

Het gebruik van *zo'n* en *zulk*

Drie hypotheses over lexicale voorkeuren, functie en register

Dirk Pijpops

Onderzoeksgroep Lilith, Universiteit van Luik

dirk.pijpops@uliege.be

Abstract

This corpus study investigates the alternation between *zo'n* and *zulk* 'such' in front of plural and uncountable singular nouns, by putting three hypotheses to the test. The first hypothesis concerns the potential influence of lectal contamination. This is an effect whereby contact between two language varieties leads to lexical biases in the determinants of language variation within each of the varieties separately. Concretely, it is hypothesized that typically Belgian phrases more often exhibit *zo'n* than typically Netherlandic phrases, both in the language use of Belgians and in the language use of Dutchmen. The second hypothesis predicts that in the Netherlands, in front of singular, uncountable nouns, *zulk* is preferred in the identifying function compared to the intensifying function. Finally, the third hypothesis holds that in Belgium, in front of plural nouns, *zulk* is preferred in formal registers. The first hypothesis is not confirmed, while the second is confirmed, pending some qualifications, and the third is also confirmed.

Keywords: corpus; contamination; such; identification; intensification

1 Inleiding

In zinnen als (1)-(4) wordt in het Nederlands zowel het voornaamwoord *zulk* als *zo'n* gebruikt, schijnbaar door elkaar (Van Olmen & Van der Auwera 2014). Dit artikel wil drie hypotheses testen die een voorspelling doen over wat die keuze tussen *zulk* en *zo'n* bepaalt. Het beperkt zich daarbij tot de keuze tussen beide voornaamwoorden voor meervoudige en niet-telbare enkelvoudige substantieven, aangezien *zulk* voor telbare enkelvoudige

substantieven bijna niet voorkomt, tenzij bij het substantief *soort* (Van Olmen & Van der Auwera 2014: 217).

- (1) Omdat je vaak uh mensen hebt van de sp die bij een andere verkiezing de volgende keer voor het Vlaams Blok stemmen. (...) kunt je dan ook afvragen of je met **zulke** mensen kunt samenwerken. (CGN-bestand: fv400036)
- (2) Daar komen dan inderdaad oudere mensen of mensen die in aanmerking komen voor zoiets. Of heel eenzamen. Dat wordt steeds moeilijker om **zo'n** mensen bij elkaar te krijgen. (CGN-bestand: fn007842)
- (3) in een tijd waar d'r eigenlijk **zulke** vreselijke dingen in de wereld gebeuren. (CGN-bestand: fn007364)
- (4) Ze bedankte hem dat ze altijd **zo'n** goeie vrienden waren geweest. (CGN-bestand: fn000461)

De eerste hypothese die aan bod komt in dit artikel, heeft betrekking op mogelijke lexicale voorkeuren in de alternantie tussen *zo'n* en *zulk*.¹ Hebben bepaalde frasen, zoals *mensen* in (2)-(1), *vreselijke dingen* in (3) en *goeie vrienden* in (4), een inherente voorkeur voor *zo'n* of *zulk*, en vooral, kunnen we die voorkeuren voorspellen?

Specifiek wordt gekeken naar een mogelijke invloed van lectale contaminatie. Lectale contaminatie is een effect waarbij taalcontact en taalverwerking in het spel zijn. Contact tussen twee taalvariëteiten zorgt ervoor dat bepaalde taalvormen of woordcombinaties vaker voorkomen, en daardoor als geheel steviger verankerd raken in het geheugen van de taalgebruikers, zodat er een lexicale voorkeur ontstaat voor een bepaalde taalvariant. Concreet voorspelt lectale contaminatie dat de woorden die vaker gebruikt worden in één taalvariëteit – oftewel 'lect' – een voorkeur ontwikkelen voor een morfosyntactische variant die typisch is voor datzelfde lect, zelfs onder sprekers van een ander lect.

Uit eerder onderzoek is al gebleken dat er bij de keuze tussen *zo'n* en *zulk* een proportioneel verschil bestaat tussen het Belgische en het Nederlandse Nederlands: bij Belgen valt de keuze vaker op *zo'n* dan bij Nederlanders (Van Olmen & Van der Auwera 2014: 217). Als zich hier lectale contaminatie

1 Ik gebruik het woord 'alternantie' hier eenvoudigweg om een onderzoeksopzet aan te duiden waarin twee taalvormen tegen elkaar afgezet worden (Pijpops 2020). Door dit woord te gebruiken, poneer ik dus niet dat beide voornaamwoorden een speciale theoretische relatie tot elkaar hebben, bijvoorbeeld als allostructies (Cappelle 2006) of via een constructionele link (Zehentner & Traugott 2020), noch dat elke individuele taalgebruiker steeds een vrije keuze heeft tussen beide voornaamwoorden (Wallis 2012).

voordoet, zou dat betekenen dat Belgen en Nederlanders elk apart meer geneigd zijn om *zo'n* te gebruiken bij meer 'Belgische' frasen, zoals *zo'n/zulke mattentaarten*, en *zulk* bij meer 'Nederlandse' frasen zoals *zo'n/zulke tompoezen*.

De tweede hypothese heeft te maken met de verschillende functies die *zo'n* en *zulk* kunnen vervullen. Er wordt met name een onderscheid gemaakt tussen de zogenaamde identificeringsfunctie, waarbij het alternerende voornaamwoord verwijst naar een referent, zoals in (1)-(2), en de intensiveringsfunctie, waarbij het voornaamwoord een bepaalde eigenschap van het bijhorende substantief versterkt, zoals in (3)-(4) (Van Olmen & Van der Auwera 2014). Die tweede functie is beperkt tot frasen met gradeerbare substantieven, zoals *zo'n/zulke schatjes*, of gradeerbare adjectieven, zoals *vreselijke* in (3) en *goeie* in (4). Voor enkelvoudige, niet-telbare substantieven zou er in Nederland bij de identificeringsfunctie vaker gekozen worden voor *zulk* (Van Olmen 2019: 218) dan bij de intensiveringsfunctie.

Tot slot kijkt de derde hypothese naar de invloed van register. In België zou het voornaamwoord *zo'n* namelijk vaker gebruikt worden in een informeel register dan in een formeel register, bij meervoudige substantieven, zoals onderstaand uitsnede van Taaladvies.net laat zien.

In België wordt *zo'n* ook veelvuldig gebruikt in combinatie met meervoudige zelfstandige naamwoorden, ook door veel standaardtaalsprekers. Toch is er een niet te verwaarlozen groep taalgebruikers die dergelijke combinaties afkeurt. Het is daarom niet duidelijk of die combinaties tot de standaardtaal in België gerekend kunnen worden. (...) Het gebruik van *zo'n* voor niet-telbare zelfstandige naamwoorden is wél standaardtaal in België. Het is er zelfs het meest normale gebruik. Toch wordt ook het gebruik van *zulk(e)* in die gevallen aanvaard. (<https://taaladvies.net/zon-of-zulke-mensen-zon-of-zulk-bier/>, bezocht op 15 september 2021).

