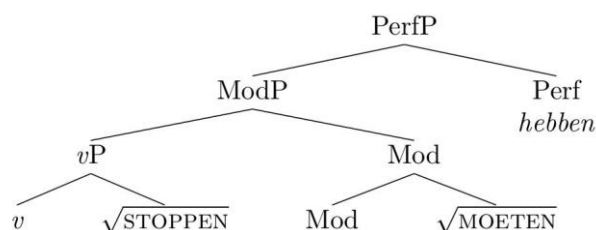
De analyse voorspelt geen invloed van de afstand tussen het hulpwerkwoord van voltooid aspect en het aspectuele werkwoord. Ik neem namelijk aan dat het perfectieve hoofd, naast naar een wortel, ook op zoek is naar een V-kenmerk (Sybesma 2002:146): het is steeds een werkwoord dat de keuze van het hulpwerkwoord bepaalt. De aanwezigheid van een bijwoord in de specificerderspositie van AspP heeft dan geen invloed op hulpwerkwoordselectie, aangezien het geen V-kenmerk bevat. Het komt dus niet in aanmerking om de kenmerken op het perfectieve hoofd weg te strepen. Dat er geen afstandseffect geattesteerd wordt in het experimentele onderzoek (§6.2.2) is in het licht van de voorgestelde analyse dus niet verrassend.

Tot slot verklaar ik de geattesteerde nationale variatie in Asp-clusters (zie §6.3.1) door een verschil in status van de aspectuele werkwoorden in Vlaanderen en Nederland. Het feit dat *hebben* meer voorkomt in Vlaanderen, impliceert dat in Vlaanderen *gaan* en *blijven* vaker voorkomen in functionele groepsvorming zoals in (117b) dan in Nederland. Het aspectuele hoofd bevat in Vlaanderen dan vaker dan in Nederland geen wortel meer. In Nederland komt nagenoeg geen hulpwerkwoordvariatie voor, dus hebben *gaan* en *blijven* in Nederland meestal een semi-lexicale status. Dat impliceert dat *gaan* en *blijven* zich in het grammaticalisatieproces van de tweede fase van semi-lexicaliteit naar functionele groepsvorming in Vlaanderen in een verder gevorderd stadium bevinden dan in Nederland.

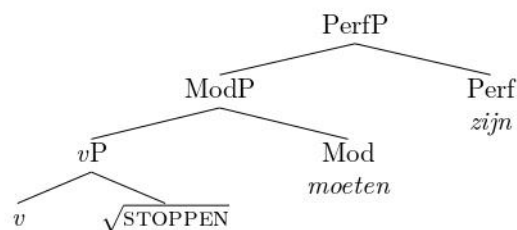
6.4.2 Asp-clusters versus Mod-clusters

De analyse die ik voorstel voor hulpwerkwoordvariatie in Asp-clusters (§6.4.2) kan worden overgenomen om de hulpwerkwoordvariatie in Mod-clusters te verklaren. Ik stel voor dat ook de modale groepsvormende werkwoorden zich in een grammaticalisatieproces bevinden tussen de tweede semi-lexicale fase van Pots (2020) en functionele groepsvorming. De structuur in (118a) geeft de tweede semi-lexicale fase voor Mod-clusters weer, die in (118b) die voor functionele groepsvorming.

(118) a.



b.

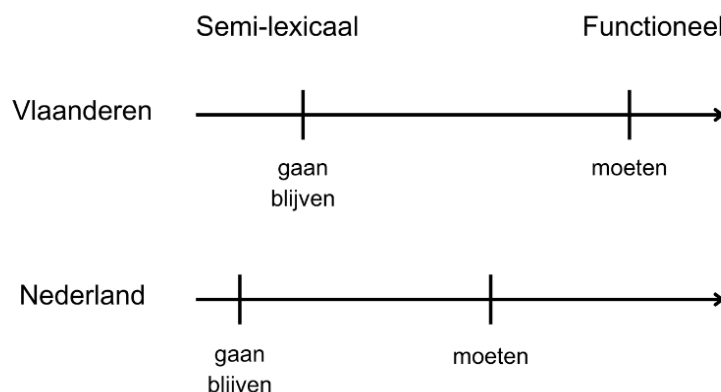


De structuren in (118) voor Mod-clusters zijn analoog aan die voor Asp-clusters. In het Mod-cluster in (118a) wordt *hebben* geselecteerd aangezien het modale hoofd een wortel bevat. Het modale hoofd kan dus het lexicale kenmerk van het perfectieve hoofd wegstrepen, waardoor het de keuze van het hulpwerkwoord van voltooid aspect bepaalt. In (118b) daarentegen is het modale werkwoord een functioneel groepsvormend werkwoord en bevat het dus geen wortel. Het modale hoofd kan het lexicale kenmerk op het perfectieve hoofd dus niet wegstrepen en daardoor de keuze van het hulpwerkwoord niet bepalen. Dat gebeurt in de eerstvolgende verbale projectie met een wortel, namelijk die van het hoofdwerkwoord. Het Z-werkwoord *stoppen* selecteert dus *zijn* als hulpwerkwoord van voltooid aspect.

Concluderend bevinden zowel de aspectuele als de modale werkwoorden zich in mijn analyse in een grammaticalisatieproces richting functionele groepsvorming. Welk hulpwerkwoord geselecteerd wordt, hangt af van de aan- of afwezigheid van een wortel in de projectie van het aspectuele of modale werkwoord. In tegenstelling tot Cardinaletti en Shlonsky (2004) en Pots (2020) ga ik er dus niet vanuit dat modale werkwoorden altijd functionele groepsvormende werkwoorden zijn. Als er een wortel aanwezig is, zijn ze semi-lexicaler dan wanneer er geen wortel aanwezig is; dan zijn ze functioneler.

Zo kan ik binnen mijn analyse het verschil in hulpwerkwoordvariatie bij Asp- en Mod-clusters verklaren. Hulpwerkwoordvariatie komt in Mod-clusters vaker voor dan in Asp-clusters. Dat impliceert dat in de projectie van de modale werkwoorden vaker geen wortel aanwezig is dan in de projectie van de aspectuele werkwoorden, die meestal wel een wortel bevat. Ik stel op basis daarvan voor dat de modale werkwoorden verder gevorderd zijn in het grammaticalisatieproces dan de aspectuele werkwoorden. Daarnaast analyseer ik de nationale variatie in Mod-clusters die aangetoond wordt in het onderzoek (zie §6.3.1) op dezelfde manier als voor de Asp-clusters. In Vlaanderen wordt *zijn* vaker geselecteerd dan in Nederland, waar het hulpwerkwoord meestal *hebben* is. Dat wil zeggen dat de modale werkwoorden in Nederland semi-lexicaler zijn dan in Vlaanderen. In Vlaanderen bevinden de modale werkwoorden zich in een verder gevorderd stadium in het grammaticalisatieproces, dus zijn ze functioneler. De projectie van het modale werkwoord bevat in Vlaanderen meestal geen wortel, in Nederland wel.

Figuur 6.1 visualiseert de verschillende positie van de aspectuele werkwoorden *gaan* en *blijven* en het modale werkwoord *moeten* in het grammaticalisatieproces van een semi-lexicale naar een functionele status, in Vlaanderen en in Nederland. De afstanden tussen de werkwoorden zijn niet gebaseerd op exacte metingen en dienen louter ter illustratie van de verschillende posities in het proces.



