

De *om*-alternantie: een verkennende casestudy

Amélie VAN BEVEREN, Timothy COLLEMAN & Gert DE SUTTER¹

Abstract

This paper focuses on the Dutch *om*-alternation, i.e. the variation between infinitival complements (=IC) with and without the prepositional complementizer *om*, as demonstrated in *Hij beloofde om op tijd te komen* vs. *Hij beloofde op tijd te komen* ('He promised to be on time'). The choice between one of the two alternating constructions is not random as it depends on different syntactic, semantic and pragmatic factors. We test the effects of a series of potential factors culled from the literature through a multifactorial corpus investigation on the basis of real-language examples from the Dutch Parallel Corpus, with special attention for register effects – the possible influence of register on this alternation has not been addressed before. This attention for register can be related to the emerging Cognitive Sociolinguistics movement, which has led to a new or renewed attention for the possible influence of all kinds of lectal factors on syntactic alternation phenomena. The analysis will reveal that both register and complexity-related factors play a significant role in the choice between the two alternants.



1. INLEIDING

Grammaticale alternanties zijn een vaak bestudeerd onderwerp binnen verschillende theoretische benaderingen van de grammatica omdat ze ons toelaten om de relatie tussen syntaxis, semantiek en pragmatiek te onderzoeken. In de constructiegrammatica worden de twee (of meer) leden van een grammaticale alternantie gezien als aparte constructies – d.w.z. afzonderlijk opgeslagen vorm/betekenis-paren – die niet alleen formeel verschillen, maar ook op semantisch en/of discourse-pragmatisch vlak (zie bv. Goldbergs (1995: 67) *Principle of No Syntactic Synonymy*). Een belangrijke onderzoekstrend in recent taalkundig onderzoek is het toenemende gebruik van multifactoriële technieken voor de statistische evaluatie van de effecten van verschillende factoren op alternanties, aangezien vaak aangenomen wordt dat de keuze tussen de twee (of meer) alternerende constructies niet vrij is maar afhankelijk van

¹ De twee eerste auteurs zijn verbonden aan de vakgroep Taalkunde van de Universiteit Gent, onderzoeksgroep GLIMS. De eerste auteur is ook verbonden aan de vakgroep Vertalen, Tolken en Communicatie, onderzoeksgroep EQTIS, net zoals de derde auteur. Het onderzoek van de eerste auteur is mogelijk gemaakt door een BOF-onderzoeksbeurs voor het onderzoeksproject 'Grammatical alternations, register variation and explicitation in original and translated Dutch', dat wordt uitgevoerd onder begeleiding van de tweede en derde auteur. Met dank aan twee anonieme reviewers voor hun opmerkingen en suggesties bij eerdere versies van dit artikel.

verschillende determinerende factoren (zie bv. Gries 2003, Grondelaers & Speelman 2007, Bresnan & Nikitina 2009, Szmrecsanyi 2010, Theijssen et al. 2013). Dergelijke technieken bieden de onderzoeker het voordeel dat de relatie sterkte van elk onderzocht effect kan worden gemeten, rekening houdend met andere variabelen. Een tweede trend die hieraan gekoppeld kan worden is de toegenomen aandacht voor lectale (d.w.z. regionale, sociale, stilistische, enz.) variatie, voornamelijk dankzij de invloed van de opkomende Cognitieve Sociolinguïstiek (Kristiansen & Dirven 2008, Geeraerts et al. 2010, Pütz et al. 2014), die beklemtoont dat lexicon en grammatica onderhevig zijn aan sociale, culturele en stilistische variatie. Vanuit dat perspectief wordt er ook binnen de constructiegrammatica steeds meer aandacht besteed aan taalexterne factoren zoals het register en het medium van de tekst (zie bv. Kerz & Wiechmann 2015, Yoon & Gries 2016). Ook Goldberg (1995: 67) vermeldt expliciet dat formele verschillen tussen al dan niet alternerende constructies in verband kunnen gebracht worden met register als een stilistisch aspect van de betekenis. Ieder register wordt namelijk gekenmerkt door specifieke talige en situationele eigenschappen en de functionele relatie tussen die twee (Biber & Conrad 2009: 6). Hiermee wordt bedoeld dat talige karakteristieken voorkomen in een bepaald register omdat ze passen in de situationele context van het register in kwestie. Anders verwoord, als blijkt dat bepaalde constructies een voorkeur vertonen voor een bepaald register, zoals krantentaal of fictie, kunnen de stilistische eigenschappen van die registers doorgetrokken worden naar de betekenis van de constructies zelf en kan zo eventueel ook voor grammaticale constructies een stilistische betekenislaag worden onderscheiden.² Zo suggereert het verkennende onderzoek van Kirsner (1988) over registereffecten in de Nederlandse datiefalternantie dat kranten, een register dat intuïtief als formeel en objectief kan worden beschouwd, relatief vaker de voorkeur geven aan de constructie met *aan* (*Jan geeft een boek aan Peter*) boven de dubbelobjectconstructie (*Jan geeft Peter een boek*) dan andere genres, zoals fictie (novelles). Kirsner brengt die tendens in verband met de grotere grammaticale explicietheid van de *aan*-constructie, die goed past bij het (nagestreefde) objectieve karakter van krantentaal. Met grammaticale explicietheid wordt dan de aanwezigheid bedoeld van een grammaticaal element dat de structuur van de zin verduidelijkt maar dat strikt genomen niet noodzakelijk is (Rodhenburg 1996). Vertaald in constructiegrammaticale termen zouden we Kirsners hypothese zo

² Dergelijke stilistische variatie is ook al aangetoond voor specifieke lexemen (zie bv. Leech 1974 die zeven verschillende types van betekenis onderscheidt waarvan stilistische betekenis er een is. Hiermee wordt bedoeld dat twee of meerdere woorden dezelfde denotatie kunnen hebben en toch kunnen verschillen in situationeel gebruik. Een eenvoudig voorbeeld is het verschil in gebruik tussen de Engelse werkwoorden *cast* en *throw*: terwijl *throw* in algemene contexten wordt gebruikt, komt *cast* meer in literaire contexten voor). Aangezien constructiegrammatica geen onderscheid maakt tussen lexemen en constructies, verwachten we dezelfde soort variatie ook bij constructies.

kunnen voorstellen dat van het semantische profiel van de *aan*-constructie ook een stilistische betekenisdimensie deel uitmaakt, namelijk een zekere associatie van de constructie met formeel en/of "objectief" taalgebruik. Indien die hypothese overeind blijft bij multifactoriële toetsing, kan ook bekeken worden of er vergelijkbare effecten spelen in andere alternanties waarin een explicietere en implicietere variant kunnen worden onderscheiden. Door de uitbreiding van dezelfde onderzoeksvragen naar andere alternanties kan dan getest worden of expliciete constructies in het algemeen geassocieerd kunnen worden met meer formele en/of objectieve registers.

In dit artikel wordt een multifactoriële analyse gepresenteerd van één zo'n Nederlandse alternantie met een verschil in grammaticale explicietheid tussen de twee leden, namelijk de *om*-alternantie, waarmee verwezen wordt naar de variatie tussen beknopte infinitiefzinnen met en zonder de complementeerder *om* zoals in: *Hij belooft om op tijd te komen* (in wat volgt: de expliciete constructie) vs. *Hij belooft op tijd te komen* (in wat volgt: de impliciete constructie). Een meer gedetailleerde omschrijving van de alternantie wordt gegeven in paragraaf 2.2. Deze casestudy over de *om*-alternantie is de eerste binnen een groter project dat verschillende zulke alternanties met een impliciete en een expliciete variant zal bestuderen. In dat grotere project wordt ook het verschil tussen origineel en vertaald Nederlands betrokken om de rol van expliciteatie in vertalingen te onderzoeken. Op die dimensie van het onderzoek gaan we hier echter niet verder in: het huidige artikel focust alleen op niet-vertaald Nederlands.

Aan de hand van een binaire logistische regressie met de afhankelijke variabele de aan- of afwezigheid van *om* en een reeks van zowel categorische als numerieke onafhankelijke variabelen worden de mogelijke beïnvloedende factoren gemodelleerd. De factoren werden verzameld uit bestaande literatuur over de *om*-alternantie (Blom 1984, ANS 1997, Bouma 2017, Jansen 1987, Pardoën 1998, Van Haaften 1991, Vliegen 2001) en aan de hand van een exploratieve blik op de dataset. Het doel van deze verkennende casestudy is een beter beeld te krijgen van de factoren die een invloed uitoefenen op het al dan niet aanwezig zijn van *om* aan het begin van de beknopte infinitiefzin, waarbij vooral gefocust zal worden op het mogelijke effect van register. Dit artikel is dan ook gedeeltelijk te beschouwen als een replicatie-onderzoek van Bouma (2017), een multifactorieel corpusonderzoek naar de *om*-alternantie in kran-tentaal. Het belangrijkste verschil met Bouma (2017) is dat onze data uit een stilistisch gedifferentieerd corpus komen, zodat we register kunnen toevoegen als een mogelijke predictorvariabele. In een later stadium van het onderzoek zal verder onderzocht worden hoe de invloed van specifieke registerkarakteristieken deel kan uitmaken van de betekenis van alternerende constructies, als een stilistische betekenislaag.

In paragraaf 2 wordt een kort overzicht gegeven van de verschillende beknopte infinitiefzinnen die voorkomen in het Nederlands, waarbij specifiek wordt ingezoomd op de *om*-alternantie. In paragraaf 3 komt de methodologie van de casestudy aan bod. Hierbij wordt aandacht besteed aan het gebruikte corpus, de zoekmethode en de statistische technieken die gebruikt zullen worden. Paragraaf 4 presenteert en interpreteert de resultaten van het onderzoek, waarbij bekeken wordt welke factoren effectief een invloed uitoefenen op de keuze tussen expliciet en impliciet *om*. Tot slot volgt er in paragraaf 5 een conclusie en een vooruitblik op het verdere verloop van het doctoraatsproject.

2. THEORETISCHE ACHTERGROND

In deze paragraaf gaan we eerst kort in op de verschillende soorten beknopte infinitiefzinnen in het hedendaagse Nederlands, waarvan de beknopte infinitiefzin met facultatief *om* er een is (zie paragraaf 2.1). Daarna volgt een bespreking van het bestaande onderzoek naar de ‘*om*-alternantie’, met extra aandacht voor de factoren die volgens die literatuur de aan- of afwezigheid van *om* mogelijk beïnvloeden.