Het artikel is opgebouwd als volgt. Sectie 2 introduceert de drie hypothesen in detail. Sectie 3 beschrijft de extractie, selectie en annotatie van de corpusdata. Deze data worden in Sectie 4 geanalyseerd, waarna Sectie 5 bekijkt in hoeverre de resultaten van die analyse de hypothesen bevestigen. Sectie 6 vat ten slotte de conclusies van het onderzoek samen, en suggereert enkele mogelijkheden voor verder onderzoek.

2 Hypotheses

In deze sectie worden de drie hypothesen voorgesteld die in dit artikel getest worden. De eerste subsectie stelt het effect van lectale contaminatie in detail voor. De tweede subsectie gaat in op de verschillende functies die zowel *zo'n* als *zulke* kunnen vervullen, en de derde subsectie kijkt naar de invloed van register.

2.1 Hypothese 1: lectale contaminatie

Lectale contaminatie doet zich voor wanneer de volgende vier condities vervuld zijn (Pijpops te versch.)

1. Een verschil in relatieve voorkeur bij het gebruik van morfosyntactische varianten tussen beide lecten
2. Verschillen in relatieve lexicale voorkeuren, oftewel proportionele verschillen tussen beide lecten bij het gebruik van woorden
3. Taalcontact tussen beide lecten
4. Cognitieve opslag van volledige taalvormen door taalgebruikers

Conditie 1 houdt in dat er een geval van morfosyntactische variatie bestaat, waarbij de taalgebruikers in het ene lect een voorkeur hebben voor één variant, terwijl taalgebruikers van het andere lect een voorkeur hebben voor een andere variant.

Een voorbeeld is te vinden bij de partitieve genitief. Dat is een constructie die zowel verschijnt in een variant met een *s*-uitgang op het adjectief, zoals in (5), als een variant zonder een *s*-uitgang, zoals in (6). De variant met *-s* is over het algemeen populairder in het Nederlandse Nederlands, terwijl de variant zonder *-s* vaker voorkomt in het Belgische Nederlands (Van der Horst 2008: 1624-1625), al lijkt ook in Belgisch taalgebruik de variant met *-s* dominant (Pijpops & Van de Velde 2018).

- (5) Maar was er **iets speciaals** of belde je zomaar?
(overgenomen uit Pijpops te versch., CGN-bestand: fno08371)
- (6) Heb je nog wat gekocht? **Niets speciaal**.
(overgenomen uit Pijpops te versch., CGN-bestand: fno08212)

Conditie 2 stelt dat *zulke* verschillen zich ook voordoen in het lexicon van beide lecten. De frase *wat bijzonder(s)* komt bijvoorbeeld vaker in het Nederlandse Nederlands voor dan in het Belgische, terwijl voor andere frasen, zoals *iets speciaal(s)*, net het omgekeerde geldt. *Zulke* lectale verschillen duiken

op doorheen de gehele Nederlandse woordenschat (Speelman, Grondelaers & Geeraerts 2008; Ruetten 2012; Daems, Heylen & Geeraerts 2015).

Uit Conditie 1 en 2 volgt dat er zich schijnbare lexicale voorkeuren zullen voordoen, wanneer men data uit beide lecten tezamen bekijkt. Woorden die vaker voorkomen in het ene lect, zullen namelijk vaker verschijnen in de variant die typisch is voor datzelfde lect. De partitieve genitief *wat bijzonder(s)* zal bijvoorbeeld vaker met -s verschijnen dan de partitieve genitief *iets speciaal(s)*, simpelweg omdat *wat bijzonder(s)* vaker door Nederlanders gebruikt wordt. Ik spreek hier nog over 'schijnbare' lexicale voorkeuren, omdat het gebruik van de frase *wat bijzonder(s)* in deze stap nog niet resulteert in een voorkeur voor de vorm met -s. Het is met andere woorden (nog) niet zo dat taalgebruikers een voorkeur hebben voor de vorm met -s omdat ze de frase *wat bijzonder(s)* gebruiken.

Conditie 3 vereist dat er zich taalcontact voordoet tussen beide lecten, d.w.z. dat er een aantal taalgebruikers zijn die blootgesteld zijn aan taalmateriaal uit beide lecten. Dat is het geval voor het Nederlandse en Belgische Nederlands (zie bv. Centraal Bureau voor de Statistiek 2020; Van Agtmaal-Wobma e.a. 2007). Hieruit volgt dat de schijnbare lexicale voorkeuren waarvan hierboven sprake is, zich voordoen in het taalmateriaal waaraan deze taalgebruikers zijn blootgesteld. Belgen die zowel taalgebruik horen van andere Belgen als van Nederlanders, zullen proportioneel vaker *wat bijzonder(s)* met een s-uitgang horen dan *iets speciaal(s)*, en hetzelfde geldt voor Nederlanders die zowel taalgebruik horen van andere Nederlanders als van Belgen.

Conditie 4 houdt in dat taalgebruikers af en toe volledige taalvormen, oftewel combinaties van woorden en morfosyntactische constructies, opslaan in hun geheugen (zie o.a. Arnon & Snider 2010; Tremblay & Baayen 2010; Tremblay e.a. 2011; Dąbrowska 2014, en de referenties die daarin aangehaald worden). Dit type opslag kan zorgen voor een toename van verwerkingssnelheid ten koste van opslagcapaciteit (vgl. Croft 2003: 61). Wanneer taalgebruikers zo'n taalvorm willen produceren, kunnen ze de vorm immers rechtstreeks uit hun geheugen opdiepen zonder hem opnieuw te hoeven samenstellen. Dat bespaart hun enerzijds tijd en moeite, maar anderzijds neemt de vorm wel geheugencapaciteit in. Des te vaker een taalgebruiker de taalvorm in kwestie nodig heeft, des te groter de tijdsinstaat van zo'n versnelde verwerking, terwijl de benodigde geheugencapaciteit dezelfde blijft. Dit soort opslag is dus vooral nuttig voor frequente taalvormen.

Neem als voorbeeld de verledentijdsvormen *gaf* en *las*. Deze vormen zouden als kant-en-klare taalvorm opgeslagen kunnen worden in het geheugen van een taalgebruiker. Daardoor hoeft hij of zij niet steeds weer

de regel *ee* → *a* toe te passen op de stamvormen *geef* en *lees*, telkens hij of zij die verledentijdsvormen wil uiten (vgl. Pijpops, Beuls & Van de Velde 2015: 86-87). Dit type opslag lijkt vooral voor te komen bij de verleden tijd van hoogfrequente werkwoordsvormen, waar de tijdsinstelling het grootst is (De Smet & Van de Velde 2019: 158). Hetzelfde geldt overigens voor complexe spierbewegingen die mensen frequent gebruiken, zoals lopen of grijpen: ook zulke bewegingen worden vaker geautomatiseerd en als geheel opgeslagen (Gazzaniga, Ivry & Mangun 2009: Hfdst.7). Het is echter nog niet duidelijk vanaf welke frequentie de opslag als een volledige taalvorm nuttig wordt, en er lijken op dat vlak zelfs verschillen tussen taalgebruikers te bestaan (Dąbrowska 2006, 2014, 2018).