Figuur 6.1: Positie aspectuele en modale werkwoorden in grammaticalisatieproces per land (Vlaanderen vs. Nederland)

Op die manier wordt mijn analyse voor het verschil tussen Asp- en Mod-clusters in Figuur 6.1 samengevat. Ik stel voor dat zowel de aspectuele werkwoorden *gaan* en *blijven* als het modale werkwoord *moeten* zich in een grammaticalisatieproces bevinden van een semi-lexicale naar een functionele status. Binnen de structuur van de tweede fase van semi-lexicaliteit bevat de projectie van het groepsvormende werkwoord een wortel en kan het dus de keuze van het hulpwerkwoord bepalen; binnen de structuur van functionele groepsvorming bevat de projectie van het groepsvormende werkwoord geen wortel en wordt het hulpwerkwoord bepaald door het lexicale hoofdwkwoord. Aangezien hulpwerkwoordvariatie in Mod-clusters meer voorkomt dan in Asp-clusters, is *moeten* verder gevorderd in het grammaticalisatieproces dan *gaan* en *blijven*.

Daarnaast blijkt uit Figuur 6.1 dat de positie van het groepsvormende werkwoord in het grammaticalisatieproces nationale variatie vertoont. Zowel de modale als de aspectuele werkwoorden blijken functioneler in Vlaanderen dan in Nederland. Zoals werd aangetoond in deze masterproef, komt *zijn* in Mod-clusters en *hebben* in Asp-clusters immers meer voor in Vlaanderen dan in Nederland. Zo sluit mijn analyse voor nationale variatie aan bij de analyse van Draye en Van der Horst (2006), die stellen dat de modale werkwoorden zich in Vlaanderen in een verder stadium van een grammaticalisatieproces bevinden (zie §2.5.1). Op basis van de verzamelde evidentie voor de hulpwerkwoordvariatie in Asp-clusters, heb ik die analyse echter uitgebreid door voor te stellen dat ook de aspectuele werkwoorden zich in Vlaanderen in een verder stadium bevinden.

De resultaten van mijn onderzoek en analyse bieden zo ondersteuning voor het bestaan van syntactische verschillen tussen Vlaanderen en Nederland, die tot voor kort grotendeels onder de radar bleven in de onderzoeksliteratuur (De Troij 2023, Grondelaers et al. 2020). Het idee dat de modale en aspectuele werkwoorden zich in Vlaanderen in een verder stadium van het proces bevinden, is echter opmerkelijk. Bij alternanties waarbij een grammaticale constructie geleidelijk wordt vervangen door een andere, blijft de oudere variant immers vaak langer bewaard in Vlaanderen (De Troij 2023:88). De voorgestelde analyse biedt daarmee een aanknopingspunt voor vervolgonderzoek naar nationale variatie.

6.4.3 Irrealis

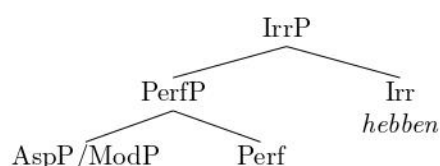
De voorgestelde analyse voor hulpwerkwoordvariatie in Asp-clusters biedt geen verklaring voor de positieve invloed van irrealis op de selectie van *hebben* die in het onderzoek werd vastgesteld (§6.2.3). Ik beargumenteer dat het effect van irrealis op hulpwerkwoordselectie losstaat van de status van het aspectuele werkwoord en stel op basis daarvan een formele analyse voor de invloed van irrealis voor.

De positieve invloed van irrealis op de selectie van *hebben* wordt in de resultaten niet enkel in Asp-clusters geattesteerd, zoals in (119a), maar ook bij de voltooide tijd van werkwoorden die *zijn* selecteren, zoals in (119b) (zie §4.2.6). Het effect van irrealis treedt dus algemener op bij hulpwerkwoordselectie dan enkel in werkwoordsclusters met een aspectueel werkwoord.

- (119) a. **Had** die man mij **blijven volgen**, voelde ik mij onveilig. E4_3_irr_inv_H
 b. Voelde mijn zus zich niet goed, ik **had** wat langer **gebleven**! E3_3_irr_H

Op basis daarvan vermoed ik dat in een Asp-cluster als (119a) de selectie van het hulpwerkwoord *hebben* toegeschreven kan worden aan de irrealis. Daarmee sluit mijn analyse aan bij het vermoeden van Coussé en Van de Velde (2014) voor Mod-clusters (zie §2.5.2). Ik ga voor de analyse van het irrealiseffect uit van een functioneel hoofd specifiek voor irrealis, zoals voorgesteld in de universele hiërarchie van Cinque (1999). Dat irrealishoofd staat hoger in de structuur dan PerfP (Cinque 1999:76), zoals geïllustreerd in (120). Het functionele irrealishoofd wordt dan uitgespeld als het werkwoord *hebben*.

(120)



Die analyse wordt ondersteund door de relatief hoge acceptabiliteitscores ($M = 3.91$, $SD = 2.20$) van de respondenten voor de irrealiszin in (121), die als filler werd opgenomen in de stimuli (zie §4.1.1.3). Hoewel ik binnen het bestek van deze masterproef niet dieper inga op de analyse van de zin in (121), lijkt er naast het hulpwerkwoord van voltooid aspect *zijn* ook een hulpwerkwoord van irrealis aanwezig te zijn, met name *had*. Dat suggereert dat *hebben* wel degelijk kan voorkomen in een irrealishoofd.

- (121) Dit **had** het **geworden kunnen zijn**, maar het is anders afgelopen. F15

6.5 Beperkingen van het onderzoek en suggesties voor vervolgonderzoek

Het onderzoek kent enkele beperkingen, in het bijzonder wat betreft het experimentele luik. Ten eerste is de taak van het experiment (zie §4.1.3) niet zuiver geformuleerd. De respondent rapporteert met de acceptabiliteitsscore immers zowel een oordeel op basis van eigen taalgevoel als op basis van het taalgebruik van zijn/haar omgeving. Op die manier is het niet erg duidelijk wat de acceptabiliteitsscore precies inhoudt. Ten tweede zijn de experimentele resultaten vermoedelijk sterk beïnvloed door keuzes die ik maakte bij het opstellen van de stimuli. Zo waren de selectiecriteria voor de hoofdwerkwoorden wellicht te strikt, waardoor bijvoorbeeld het werkwoord *winnen* werd opgenomen als HZ-werkwoord terwijl dat achteraf gezien niet optimaal was (zie §6.2.1). Een vroeger en beter inhaken van het corpusonderzoek op de experimentele stimuli had dat wellicht kunnen vermijden.

Daarnaast opent deze masterproef verschillende paden voor vervolgonderzoek. In de eerste plaats verdient het afstandseffect aandacht in toekomstig onderzoek, aangezien de resultaten van het experiment en het corpusonderzoek wat betreft de invloed van afstand divergeren. Ook de registervariatie kan in vervolgonderzoek uitgediept worden: in een experimentele studie kan het interessant zijn om met gesproken stimuli te werken, en in corpusonderzoek biedt een diepgaandere analyse van de invloed van register op het effect van andere factoren mogelijk interessante inzichten.