2.1. Beknopte infinitiefzinnen: een kort overzicht

In de grammatica van het Nederlands worden drie verschillende hoofdtypes van beknopte infinitiefzinnen onderscheiden afhankelijk van het bindwoord, nl. met een logisch (1), vragend (2) of grammaticaal bindwoord (3) (Vandeweghe 2004: 329; 332-335; ANS 1997: 548-549; 557-559; 856-857; 899-900; 1110-1112; SoD 2015: 610-623).

- (1) Hij beslist zonder na te denken over de consequenties.
- (2) Ik vraag me af wat te doen op zondagavond.
- (3) Hij belooft om thuis te blijven dit weekend.

Logische bindwoorden drukken een semantisch verband uit tussen de matrixzin en de beknopte bijzin, zoals finaliteit (bv. *om*), omstandigheid (bv. *zonder*), oorzaak (bv. *door*), enzovoort. Vragende bindwoorden kunnen bestaan uit een zelfstandig vragend voornaamwoord (of een naamwoordelijke constituent met een bijvoeglijk vragend voornaamwoord als deel), een vragend bijwoord of voornaamwoordelijk bijwoord, maar ook uit een voorzetselconstituent met een vraagwoord als deel en bevragen een bepaald aspect van de zin. Ten slotte is er het grammaticale bindwoord *om*. In tegenstelling tot bij de twee voorgaande types is de betekenisbijdrage daarvan minimaal: *om* dient enkel als verbindend element tussen matrixzin en bijzin. *Om* kan dus zowel een logisch bindwoord zijn als een grammaticaal. In dat laatste geval is het weg-

laatbaar, waardoor er sprake is van de *om*-alternantie, die in paragraaf 2.2 hieronder nader aan bod komt. Naast beknopte infinitiefzinnen die verplicht ingeleid worden door *om* en zinnen waarbij *om* facultatief is, zijn er ook infinitiefzinnen waarbij toevoeging van *om* niet mogelijk is, namelijk als de beknopte bijzin lijdend voorwerp is bij werkwoorden met de betekenis '(een vorm van) zeggen' of '(een vorm van) bemerken' zoals in *De minister verklaarde de motie naast zich neer te zullen leggen* en *Hij voelde een fout begaan te hebben* (ANS 1997: 1111). Beknopte infinitiefzinnen waarin *om* als logisch bindwoord functioneert en dus verplicht is of beknopte infinitiefzinnen waarin *om* onmogelijk is, laten we hier buiten beschouwing.

2.2. Om-alternantie: beknopte infinitiefzinnen met facultatief *om*

2.2.1. Algemeen

We onderscheiden drie verschillende types van de constructie met facultatief *om*, afhankelijk van het hoofd van de beknopte infinitiefzin (=BIZ). Het hoofd is het element in de matrixzin waarvan de beknopte infinitiefzin syntactisch afhangt en dat kan een werkwoord (4), een substantief (5) of een adjectief (6) zijn.

- (4) Ik beloof (*om*) op tijd te komen.
- (5) Ik heb een hekel aan zijn neiging (*om*) altijd alles uit te stellen.
- (6) Ik ben blij (*om*) je te zien.

In bovenstaande voorbeelden van de *om*-alternantie staat het hoofd van de beknopte infinitiefzin steeds onderstreept. Bouma (2017) onderscheidt nog een vierde categorie, namelijk beknopte infinitiefzinnen die afhangen van een voorzetsel. In bijna alle gevallen gaat het daarbij om voorzetsels die worden geselecteerd door een werkwoord, zoals *houden van*, *aanzetten tot* en *slagen in*. In deze paper opteren we er dan ook voor om zulke voorbeelden te categoriseren onder de beknopte infinitiefzinnen die afhangen van een werkwoord, d.w.z. waarbij de infinitiefzin fungeert als complement van dat werkwoord (direct object dan wel voorzetselobject). Enkele uitzonderingen zijn de hoofden *bevoegd tot* en *gehouden tot* die dan als adjectieven worden gecategoriseerd. Van Haaften (1991) geeft in zijn dissertatie over de interpretatie van verzwegen subjecten een uitgebreide opsomming van de verschillende substantieven, adjectieven en werkwoorden die als hoofd van een BIZ kunnen fungeren. De lijst is niet exhaustief, maar toont duidelijk aan dat de lexicale mogelijkheden erg groot zijn. Ook bij Van Haaften (1991) wordt er overigens geen bijkomend onderscheid gemaakt tussen werkwoorden en voorzetsels als hoofd van de BIZ.

Als *om* aanwezig is, heeft het een grensmarkerende functie. In enkele zeldzame gevallen, zoals in (7) hieronder, kan de aanwezigheid van *om* de structuur van de zin desambigueren.³

- (7) Maarten belooft Nora te bellen.
- (8) Maarten belooft om Nora te bellen.
- (9) Maarten belooft Nora om te bellen.

In zin (7) is het namelijk niet duidelijk of Maarten aan Nora belooft dat hij iemand zal bellen of dat hij aan iemand belooft dat hij Nora zal bellen. De plaatsing van *om* verheldert in dat geval de betekenis van de zin. In de meeste gevallen is dat echter niet nodig, aangezien er geen ambiguïteit optreedt.

Er bestaan ook semantische restricties voor de werkwoorden, adjectieven of substantieven die als hoofd van de BIZ fungeren en de inhoud van de BIZ zelf. De propositionele inhoud van de BIZ geldt steeds als nog te verwezenlijken maar nog niet verwezenlijkt op het moment van de propositie (Vliegen 2001: 118). Facultatief *om* verschijnt dus alleen bij BIZ's met een irrealismodaliteit waarin een gebeurtenis beschreven wordt die nog moet plaatsvinden zoals bij *beloven* en *besluiten*. Volgens Bouma (2017) staan ook factieve werkwoorden – d.w.z. werkwoorden die een feit in het hier en nu beschrijven zoals *presteren* en *ophouden* – optioneel *om* toe.

In Vlaanderen wordt in het informele gesproken taalgebruik *om* vaak vervangen door *van* (Van Craenenbroeck 2000: 134), een verschijnsel dat in Nederland volledig onbekend is. Complementerend *van* komt in meerdere syntactische contexten voor en fungeert dus niet uitsluitend als vervanger van *om* als inleidend grammaticaal bindwoord van de beknopte infinitiefzin. Zo komt *van* ook voor bij werkwoorden als *menen* en *zeggen*, waarbij *om* onmogelijk is in de standaardtaal. In deze paper laten we de niet-standaardtalige vorm *van* buiten beschouwing.

2.2.2. Factoren die de keuze tussen impliciet en expliciet *om* beïnvloeden

Zoals bij de meeste alternanties is de keuze tussen expliciet en impliciet *om* niet geheel vrij, maar afhankelijk van verschillende taalinterne en -externe factoren. Er zijn verschillende bronnen waarin hypothesen worden geformuleerd – en in sommige gevallen ook getoetst – over de precieze factoren die de keuze zouden bepalen (Blom 1984, 1990, ANS 1997, Jansen 1987, Van Haften

³ In gesproken taal kan bij onduidelijkheid ook een beroep gedaan worden op het zinsaccent: accentueert de spreker 'Nora' dan wordt de eerste betekenis geselecteerd: Maarten belooft dat hij Nora zal bellen; ligt de nadruk op 'bellen' dan is de tweede betekenis aan de orde: Maarten belooft Nora dat hij iemand zal bellen. Het toevoegen van *om* doet hier dan dienst als extra steun (ANS 1997: 1111).

1991, Pardoën 1998, Vliegen 2001, Bouma 2017). Hieronder geven we een overzicht van de in de literatuur voorgestelde factoren en welke ervan onderzocht worden in dit artikel.

2.2.2.1. Taalinterne factoren

Binnen de taalinterne factoren maken we een onderscheid tussen syntactische (2.2.2.1.1) en semantische factoren (2.2.2.1.2). De meeste semantische factoren komen uit het onderzoek van Vliegen (2001). De syntactische factoren werden voor het grootste deel ontleend aan het onderzoek van Bouma (2017); daarnaast zijn er een paar factoren die we zelf hebben toegevoegd, op basis van analogie met factoren die in de literatuur zijn voorgesteld.

2.2.2.1.1. Syntactische factoren

Het merendeel van de factoren die in de literatuur worden genoemd als potentieel relevant voor het al dan niet aanwezig zijn van *om* als inleidend voegwoord van de beknopte infinitiefzin is syntactisch van aard. Een eerste factor die in de studie van Bouma (2017) aangehaald wordt, heeft betrekking op het hoofd van de beknopte infinitiefzin. Bouma (2017) is het enige grootschalige corpusonderzoek dat niet alleen rekening houdt met werkwoorden als mogelijk hoofd van de beknopte infinitiefzin en dat duidelijk maakt dat de voorkeur voor expliciet of impliciet *om* verschilt naargelang de woordsoort van het hoofd van de beknopte infinitiefzin. Aan de hand van het Twente Nieuws Corpus en het Corpus Gesproken Nederlands onderzoekt hij met logistische regressie en *mixed effects modeling* de mogelijke effecten van het hoofd van de beknopte infinitiefzin, de semantiek van het matrixwerkwoord en de complexiteit van de zinnen op het al dan niet aanwezig zijn van *om* met het concrete verbale hoofd als *random* effect. Bouma toont ten eerste aan dat de voorkeur voor *om* het grootst is wanneer het hoofd een substantief is. Adjectivale hoofden hebben nog steeds een zekere voorkeur voor de expliciete variant, maar die is minder groot dan bij de substantieven. Werkwoorden daarentegen tonen een voorkeur voor de impliciete variant van de constructie. Bouma (2017) onderscheidt zoals hierboven al opgemerkt nog een vierde hoofdcategory, nl. voorzetsels, maar betreft die niet in zijn analyse over de verschillende hoofden. Hij vermeldt hiervoor geen reden, maar vermoedelijk heeft dat met een te lage frequentie van die categorie te maken. Zo goed als alle hoofden die als prepositioneel kunnen worden beschouwd, zijn vaste voorzetsels bij een werkwoord, bv. *houden van*, *slagen in*, *verplichten tot* enz. en om die reden categoriseren we de betreffende BIZ's als zinnen met een verbaal hoofd, met uitzondering van de BIZ's bij adjectieven met een vast voorzetsel zoals

bevoegd tot en *gehouden tot*, die uiteraard in de adjectivische categorie werden ondergebracht. Om de eventuele invloed van de aanwezigheid van een voorzetsel toch te kunnen testen, hebben we echter een afzonderlijke factor toegevoegd die daar gedeeltelijk bij aansluit, nl. de aan- of afwezigheid van een voornaamwoordelijk bijwoord bij een verbaal of adjectivaal hoofd. De variabele werd toegevoegd omdat het ons na een exploratief onderzoek van de verschillende mogelijk beïnvloedende factoren opviel dat de aanwezigheid van dergelijke bijwoorden gepaard ging met de aanwezigheid van *om* aan het begin van de BIZ. Een voorbeeld hiervan volgt in zin (10) hieronder.