Op dezelfde manier zouden ook *wat bijzonders* en *iets speciaal* als volledige taalvormen in het geheugen opgeslagen kunnen worden. Dat betekent dat, telkens taalgebruikers *wat bijzonders* horen, deze hele vorm dieper in hun geheugen ingeslepen raakt. Als gevolg daarvan zou de vorm met *s*-uitgang van de frase *wat bijzonder(s)* bij Belgen die in contact staan met Nederlanders en omgekeerd, proportioneel dieper in het geheugen ingeslepen worden dan de vorm met *s*-uitgang van *iets speciaal(s)*. Deze taalgebruikers hebben *wat bijzonder(s)* immers proportioneel vaker met *-s* gehoord dan *iets speciaal(s)*.

Tot slot volgt uit bovenstaande paragraaf dat wanneer de Belgen of Nederlanders in dit voorbeeld zelf een partitieve genitief met het voornaamwoord *wat* en het adjectief *bijzonder* uitdrukken, de kans groter zal zijn dat ze de vorm met *-s* gebruiken dan wanneer ze een partitieve genitief met het voornaamwoord *iets* en het adjectief *speciaal* uitdrukken. Voor *wat bijzonder(s)* is de vorm met *-s* immers dieper in hun geheugen ingeslepen dan voor *iets speciaal(s)*. Nu kan men wel spreken over 'echte' lexicale voorkeuren. Het gebruik van de frase *wat bijzonder(s)* leidt immers rechtstreeks tot een voorkeur voor de vorm met *-s* bij de taalproducent. Meer algemeen gesteld zouden de woorden die typisch zijn voor één lect een voorkeur vertonen voor de morfosyntactische variant van datzelfde lect, onder taalgebruikers van één en hetzelfde lect. De lexicale voorkeuren zijn dus binnengedrongen in de lecten zelf, of met andere woorden; de lecten hebben elkaar gecontamineerd.

Eenmaal dit effect zich voordoet bij Belgen die wel eens Nederlands Nederlands gehoord hebben, kan het zich ook verspreiden naar andere Belgen die nog nooit een Nederlander horen praten hebben, maar wel blootgesteld zijn aan het taalgebruik van deze 'gecontamineerde' Belgen. Hetzelfde zou natuurlijk gelden voor Nederlanders die nooit een Belg hebben horen praten. Conditie 3 stelt dus niet dat de onderzochte taalgebruikers rechtsteeks blootgesteld moeten zijn aan het taalgebruik van de andere groep – blootstelling via tussenpersonen volstaat.

Merk op dat het niet noodzakelijk is voor lectale contaminatie dat taalgebruikers de frasen die typisch zijn voor een bepaald lect ook zelf aanvoelen als typisch voor dat lect. Het is met andere woorden enkel noodzakelijk dat bijvoorbeeld *wat bijzonder(s)* vaker gebruikt wordt in het Nederlandse Nederlands dan in het Belgische, maar de taalgebruikers zelf hoeven zich daar niet van bewust te zijn.

In Pijpops (te versch.) is dit mechanisme op twee manieren getest. Ten eerste werd een corpusonderzoek naar de partitieve genitief uitgevoerd op het Sonar-corpus (Oostdijk e.a. 2013) en het Corpus Gesproken Nederlands (CGN, Oostdijk e.a. 2002).² Van Sonar is enkel het materiaal gebruikt waarvan auteursinformatie beschikbaar is, terwijl van het CGN al het materiaal gebruikt is. Voor dat corpus is immers is voor al het materiaal sprekersinformatie beschikbaar. Hieruit bleek dat partitievegenitieffrasen die vaker door Nederlanders gebruikt werden, inderdaad door Belgen vaker van een *s*-uitgang voorzien werden. Hetzelfde bleek te gelden voor Nederlanders: ook zij gebruikten vaker de *-s* bij partitievegenitieffrasen die vaker door Nederlanders gebruikt werden, en minder vaak bij partitievegenitieffrasen die vaker door Belgen gebruikt werden.

Ten tweede is in Pijpops (te versch.) een agent-gebaseerde computersimulatie gebouwd waarin de vier bovengenoemde condities geïmplementeerd werden. In die simulatie bleek het effect van lectale contaminatie consistent te ontstaan. Wanneer echter aan één van de vier condities niet voldaan werd, verscheen het effect niet.

Wat betreft de alternantie tussen *zo'n* en *zulk* lijken de Condities 1, 2 en 3 vervuld. Voor Conditie 4 is dat minder zeker. Enerzijds zijn bij de partitieve genitief de individuele taalvormen zoals *wat bijzonders*, *iets speciaal*, *iets leuks*, enzovoort, vrij frequent, wat het nuttig maakt deze vormen volledig in het geheugen op te slaan. Het is vooralsnog niet duidelijk of dat ook het geval zal zijn voor de frasen waarin *zo'n* en *zulk* verschijnen. Bovendien bestonden er voor de partitieve genitief onafhankelijke aanwijzingen dat taalgebruikers tenminste een aantal partitievegenitiefvormen volledig opslaan in hun geheugen (zie Pijpops & Van de Velde 2016). Voor *zo'n* en *zulk* beschikken we niet over zulke aanwijzingen.

Anderzijds zijn er wel duidelijke overeenkomsten tussen de partitieve genitief en de *zo'n-zulke* alternantie. Zo betreffen ze beide variatie binnen de nominale constituent en zijn ze beide productief. Bij de partitievegenitiefconstructie is de voornaamwoordelijke kavel weliswaar min of meer beperkt tot een opsombaar aantal onbepaalde voornaamwoorden, zoals *iets*,

2 Beide corpora zijn intussen trouwens toegankelijk via het online platform OpenSonar.

niets, wat, veel, weinig, enzovoort, maar het adjectivisch slot kan door zowat elke adjectivische constituent ingevuld worden, inclusief eigen voor- en nabepalingen.³ En van die mogelijkheid maken taalgebruikers in de praktijk ook gretig gebruik, zoals (7)-(8) laten zien.