Het geattesteerde irrealiseffect verdient daarnaast ook verdere aandacht. Aangezien ik vermoed dat de bandbreedte van het irrealiseffect niet volledig in kaart werd gebracht (zie §6.2.3), is het interessant om de invloed van irrealis op hulpwerkwoordselectie te onderzoeken zonder inversie als beïnvloedende factor. Daarnaast werd de formele analyse wat betreft de invloed van irrealis op hulpwerkwoordselectie nog niet uitgebreid verkend en biedt die dus mogelijkheid tot verder onderzoek.

Daarnaast biedt de geattesteerde nationale variatie en de voorgestelde analyse voor die variatie mogelijkheden tot vervolgonderzoek, zoals beargumenteerd in §6.4.2. Ook de rol van regionale variatie en dialect verdient verdere aandacht. Het vergelijken van de oordelen van dialectsprekers en niet-dialectsprekers is een interessante invalshoek voor vervolgonderzoek aangezien de hulpwerkwoordselectie bij voltooid aspect kan verschillen bij dialectsprekers, met name in de West-Vlaamse dialecten (zie §2.4.2.2). Wellicht biedt zulk onderzoek ook meer inzicht in de positie van West-Vlaanderen wat hulpwerkwoordselectie betreft, die in deze masterproef niet eenduidig kon worden aangetoond. Vervolgonderzoek met een meer gelijke verdeling van de datapunten over de regio's zou de positie van West-Vlaanderen en de andere regionale variatie nog verder kunnen bestuderen.

Ook de invloed van het afzonderlijke aspectuele werkwoord is een belangrijke piste voor vervolgonderzoek. In deze masterproef werd gefocust op de werkwoorden *gaan* en *blijven*, maar vermoedelijk nemen de werkwoorden *komen* en *beginnen* een andere positie in binnen het geschetste grammaticalisatieproces.

Tot slot, aangezien het corpusonderzoek geen alternantiestudie betrof, kan er op basis van de resultaten geen verhouding tussen *hebben* en *zijn* worden vastgesteld in Asp-clusters. Het vaststellen van die verhouding is een interessante volgende stap in het in kaart brengen van de hulpwerkwoordvariatie in Asp-clusters. De dataset van deze masterproef voor Asp-clusters met *hebben* zou in dat opzicht uitgebreid kunnen worden met de *zijn*-varianten.

7 Conclusie

In deze masterproef werd onderzocht welke factoren de hulpwerkwoordselectie in drieledige werkwoordsclusters met een aspectueel werkwoord (Asp-clusters) bepalen. Daarvoor werd de invloed van vijf taalinterne en -externe factoren geanalyseerd: het type hoofdwerkwoord, de afstand tussen het hulpwerkwoord en het aspectuele werkwoord, irrealis, geografische regio en register. De selectie van die factoren gebeurde op basis van onderzoek naar hulpwerkwoordvariatie in drieledige werkwoordsclusters met modale werkwoorden (Mod-clusters) en het beperkte beschikbare onderzoek naar hulpwerkwoordvariatie in Asp-clusters (Hofmans 1981, Van Eynde et al. 2016, Dhont 2024).

De invloed van de factoren werd onderzocht aan de hand van een experiment en een corpusonderzoek. In het experimentele onderzoek beoordeelden 1105 Vlaamse en Nederlandse respondenten zinnen met Asp-clusters. De acceptabiliteitsoordelen werden geanalyseerd met lineaire regressieanalyses. Het corpusonderzoek gebeurde in het SoNaR-corpus en het Corpus Gesproken Nederlands, en resulteerde in een dataset van 346 zinnen die Asp-clusters met *hebben* bevatten. Met frequentieanalyses werd het voorkomen van de taalinterne en -externe factoren in de dataset in kaart gebracht.

Het onderzoek bevestigt dat hulpwerkwoordselectie in Asp-clusters beïnvloed wordt door irrealis, geografische regio en register. Er wordt aangetoond dat irrealis een significant positief effect heeft op de aanvaardbaarheid van *hebben*, zowel in Asp-clusters als bij hoofdwerkwoorden die *zijn* selecteren. Daarnaast tonen de resultaten aan dat het hulpwerkwoord *hebben* in Vlaanderen, en vooral in West-Vlaanderen, meer aanvaard wordt en meer voorkomt dan in Nederland. Tot slot bevestigen de resultaten dat Asp-clusters met *hebben* het vaakst voorkomen in informeel taalgebruik, zoals in spontane conversaties en discussiefora, en minder in formele taal. Het type hoofdwerkwoord blijkt geen significante invloed te hebben op hulpwerkwoordselectie in Asp-clusters, en de invloed van afstand wordt in de resultaten niet eenduidig aangetoond.

De resultaten van het onderzoek leiden tot een formele analyse waarin hulpwerkwoordselectie in drieledige werkwoordsclusters bepaald wordt door de aan- of afwezigheid van een wortel in de projectie van het groepsvormende werkwoord. De hulpwerkwoordvariatie wordt verklaard doordat het aspectuele werkwoord zich in een grammaticalisatieproces van semi-lexicaliteit naar functionaliteit bevindt. Gebaseerd op analyses voor hulpwerkwoordvariatie in Mod-clusters (Cardinaletti & Shlonsky 2004, Draye en Van der Horst 2006, Pots 2020) wordt het verschil tussen Asp- en Mod-clusters en de nationale variatie bij hulpwerkwoordselectie verklaard door een verschillende positionering binnen het grammaticalisatieproces. Daarnaast suggereert de voorgestelde analyse dat de geattesteerde invloed van irrealis op hulpwerkwoordselectie verklaard wordt door een specifiek irrealishoofd.

Deze masterproef brengt voor het eerst de hulpwerkwoordvariatie in drieledige werkwoordsclusters met een aspectueel werkwoord in Vlaanderen en Nederland uitgebreid in kaart. De resultaten en de voorgestelde analyse openen zo verschillende perspectieven voor vervolgonderzoek naar de vraag of we zijn (of hebben?) gaan twijfelen over het hulpwerkwoord.

8 Referenties

- Ackema, Peter, en Antonella Sorace. 2017. Auxiliary Selection. In *The Wiley Blackwell Companion to Syntax*, ed. Martin Everaert en Henk van Riemsdijk, 1-32. Wiley Blackwell.
- Augustinus, Liesbeth. 2015. Complement raising and cluster formation in Dutch: A treebank-supported investigation. PhD-dissertatie, Katholieke Universiteit Leuven.
- Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In *Syntax and variation. Reconciling the biological and the social*, ed. Leonie Cornips en Karen P. Corrigan, 233–264. Amsterdam: John Benjamins.
- Bartoń, Kamil. 2024. *MuMIn: Multi-Model Inference*. R package version 1.48.4. <https://CRAN.R-project.org/package=MuMIn>
- Broekhuis, Hans, en Norbert Corver. 2016. *Syntax of Dutch: Verbs and Verb Phrases*. Amsterdam: Amsterdam University Press.
- Cardinaletti, Anna, en Ur Shlonsky. 2004. Clitic positions and restructuring in Italian. *Linguistic Inquiry* 35(4):519-557.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.
- Cinque, Guglielmo. 2001. “Restructuring” and the order of aspectual and root modal heads. In *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, ed. Guglielmo Cinque en Giampaolo Salvi, 137-155. Amsterdam: Elsevier.
- Coussé, Evie. 2014. Lexical expansion in the HAVE and BE perfect in Dutch: A constructionist prototype account. *Diachronica* 31(2):159-191.
- Coussé, Evie, en Freek Van de Velde. 2014. Hulpwerkwoordselectie in drieledige perfecta met een modaal: Een alternatieve historische verklaring. In *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, ed. Freek Van de Velde, Hans Smessaert, Frank Van Eynde en Sara Verbrugge, 349-364. Leuven: Universitaire Pers Leuven.
- De Rooij, Jaap. 1988. *Van hebben naar zijn: het gebruik van hebben en zijn in de voltooide tijden (actief) van zijn, gaan, vergeten en verliezen in standaardtaal, ouder Nederlands en dialect*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- De Troij, Robbert. 2023. Natiolectal variation in Dutch grammar. A data-driven approach. PhD-dissertatie, Radboud Universiteit Nijmegen/Katholieke Universiteit Leuven.
- De Voofs, Cornelis G. N. 1967. *Nederlandse Spraakkunst*. Groningen: Wolters.
- Dhont, Arne. 2024. Pluricentrisme in de praktijk. Naar een methode voor de codificatie van grammaticale verschillen tussen Belgisch Nederlands en Nederlands Nederlands. PhD-dissertatie, Universiteit Gent.