- (10) Blijkbaar zette hij Van Gogh ertoe aan om zijn aandacht te richten op de historische plekken van Arles. (DPC)

Een andere groep van syntactische factoren heeft specifiek te maken met het geïmpliceerde subject van de BIZ. Syntactische factoren die we daaraan kunnen koppelen zijn de grammaticale persoon en de syntactische functie in de matrixzin van het geïmpliceerd subject van de BIZ. Vanuit zijn algemene hypothese over sprekerssubjectiviteit – waarover meer in de volgende paragraaf – stelt Vliegen (2001) dat een geïmpliceerd subject in de eerste persoon het gebruik van *om* bevordert. De factor controlerend zinsdeel (functie) werd toegevoegd naar analogie met de factor semantische rol van het geïmpliceerde subject. Het geïmpliceerde subject van de BIZ kan namelijk verschillende syntactische functies bekleden in de matrixzin. In deze studie hebben we geopteerd voor een binaire variabele: het geïmpliceerde subject van de BIZ verschijnt als subject in de matrixzin zoals in (11) of niet. In de tweede categorie zitten verschillende mogelijkheden vervat het geïmpliceerde subject kan namelijk ook verschijnen als direct object (12), indirect object (13) of het heeft geen equivalent in de matrixzin (14):

- (11) De laatste keer dat ik hem zag, stond hij erop me een plek te laten zien die hij op een van zijn wandelingen had ontdekt. (DPC)
- (12) Het onderzoek van deze reus zal ons in staat stellen een paar van de mysteries van de diepzee te ontrafelen. (DPC)
- (13) Wij raden u aan de filters elk jaar te vervangen. (DPC)
- (14) De verwerkingstechnologie, zoals van de halfgeleidertechnologie bekend is, laat toe om heel kleine sensoren te maken. (DPC)

De voorlaatste groep van syntactische factoren die wordt meegenomen in de analyse heeft te maken met complexiteit. Volgens het complexiteitsprincipe wordt aangenomen dat als de taalgebruiker de keuze krijgt tussen altemnerende constructies met een impliciete of expliciete variant, de meer expliciete optie de voorkeur krijgt in cognitief meer complexe omgevingen (Rohdenburg 1996 Hawkins 2004, Mondorf 2014). In het geval van de *om*-alternantie wordt dat

als volgt vertaald: hoe complexer de zin is, hoe groter de kans op de aanwezigheid van *om*, aangezien *om* de verwerking van de zin kan vergemakkelijken (Bouma 2017: 69). Hetzelfde fenomeen doet zich voor bij de Engelse *that*-alternantie, een alternantie die in zekere zin lijkt op de Nederlandse *om*-alternantie omdat er in beide gevallen een complementzin aanwezig is die optioneel ingeleid wordt door een functiewoord, respectievelijk *that* en *om*. De aanwezigheid van *that* in zinnen als *I believe (that) he was a genius of his time* maakt de complementatieconstructie wat doorzichtiger en dus makkelijker te verwerken (zie bv. Wasow et al. 2011). Complexiteit vertalen we in deze studie in eerste instantie naar twee numerieke factoren en één categoriale factor. Daarnaast zijn er nog een paar factoren die te maken hebben met de vorm van het werkwoord van de matrixzin en die eveneens met complexiteit in verband kunnen worden gebracht.

In de eerste plaats bekijken we de afstand tussen het hoofd van de beknopte infinitiefzin en het effectieve begin ervan, uitgedrukt in het aantal tussenliggende woorden. Ten tweede tellen we, net als Bouma (2017), het aantal woorden van de BIZ zelf (zonder *om* zelf mee te tellen).

Het voorkomen van *meerdere* beknopte infinitiefzinnen in één zin kan eveneens bijdragen tot de complexiteit van de zin (Bouma 2017: 69). We operationaliseren dat als een binaire variabele, nl. de BIZ maakt wel of niet deel uit van een globale zin die meer dan één BIZ bevat. Voorlopig maken we daarbij geen nader onderscheid tussen voorbeelden met meer dan één infinitiefzin in nevenschikkend verband tot elkaar (15) en voorbeelden met meer dan één infinitiefzin die in onderschikkend verband staan ten opzichte van elkaar (16). Noch maken we een onderscheid tussen verschillende niveaus van inbedding.

- (15) Ze glimlachte, want we waren net in een gang gekomen die zo smal was dat we gedwongen waren onze rokken bijeen te nemen en achter elkaar te lopen. (DPC)
- (16) Uw taak is die kloof bloot te leggen en te tonen hoe uw product of dienst hem kan helpen om ze te overbruggen. (DPC)

Ten slotte zijn er drie potentieel relevante syntactische factoren die specifiek te maken hebben met het werkwoord van de matrixzin, en die ook gelinkt kunnen worden aan de hierboven vermelde complexiteit-gerelateerde factoren. In eerste instantie wordt er voor de subcategorie van de BIZ's die afhangen van een werkwoord een onderscheid gemaakt tussen finiete en infiniete werkwoorden. Een matrixzin met een infiniet werkwoord is in zekere zin complexer dan een finiet werkwoord, omdat het doorgaans een hulpwerkwoord bij zich krijgt en vaak voorkomt in een tangconstructie waarbij er mogelijk meerdere woorden tussen de beide delen kunnen verschijnen. Ook Bouma (2017: 66) maakt een

onderscheid tussen finiete werkwoorden en infinitieve werkwoorden, maar poneert een andere hypothese dan bovenstaande. Volgens hem is de verwerking van de beknopte infinitiefzin net gemakkelijker als het werkwoord van de matrixzin infinitief is en dus meer naar achteren in de matrixzin staat. De grotere afstand tussen de beknopte infinitiefzin en een finiet werkwoord als matrixwerkwoord zorgt volgens hem dus voor een grotere complexiteit. In elk geval zijn er dus redenen om de factor finietheid van het matrixwerkwoord mee te nemen: aan de hand van de richting van het eventuele effect ervan kunnen we nagaan welke hypothese de meeste plausibele is voor de *om*-alternantie.

Een onderscheid dat hier bij aansluit, is dat tussen actieve en passieve werkwoorden. Op basis van de annotatie van een kleiner sample als exploratief vooronderzoek⁴ werd duidelijk dat passieve werkwoorden meestal met *om* voorkwamen. Dat is ook in lijn met de visie van Rohdenburg (1996) die passief als een complexere context beschouwd dan actief en cognitief complexere omgevingen worden geassocieerd met een frequenter voorkomen van expliciete constructies.

Een meer specifieke factor die opnieuw gelinkt kan worden aan de vorm van het werkwoord en complexiteit is de aanwezigheid van een gescheiden scheidbaar samengesteld werkwoord. Wanneer een scheidbaar samengesteld werkwoord gescheiden voorkomt in de zin, vermoeden we dat die zin complexer is dan wanneer het werkwoord niet gescheiden is, aangezien het verbaal hoofd van de zin gesplitst voorkomt en daartussen verschillende andere elementen aanwezig kunnen zijn. Twee voorbeelden:

- (17) Als ik het bij een coaching nodig vind, stel ik aan de persoon voor een vragenlijst in te vullen die wordt geleverd en verwerkt door Kahler Communication. (DPC)
- (18) "Een aantal medewerkers bood spontaan aan om collega's te coachen", herinnert Daniel Kurgan zich. (DPC)

Ten slotte voegden we ook nog een factor toe van een iets andere aard, namelijk de aanwezigheid van een andere *om* in de zin, zowel voor de beknopte infinitiefzin als erin. Onder een andere *om* wordt dan zowel een *om* verstaan die een andere BIZ in dezelfde zin inleidt of een *om* met een andere functie, zoals een afscheidbaar deel van een werkwoord of als logisch bindwoord met doelbetekenis in bijvoorbeeld *De regering vraagt aan deze Conferentie om ook reeds na te denken over nieuwe werven die we moeten openen om in de toekomst meer tewerkstelling te scheppen* (DPC). De aanwezigheid van die

⁴ Voor het exploratieve onderzoek werden alle relevante zinnen random gesorteerd en op basis van de eerste 500 resultaten werden de in de literatuur voorgestelde factoren getest. Bij het annoteren vielen enkele zaken op zoals het verschil tussen actieve en passieve werkwoorden en de aanwezigheid van voornaamwoordelijke bijwoorden als mogelijk relevante factoren.

andere *om* in de voorafgaande context is mogelijk een trigger om opnieuw een *om* te realiseren in de BIZ die volgt, als een soort primingeffect.

2.2.2.1.2. *Semantische factoren*

In de literatuur omtrent de *om*-alternantie zijn er ook een aantal hypotheses geformuleerd over *semantische* effecten op de alternantie. In tegenstelling tot de net besproken syntactische factoren laten de meeste van die geponeerde semantische factoren zich moeilijk operationaliseren in een brede en diverse dataset. Om die reden hebben we er dan ook voor gekozen om ons in dit verkennende onderzoek voorlopig voornamelijk te focussen op syntactische en taalexterne factoren (op die laatst genoemde factoren wordt ingegaan in 2.2.2.2), in afwachting van verdere reflectie op de optimale operationalisering van een aantal andere potentieel relevante semantische factoren.

We nemen wel al een aantal semantische factoren mee die te maken hebben met de agentiviteit van het geïmpliceerde subject in de beknopte infinitiefzin, voortbouwend op een suggestie van Vliegen (2001), een onderzoek over facultatief *om* bij illocutionaire werkwoorden, werkwoorden met een 'communicatief betekenismoment' (Verhagen 1996: 256) dat hoofdzakelijk aandacht besteedt aan semantische factoren. Vliemens centrale hypothese is dat het gebruik van *om* in door *te* ingeleide infinitiefcomplementen sprekerssubjectiviteit aanduidt, voortbouwend op ideeën van Blom (1984, 1990) en Pardoën (1998). Aangenomen wordt dat de aanwezigheid van *om* duidt op de betrokkenheid van de spreker: door *om* tussen de matrixzin en de beknopte infinitiefzin te plaatsen, wil hij zijn verwachting ten opzichte van de realisering van de stand van zaken beschreven in de BIZ expliciet maken. Een positieve verwachting ten opzichte van de realisering van de in de BIZ beschreven stand van zaken bevordert volgens Vliegen (2001: 129) het gebruik van *om*, zonder dat daarbij de (on)wenselijkheid van de realisering ervan vanuit diezelfde spreker aangegeven wordt waarover meer in het vervolg van deze paragraaf.