- (7) Ik kan het horen aan je stem. **Iets, instinctief beschermends.**
(Sonar-id: WR-P-E-G-0000011094.p.913.s.1)
- (8) Ik stel me voor dat ze zongen over **iets zo onvoorstelbaar moois ... dat je hart er pijn van doet.**
(Sonar-id: WR-P-E-G-0000011257.p.624.s.1)

Bovendien bestaat er een uitgebreide literatuur die aantoont dat taalgebruikers best in staat zijn woordcombinaties van vergelijkbare complexiteit als geheel in hun geheugen op te slaan (Bod 1995; Arnon & Snider 2010; Tremblay & Baayen 2010; Tremblay e.a. 2011; Dąbrowska 2014; Diessel 2019: 113-195, alsook de referenties die daarin aangehaald worden). Het is daarom aannemelijk – zij het niet zeker – dat Conditie 4 ook vervuld is voor de frasen met *zo'n* en *zulk*. Dat leidt tot de volgende hypothese.

Hypothese 1: Frasen die vaker door Belgen gebruikt worden, zullen vaker voorkomen met *zo'n* dan frasen die vaker door Nederlanders gebruikt worden, zowel in het taalgebruik van Belgen als in het taalgebruik van Nederlanders.

2.2 Hypothese 2: De functies van *zo'n* en *zulk*

In de literatuur over *zo'n* en *zulk* worden de onderstaande functies onderscheiden (o.a. Ghesquière & Van de Velde 2011; Van Olmen & Van der Auwera 2014; Van Olmen 2019).

1. Identificeringsfunctie: exofoor, bv. (9), katafoor, bv. (10), of anafoor, bv. (11)
 2. Intensiveringsfunctie, bv. (12)
 3. Herkenningsfunctie, bv. (13)
 4. Benaderingsfunctie, bv. (14)
- (9) a. N: Ik kwam even langs kantoor omdat ik iets was vergeten. B: In **zo'n** avondjurk?
(overgenomen uit Van Olmen 2019: 226; 1989, Annie M.G. Schmidt, *En ik dan?*)

3 Ik ga hier uit van de traditionele analyse van de partitievegenitiefconstructie. Zie Pijpops (2018: 103-106) voor een overzicht van de verschillende analyses.

- b. Kreeg een rekenregel om te zeggen van **zulke** theorieën mogen wel en **zulke** theorieën mogen niet.
(CGN-bestand: fn007583)
- (10) a. Nou was me vader van **zoo'n** natuur, dat as ie, met permissie, maar aan 't zweeten kommen kon, dan was ie weer klaar.
(overgenomen uit Van Olmen 2019: 226, 1839, Nicolaas Beets, *Camera Obscura*)
- b. Algebraïsche meetkunde neemt een centrale plaats in de moderne wiskunde in en heeft meerdere conceptuele verbindingen met **zulke** uiteenlopende gebieden als complexe analyse, topologie en getaltheorie.
(Ghesquière & Van de Velde 2011: 774, http://nl.wikipedia.org/wiki/Algebraïsche_meetkunde, bezocht in juli 2010)
- (11) a. De jas vind ik echt prachtig, **zo'n** jas stond nou net zelf ook op mijn wishlist.
(overgenomen uit Van Olmen 2019: 217, <http://mounirasmansion.com/herfst-look>, bezocht in oktober 2016)
- b. Ik zeg niet dat een schrijver nooit tweemaal of tienmaal eenzelfde formule mag hanteren, indrukwekkende oeuvres werden op **zulke** herhaling opgebouwd.
(overgenomen uit Ghesquière & Van de Velde 2011: 771, DBNL 1969)
- (12) a. Sinds drie weken heb ik een puppy. Het is **zo'n** schatje!
(overgenomen uit Van Olmen & Van der Auwera 2014: 216, <http://www.hoewatwaarom.nl/dier-en-natuur/welk-voedsel-mag-pupabsoluut-eten.html>, bezocht op 28 december 2013)
- b. We zijn allemaal **zulke** afschuwelijke hypocrieten.
(overgenomen uit Ghesquière & Van de Velde 2011: 768, DBNL 1969)
- (13) Ik was toen zwanger van wie dat weet ik niet meer ... Nou weet ik het ik had **zo'n** colbertjas aan en ik had een dik buik nou weet ik 't.
(overgenomen uit Van Olmen & Van der Auwera 2014: 219, CGN-bestand: fn000773)
- (14) **Zo'n** twintig kilometer fietsen daar efkens een terrasje doen en dan terugkomen.
(overgenomen uit Van Olmen & Van der Auwera 2014: 219, CGN-bestand: fv400292)

In de identificeringsfunctie drukt het voornaamwoord een gelijkenis uit tussen het hoofdstantief van de nominale constituent en een referent. Bij exofore identificering bevindt die referent in de buitentalige context, zoals in (9). Bij katafore identificering staat de referent in de talige context

achter het voornaamwoord, zoals in (10), en bij anafore identificering staat hij voor het voornaamwoord, zoals in (11). Daarnaast vertonen zowel *zo'n* als *zulk* ook een intensiveringsfunctie, zoals in (12). Hierbij wordt geen gelijkenis met een referent uitgedrukt, maar wordt een zekere evaluatieve of scalaire eigenschap van het substantief verhoogd of gemeten (Ghesquière & Van de Velde 2011: 782).

In de herkenningfunctie wordt *zo'n* gebruikt 'ter aanduiding van iets dat bij de toegesprokene als bekend verondersteld wordt', zoals in (13) (Den Boon & Geeraerts 2005, s.v. *zo'n*). Deze functie zou nooit of bijna nooit voorkomen bij *zulk* (Van Olmen & Van der Auwera 2014: 219).⁴ In tegenstelling tot de identificeringsfunctie wordt hier niet verwezen naar een referent uit de context, maar naar een type dat bij de gesprekspartner(s) wordt verondersteld bekend te zijn en waartoe het substantief behoort. Tot slot komt ook de benaderingsfunctie, zoals in (14), enkel voor bij *zo'n* (Van Olmen 2019: 217-218). In deze functie gaat *zo'n* vooraf aan een hoofdtelwoord en kan het omschreven worden als 'ongeveer'. De voorkomens van de herkenningfunctie en de benaderingsfunctie zullen uitgesloten worden van het onderzoek, omdat het gebruik van *zulk* in die voorkomens niet of nauwelijks zou voorkomen.

Van Olmen (2019: 218) stelt dat in Nederland, bij de enkelvoudige niet-telbare substantieven, *zo'n* gebruikt wordt in de intensiveringsfunctie, terwijl *zulk* verkozen wordt voor identificering (vgl. ook Haeseryn e.a. 1997: Sectie 5.6.6). Dat vormt Hypothese 2:

Hypothese 2: In de Nederlandse data zal er bij de enkelvoudige, niet-telbare substantieven een verschil zijn tussen de functies. In de identificeringsfunctie zal vaker gekozen worden voor *zulk*, terwijl in de intensiveringsfunctie vaker gekozen wordt voor *zo'n*.