- Draye, Luk, en Joop Van der Horst. 2006. Hebben/haben versus zijn/sein in drieledige werkwoordsgroepen in het Duits en het Nederlands. In *Nederlands tussen Duits en Engels. Handelingen van de workshop op 30 september en 1 oktober 2005 aan de Freie Universität Berlin*, ed. Matthias Hüning, Ulrike Vogl, Ton van der Wouden en Arie Verhagen, 229-242. Leiden: Stichting Neerlandistiek.
- Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8(15): 1-27.
- Fox, John, en Sanford Weisberg. 2019. *An R Companion to Applied Regression*. 3de editie. Thousand Oaks: Sage.
- Gries, Stefan Th. 2021. *Statistics for Linguistics with R: A practical introduction*. Berlijn: De Gruyter.
- Grondelaers, Stefan, Robbert De Troij, Dirk Speelman, en Antal Van den Bosch. 2020. Vissen naar variatie: Digitaal op zoek naar onbekende Noord/Zuid-verschillen in de grammatica van het Nederlands. *Nederlandse Taalkunde* 25(1):73-99.
- Grondelaers, Stefan, Roeland van Hout, Hans van Halteren, en Esther Veerbeek. 2023. Why do we say them when we know it should be they? Twitter as a resource for investigating nonstandard syntactic variation in The Netherlands. *Language Variation and Change* 35:223-245.
- Haeseryn, Walter, Kristin Romijn, Guido Geerts, Jaap de Rooij, en Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen: Martinus Nijhoff Uitgevers.
- Halle, Morris, en Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20*, ed. Kenneth Hale en Samuel Jay Keyser, 111-176. Cambridge: MIT Press.
- Hofmans, Mark. 1981. Hebben of zijn: een enquête naar het gebruik van “hebben” of “zijn” in de konstruktie Thww + Mhww + HWW in Nederland en Vlaanderen. *Tijdschrift van de Vrije Universiteit Brussel* 1: 62-114.
- Hofmans, Mark. 1982a. Hebben of zijn en de deverbalisering van de modale werkwoorden in het Nederlands. *Linguistics in Belgium* 5: 81-109.
- Hofmans, Mark. 1982b. Aspects of deverbalsation: the semi-modals. *Linguistics in the Netherlands*: 83-95.
- Hofmans, Mark. 1985. Over demodalisering bij semi-modalen. *Linguistics in Belgium* 6: 72-83.
- Hothorn Torsten, Frank Bretz, en Peter Westfall. 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50(3): 346-363.
- Kern, Johan Hendrik. 1912. *De met het participium praeteriti omschreven werkwoordsvormen in 't Nederlands*. Amsterdam: Müller.
- Keuleers, Emmanuel, Marc Brysbaert, en Boris New. 2010. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods* 42:643-650.

- Kuznetsova, Alexandra, Per B. Brockhoff, en Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13):1-26.
- Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Lüdecke, Daniel. 2024. *sjstats: Statistical Functions for Regression Models* (Version 0.19.0). <https://CRAN.R-project.org/package=sjstats>
- Lüdecke, Daniel. 2024. *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.16. <https://CRAN.R-project.org/package=sjPlot>
- Meyer, David, Achim Zeileis, Michael Friendly, en Kurt Hornik. 2023. *vcd: Visualizing Categorical Data*. R package version 1.4-12. <https://CRAN.R-project.org/package=vcd>
- Onze Taal. 2024. *Is het had of was in de zin: 'Ik stond in de file, anders had/was ik wel eerder thuis geweest'?*. <https://onzetaal.nl/taalloket/had-geweest-was-geweest>
- Onze Taal. 2024. *Wat is juist: 'Ze hebben daar opgetreden' of 'Ze zijn daar opgetreden'?*. <https://onzetaal.nl/taalloket/opgetreden-hebben-zijn>
- Oostdijk, Nelleke. 2002. The design of the Spoken Dutch Corpus. In *New Frontiers of Corpus Research*, ed. Pam Peters, Peter Collins, en Adam Smith, 105-112. Amsterdam: Brill.
- Oostdijk, Nelleke. 2004. *Het project Corpus Gesproken Nederlands*. https://taalmaterialen.ivdnt.org/wpcontent/uploads/documentatie/cgn_website/doc_Dutch/to_pics/index.htm
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste, en Ineke Schuurman. 2013. The construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In *Essential Speech and Language Technology for Dutch: Results by the STEVIN-programme*, ed. Peter Spyns en Jan Odijk, 219-247. Springer.
- Paardekooper, Petrus Cornelis. 1978. *Beknopte ABN-syntaksis*. 5^{de} druk. Eindhoven: Paardekooper.
- Posit Team. 2024. *RStudio: Integrated Development Environment for R*. Posit Software. Boston. <http://www.posit.co/>
- Pots, Cora. 2020. *Roots in Progress. Semi-lexicity in the Dutch and Afrikaans verbal domain*. PhD-dissertatie, Katholieke Universiteit Leuven.
- R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Wenen. <https://www.R-project.org/>
- Reynaert, Martin, Nelleke Oostdijk, Orphée De Clercq, Henk Van den Heuvel, en Franciska de Jong. 2010. Balancing SoNaR: IPR versus processing issues in a 500-million-word written Dutch Reference Corpus. In *Seventh conference on International Language Resources and Evaluation (LREC)*, 2693-2698. European Language Resources Association.