Een van de centrale factoren in het onderzoek van Vliegen is zoals hierboven al vermeld de agentiviteit van het geïmpliceerde subject in de beknopte infinitiefzin. Het verzwegen subject van de BIZ correspondeert namelijk met een syntactisch al dan niet gerealiseerd argument van het matrixwerkwoord (Vliegen 2001: 114). Vliegen (2001: 125) stelt dat de semantische rol en meer bepaald de agentiviteit van de referent van het geïmpliceerde subject in de matrixzin invloed kan hebben op de keuze tussen een van de varianten, omdat hogere agentiviteit kan wijzen op een grotere mate van sprekerssubjectiviteit. Wanneer het geïmpliceerde subject agens is (en voornamelijk als het over de eerste persoon gaat) bevordert dat het gebruik van *om*. Vliegen haalt alleen de semantische rol van het geïmpliceerd subject van de BIZ in de matrixzin aan,

maar het geïmpliceerde subject van de BIZ kan natuurlijk ook in de BIZ zelf zowel agens als niet-agens zijn. Zie het verschil tussen zin (19) en (20): in zin (19) is het geïmpliceerde subject agens in de beknopte infinitiefzin en niet-agens in de matrixzin, in zin (20) is dat andersom.

(19) Hij beval mij de afwas te doen.

(20) Ik eis op de hoogte te worden gehouden.

Daarom hebben we de zinnen niet enkel gecodeerd op [+/- agentiviteit] van het geïmpliceerde subject in de matrixzin, maar ook op [+/- agentiviteit] in de BIZ zelf.

Naast agentiviteit zijn er nog een aantal andere potentieel relevante semantische factoren die verband houden met Vliemens algemene hypothese in termen van sprekerssubjectiviteit of betrokkenheid, maar die we voorlopig niet meenemen in het onderzoek. Zoals hierboven al even aangehaald wordt het gebruik van *om* in verschillende onderzoeken geassocieerd met een zekere distantiëring van de spreker ten opzichte van de infinitiefzin. Zowel Blom (1984: 175) als Pardoën (1998: 428) koppelen dit resoluut aan een negatieve houding ten opzichte van de realisering van de inhoud van de beknopte infinitiefzin, terwijl Vliemen (2001: 115) terecht opmerkt dat distantiëring niet altijd negatief is en dat de aanwezigheid van *om* niet per se distantiëring impliceert. Vliemen (2001) poneert zelfs de omgekeerde hypothese: een positieve houding ten opzichte van de realisering van de in het infinitiefcomplement uitgedrukte stand van zaken bevordert het gebruik van *om*. Een positieve houding komt volgens hem voor bij werkwoorden als *beloven* en *toezeggen*: in beide gevallen gaat de spreker ervan uit dat de stand van zaken beschreven in de beknopte infinitiefzin effectief gerealiseerd zal worden. Een negatieve houding ten opzichte van het uitvoeren van de inhoud van de beknopte infinitiefzin is dan weer terug te vinden bij onder andere *weigeren*, aangezien de stand van zaken beschreven in de bijzin niet zal plaatsvinden. Bij *beloven* en *toezeggen* is er volgens Vliemen dan ook een relatief grotere kans op *om* dan bij *weigeren*. Het is echter niet meteen duidelijk hoe de veronderstelde factor van een positieve dan wel negatieve houding van de spreker op een betrouwbare manier geoperationaliseerd kan worden. Het onderzoek van Vliemen (2001) is gebaseerd op slechts 11 geselecteerde illocutionaire werkwoorden, die zich relatief makkelijk op die manier laten klasseren. Bij veel andere werkwoorden, is er echter nauwelijks sprake van een positieve of negatieve houding van de spreker ten opzichte van de stand van zaken in de beknopte infinitiefzin.

Een andere semantische factor die voorlopig niet wordt meegenomen in de multifactoriële analyse van deze casestudy is de semantische categorie van het werkwoord in de matrixzin. In het onderzoek van zowel Vliemen (2001) als Bouma (2017) – die naast de semantiek van het controlerende werkwoord

voornamelijk syntactische factoren bespreekt, supra – wordt duidelijk dat predicaten die een doelbetekenis in zich dragen zoals *op het punt staan* significant vaker *om* bij zich zullen krijgen dan werkwoorden zonder doelbetekenis. Bouma (2017) neemt het verbale hoofd mee als random effect in zijn model maar voegt daarnaast ook een semantische factor toe als *fixed* effect, nl. het verschil tussen BIZ's met en zonder doelbetekenis. Historisch gezien is de BIZ met facultatief *om* afkomstig van beknopte infinitiefzinnen met verplicht *om* die een doel uitdrukken (IJbema 2002). De hypothese die Bouma (2017) hieruit afleidt, is dat beknopte infinitiefzinnen met facultatief *om* die een doel uitdrukken meer kans hebben op het expliciteren van *om*. Bouma stelt dat het moeilijk is om consistent te bepalen of een bepaalde BIZ een doel uitdrukt of niet. Aan de hand van zelf samengestelde lijsten van werkwoorden die enerzijds typisch zijn voor *complement clauses* – BIZ's met facultatief *om* – zoals *beloven* en *beslissen* en die anderzijds typisch zijn voor *modifier clauses* – BIZ's met verplicht *om* en doelbetekenis – zoals *promoten* en *voorkomen*, maakt hij een inschatting van hoe goed een BIZ met facultatief *om* nog lijkt op een BIZ met verplicht *om* en dus een doelbetekenis heeft. Ten eerste neemt hij aan dat alle zinnen in de dataset met expliciet *om* een doel uitdrukken en die gaat hij als basis gebruiken om een vergelijking te maken met werkwoorden die typisch zijn voor *modifier clauses*, die hij via een automatische syntactische analyse haalt uit het Twente Nieuws Corpus. Hij gebruikt hiervoor de statistische methode *pointwise mutual information* (Church & Hanks 1990). Indien zo'n typisch werkwoord voor doelzinnen voorkomt als verbaal hoofd van de BIZ, concludeert Bouma dat de BIZ in kwestie lijkt op een *modifier clause* en dus een doelbetekenis heeft. Het omgekeerde test hij eveneens: Bouma onderzoekt hoe sterk een ingebedde infinitief geassocieerd is met het hoofd van de BIZ. Hij gaat ervan uit dat verbale hoofden en infinitieven in een *complement clause* een hogere associatiescore hebben dan verbale hoofden en infinitieven in *modifier clause*, aangezien complementen effectief geselecteerd worden door het verbale hoofd. De analyses van Bouma tonen aan dat beknopte infinitiefzinnen met een doelbetekenis inderdaad een significante voorkeur hebben voor de expliciete constructie. Bouma (2017) concludeert dat complexiteit-gerelateerde factoren een belangrijkere rol spelen dan semantiek (doelbetekenis of niet), maar dat de combinatie van beide soorten factoren het beste model oplevert. Doordat een automatische syntactische analyse echter niet mogelijk is met het Dutch Parallel Corpus (waarover meer in § 3.1) en Bouma in zijn studie geen verdere voorbeelden geeft van de werkwoorden die deel uitmaken van zijn beide lijsten, kunnen wij deze semantische factor niet op een gelijkaardige manier operationaliseren.

Een derde semantische factor die we voorlopig niet meenemen, is de aanwezigheid van zinsevaluerende bijwoorden/bepalingen. Volgens Pardoën (1998:

428) bevorderen zinsbepalingen in de matrixzin die naar een evaluerend oordeel van de spreker verwijzen het gebruik van *om*, aangezien die op een zekere sprekergeoriënteerdheid wijzen en ze de evaluatie van de spreker expliciteren. Enkele voorbeelden hiervan zijn *tot mijn grote ergernis*, *gelukkigerwijs* en *he-laas*. Opnieuw is dit een factor die bij gebrek aan welomschreven herkenningcriteria voor evaluerende bepalingen niet makkelijk te operationaliseren is.

2.2.2.2. Taalexterne factoren

Een meermaals aangehaalde taalexterne factor is het medium van de tekst, het verschil tussen gesproken en geschreven taal. In verschillende bronnen (Blom 1984, Jansen 1987, ANS 1997: 1110-1112, Bouma 2017) wordt vermeld dat er in de gesproken taal een algemene voorkeur heerst voor de expliciete vorm en in geschreven taal voor de impliciete constructie. Een exploratief onderzoek⁵ aan de hand van geschreven data uit het Dutch Parallel Corpus en gesproken taal uit het Corpus Gesproken Nederlands suggereert dat expliciet *om* inderdaad relatief vaker voorkomt in gesproken taalgebruik. De *om*-alternantie is in dat opzicht bijzonder, want de aangetroffen distributie staat haaks op wat door de meeste onderzoeken wordt beweerd over de relatie tussen explicietheit en formaliteit, namelijk dat de meer expliciete optie geassocieerd wordt met meer formeel taalgebruik (Biber et al. 1999: 680, Tagliamonte & Smith 2005, Torres Cacoullos & Walker 2009).

Een andere taalexterne factor die de aan- of afwezigheid van *om* mogelijk kan verklaren is het verschil tussen Belgisch Nederlands en Nederlands Nederlands. Pardoën (1998: 427) vermeldt in haar studie over de relatie tussen woordvolgorde en zinsbetekenis dat in teksten van Zuidnederlandse auteurs opvallend minder gebruik gemaakt wordt van *om* in vergelijking met teksten van Noordnederlandse auteurs, maar ze test dat eventuele verschil niet op een systematische manier.