2.3 Hypothese 3: De invloed van register

In België gaat in een formeler register vaak de voorkeur uit naar typisch Nederlandse varianten (zie o.a. Tummers, Speelman & Geeraerts 2005; Grondelaers, Speelman & Geeraerts 2008; Speelman & Geeraerts 2009). Dat is bijvoorbeeld te zien bij de partitieve genitief, waarbij in formelere registers vaker gebruik wordt gemaakt van de 'Nederlandse' variant met *-s* (Pijpops & Van de Velde 2018), of bij de keuze tussen de voornaamwoorden *ijj* en *gij*, tussen de verkleinsuffixen *-je* en *-ke*, enzovoort (zie het overzicht

4 Het is bij mijn weten nog niet onderzocht wanneer of waarom *zulk* wel gebruikt zou worden in de herkenningfunctie.

in Geeraerts & Van de Velde 2013). Het zou ook gelden voor *zo'n* en *zulk*, maar enkel in nominale constituenten met meervoudige substantieven (Taaladvies.net, bezocht op 15 september 2021). Bij enkelvoudige, niet-telbare substantieven zou *zo'n* wel als standaardtalig gelden in België. Dat leidt tot de laatste hypothese:

Hypothese 3: In de Belgische data zal er bij de meervoudige substantieven een verschil zijn tussen de registers. In een formeel register zal *zulk* vaker voorkomen, terwijl in een informeel register *zo'n* vaker zal opduiken.

Bij deze hypothese wordt voorlopig in het midden gelaten of dit effect ontstaat omdat een schrijver van een formele tekst in de eerste versie van die tekst onmiddellijk voor *zulk* kiest, of omdat in een latere redactieronde een oorspronkelijk *zo'n* wordt aangepast naar *zulk* – zij het door de schrijver zelf, zij het door een redacteur. Hoewel een groot deel van de data die gebruikt zal worden om deze hypothese te testen gesproken taalgebruik betreft, is enige redactie ook daar niet uit te sluiten, aangezien het deels gaat om voorbereide lezingen en voorgelezen teksten (zie hieronder, Van Eerten 2007: 196).

3 Data

Als bron van data is gebruik gemaakt van hetzelfde corpusmateriaal als in Pijpops (te versch.), namelijk het Corpus Gesproken Nederlands (CGN), waarin voor al het materiaal sprekersinformatie beschikbaar is, en het materiaal uit het Sonar corpus waarvan auteursinformatie beschikbaar is (Oostdijk e.a. 2002, 2013). Op basis van deze informatie is de indeling gemaakt in een Belgische en Nederlandse dataset. De Belgische dataset bevat dus uitsluitend taalmateriaal dat geproduceerd is door Belgen, en de Nederlandse data uitsluitend taalmateriaal dat geproduceerd is door Nederlanders.

Eerst is uit het Referentiebestand Nederlands een lijst gehaald van alle substantieven die in een niet-telbaar gebruik kunnen voorkomen (Van der Vliet 2007; Referentiebestand Nederlands – RBN 2014). Vervolgens is in het corpusmateriaal gezocht naar alle voorkomens van *zo'n*, *zo een*, *zulk een*, *zulk* en *zulk* waarop binnen de 5 volgende woorden een substantief volgde dat in de meervoudsvorm stond of dat voorkwam in de lijst met mogelijks niet-telbare substantieven. Voorkomens waarbij het enkelvoudige substantief in de verkleinvorm verscheen, zijn overgeslagen. In zulke gevallen, zoals (15), ging het immers steeds over een telbaar gebruik. Dit leverde een dataset op van 9206 voorkomens, die vervolgens manueel gecontroleerd zijn.

- (15) en een keer bloemkool uh met **zo'n** uh sausje
(CGN-bestand: fn006792)

Alle voorkomens met enkelvoudige substantieven die toch telbaar gebruikt werden, zoals (16), zijn uitgesloten, net als twee voorkomens waar beide varianten gebruikt zijn, namelijk (17). Voorts zijn een aantal dubbele voorkomens uit Sonar die duidelijk gevallen van kopiëring zijn, uitgesloten, en zo ook alle voorkomens in het CGN-extract in (18), waar metalinguïstische commentaar gegeven wordt, alsook de voorkomens van *zo'n* met *mogelijk* zoals (19), die geherformuleerd kunnen worden als *een zo X mogelijk*. Daarnaast zijn nog de voorkomens verwijderd waar *zo'n* of *zulk* zelfstandig gebruikt zijn, of alleszins niet als determinator bij een substantief voorkomen, zoals (20). Tot slot zijn de voorkomens uitgesloten waarbij het hoofds substantief meer dan vier woorden verwijderd is van *zo'n* of *zulk*, maar die vanwege een eerder substantief toch geëxtraheerd zijn, zoals (21).

- (16) weet nu toch dat het drie maanden kost om **zo'n** papier in handen te krijgen
(Sonar-id: WR-P-P-B-000000237.p.217.s.4)
- (17) a. merkwaardig dat uh dat dat dat **zo'n** uh **zo'n** **zo'n** Hans Simons bijvoorbeeld **zo'n zulke** etentjes niet zelf betaalt.
(CGN-bestand: fn000971)
- b. goed jongen ja*x dat kan ook wel maar ik maar z*a maar **zo'n zulke** mensen gaan vaak
(CGN-bestand: fn000971)
- (18) dat licht is wel heel mooi als je dat dan vergelijkt met uh met met met **zo'n** kippen ... dat is dan heel vies eigenlijk hè al die uh kuikens bij mekaar... **zo'n** kippengedoe en met **zulke** kippen niet met **zo'n** kippen. is niks met **zo'n** kippen. hoort bij enkelvoud... dat komt omdat ik de neiging... de hoe heet dat uh uh de aandrang heb ja de de de drang heb om uhm om verkeerde uitdrukkingen over te nemen en toe te passen. ggg. ja? ja... zit ik niet expres uh te bedenken of te doen... dat vind ik leuk **zo'n zo'n** kippen niet **zulke** kippen maar **zo'n** kippen maar andere m*a uh jij vind daar... . dat vind ik echt verschrikkelijk
(CGN-bestand: fn008515)
- (19) Heb ik altijd film willen maken voor **zo'n** groot mogelijk publiek.
(Sonar-id: WR-P-P-H-0000125566.p.69.s.2)
- (20) a. Waren net **zulke** als die de nieuwe tempeliers op hun mantels droegen
(Sonar-id: WR-P-P-B-000000198.p.2041.s.8)

- b. ook dat is een beetje bluf hè als derdejaars over beelden die ze op TV hebben gezien in het nieuws dan van van terechtstellingen (...) dat uh leerlingen op dit moment **zo een** een beter zicht hebben op wat er in de wereld gebeurt dan dan u vroeger
(CGN-bestand: fv400145)
- (21) P.N. Muller werd in de jaren tachtig een uitgesproken pleitbezorger van **zulk** door ondernemingszin en nationaal prestige gemotiveerd imperialisme.
(Sonar-id: WR-P-P-B-0000000370.p.1316.s.4)

De overige observaties zijn vervolgens zo consistent mogelijk manueel geannoteerd op hun functie. Hierbij is gebruik gemaakt van de criteria genoemd in Altenberg (1994: 234) en Van Olmen (2019: 225, 227, 232, 235), die hieronder opgesomd staan.