- Speelman, Dirk, en Mariana Montes. 2022. *mclm: Mastering Corpus Linguistics Methods*. R package version 0.2.7. <https://CRAN.R-project.org/package=mclm>.
- Sybesma, Rint. 2002. *Syntaxis: een generatieve inleiding*. Bussum: Coutinho.
- Taaladvies. 2024. *Beginnen + infinitief*. <https://taaladvies.net/beginnen-en-infinitief/>
- Taaladvies. 2024. *Gaan / zullen (algemeen)*. <https://taaladvies.net/gaan-of-zullen-algemeen/>
- Taaladvies. 2024. *Gevolgd (ik ben / ik heb-)*. <https://taaladvies.net/ik-ben-of-heb-hem-gevolgd/>
- Taaladvies. 2024. *Gewonnen (ik ben / ik heb -)*. <https://taaladvies.net/ik-ben-of-ik-heb-gewonnen/>
- Van Craenenbroeck, Jeroen, en Marjo Van Koppen. 2017. Look! Kaatje is still not heard sing a song (even without government). In *GLOW workshop What happened to government?*.
- Van Craenenbroeck, Jeroen. 2019. Dialectologie en Nederlandse zinsbouw. *Nederlandse Taalkunde* 24(1):39-68.
- Van der Horst, Joop. 1988. De Croma-verandering. *Onze Taal* 57:172-174.
- Van der Horst, Joop. 2005. Hoe is dat kunnen gebeuren? Vlaamse zinswending als toekomstig Nederlands. *Onze Taal* 74:100-103.
- Van der Horst, Joop. 2008. *Geschiedenis van de Nederlandse syntaxis*. Leuven: Universitaire Pers Leuven.
- Van Eerten, Laura. 2007. Over het Corpus Gesproken Nederlands. *Nederlandse Taalkunde* 12(3):194-215.
- Van Eynde, Frank, Liesbeth Augustinus, Ineke Schuurman, en Vincent Vandeghinste. 2016. Hebben of zijn bij IPP's. *Leuvense Bijdragen* 100:11-28.
- Van Leeuwen, Anne. 2009. Hij was/had moeten gaan...?. Een onderzoek naar de keuze van het hulpwerkwoord in clusters met een modaal-, aspectueel- en hoofdwerkwoord. Bachelorscriptie, Universiteit Utrecht.
- Van Noord, Gertjan, Gosse Bouma, Frank Van Eynde, Daniël de Kok, Jelmer Van der Linde, Ineke Schuurman, en Vincent Vandeghinste. 2013. Large scale syntactic annotation of written Dutch: Lassy. In *Essential Speech and Language Technology for Dutch: Results by the STEVIN-programme*, ed. Peter Spyns en Jan Odijk, 147-164. Springer.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana

Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, en Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43):1686.

Wickham, Hadley, en Jennifer Bryan. 2023. *readxl: Read Excel Files*. R package version 1.4.3, <https://CRAN.R-project.org/package=readxl>

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, en Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. <https://CRAN.R-project.org/package=dplyr>

Wurmbrand, Susi. 2001. *Infinitives: Restructuring and clause structure*. Mouton de Gruyter.

Appendix

Appendix A. Overzicht stimuli

Experiment 1

Code	Variant
E1_1_A0_Z	De Argentijnen zijn blijven winnen en kwalificeerden zich zo voor het WK.
E1_1_A0_H	De Argentijnen hebben blijven winnen en kwalificeerden zich zo voor het WK.
E1_1_A1_Z	De Argentijnen zijn toen blijven winnen en kwalificeerden zich zo voor het WK.
E1_1_A1_H	De Argentijnen hebben toen blijven winnen en kwalificeerden zich zo voor het WK.
E1_1_A2_Z	De Argentijnen zijn toen maar blijven winnen en kwalificeerden zich zo voor het WK.
E1_1_A2_H	De Argentijnen hebben toen maar blijven winnen en kwalificeerden zich zo voor het WK.
E1_2_A0_Z	De turnster is blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_2_A0_H	De turnster heeft blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_2_A1_Z	De turnster is daar blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_2_A1_H	De turnster heeft daar blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_2_A2_Z	De turnster is daar toen blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_2_A2_H	De turnster heeft daar toen blijven winnen, dus nu mag ze naar de Olympische Spelen.
E1_3_A0_Z	De pianist is blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_3_A0_H	De pianist heeft blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_3_A1_Z	De pianist is toen blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_3_A1_H	De pianist heeft toen blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_3_A2_Z	De pianist is toen maar blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_3_A2_H	De pianist heeft toen maar blijven winnen en zit nu in de finale van de muziekwedstrijd.
E1_4_A0_Z	Ella is blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_4_A0_H	Ella heeft blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_4_A1_Z	Ella is toen blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_4_A1_H	Ella heeft toen blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_4_A2_Z	Ella is toen lang blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_4_A2_H	Ella heeft toen lang blijven winnen en zit nu in de finale van De Slimste Mens ter Wereld.
E1_5_A0_Z	De wielrenner is blijven winnen en mag dus meedoen aan de Tour de France.
E1_5_A0_H	De wielrenner heeft blijven winnen en mag dus meedoen aan de Tour de France.
E1_5_A1_Z	De wielrenner is maar blijven winnen en mag dus meedoen aan de Tour de France.
E1_5_A1_H	De wielrenner heeft maar blijven winnen en mag dus meedoen aan de Tour de France.
E1_5_A2_Z	De wielrenner is toen maar blijven winnen en mag dus meedoen aan de Tour de France.
E1_5_A2_H	De wielrenner heeft toen maar blijven winnen en mag dus meedoen aan de Tour de France.
E1_6_A0_Z	De acteurs zijn blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.
E1_6_A0_H	De acteurs hebben blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.
E1_6_A1_Z	De acteurs zijn daar blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.
E1_6_A1_H	De acteurs hebben daar blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.
E1_6_A2_Z	De acteurs zijn daar maar blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.
E1_6_A2_H	De acteurs hebben daar maar blijven winnen, dus nu mogen ze naar een theaterfestival in Japan.

Experiment 2A

Code	Variant
E2A_1_sla_H_Z	Alex is vroeg gaan slapen, want hij moet morgen om zes uur op.
E2A_1_sla_H_H	Alex heeft vroeg gaan slapen, want hij moet morgen om zes uur op.
E2A_2_lez_H_Z	Leyla is veel gaan lezen toen ze naar de universiteit ging.
E2A_2_lez_H_H	Leyla heeft veel gaan lezen toen ze naar de universiteit ging.
E2A_3_win_HZ_Z	Hasan is meer gaan winnen toen hij een andere trainer kreeg.
E2A_3_win_HZ_H	Hasan heeft meer gaan winnen toen hij een andere trainer kreeg.
E2A_4_vol_HZ_Z	De politie is hem gaan volgen, maar ze kregen hem niet te pakken.
E2A_4_vol_HZ_H	De politie heeft hem gaan volgen, maar ze kregen hem niet te pakken.

Experiment 2B

Code	Variant
E2B_1_sto_Z_Z	Lotte is toen moeten stoppen met studeren omdat haar vader erg ziek werd.
E2B_1_sto_Z_H	Lotte heeft toen moeten stoppen met studeren omdat haar vader erg ziek werd.
E2B_2_tro_Z_Z	Michelle is snel moeten trouwen, want ze bleek zwanger.
E2B_2_tro_Z_H	Michelle heeft snel moeten trouwen, want ze bleek zwanger.
E2B_3_win_HZ_Z	De dansers zijn toen moeten winnen om door te kunnen gaan naar de finale.
E2B_3_win_HZ_H	De dansers hebben toen moeten winnen om door te kunnen gaan naar de finale.
E2B_4_vol_HZ_Z	Zijn vader is hem moeten volgen tot aan de deur, anders durfde hij niet naar binnen.
E2B_4_vol_HZ_H	Zijn vader heeft hem moeten volgen tot aan de deur, anders durfde hij niet naar binnen.