In deze casestudy focussen we verder echter alleen op stilistische variatie tussen vijf verschillende registers binnen de geschreven taal en maken we voorlopig abstractie van het onderscheid tussen Belgisch Nederlands en Nederlands Nederlands. In de bestaande literatuur over de *om*-alternantie wordt nergens gewag gemaakt van zulke meer specifieke registersverschillen, terwijl

⁵ Met een selectie van vijf werkwoorden (*proberen* (876), *slagen in* (463), *vragen* (419), *beslissen* (323), *toelaten* (294)), vijf substantieven (*mogelijkheid* (448), *bedoeling* (294), *doel* (276), *poging* (257), *kans* (240)) en vijf adjectieven (*mogelijk* (376), *moelijk* (329), *in staat* (285), *bereid* (217), *belangrijk* (163)), gekozen op basis van hun frequentie in het Dutch Parallel Corpus, werd bekeken hoe vaak *om* aan- of afwezig is in spontane conversatie met tekstmateriaal uit het Corpus Gesproken Nederlands. De relatieve frequenties van alle geselecteerde woorden behalve *proberen* en *doel* tonen aan dat *om* in gesproken taal relatief vaker voorkomt dan de constructie zonder *om* in vergelijking met het geschreven taalgebruik in het Dutch Parallel Corpus: in gesproken data komt de expliciete constructie in 53,49% van de gevallen voor en in geschreven data gaat het om 45,81% ($\chi^2 = 23,32$; $df = 1$; $p < .001$).

voor andere alternanties al aangetoond is dat register wel degelijk een invloed kan uitoefenen (De Sutter et al. 2005, Kolbe-Hanna & Szmrecsanyi 2015, Tagliamonte & Smith 2005, Torres Cacoullos & Walker 2009). Meer details over de verschillende registers worden gegeven in paragraaf 3.1.

3. METHODOLOGIE

Om het effect van de verschillende in de literatuur voorgestelde factoren te bepalen naast dat van register moet er eerst een relevant sample worden samengesteld. Hierbij is het belangrijk dat de voorbeelden representatief zijn voor het natuurlijke taalgebruik en dat de dataset ook voldoende groot is, zodat er statistisch relevante conclusies uit getrokken kunnen worden. In deze paragraaf worden de keuze van het corpus, de zoekmethode en de gebruikte statistische techniek in detail toegelicht.

3.1. Corpus

Aangezien deze casestudy deel uitmaakt van een onderzoek waarin ook vertaald Nederlands centraal staat en waarin gefocust wordt op het effect van registervariatie, kozen we ervoor om te werken met het Dutch Parallel Corpus (=DPC) (Macken et al. 2011). Het corpus bevat ongeveer 10 miljoen woorden origineel Nederlands en vertaald Nederlands uit het Engels en het Frans. Het corpus is verdeeld in vijf verschillende registers die op hun beurt verder onderverdeeld kunnen worden op basis van de prototypebenadering van Lee (2001): literatuur, opgedeeld in fictie en non-fictie, journalistieke teksten, instructieve teksten, administratieve teksten en externe communicatie. Een schematische weergave van de verschillende registers wordt hieronder in Tabel 1 gegeven.

Tabel 1: Overzicht van registers en tekstvoorbeelden in DPC

| Register | | Tekstvoorbeelden |
|------------------------|---------------------|-------------------|
| Literatuur | Fictie ⁶ | Novelles |
| | | Kortverhalen |
| | Non-fictie | Essays |
| | | (Auto)biografieën |
| | Naslagwerken | |
| Journalistieke teksten | | Krantenartikelen |
| | | Krantencolumns |
| | | Opiniestukken |

⁶ Fictie zal in dit artikel niet worden meegenomen omdat het huidige sample slechts drie relevante hits uit dat register bevatte en dat geen statistisch relevante resultaten kan leveren.

Tabel 1: Overzicht van registers en tekstvoorbeelden in DPC (vervolg)

| Register | Tekstvoorbeelden |
|-------------------------|--|
| Instructieve teksten | Handleidingen |
| | Interne wettelijke teksten |
| | Beschrijving van procedures |
| Administratieve teksten | Wetgeving |
| | Notulen van parlementaire debatten |
| | Notulen van vergaderingen |
| | Jaarrapporten |
| | Correspondentie |
| | Officiële speeches |
| Externe communicatie | Presentaties van organisaties, projecten, events |
| | Promoties en reclamemateriaal |
| | Informatiebrochures |
| | Wetenschappelijke teksten |

Aangezien we voorlopig alleen een beeld willen krijgen van de factoren die het al dan niet realiseren van *om* aan het begin van de beknopte infinitiefzin beïnvloeden, werken we alleen met niet-vertaald Nederlands. In een latere fase zal een vergelijking worden gemaakt tussen vertaald en origineel Nederlands en wordt gekeken of er een verschil is tussen beide. Dat zou kunnen bijdragen tot het karakteriseren van vertalingen als een (geheel van) tekstsoort(en) met zijn/hun eigen bijzondere eigenschappen.

3.2. Zoekmethode

Het DPC is via de online interface volgens verschillende zoekmethoden te doorzoeken.⁷ Voor de huidige casestudy werd gewerkt met een zoekopdracht die bestaat uit vier wildcards en drie concrete deelzoekopdrachten.⁸ We selecteerden zowel de taalvariëteiten Nederlands Nederlands als Belgisch Nederlands, zodat we in een latere fase van het onderzoek ook hierin een onderscheid kunnen maken. De interface van het DPC biedt de mogelijkheid om te zoeken op woord, lemma of woordsoort.⁹ De bedoeling was uiteraard om zowel expliciete als impliciete constructies te verzamelen. We verwerkten dus geen *om* in de zoekopdracht, maar wel elementen die voorkomen in zowel de

⁷ Via <http://dpcserv.ugent.be/comure> kan het DPC geraadpleegd worden met login en wachtwoord.

⁸ Het gebruik van wildcards maakt het mogelijk de zoekvraag iets ruimer te definiëren zodat enige variatie in de concrete voorkomens kan worden afgedekt, maar brengt uiteraard wel meer ruis in de zoekresultaten mee.

⁹ Bovendien kan aan de hand van zgn. *flags* gerichter worden gezocht op woordsoorten. Bijvoorbeeld: binnen de woordsoort substantief kan aan de hand van een *flag* nader gezocht worden naar substantieven in het enkelvoud of in het meervoud (Macken et al 2011).

expliciete constructie als de impliciete: concreet werd gezocht naar zinnen waarin een werkwoord, substantief of adjectief gevolgd wordt door de woordvorm *te* binnen een afstand van nul tot tien woorden en waarin *te* op zijn beurt binnen een afstand van nul tot twee woorden gevolgd wordt door een infinitief. De gebruikte zoekopdracht begint en eindigt met een wildcard, omdat uiteraard ook nog andere woorden voor en na de string in kwestie kunnen voorkomen. Schematisch ziet de zoekopdracht er als volgt uit:

| Wildcard | Deel 1 | Wildcard | Deel 2 | Wildcard | Deel 3 | Wildcard |
|----------|------------|----------|--------|----------|---------------|----------|
| / | Ww, N, Adj | 0-10 | 'te' | 0-2 | Ww, flag: inf | / |

De hierboven beschreven zoekopdracht genereerde in totaal 40.100 te onderzoeken resultaten. Die werden eerst manueel gefilterd op bruikbaarheid. Voor het onderzoek zijn enkel die beknopte infinitiefzinnen relevant waarin *om* zou kunnen worden weggelaten of waarin het, als het er niet staat, zou kunnen worden toegevoegd (zonder dat de betekenis van de zin verandert of de zin ongrammaticaal wordt). In sommige gevallen was niet meteen duidelijk of *om* effectief weggelaten of toegevoegd kon worden. Bij twijfel werd Google geraadpleegd om te controleren of op het Internet wel degelijk voorkomen van het hoofd in kwestie met of zonder *om* voorkomen. Als op de eerste vijf zoekpagina's meer dan 20 voorbeelden werden gevonden, werd besloten de hoofden in kwestie mee te nemen in de analyse. De werkwoorden *wensen* en *hopen* bijvoorbeeld komen in de gehele dataset geen enkele keer mét *om* voor, tegenover respectievelijk 14 en 8 keer zonder *om*. Via Google zijn echter tientallen voorbeelden te vinden waarin een beknopte infinitiefzin bij het matrixwerkwoord *hopen* en *wensen* wel degelijk met *om* wordt ingeleid.

3.3. Statistische analyse

De relevante corpusdata werden geannoteerd op de 14 variabelen die in de vorige paragraaf werden gepresenteerd en dienden als input voor een binaire logistische regressieanalyse.

3.3.1. Binaire logistische regressie

We hebben ervoor gekozen om te werken met binaire logistische regressie omdat we te maken hebben met een dichotome afhankelijke variabele: expliciet vs. impliciet. Op die manier kunnen we bekijken welke factoren een significante invloed uitoefenen op de afhankelijke variabele als er rekening gehouden wordt met meerdere onafhankelijke factoren. Een overzicht van de

onafhankelijke variabelen die uiteindelijk werden meegenomen in het model volgt in paragraaf 3.3.2. Voor dit artikel werden twee verschillende regressiemodellen opgesteld met behulp van het statistische programma R. Het eerste model bevat alle data van het sample, zowel de zinnen met een werkwoord als hoofd als de zinnen met een nominaal of adjectivaal hoofd. De variabelen die alleen toepasbaar zijn op zinnen met een werkwoord als hoofd werden hierin niet meegenomen. Vervolgens werd een afzonderlijk model opgesteld enkel voor de zinnen met een werkwoord als hoofd. We voerden een *stepwise* logistische regressie uit met enkel hoofdeffecten. Zoals gebruikelijk is in dit soort van statistische modellen, hebben we alle numerieke factoren, die erg scheef verdeeld bleken, logaritmisch getransformeerd. Er werd bewust gekozen om in deze paper nog geen rekening te houden met eventuele interacties van de variabelen, aangezien we eerst een algemeen beeld willen krijgen van de significante variabelen, zonder het beeld complexer te maken door ook mogelijke interacties aan het model toe te voegen. De drempel voor significantie werd gelegd op 005. De modellen werden steeds gecontroleerd op multicollineariteit. In deze paper zullen we alleen de resultaten van de binaire logistische regressie van de zinnen met verbale hoofden in detail bespreken. Een eerste reden hiervoor is dat de meeste onderzochte factoren vaker van toepassing zijn op de groep van beknopte infinitiefzinnen met een verbaal hoofd. Daarnaast heeft het model van de werkwoorden de hoogste classificatiescore, i.e. de score die een index geeft voor de correlatie tussen de voorspelde probabiliteiten en de geobserveerde waarden van de responsvariabele.