1. Als er zich een referent in de context bevindt, zoals *de jas* in (11a) of *tweemaal oftienmaal eenzelfde formule mag hanteren* in (11b): identificeringsfunctie. Afhankelijk van waar die referent zich bevindt, wordt vervolgens een onderscheid gemaakt tussen exofore, katafore en anafore identificering.
 2. Als er zich geen referent in de context bevindt, en er in de naamwoordelijke constituent een concept wordt uitgedrukt dat gradeerbaar is, d.w.z. dat in verscheidene groottes of schakeringen kan voorkomen, zoals *schatje* in (12a) en *afschuwelijke* in (12b): intensiveringsfunctie.
 3. Als er zich geen referent in de context bevindt, maar er in de context elementen staan die op herkenning duiden, zoals *ja ja ja* in (22), of als *zo'n* of *zulk* weggelaten kan worden of vervangen kan worden zonder een groot betekenisverschil te veroorzaken, zoals in (13): herkenningsfunctie.
 4. Als er op het voornaamwoord een hoofdtelwoord volgt, en het geparafraseerd kan worden als 'ongeveer', zoals in (14): benaderingsfunctie.
- (22) denk dat 't die mensen zijn uh ik ken ze van vroeger zuh. den*d Bahwah*x Bahwan*x of zoiets. ja ja ja. **zo een** indische ja een religieuze groeperingen denk 'k. ja ja. ja (CGN-bestand: fv701052)

Er is gekozen om geen aparte categorie op te nemen voor de ambigue gevallen, om twee redenen. Ten eerste vermelden zowel Ghesquière & Van de Velde (2011: 786) als Van Olmen (2019: 225, 243) dat het aantal ambigue gevallen erg klein is. Ten tweede verschuift *zo'n* aparte categorie het probleem alleen maar. Ook in dat geval zullen er immers twijfelgevallen zijn die eerder een bepaalde functie lijken te hebben, maar waarbij een alternatieve interpretatie

ook niet categorisch uitgesloten is, zodat ze zich op de scheidingslijn bevinden tussen de ambigue en de niet-ambigue gevallen.

Ten slotte zijn de observaties in de herkenningsfunctie, zoals (22), en de benaderingsfunctie, zoals (23), verwijderd uit de dataset. Bij die eerste groep bevonden zich ook een aantal voorkomens van *zulk*, zoals (24). Tijdens deze annotatie zijn ook alle waarden van de variabelen die in de analyse opgenomen zullen worden (zie Sectie 5), manueel gecontroleerd en zo nodig gecorrigeerd.

- (23) dus er zijn toch wel zo veel gebruikers van dat pakket ja . da's wel lekker . hebben al zo'n uh poeh twintig vijftwintig gemeentes.
(CGN-bestand: fn008213)
- (24) ... ja ik heb me nooit verveeld. maar hier heb de*d veel*z meer vrijigheid. dat wel hoor. **zulke** dorpjes vroeger en zo... vandaag ja nog wel. genoeg. want nou heb de*d toch veel*z meer vrijigheid
(CGN-bestand: fn000973)

4 Analyse

De overgehouden observaties werden opgedeeld in een Belgische en een Nederlandse dataset. Voor de analyse is gebruik gemaakt van de onderstaande variabelen. De eerste variabele is de afhankelijke variabele, de overige zijn onafhankelijke variabelen. BELGISCHHEID FRASE EN LENGTE CONSTITUENT zijn numerieke variabelen, de overige categoriaal.

1. VARIANT: *zulk*, *zo'n*
2. BELGISCHHEID FRASE: zie Vergelijking 1
3. REGISTER: formeel, informeel
4. CORPUS: sonar, cgn
5. FUNCTIE: intensivering, exofoor, katafoor, anafoor
6. TYPE SUBSTANTIEF: meervoudig, enkelvoudig niet-telbaar
7. COMPLEXITEIT FRASE: natuurlijk logaritme van het aantal woorden van de naamwoordelijke constituent
8. FRASE: aparte waarde voor elke unieke lexicale invulling van de nominale constituent, bv. *zo'n/zulke* belangrijke mensen, *zo'n/zulke* mensen, *zo'n/zulk* gedrag.

De variabele BELGISCHHEID FRASE wordt gebruikt om Hypothese 1 te testen, en kwantificeert hoe 'Belgisch' elke frase is. Om *zo'n* maatstaf betrouwbaar te laten zijn, moet je beschikken over voldoende voorkomens per frase. Daarom is ervoor gekozen in dit onderzoek BELGISCHHEID FRASE te berekenen op

basis van het aantal voorkomens van een frase in de Sonar-component WRPPG, d.i. het krantenmateriaal. WRPPG is de grootste component van Sonar en bevat ook het meeste materiaal uit België en Nederland. De overige componenten zijn niet gebruikt omdat de verhouding tussen het Belgische en het Nederlandse materiaal niet consistent is doorheen het corpus. Voor het materiaal in WRPPG is geen auteursinformatie beschikbaar. De indeling tussen Belgisch en Nederlands materiaal is dan ook gebeurd op basis van de origine van de tekst en is dus mogelijk minder zuiver, maar voor de berekening van BELGISCHHEID FRASE volstaat het wanneer het meeste Belgische materiaal inderdaad afkomstig is van Belgen en het meeste Nederlandse materiaal van Nederlanders.

Zoals te zien in Vergelijking 1 wordt BELGISCHHEID FRASE voor de frase f berekend als het logaritme met grondtal 10 van de deling van de kansverhouding van frase f in het Belgische WRPPG door de kansverhouding van frase f in het Nederlandse WRPPG.⁵ Om deling door 0 te voorkomen, wordt 1 opgeteld bij de Nederlandse frequentie, en hetzelfde gebeurt bij de Belgische frequentie om het evenwicht te bewaren. De resulterende waarde is hoger naarmate de frase vaker voorkomt in Belgisch taalgebruik en lager naarmate het vaker voorkomt in Nederlands taalgebruik.

BELGISCHHEID FRASE _{f} =

$$\log \left(\frac{(\text{freq. van } f \text{ in het Belg. WRPPG} + 1) / (\text{grootte van het Belg. WRPPG} - (\text{freq. van } f \text{ in het Belg. WRPPG} + 1))}{(\text{freq. van } f \text{ in het Ned. WRPPG} + 1) / (\text{grootte van het Ned. WRPPG} - (\text{freq. van } f \text{ in het Ned. WRPPG} + 1))} \right)$$

Formule 1: berekening van BELGISCHHEID FRASE.