Experiment 3

Code	Variant
E3_1_rea_Z	Ik maakte me veel zorgen, dus ik ben snel naar huis gegaan.
E3_1_rea_H	Ik maakte me veel zorgen, dus ik heb snel naar huis gegaan.
E3_1_irr_Z	Maakte ik me zo veel zorgen, ik was snel naar huis gegaan!
E3_1_irr_H	Maakte ik me zo veel zorgen, ik had snel naar huis gegaan!
E3_2_rea_Z	Ik moest alles nog klaarzetten, dus ik ben wat vroeger gekomen.
E3_2_rea_H	Ik moest alles nog klaarzetten, dus ik heb wat vroeger gekomen.
E3_2_irr_Z	Moest ik alles nog klaarzetten, ik was wat vroeger gekomen!
E3_2_irr_H	Moest ik alles nog klaarzetten, ik had wat vroeger gekomen!
E3_3_rea_Z	Mijn zus voelde zich niet goed, dus ik ben wat langer gebleven.
E3_3_rea_H	Mijn zus voelde zich niet goed, dus ik heb wat langer gebleven.
E3_3_irr_Z	Voelde mijn zus zich niet goed, ik was wat langer gebleven!
E3_3_irr_H	Voelde mijn zus zich niet goed, ik had wat langer gebleven!
E3_4_rea_Z	Ik wist dat van tevoren, dus ik ben pas later begonnen.
E3_4_rea_H	Ik wist dat van tevoren, dus ik heb pas later begonnen.
E3_4_irr_Z	Wist ik dat van tevoren, ik was pas later begonnen!
E3_4_irr_H	Wist ik dat van tevoren, ik had pas later begonnen!

Experiment 4

Code	Variant
E4_1_rea_ninv_Z	De spion was haar blijven volgen, hij kende nu haar geheim.
E4_1_rea_ninv_H	De spion had haar blijven volgen, hij kende nu haar geheim.
E4_1_irr_ninv_Z	Was de spion haar blijven volgen, hij kende nu haar geheim.
E4_1_irr_ninv_H	Had de spion haar blijven volgen, hij kende nu haar geheim.
E4_1_irr_inv_Z	Was de spion haar blijven volgen, kende hij nu haar geheim.
E4_1_irr_inv_H	Had de spion haar blijven volgen, kende hij nu haar geheim.
E4_2_rea_ninv_Z	De politie was hem blijven volgen, hij werd op heterdaad betrapt.
E4_2_rea_ninv_H	De politie had hem blijven volgen, hij werd op heterdaad betrapt.
E4_2_irr_ninv_Z	Was de politie hem blijven volgen, hij werd op heterdaad betrapt.
E4_2_irr_ninv_H	Had de politie hem blijven volgen, hij werd op heterdaad betrapt.
E4_2_irr_inv_Z	Was de politie hem blijven volgen, werd hij op heterdaad betrapt.
E4_2_irr_inv_H	Had de politie hem blijven volgen, werd hij op heterdaad betrapt.
E4_3_rea_ninv_Z	Die man was mij blijven volgen, ik voelde mij onveilig.
E4_3_rea_ninv_H	Die man had mij blijven volgen, ik voelde me onveilig.
E4_3_irr_ninv_Z	Was die man mij blijven volgen, ik voelde mij onveilig.
E4_3_irr_ninv_H	Had die man mij blijven volgen, ik voelde mij onveilig.
E4_3_irr_inv_Z	Was die man mij blijven volgen, voelde ik mij onveilig.
E4_3_irr_inv_H	Had die man mij blijven volgen, voelde ik mij onveilig.
E4_4_rea_ninv_Z	Dirk was haar blijven volgen, hij wist nu waar ze woonde.
E4_4_rea_ninv_H	Dirk had haar blijven volgen, hij wist nu waar ze woonde.
E4_4_irr_ninv_Z	Was Dirk haar blijven volgen, hij wist nu waar ze woonde.
E4_4_irr_ninv_H	Had Dirk haar blijven volgen, hij wist nu waar ze woonde.
E4_4_irr_inv_Z	Was Dirk haar blijven volgen, wist hij nu waar ze woonde.

E4_4_irr_inv_H	Had Dirk haar blijven volgen, wist hij nu waar ze woonde.
E4_5_rea_ninv_Z	Die vrouw was mij blijven volgen, ik werd er heel nerveus van.
E4_5_rea_ninv_H	Die vrouw had mij blijven volgen, ik werd er heel nerveus van.
E4_5_irr_ninv_Z	Was die vrouw mij blijven volgen, ik werd er heel nerveus van.
E4_5_irr_ninv_H	Had die vrouw mij blijven volgen, ik werd er heel nerveus van.
E4_5_irr_inv_Z	Was die vrouw mij blijven volgen, werd ik er heel nerveus van.
E4_5_irr_inv_H	Had die vrouw mij blijven volgen, werd ik er heel nerveus van.
E4_6_rea_ninv_Z	Die boze hond was mij blijven volgen, ik raakte in paniek!
E4_6_rea_ninv_H	Die boze hond had mij blijven volgen, ik raakte in paniek!
E4_6_irr_ninv_Z	Was die boze hond mij blijven volgen, ik raakte in paniek!
E4_6_irr_ninv_H	Had die boze hond mij blijven volgen, ik raakte in paniek!
E4_6_irr_inv_Z	Was die boze hond mij blijven volgen, raakte ik in paniek!
E4_6_irr_inv_H	Had die boze hond mij blijven volgen, raakte ik in paniek!

Fillers

Code	Filler
F1	Rayan heeft niet kunnen komen helpen omdat hij zich gisteren geblesseerd had tijdens de wedstrijd.
F2	Kaat is lang moeten blijven wachten omdat de trein vertraging had.
F3	Sara heeft niet mogen gaan zwemmen, want haar huiswerk was niet af.
F4	Onze ploeg is gisteren gewonnen omdat de keeper van de andere ploeg ziek was.
F5	Ik had mijn handschoenen verloren, maar ik heb ze inmiddels teruggevonden.
F6	Kijk nog eens na of je niets vergeten bent, want we komen niet meer terug!
F7	Dat kindje heeft me gevolgd tot aan de straat, maar liep daarna terug naar haar mama.
F8	Hamza moet dat boek toch hebben gelezen, hij is leerkracht Nederlands!
F9	Het is al even geleden dat Bart van de trap gevallen is, maar hij heeft nog steeds pijn.
F10	Meyra heeft het eindelijk aan Tom durven vertellen, maar hij geloofde haar niet.
F11	Oma kan hier rustig zitten te lezen, maar ik moet studeren.
F12	Simon heeft die jas proberen te wassen, maar die was veel te vuil.
F13	Ik had echt niet verwacht dat Vic een slang als huisdier heeft, ik kan het bijna niet geloven.
F14	Ik had moeten meegaan hebben, dan leefde Eva nu nog!
F15	Dit had het geworden kunnen zijn, maar het is anders afgelopen.
F16	Had ik geweten dat het ging regenen, had ik een paraplu meegenomen.
F17	Had ze maar een goede danseres geweest, dan kon ze zeker meedoen aan de musical.
F18	Had ik meer tijd, ik kon dat boek eindelijk eens lezen.
F19	Als ze hier zo'n dingen zouden zeggen, was ik al lang vertrokken.
F20	De hamburgers roken heel lekker, dus van alle kanten kwamen mensen aanlopen.
F21	Yasmin is al in veel landen geweest, dus ze weet veel over verschillende culturen.
F22	Kristien weet alles wat ze hoeft te weten, ze kan nu aan de slag.
F23	Ik vind zwemmen niet leuk meer, dus ik ga niet meer.
F24	Emma begon meteen te huilen toen ik achter haar gezondheid vroeg.