3.3.2. Schematisch overzicht van de onafhankelijke variabelen

In de vorige paragraaf werd al uitgebreid besproken welke factoren een mogelijke invloed uitoefenen op de keuze van de taalgebruiker tussen expliciet of impliciet *om*. Tabel 2 hieronder geeft nog een schematisch overzicht van de variabelen die worden meegenomen in de regressieanalyse, de respectievelijke waarden die voor categoriale variabelen werden onderscheiden en de bron waaraan de variabele in kwestie werd ontleend. Alle variabelen werden handmatig geannoteerd, aangezien een automatische syntactische analyse niet mogelijk is met het DPC.

Tabel 2: Overzicht van de te onderzoeken variabelen in de regressieanalyse

| Variabele | Waarden | Bron |
|---|---|------------------------------------|
| Register | non-fictie, externe communicatie, journalistieke, instructieve en administratieve teksten | DPC verdeling |
| Hoofd BIZ | Ww, N, Adj | Bouma 2017 |
| Meerdere BIZs | ja, nee | Bouma 2017 |
| Finietheid werkwoord | finiet, infinit | Bouma 2017 |
| Wijze werkwoord | actief, passief | Exploratief onderzoek |
| Scheidbaar samengesteld werkwoord | gescheiden, niet gescheiden | Bouma 2017 |
| Grammaticale persoon van het geïmpliceerd S | 1 ^{ste} persoon, andere persoon | Vliegen 2001 |
| Semantische rol van het geïmpliceerd S in matrixzin | agens, niet-agens | Vliegen 2001 |
| Semantische rol van het geïmpliceerd S in de BIZ | agens, niet-agens | Naar analogie met de vorige factor |
| Functie van het controlerende zinsdeel | subject, anders | Vliegen 2001 |
| Aantal woorden tussen hoofd en BIZ | numerieke variabele | Rohdenburg 1996 |
| Aantal woorden van de BIZ | numerieke variabele | Bouma 2017 |
| Aanwezigheid van ander <i>om</i> | ja, nee | Analogie met factor meerdere BIZs |
| Voornaamwoordelijk bw | ja, nee | Analogie met Bouma 2017 |

4. RESULTATEN

Hieronder wordt een overzicht gegeven van de belangrijkste resultaten van de casestudy. Eerst presenteren we algemene frequentie-informatie en daarna volgt een olijsting en bespreking van de factoren die een significante invloed uitoefenen volgens de binaire logistische regressieanalyse.

4.1. Frequenties

Na de manuele filtering van de gehele dataset werden er uiteindelijk 12.650 beknopte infinitiefzinnen ingeleid met of zonder *om* gevonden, nl. 6.301 voorbeelden van de expliciete en 6.349 van de impliciete constructie. De verzameling bevat zowel vertaalde zinnen als niet-vertaalde zinnen. Voor de huidige casestudy focussen we alleen op de niet-vertaalde zinnen en daarvoor werd

een random sample van 1.138 zinnen samengesteld. De verdeling over expliciet/impliciet en verbaal hoofd/nominaal en adjectivaal hoofd¹⁰ van dat sample wordt in Tabel 3 hieronder weergegeven.

Tabel 3: Verhouding expliciete en impliciete constructies t.o.v. syntactische categorie van het hoofd in het random sample van 1138 attestaties

| Hoofdtype | Expliciet | Impliciet | Totaal |
|-----------------------|--------------|--------------|--------|
| Werkwoord | 242 (45,49%) | 290 (54,51%) | 532 |
| Substantief/Adjectief | 392 (64,69%) | 214 (35,31%) | 606 |
| Totaal | 634 (55,71%) | 504 (44,29%) | 1138 |

We zien een hoogst significant verschil tussen de twee types constructies. De impliciete constructie komt significant vaker voor in BIZ's die afhangen van een werkwoord dan in BIZ's die afhangen van een adjectief of substantief ($\chi^2=41,54$, $df=1$, $p<.001$). Die frequenties laten dus al vermoeden dat het hoofd van de BIZ een invloedrijke factor zal zijn in de keuze tussen de expliciete of de impliciete constructie.

4.2. *Significante factoren: bespreking en interpretatie*

Zoals al vermeld in paragraaf 3.3.1. werden er twee verschillende regressie-analyses uitgevoerd. In het model voor de gehele dataset is het hoofd van de beknopte infinitiefzin de meest significante factor (zie Appendix A). We focussen in dit artikel echter hoofdzakelijk op het afzonderlijke model voor de 532 zinnen met een verbaal hoofd (zie paragraaf 3.2). Hieronder in Tabel 4 wordt een overzicht gegeven van de significante factoren, de schatter die de richting van de voorkeur aangeeft (de referentiewaarde is +om, dus positieve schatters duiden op een relatief grotere kans op de impliciete constructie t.o.v. het intercept en negatieve op een relatief kleinere kans op de impliciete constructie en de significantie van het effect).

Tabel 4: Voorkeur en significantiegraad van de significante variabelen

| Variabele | Schatter | Voorkeur | Sign. |
|--------------------------------------|----------|-----------|-------|
| Intercept= 0.27537 | | | |
| Controlerend zinsdeel _S | 0.91 | Impliciet | *** |
| Gescheiden werkwoord_niet gescheiden | 0.96 | Impliciet | ** |

¹⁰ De beknopte infinitiefzinnen met een substantief en adjectief als hoofd werden samengenomen omdat ze een vergelijkbaar gedrag vertonen. Niet alle geselecteerde factoren zijn namelijk toepasbaar op de zinnen met een substantief of adjectief als hoofd, maar wel op de zinnen met een verbaal hoofd. Enkele voorbeelden: semantische rol van het geïmpliceerde subject in zowel de matrixzin als de BIZ, controlerend zinsdeel enz. Voorbeelden van factoren die op alle beknopte infinitiefzinnen van toepassing zijn, zijn hoofd van de beknopte infinitiefzin, aanwezigheid van andere om enz.

Tabel 4: Voorkeur en significantiegraad van de significante variabelen (vervolg)

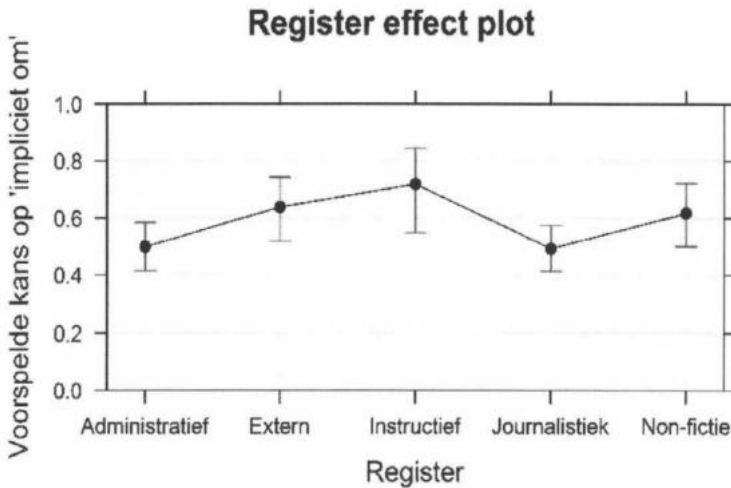
| Variabele | Schatter | Voorkeur | Sign. |
|-------------------------------|----------|-----------|-------|
| Finietheid werkwoord_infiniet | -0.92 | Expliciet | *** |
| Elementen_hoofd_BIZLog | -0.34 | Expliciet | ** |
| Register_extern | 0.57 | Impliciet | . |
| Register_instructief | 0.94 | Impliciet | * |
| Andere om_nee | -0.87 | Expliciet | . |

Bij het interpreteren van Tabel 4 moet er rekening gehouden worden met de referentiewaarde van iedere variabele. Dat betekent dat binnen iedere (categorische) variabele automatisch één waarde als defaultniveau gekozen wordt, en dat alle andere waarden daartegen afgezet worden. Een voorbeeld: in vergelijking met zinnen die een finiet matrixwerkwoord hebben, is er in zinnen met een infiniet matrixwerkwoord een significant kleinere kans op impliciet *om*. De positieve *intercept* volgt uit het feit dat het merendeel van de onderzochte zinnen impliciet is. Zie de effectplots in Appendix C voor een visuele weergave van de effecten.

Een eerste vaststelling is dat er slechts vijf van de 13 onderzochte factoren een significante invloed hebben als hoofdeffect op de aan- of afwezigheid van *om*: register (meer bepaald is er in de instructieve teksten een significant kleinere kans op *om* dan in het referentieniveau, de administratieve teksten), controlerend zinsdeel van het geïmpliceerd subject in de BIZ, de aanwezigheid van een gescheiden werkwoord, de finietheid van het werkwoord en het aantal woorden tussen het matrixwerkwoord en het begin van de beknopte infinitiefzin. Twee andere effecten zitten net boven de significantiedrempel van 0,05: opnieuw register maar dan externe communicatie, waarover hieronder meer, en de aanwezigheid van een ander *om*. Wanneer in dezelfde zin géén ander *om* aanwezig is, is er een relatief grotere kans op de expliciete constructie, al is dat effect dus net niet significant. Als er wel al een ander *om* aanwezig is – hetzij als inleidend voegwoord van een andere BIZ (nevengeschild of ondergeschild), hetzij als een syntactisch ander soort *om* zoals eerste deel van een scheidbaar samengesteld werkwoord – dan is er dus een relatief kleinere kans op *om* aan het begin van de BIZ in kwestie. Omdat dit effect an sich moeilijk te interpreteren is, zal in een volgende fase zal een onderscheid gemaakt worden tussen de verschillende soorten *om* en hun positie in de zin.

Een belangrijk resultaat van deze casestudy is de bevinding dat register een significant effect uitoefent. Instructieve teksten tonen in vergelijking met teksten van het register administratieve teksten, de referentiewaarde, een significant grotere kans op de impliciete constructie; externe communicatie een bijna significante voorkeur. Journalistieke teksten en non-fictie tonen dan weer geen significant verschil met de administratieve teksten. Zoals ook in grammatica's

zoals de ANS en de SoD al wordt vermeld, is het weglaten van *om* eigen aan geschreven en formeel taalgebruik. De resultaten van de logistische regressie lijken dat deels te bevestigen. Instructieve teksten en externe communicatie kunnen inderdaad als vrij formele registers beschreven worden; het is echter niet zo dat de andere registers per definitie *minder* formeel zijn dan instructieve teksten en externe communicatie. Teruggrijpend naar de resultaten van het onderzoek van Kirsner (1988) over de datiefalternantie in krantentaal, hadden we ook kunnen verwachten dat instructieve teksten eerder vaker expliciet *om* bij zich zou krijgen, aangezien aangenomen mag worden dat dat register "duidelijkheid" vraagt en het expliciteren van *om* daarvoor een hulpmiddel kan zijn. Het effectplot in Figuur 1 hieronder laat zien dat het journalistieke register – hoewel niet significant – op hetzelfde niveau zit als administratieve teksten, de referentiewaarde en dus geen voorkeur heeft voor de impliciete constructie. Kirsner (1988) suggereerde in zijn onderzoek over de datiefalternantie een voorkeur voor de explicietere constructie in journalistiek taalgebruik maar wij treffen dus voor de *om*-alternantie geen vergelijkbare voorkeur voor de explicietere variant aan in journalistieke teksten.