Tabel 1 en Tabel 2 geven een overzicht van de 10 frasen die respectievelijk het laagst en het hoogst scoren voor deze maatstaf. Er zijn natuurlijk tal van andere manieren om te berekenen hoe typisch een bepaalde woordcombinatie is voor een subcorpus of een bepaalde (talige) context (zie o.a. Stefanowitsch & Gries 2003; Stefanowitsch & Flach 2016: 115-116; Spielman e.a. 2020). Er is voor deze maatstaf gekozen vanwege zijn eenvoud en om zo consistent mogelijk te blijven met Pijpops (te versch.).

⁵ Deze maatstaf is anders berekend in Pijpops (te versch.). Daar werd het logaritme berekend van de deling van de kansverhouding van een frase in de Belgische dataset door de kansverhouding van de frase in de Nederlandse dataset. Dat is haalbaar als je beschikt over in totaal 9984 observaties van 173 unieke frasen, zoals Pijpops (te versch.). Het huidige onderzoek beschikt echter slechts over 5827 observaties van 3120 unieke frasen. De tellingen voor elke unieke frase zouden dus erg laag zijn, wat zo'n maatstaf minder betrouwbaar maakt.

Tabel 1: 10 meest 'Nederlandse' frasen, oftewel 10 frasen met de laagste waarde voor BELGISCHHEID FRASE, geordend in stijgende volgorde.

Frase	Freq. in het Belg. WRPPG	Freq. in het Ned. WRPPG	BELGISCHHEID FRASE
zo'n/zulke geluiden	1	13	-1,25
zo'n/zulke grote aantallen	0	6	-1,25
zo'n/zulke hoge cijfers	0	6	-1,25
zo'n/zulk ongemak	0	6	-1,25
zo'n/zulke aantallen	1	11	-1,19
zo'n/zulke regelingen	0	5	-1,19
zo'n/zulke overwegingen	1	10	-1,15
zo'n/zulke bezwaren	0	4	-1,11
zo'n/zulke diversiteit	0	4	-1,11
zo'n/zulke gemeenschappen	0	4	-1,11

Tabel 2: 10 meest 'Belgische' frasen, oftewel 10 frasen met de hoogste waarde voor BELGISCHHEID FRASE, geordend in aflopende volgorde.

Frase	Freq. in het Belg. WRPPG	Freq. in het Ned. WRPPG	BELGISCHHEID FRASE
zo'n/zulke matchen	55	0	1,34
zo'n/zulke toestanden	46	0	1,26
zo'n/zulke drama's	16	0	0,82
zo'n/zulke kansen	15	0	0,80
zo'n/zulke borden	14	0	0,77
zo'n/zulke fases	13	0	0,74
zo'n/zulke dossiers	13	0	0,74
zo'n/zulke personen	12	0	0,70
zo'n/zulke wedstrijden	123	10	0,64
zo'n/zulke spelers	49	4	0,59

De variabele REGISTER maakt een onderscheid tussen een formeel en een informeel register. Voor het CGN is gebruik gemaakt van de indeling voorgesteld in Plevoets (2008: 80). Dit houdt in dat de spontane conversaties en de telefoondialogen beschouwd worden als een informeel register, en de overige componenten als een formeel register. Voor Sonar wordt een onderscheid gemaakt tussen de componenten die eerder kort, niet geredigeerd taalmateriaal bevatten, en diegene die eerder langere, wel geredigeerde teksten bevatten. Concreet worden de internetfora, de tweets, de chatgesprekken en de SMS-berichten als een informeel register beschouwd, en de websites, boeken, tijdschriften en magazines, rapporten en geschreven opdrachten

als een formeel register. Geen van de overige Sonar-componenten bevat materiaal waarvan auteursinformatie beschikbaar is.

De variabele CORPUS onderscheidt de data uit het CGN van de data uit Sonar, en de variabele FUNCTIE is hierboven reeds geïntroduceerd. TYPE SUBSTANTIEF maakt een onderscheid tussen de voorkomens met een enkelvoudig, niet-telbaar substantief en de voorkomens met een meervoudig substantief. COMPLEXITEIT FRASE is overgenomen uit Pijpops e.a. (2018: 524) en wordt berekend als het natuurlijk logaritme van het aantal woorden van de frase, d.i. van de naamwoordelijke constituent waartoe *zo'n* of *zulk* behoort. Zowel *zulk(e)* als *zo'n*, *zo een* en *zulk een* worden hierbij geteld als slechts één woord.

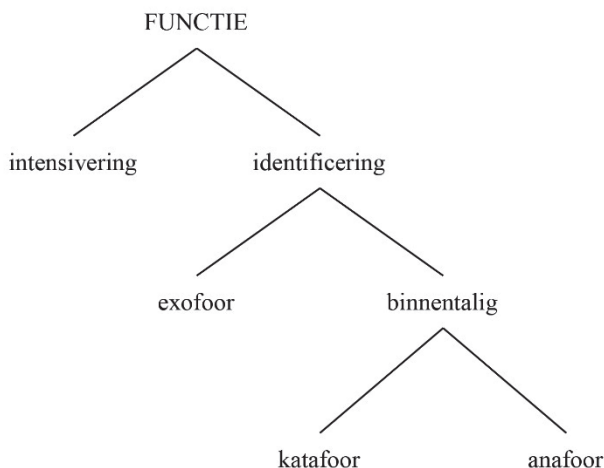
Deze laatste variabele dient als controlevariabele bij Hypothese 2. De intensiveringsfunctie zou immers vaker voorkomen bij frasen met adjectieven (Van Olmen 2019: 229), en zulke frasen zijn natuurlijk langer dan frasen zonder adjectieven. Een eventueel verschil tussen de voorkomens met intensiveringsfunctie en identificeringsfunctie zou dus te wijten kunnen zijn aan een onderliggend verschil in COMPLEXITEIT FRASE, als daarvoor niet gecontroleerd wordt.

Tot slot heeft de variabele FRASE een aparte waarde voor elke unieke frase, d.i. elke unieke lexicale invulling van de naamwoordelijke constituent. Omdat de afhankelijke variabele VARIANT binair is, zijn beide datasets geanalyseerd aan de hand van gemengde logistische regressiemodellering. De variant *zo'n* is de succeswaarde.

De analyse van de Nederlandse dataset komt eerst aan bod. Deze dataset bevat 1115 voorkomens van de minst frequente variant, namelijk *zo'n*. Bij regressiemodellering geldt als vuistregel niet meer parameters op te nemen dan dit getal gedeeld door 20 (Speelman 2014: 530). Dat betekent dat er maximaal 55 parameters opgenomen kunnen worden, wat ruim voldoende is voor de voorziene analyse. De numerische variabelen BELGISCHHEID FRASE en COMPLEXITEIT FRASE kunnen rechtstreeks als parameters in het model opgenomen worden. Omdat deze variabelen een identieke waarde hebben voor alle observaties van dezelfde frase, wordt ook de variabele FRASE aan het model toegevoegd als *random effect* met *random intercepts* (Speelman, Heylen & Geeraerts 2018: 2).