Appendix B. Condities experiment

Conditie 1		Conditie 2		Conditie 3	
1	E1_1_A0_Z	1	E1_1_A0_H	1	E1_1_A1_Z
2	E1_2_A2_H	2	E1_2_A0_Z	2	E1_2_A0_H
3	E1_3_A2_Z	3	E1_3_A2_H	3	E1_3_A0_Z
4	E1_4_A1_H	4	E1_4_A2_Z	4	E1_4_A2_H
5	E1_5_A1_Z	5	E1_5_A1_H	5	E1_5_A2_Z
6	E1_6_A0_H	6	E1_6_A1_Z	6	E1_6_A1_H
7	E2A_1_sla_H_Z	7	E2A_1_sla_H_H	7	E2A_1_sla_H_Z
8	E2A_2_lez_H_H	8	E2A_2_lez_H_Z	8	E2A_2_lez_H_H
9	E2A_3_win_HZ_Z	9	E2A_3_win_HZ_H	9	E2A_3_win_HZ_Z
10	E2A_4_vol_HZ_H	10	E2A_4_vol_HZ_Z	10	E2A_4_vol_HZ_H
11	E2B_1_sto_Z_H	11	E2B_1_sto_Z_Z	11	E2B_1_sto_Z_H
12	E2B_2_tro_Z_Z	12	E2B_2_tro_Z_H	12	E2B_2_tro_Z_Z
13	E2B_3_win_HZ_H	13	E2B_3_win_HZ_Z	13	E2B_3_win_HZ_H
14	E2B_4_vol_HZ_Z	14	E2B_4_vol_HZ_H	14	E2B_4_vol_HZ_Z
15	E3_1_rea_Z	15	E3_1_rea_H	15	E3_1_irr_Z
16	E3_2_rea_H	16	E3_2_rea_Z	16	E3_2_rea_H
17	E3_3_irr_Z	17	E3_3_irr_H	17	E3_3_rea_Z
18	E3_4_irr_H	18	E3_4_irr_Z	18	E3_4_irr_H
19	E4_1_rea_ninv_Z	19	E4_1_rea_ninv_H	19	E4_1_irr_ninv_Z
20	E4_2_irr_inv_H	20	E4_2_rea_ninv_Z	20	E4_2_rea_ninv_H
21	E4_3_irr_inv_Z	21	E4_3_irr_inv_H	21	E4_3_rea_ninv_Z
22	E4_4_irr_ninv_H	22	E4_4_irr_inv_Z	22	E4_4_irr_inv_H
23	E4_5_irr_ninv_Z	23	E4_5_irr_ninv_H	23	E4_5_irr_inv_Z
24	E4_6_rea_ninv_H	24	E4_6_irr_ninv_Z	24	E4_6_irr_ninv_H
25 - 48	F1 – F24	25 - 48	F1 – F24	25 - 48	F1 – F24
Conditie 4		Conditie 5		Conditie 6	
1	E1_1_A1_H	1	E1_1_A2_Z	1	E1_1_A2_H
2	E1_2_A1_Z	2	E1_2_A1_H	2	E1_2_A2_Z
3	E1_3_A0_H	3	E1_3_A1_Z	3	E1_3_A1_H
4	E1_4_A0_Z	4	E1_4_A0_H	4	E1_4_A1_Z
5	E1_5_A2_H	5	E1_5_A0_Z	5	E1_5_A0_H
6	E1_6_A2_Z	6	E1_6_A2_H	6	E1_6_A0_Z
7	E2A_1_sla_H_H	7	E2A_1_sla_H_Z	7	E2A_1_sla_H_H
8	E2A_2_lez_H_Z	8	E2A_2_lez_H_H	8	E2A_2_lez_H_Z
9	E2A_3_win_HZ_H	9	E2A_3_win_HZ_Z	9	E2A_3_win_HZ_H
10	E2A_4_vol_HZ_Z	10	E2A_4_vol_HZ_H	10	E2A_4_vol_HZ_Z
11	E2B_1_sto_Z_Z	11	E2B_1_sto_Z_H	11	E2B_1_sto_Z_Z
12	E2B_2_tro_Z_H	12	E2B_2_tro_Z_Z	12	E2B_2_tro_Z_H
13	E2B_3_win_HZ_Z	13	E2B_3_win_HZ_H	13	E2B_3_win_HZ_Z
14	E2B_4_vol_HZ_H	14	E2B_4_vol_HZ_Z	14	E2B_4_vol_HZ_H
15	E3_1_irr_H	15	E3_1_rea_Z	15	E3_1_rea_H
16	E3_2_irr_Z	16	E3_2_irr_H	16	E3_2_rea_Z
17	E3_3_rea_H	17	E3_3_irr_Z	17	E3_3_irr_H
18	E3_4_rea_Z	18	E3_4_rea_H	18	E3_4_irr_Z
19	E4_1_irr_ninv_H	19	E4_1_irr_inv_Z	19	E4_1_irr_inv_H
20	E4_2_irr_ninv_Z	20	E4_2_irr_ninv_H	20	E4_2_irr_inv_Z
21	E4_3_rea_ninv_H	21	E4_3_irr_ninv_Z	21	E4_3_irr_ninv_H
22	E4_4_rea_ninv_Z	22	E4_4_rea_ninv_H	22	E4_4_irr_ninv_Z
23	E4_5_irr_inv_H	23	E4_5_rea_ninv_Z	23	E4_5_rea_ninv_H
24	E4_6_irr_inv_Z	24	E4_6_irr_inv_H	24	E4_6_rea_ninv_Z
25 - 48	F1 – F24	25 - 48	F1 – F24	25 - 48	F1 – F24

Appendix C. Demografische vragen

Code	Vraag met antwoordmogelijkheden
D1	Hoe oud ben je?
D2	Ben je... ? ○ Man ○ Vrouw ○ Anders / dat zeg ik liever niet
D3	Wat is je hoogst behaalde diploma? ○ Lager onderwijs ○ Middelbaar onderwijs ○ Hoger/universitair onderwijs
D4	Is het Nederlands je moedertaal? ○ Ja ○ Nee
D5	Indien nee : Welke taal is je moedertaal?
D6	Spreek je een dialect? ○ Ja ○ Nee
D7	Zo ja: Welk dialect spreek je?
D8	Waar ben je geboren? ○ In Vlaanderen ○ In Nederland ○ Elders
D9	In welke provincie heb je tot je achttiende het langst gewoond? ○ Antwerpen (BE) ○ Brussel (BE) ○ Limburg (BE) ○ Oost-Vlaanderen (BE) ○ Vlaams-Brabant (BE) ○ West-Vlaanderen (BE) ○ Drenthe (NL) ○ Flevoland (NL) ○ Friesland (NL) ○ Gelderland (NL) ○ Groningen (NL) ○ Limburg (NL) ○ Noord-Brabant (NL) ○ Noord-Holland (NL) ○ Overijssel (NL) ○ Utrecht (NL) ○ Zeeland (NL) ○ Zuid-Holland (NL) ○ Elders
D10	In welke provincie woon je nu? ○ Antwerpen (BE) ○ Brussel (BE) ○ Limburg (BE) ○ Oost-Vlaanderen (BE) ○ Vlaams-Brabant (BE) ○ West-Vlaanderen (BE)

	○ Drenthe (NL) ○ Flevoland (NL) ○ Friesland (NL) ○ Gelderland (NL) ○ Groningen (NL) ○ Limburg (NL) ○ Noord-Brabant (NL) ○ Noord-Holland (NL) ○ Overijssel (NL) ○ Utrecht (NL) ○ Zeeland (NL) ○ Zuid-Holland (NL) ○ Elders
--	---

Appendix D. Introductie experiment

Inleidende tekst experiment

Beste deelnemer

In het kader van mijn masterproef aan de KU Leuven voer ik een onderzoek uit naar de aanvaardbaarheid van bepaalde zinnen bij moedertaalsprekers van het Nederlands.