Figuur 1: Effectplot van register als significante factor

Onze resultaten tonen aan dat er bepaalde kenmerken aanwezig moeten zijn – los van de formaliteit van de tekst – in externe communicatie en instructieve teksten die niet of minder sterk aanwezig zijn in de overige drie onderzochte registers en die relevant zijn voor de keuze tussen de constructie met en die zonder *om*. Welke situationele kenmerken dat precies zijn, is op dit moment niet te zeggen, maar in elk geval laat het model dus zien dat de alternantie wel degelijk gevoelig is voor registreffecten, ook binnen de geschreven taal.

Als we terugkeren naar de significante hoofdeffecten, dan zien we dat ook de factor 'controleerend zinsdeel' de keuze beïnvloedt. Zinnen waarin het geïmpliceerde subject van de BIZ correspondeert met het subject van de matrixzin vertonen een relatief sterkere voorkeur voor impliciet *om*. Op basis van dit resultaat alleen kunnen we geen vergelijking maken met de resultaten van Vliegen (2001). Daarvoor zullen we in de toekomst enkele interacties toevoegen aan het regressiemodel, zoals de interacties tussen controleerend zinsdeel, persoon van het geïmpliceerd subject en de semantische rol van het geïmpliceerd subject in zowel de matrixzin als de beknopte infinitiefzin. Het niet toevoegen van dergelijke interacties kan mogelijk verklaren waarom factoren als de persoon van het geïmpliceerd subject en de semantische rol van datzelfde subject in de matrixzin geen significante rol spelen; Vliegen suggereert namelijk een zekere interactie tussen de persoon van het geïmpliceerde subject en de agentiviteit ervan. De aanwezigheid van een geïmpliceerd subject in de eerste persoon dat ook agens is in de matrixzin zou de aanwezigheid van *om* bevorderen.

De overige drie significante factoren hebben allemaal in zekere zin te maken met complexiteit. Alle drie suggereren ze dat hoe complexer de zin wordt, hoe groter de kans is op expliciet *om*. In de eerste plaats vertonen niet gescheiden werkwoorden een voorkeur voor impliciet *om* en gescheiden samengestelde werkwoorden voor expliciet *om*. Ook de finietheid van het werkwoord speelt een significante rol in de keuze tussen een van de alternerende constructies. Als het werkwoord infinitief is, en dus volgens onze hypothese complexer, is er een significant grotere voorkeur voor expliciet *om*. De bevindingen van Bouma (2017) komen hiermee slechts gedeeltelijk overeen: hij constateert weliswaar dat het hoogste percentage van expliciet *om* gevonden wordt bij finiete werkwoorden, maar het verschil met matrixzinnen met een infinitief werkwoord is niet significant. Mogelijk heeft dat iets te maken met het register waaruit de data voor het onderzoek van Bouma komen. Hij onderzoekt alleen journalistieke teksten en dat register heeft volgens de resultaten van het huidige onderzoek in het algemeen een zekere voorkeur voor het expliciteren van *om*. De laatste complexiteit-gerelateerde significante factor laat zien dat hoe meer woorden er zich tussen het matrixwerkwoord en het begin van de beknopte infinitiefzin bevinden, hoe groter de kans dat de taalgebruiker *om* zal gebruiken, zoals te verwachten was volgens het complexiteitsprincipe (Rohdenburg 1996). Deze interpretatie van complexiteit staat haaks op de manier waarop De Sutter (2007) in zijn onderzoek over de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen de afstand tussen het matrixwerkwoord en het begin van de bijzin interpreteert. Meer tussenliggende woorden vergemakkelijken de verwerking van de werkwoordelijke eindgroep en bieden de toehoorder meer informatie om de inhoud van de werkwoordelijke eind-

groep zelf te kunnen voorspellen. Met andere woorden: hoe hoger het aantal tussenliggende woorden, hoe lager de complexiteit van de zin. Het verschil tussen beide interpretaties zit vooral in het perspectief: in De Sutter (2007) staat het verwerken van de volgende zin centraal, terwijl in deze paper de focus meer op het al dan niet produceren van *om* ligt. Daarnaast staan, in het geval van de *om*-alternantie, de hoofdzin en de bijzin relatief los van elkaar. Meer woorden in de hoofdzin zullen dus niet noodzakelijk helpen om de inhoud van de BIZ – met zijn eigen zinswerkwoord – te kunnen voorspellen, terwijl dat bij de werkwoordelijke eindgroep wel het geval is aangezien het steeds gaat over het zelfstandige werkwoord van de hoofdzin dat nog moet volgen.

De syntactische complexiteit van de beknopte infinitiefzin zelf lijkt geen rol te spelen, wat ook bij Bouma (2017) het geval was. Het is ook intuïtief plausibel dat de lengte van de beknopte infinitiefzin als zodanig eigenlijk niet van belang is voor het al dan niet realiseren van *om*, aangezien die keuze meteen aan het begin van de BIZ gemaakt moet worden.

Samengevat kunnen we op basis van ons model stellen dat de voornaamste factoren die invloed uitoefenen op de keuze tussen expliciet en impliciet *om* te maken lijken te hebben met registervariatie en met de complexiteit van de matrixzin van de BIZ. Hieruit concluderen we dat zowel taalinterne als taalexterne variabelen een belangrijke rol spelen in de keuze tussen twee alternerende constructie. Bij de taalinterne factoren gaat het grotendeels om syntactische factoren, aangezien de meerderheid van de onderzochte factoren syntactisch van aard waren. De lage C- en vooral lage R²-score (respectievelijk 0,72 en 0,19) wijzen erop dat er nog andere factoren een rol moeten spelen. Het huidige opgestelde model kan de variatie tussen de impliciete en expliciete constructie niet voldoende verklaren. Het toevoegen van onder meer enkele semantische factoren zoals de semantische categorie van het controlerende werkwoord en het toepassen van een *mixed effects* regressie-analyse met het specifieke hoofd van de BIZ als random effect kan hiervoor een mogelijke oplossing bieden.

5. CONCLUSIES

In deze paper stonden de factoren die de keuze bepalen tussen het al dan niet expliciteren van *om* aan het begin van de beknopte infinitiefzin centraal. We hebben daarvoor een multifactorieel corpusonderzoek uitgevoerd met behulp van het Dutch Parallel Corpus. Een eerste belangrijke vaststelling is dat vele factoren die in de literatuur werden voorgesteld als mogelijke determinerende variabelen geen effect blijken uit te oefenen wanneer ze simultaan onderzocht worden. In de meeste aangehaalde studies over de *om*-alternantie werden dan ook geen statistische controles doorgevoerd en werd er veelal geen rekening

gehouden met de simultane effecten van de verschillende mogelijk beïnvloedende factoren; Bouma (2017) is daarop de grote uitzondering. In zijn onderzoek blijken voornamelijk complexiteit-gerelateerde factoren een significante rol te spelen, naast een semantische factor, namelijk het verschil tussen beknopte infinitiefzinnen met of zonder doelbetekenis.

De factoren die volgens ons onderzoek een rol spelen in de keuze tussen expliciet en impliciet *om* kunnen grosso modo in twee groepen worden verdeeld. Ten eerste is er een groep met complexiteit-gerelateerde variabelen. Duidelijk is dat hoe complexer de (matrix)zin is, hoe groter de kans dat de taalgebruiker *om* realiseert. In de literatuur wordt complexiteit vaak als een belangrijke factor gezien om te expliciteren, d.w.z. bij alternanties waar de ene variant explicieter is dan de andere wordt er relatief vaker gekozen voor de meer expliciete constructie in cognitief meer complexe omgevingen. Door een hogere complexiteit van de zin moet de taalgebruiker meer informatie opnemen en dat verhoogt de noodzakelijkheid van expliciete constructies om de taalgebruiker te helpen bij het verwerken en interpreteren van de zin. Deze resultaten sluiten goed aan bij het onderzoek van Bouma (2017).

Terwijl de eerste groep van significante factoren taalintern was, is de tweede taalextern, namelijk het register van de tekst. Instructieve teksten en externe communicatie hebben een voorkeur voor de impliciete constructie in vergelijking met administratieve teksten als referentiewaarde. Hieruit kunnen we besluiten dat er steeds rekening moet gehouden worden met taalexterne factoren. In het onderzoek van Bouma (2017) werd alleen gewerkt met krantentaal en zelfs daar werd duidelijk dat de keuze tussen expliciet of impliciet *om* significant verschilde afhankelijk van de precieze krant. Zoals hierboven al vermeld zullen in een volgende fase de registers zelf in meer detail onderzocht worden met aandacht voor de stilistische karakteristieken waarin ze verschillen van elkaar. Op die manier willen we concrete stilistische aspecten toevoegen aan de constructionele betekenis van zowel de expliciete constructie als de impliciete. Deze casestudy was een eerste poging om na te gaan of registerverschillen inderdaad een invloed uitoefenen op de grammaticale keuze van de taalgebruiker wanneer ook rekening gehouden wordt met andere taalinterne factoren. De lage C- en R²-scores van het in dit artikel opgestelde model wijzen erop dat nog andere factoren een rol moeten spelen in het al dan niet expliciteren van *om*. In de toekomst zullen we dan ook rekening houden met de verschillende niet meegenomen semantische factoren zoals de semantische categorie van het werkwoord.