In deze enquête krijg je verschillende Nederlandse zinnen te zien. Duid je alsjeblieft voor elke zin aan in welke mate die voor jou aanvaardbaar is? Kan je je voorstellen dat jij of iemand uit je omgeving die zin zou gebruiken? Ga daarbij op je taalgevoel af en antwoord zo eerlijk mogelijk; er zijn geen juiste of foute antwoorden.

Om de enquête in te vullen moet je moedertaalspreker van het Nederlands en minstens 18 jaar oud zijn.

Heel erg bedankt om me verder te helpen met mijn onderzoek! Als je vragen hebt, kan je me steeds bereiken op fien.croux@student.kuleuven.be.

Fien Croux

Geïnformeerde toestemming

Voor we naar de enquête gaan, moet je nog het volgende weten:

- Je deelname aan dit onderzoek is geheel vrijwillig.
- Je kan op elk moment de enquête beëindigen door het browservenster te sluiten.
- Je antwoorden worden volledig anoniem verwerkt en de vertrouwelijkheid van je gegevens is gewaarborgd.
- De resultaten van deze enquête worden enkel gebruikt voor wetenschappelijk onderzoek.

Appendix E. Corpusontwerp SoNaR-corpus

Onderstaande tabel is gebaseerd op de tabel in Oostdijk et al. (2013:5) en Reynaert et al. (2010:2695).

WR/WS	P/U	E/P/T	Text type
Written to be read (WR)	Published (P)	Electronic (E)	Discussion lists
			E-books
			E-magazines
			E-mail (spam)
			Newsletters
			Press releases
			Subtitles
			Teletext pages
			Tweets
			Websites
			Wikipedia
		Printed (P)	Abstracts, summaries
			Books
			Brochures
			Newsletters
			Guides, manuals
			Legal texts
			Newspapers
			Periodicals, magazines
	Unpublished (U)	Electronic (E)	Policy documents
			Proceedings
			Reports
			Surveys
			Theses
			Poems
	Printed (P)	Chats	
		E-mail (non-spam)	
		Minutes	
	Typed (T)	SMS	
		Written assignments	
Written to be spoken (WS)	Unpublished (U)	Electronic (E)	Blogs
		Printed (P)	Theses
		Typed (T)	Minutes
			Written assignments
		Electronic (E)	Autocues
		Typed (T)	New scripts
			Texts for the visually impaired

Appendix F. Hoofdwerkwoorden corpusonderzoek

HWW	abs	rel	HWW	abs	rel	HWW	abs	rel
halen	32	9.2	afnemen	1	0.3	paintballen	1	0.3
kijken	23	6.6	afschaffen	1	0.3	panikeren	1	0.3
zien	17	4.9	bepalen	1	0.3	plukken	1	0.3
doen	10	2.9	bestellen	1	0.3	praten	1	0.3
bekijken	8	2.3	betalen	1	0.3	rechtstaan	1	0.3
eten	8	2.3	bevestigen	1	0.3	regenen	1	0.3
staan	8	2.3	bezien	1	0.3	repeteren	1	0.3
vragen	8	2.3	bezoeken	1	0.3	roepen	1	0.3
zoeken	7	2.0	bonken	1	0.3	rondvertellen	1	0.3
kopen	6	1.7	branden	1	0.3	rondwandelen	1	0.3
spelen	5	1.4	combineren	1	0.3	rusten	1	0.3
brengen	4	1.2	degusteren	1	0.3	samenwerken	1	0.3
geven	4	1.2	doorknippen	1	0.3	scouten	1	0.3
krijgen	4	1.2	doorspelen	1	0.3	solliciteren	1	0.3
opzoeken	4	1.2	doortrekken	1	0.3	sporten	1	0.3
volhouden	4	1.2	drijven	1	0.3	steken	1	0.3
zitten	4	1.2	duren	1	0.3	stelen	1	0.3
drinken	3	0.9	feliciteren	1	0.3	stilstaan	1	0.3
helpen	3	0.9	filmen	1	0.3	stretchen	1	0.3
maken	3	0.9	financieren	1	0.3	studeren	1	0.3
slapen	3	0.9	geloven	1	0.3	tatoeëren	1	0.3
wandelen	3	0.9	gieten	1	0.3	tegensteken	1	0.3
beweren	2	0.6	gooien	1	0.3	tekenen	1	0.3
boeien	2	0.6	hebben	1	0.3	toeren	1	0.3
fietsen	2	0.6	horen	1	0.3	toeteren	1	0.3
hangen	2	0.6	ijveren	1	0.3	tonen	1	0.3
houden	2	0.6	indienen	1	0.3	twijfelen	1	0.3
huren	2	0.6	inschrijven	1	0.3	uitbouwen	1	0.3
installeren	2	0.6	insteken	1	0.3	uitmaken	1	0.3
jagen	2	0.6	investeren	1	0.3	uitmelken	1	0.3
kuisen	2	0.6	klagen	1	0.3	uitstorten	1	0.3
liggen	2	0.6	kneden	1	0.3	uittrekken	1	0.3
lopen	2	0.6	knokken	1	0.3	uitzoeken	1	0.3
nemen	2	0.6	koersen	1	0.3	vergaderen	1	0.3
noemen	2	0.6	kotsen	1	0.3	verneuken	1	0.3
ophalen	2	0.6	laden	1	0.3	versoepelen	1	0.3
pikken	2	0.6	leven	1	0.3	verwittigen	1	0.3
poolen	2	0.6	lezen	1	0.3	vloeren	1	0.3
stemmen	2	0.6	luisteren	1	0.3	voegen	1	0.3
steunen	2	0.6	luistervinken	1	0.3	volkladden	1	0.3
uitoefenen	2	0.6	meefeesten	1	0.3	volleyballen	1	0.3
verdienen	2	0.6	melden	1	0.3	voorleggen	1	0.3
verkennen	2	0.6	nadenken	1	0.3	voorstellen	1	0.3
vertellen	2	0.6	nazinderen	1	0.3	voortdoen	1	0.3
volgen	2	0.6	neuzen	1	0.3	waarderen	1	0.3
wegdoen	2	0.6	onderhouden	1	0.3	weghalen	1	0.3
werken	2	0.6	ondervragen	1	0.3	weigeren	1	0.3
zeggen	2	0.6	ontkennen	1	0.3	winnen	1	0.3
zingen	2	0.6	opbouwen	1	0.3	wonen	1	0.3
aanbieden	1	0.3	opdissen	1	0.3	zagen	1	0.3
aanplanten	1	0.3	opdoen	1	0.3	zich ophopen	1	0.3
afdreigen	1	0.3	optekenen	1	0.3	zich vergapen	1	0.3
afgeven	1	0.3	optreden	1	0.3	zich verontschuldigen	1	0.3
afhalen	1	0.3	opzetten	1	0.3			