In toekomstig onderzoek willen we als extra buitentalige factor zeker ook het onderscheid tussen Belgisch Nederlands en Nederlands Nederlands meenemen in het multifactoriële model, zoals ook al vermeld in 2.2.2.2. Pardoën (1998: 427) vermeldt in haar studie over de relatie tussen woordvolgorde en

zinsbetekenis dat in teksten van Zuidnederlandse auteurs opvallend minder gebruik gemaakt wordt van *om* in vergelijking met teksten van Noordnederlandse auteurs. De dataset zal ook worden uitgebreid met materiaal uit het Corpus Gesproken Nederlands om op die manier een onderscheid te kunnen maken tussen geschreven en gesproken taalgebruik. Exploratief onderzoek heeft al aangetoond dat in gesproken taal beduidend vaker expliciet *om* lijkt voor te komen dan in geschreven taal. Door het toevoegen van een gesproken component aan de dataset kan dat ook statistisch in meer detail worden onderzocht.

Nadat we een beter beeld gekregen hebben van de factoren die een invloed uitoefenen op het al dan niet gebruiken van *om* in niet-vertaald Nederlands, zullen we ook vertaalde teksten betrekken in het onderzoek. We zullen daarbij voornamelijk bestuderen of er gelijkenissen en/of verschillen zijn tussen vertaalde en niet-vertaalde teksten in de factoren die invloed uitoefenen op het al dan niet realiseren van *om*. Op die manier kunnen we het specifieke karakter van vertaalde teksten in kaart brengen. In een nog later stadium van het onderzoeksproject zullen er meerdere alternanties betrokken worden, zoals de datiefalternantie en de variatie tussen het semi-copulatieve gebruik van de evidentieële werkwoorden *blijken*, *lijken* en *schijnen* en hun gebruik als hulpwerkwoord bij de infinitief *te zijn* (dus: *Hij blijkt afwezig* vs. *Hij blijkt afwezig te zijn*). Door verschillende alternanties te betrekken in het onderzoek zullen we beter in staat zijn de resultaten door te trekken naar overkoepelende uitspraken over alternanties, registervariatie en expliciteer in het algemeen.

BIBLIOGRAFIE

- ANS = Haeseryn, W. et al. 1997. *Algemene Nederlandse Spraakkunst*. Groningen: Nijhoff/Deurne: Wolters Plantyn.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Pearson Education.
- Biber, D. & S. Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.
- Blom, A. 1984. Enkele opmerkingen over *te* en *om*. *Spektator*, 170-176.
- Blom, A. 1990. *Om* voor de infinitief in *Eline Vere*: een weloverwogen keuze. *Spektator*, 9-18.
- Bouma, G. 2017. Om-omission. In: Wieling, M., M. Kroon, G. van Noord & G. Bouma (eds.). *From Semantics to Dialectometry: Festschrift in honor of John Nerbonne*. (Tributes; Vol. 32). UK: College Publications, 65-73.
- Bresnan, J. & T. Nikitina. 2009. The gradience of the Dative Alternation. In: Uyechi, L. & L.H. Wee (eds.). *Reality exploration and discovery: Pattern interaction in language and life*. Stanford: CSLI, 161-184.
- Church, K. W. & P. Hanks. 1990. Word association norms, mutual information and lexical production. *Computational Linguistics* 16 (1), 22-29.

- De Sutter, G., D. Speelman & D. Geeraerts. 2005. Regionale en stilistische effecten op de woordvolgorde in werkwoordelijke eindgroepen. *Nederlandse Taalkunde* 10 (2), 97-128.
- De Sutter, G. 2007. "Naar een corpusgebaseerde, cognitief-functionele verklaring van de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen." *Nederlandse Taalkunde* 12 (4), 302-330.
- Geeraerts, D., G. Kristiansen & Y. Peirsman (eds.). 2010. *Advances in Cognitive Sociolinguistics*. Berlin: de Gruyter Mouton.
- Goldberg, A. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gries, S. Th. 2003. *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. London: Continuum.
- Grondelaers, S. & D. Speelman. 2007. A variationist account of constituent ordering in presentative sentences in Belgian Dutch. *CL & LT* 3, 161-193.
- Hawkins, J. A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Ijbema, A. 2002. *Grammaticalization and infinitival complements in Dutch*. Leiden University.
- Jansen, F. 1987. Omtrent de om-trend. *Spektator* 17, 83-98.
- Kerz, E. & D. Wiechmann, 2015. Register-contingent entrenchment of constructional patterns: Causal and concessive adverbial clauses in academic and newspaper writing. *English Linguist* 43, 61-85.
- Kirsner, R.S. 1988. Prepositional versus 'bare' indirect objects in the written Dutch of novels and newspapers. In: Broos, T. (eds.). *Papers from the Third Interdisciplinary Conference on Netherlandic Studies*. Lanham: University Press of America, 279-296.
- Kolbe-Hanna, D. & B. Szmrecsanyi. 2015. Grammatical Variation. In: Biber, D. & R. Reppen (eds.). *The Cambridge Handbook of English Corpus linguistics*. Cambridge: Cambridge University Press, 161-179.
- Lee, D. 2001. Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology* 5 (3), 37-72.
- Macken, L., O. De Clercq & H. Paulussen. 2011. Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus. *Meta* 56 (2), 374-390.
- Mondorf, B. 2014. (Apparently) Competing Motivations in Morpho-Syntactic Variation. In: MacWhinney, B., A. Malchukov & E. Moravcsik (eds.). *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press, 209-228.
- Pardoën, J.A. 1998. *Interpretatiestructuur. Een onderzoek naar de relatie tussen woordvolgorde en zinsbetekenis in het Nederlands*. Diss. Vrije Universiteit Amsterdam. Uitgaven Stichting Neerlandistiek VU, 25.
- Pütz, M., J.A. Robinson & M. Reif (eds.). 2014. *Cognitive Sociolinguistics: Social and cultural variation in cognition and language use*. Amsterdam/Philadelphia: John Benjamins.

- Rohdenburg, G. 1996. Cognitive complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics* 7 (2), 149-182.
- SoD = Broekhuis, H., N. Corver, R. Vos & H. Bennis. 2015. *Syntax of Dutch: verbs and verb phrases*. Amsterdam: Amsterdam University Press.
- Szmrecsanyi, B. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In: Geeraerts, D., G. Kristiansen & Y. Peirsman (eds.), 141-166.
- Tagliamonte, S. & J. Smith. 2005. No Momentary Fancy! The Zero "Complementizer" in English Dialects. *English Language and Linguistics* 9 (2), 289-309.
- Theijssen, D., L. ten Bosch, L. Boves, B. Cranen & H. van Halteren. 2013. Choosing alternatives: Using Bayesian Networks and memory-based learning to study the dative alternation. *CL & LT* 9, 227-262.
- Torres Cacoullous, R. & J. A. Walker. 2009. On the Persistence of Grammar in Discourse Formulas: A Variationist Study of *that*. *Linguistics* 47 (1), 1-43.
- Van Craenenbroeck, J. 2000. 'Complementierend *van*. Een voorbeeld van syntactische variatie in het Nederlands'. *Nederlandse Taalkunde* 5 (2), 133-163.
- Vandeweghe, W. 2004. *Grammatica van de Nederlandse zin*. Antwerpen /Apeldoorn: Garant.
- Van Haaften, T. 1991. *De interpretatie van verzwegen subjecten*. Dordrecht: ICG printing.
- Verhagen, A. 1996. Tekstsegmentatie, onderschikking en subjectiviteit. *Gamma/TTT* 5, 249-268.
- Vliegen, M. 2001. Het facultatieve om na illocutionaire werkwoorden. *Nederlandse Taalkunde* 6, 112-132.
- Yoon, J., & S. Th. Gries. 2016. *Corpus-based approaches to construction grammar*. Amsterdam: John Benjamins Publishing Company.
- Wasow, T., T. F. Jaeger & D. Orr. 2011. *Lexical variation in relativizer frequency. Expecting the unexpected: Exceptions in grammar*, 175-196.

APPENDIX

Appendix A: Binaire logistische regressie: model gehele sample in niet-vertaald Nederlands

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -1.6057 | -1.0114 | -0.8309 | 1.1538 | 1.7906 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|---------------------------------------|----------|------------|---------|--------------|
| (Intercept) | -1.08261 | 0.28246 | -3.833 | 0.000127 *** |
| Hoofd_BIZN | 0.13247 | 0.17247 | 0.768 | 0.442458 |
| Hoofd_BIZV | 0.97874 | 0.15956 | 6.134 | 8.57e-10 *** |
| Herverdeling3_registerPolitical_texts | 0.33893 | 0.17583 | 1.928 | 0.053903 |
| Herverdeling3_registerSpecial_special | -0.34676 | 0.15998 | -2.167 | 0.030200 * |
| Vorm_wwl infinit | -0.43538 | 0.15300 | -2.846 | 0.004433 ** |
| Synt_complexiteit_BIZLog | 0.14638 | 0.08145 | 1.797 | 0.072308 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1567.6 on 1140 degrees of freedom

Residual deviance: 1502.2 on 1134 degrees of freedom

AIC: 1516.2

Number of Fisher Scoring iterations: 4

| | |
|----------------------------|------------|
| C-score | 0.63458116 |
| R²-score | 0.07458084 |

Appendix B: Binaire logistische regressie: model verbaal hoofd in niet-vertaald Nederlands

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.9320 | -1.0619 | 0.6186 | 0.9404 | 1.9747 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|------------------------------|----------|------------|---------|--------------|
| (Intercept) | 0.27537 | 0.61945 | 0.445 | 0.65665 |
| Controlerende_functieS | 0.91018 | 0.20142 | 4.519 | 6.22e-06 *** |
| Gescheiden_wwniet_gescheiden | 0.95540 | 0.29484 | 3.240 | 0.00119 ** |
| Vorm_wwlinfiniet | -0.92417 | 0.22398 | -4.126 | 3.69e-05 *** |
| Elementen_hoofd_BIZLog | -0.33524 | 0.12634 | -2.653 | 0.00797 ** |
| RegisterExtern | 0.56916 | 0.30403 | 1.872 | 0.06120 |
| RegisterInstructief | 0.94070 | 0.41448 | 2.270 | 0.02323 * |
| RegisterJournalistiek | -0.02502 | 0.24333 | -0.103 | 0.91810 |
| RegisterNon-fictie | 0.48137 | 0.30237 | 1.592 | 0.11138 |
| Andere_omnee | -1.06577 | 0.55238 | -1.929 | 0.05368. |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 733.17 on 531 degrees of freedom

Residual deviance: 653.21 on 522 degrees of freedom

AIC: 673.21

Number of Fisher Scoring iterations: 4

| | |
|-----------------------|-----------|
| C-score | 0.7194856 |
| R ² -score | 0.1865867 |

Appendix C: Effectplots bij binaire logistische regressie van de verbale hoofden