Voor de implementatie van de categoriale variabelen is gebruik gemaakt van gebruikersgedefinieerde contrastcodering (Heller 2018: 85-88). Met dit type codering is het mogelijk voor de onderzoeker om zelf theoretisch gemotiveerde vergelijkingspunten te kiezen die in het regressiemodel geïmplementeerd worden. De mogelijkheid zelf de vergelijkingspunten te kiezen is nuttig voor de variabele FUNCTIE, die geïmplementeerd is volgens het schema in Figuur 1. Eerst wordt de intensiveringsfunctie afgezet tegen

de overige drie waarden, die elk een vorm van identificering betreffen. Vervolgens wordt de buitentalige identificering afgezet tegen de binnentalige identificering, en tot slot de katafoor tegen de anafoor. De overige categoriale variabelen hebben slechts twee waarden, dus daarbij kunnen beide waarden sowieso enkel tegen elkaar afgezet worden.



Figuur 1: Implementatieschema voor de variabele FUNCTIE.

Hypothese 1 voorspelt een significant positief effect van BELGISCHHEID FRASE op het gebruik van *zo'n*. Hypothese 2 voorziet dat de invloed van de variabele FUNCTIE afhankelijk is van de waarde van de variabele TYPE SUBSTANTIEF: ze stelt dat er bij de enkelvoudige, niet-telbare substantieven een verschil zal zijn tussen de identificeringsfunctie en de intensiveringsfunctie, waarbij de observaties in de identificeringsfunctie een voorkeur hebben voor *zulk* en de observaties in de intensiveringsfunctie een voorkeur voor *zo'n*. Bij de meervoudige substantieven wordt geen verschil verwacht. Hiermee kan het model rekening houden door een interactie in te bouwen tussen de variabelen FUNCTIE en TYPE SUBSTANTIEF. Dit model is vervolgens op de Nederlandse dataset toegepast.

Om het te model te laten convergeren, bleek het nodig om alle frasen die slechts éénmaal in de dataset voorkwamen, samen te nemen in een restcategorie voor het random effect FRASE (Wolk e.a. 2013: 399). Multicol-lineariteit lijkt geen probleem: de variantie-inflatiefactoren van het model zijn alle lager dan 5 (Levshina 2015: 160).⁶ Het model heeft een uitstekende

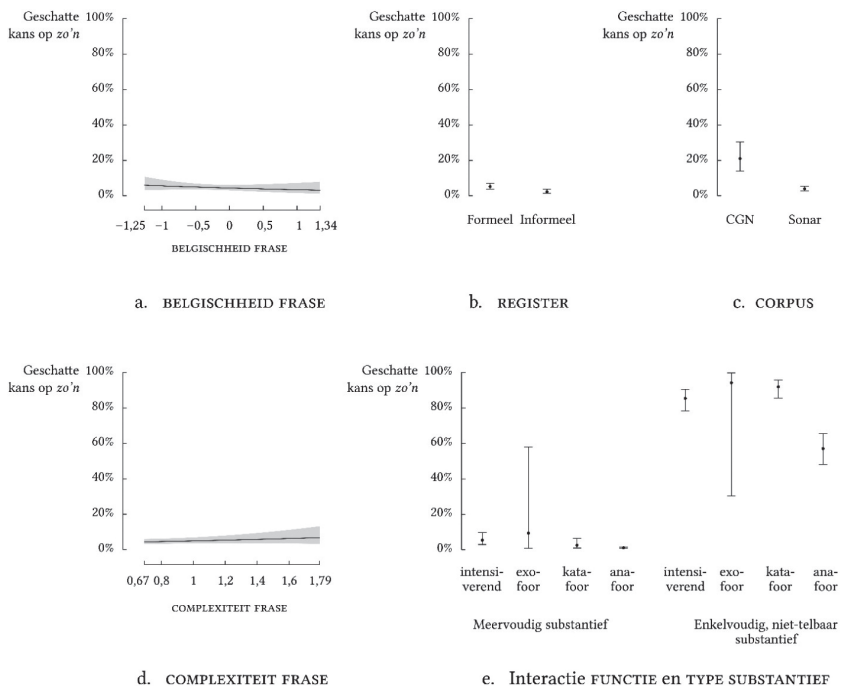
6 Deze variantie-inflatie factoren zijn berekend op hetzelfde model, zonder de interactie tussen FUNCTIE en TYPE SUBSTANTIEF.

voorspellende kwaliteit, met een C-index van 0,974 (Hosmer & Lemeshow 2000: 162). Tabel 3 bevat de specificaties van het Nederlandse model, en de effectplots zijn te vinden in Figuur 2.

Tabel 3: Specificaties van het Nederlandse model.

AIC: 2057,5 Aantal voorkomens van *zulk*: 4415
 C-index: 0,974 Aantal voorkomens van *zo'n* (succeswaarde): 1115

Fixed effects	Waarde	Schatting coëfficiënt	Standaard- fout	Z-waarde	P-waarde
	intercept	-0.74	0.43	-1.70	0.0900
BELGISCHHEID FRASE		-0.28	0.29	-0.98	0.3295
REGISTER	formeel (vs. informeel)	0.43	0.12	3.60	0.0003
CORPUS	cgn (vs. Sonar)	0.93	0.12	7.76	< 0.0001
FUNCTIE	intensivering (vs. identificering)	0.20	0.31	0.64	0.5207
	exofoor (vs. binnentalig)	1.08	0.76	1.42	0.1567
	katafoor (vs. anafoor)	0.75	0.14	5.20	< 0.0001
TYPE SUBSTANTIEF	meervoudig (vs. enkelvoudig niet-telbaar)	-2.56	0.30	-8.58	< 0.0001
COMPLEXITEIT FRASE		0.42	0.34	1.25	0.2131
Interactie FUNCTIE en TYPE SUBSTANTIEF	intensivering & meervoudig	0.25	0.31	0.81	0.4182
	exofoor & meervoudig	0.13	0.76	0.17	0.8680
	katafoor & meervoudig	-0.32	0.14	-2.23	0.0259
Random effect	Aantal waarden	Variantie	Standaard- afwijking		
FRASE	540	2.77	1.67		



Figuur 2: Effectplots van het Nederlandse model.

Nu volgt de analyse van de Belgische dataset. Deze dataset bevat slechts 88 voorkomens van de minst frequente variant, nl. *zulk*. De Belgische dataset is immers kleiner, omdat Sonar meer Nederlandse data bevat waarvan auteursinformatie beschikbaar is dan Belgische data. Overeenkomstig met de hierboven vermelde vuistregel, kunnen er dan ook niet meer dan 4 parameters opgenomen worden in het Belgische model, wat nipt volstaat om beide hypotheses te testen die betrekking hebben op de Belgische data (Speelman 2014: 530).

Hypothese 1 voorziet een significant positief effect van BELGISCHHEID FRASE, net als in de Nederlandse data. Hypothese 3 houdt in dat het effect van REGISTER afhankelijk is van de variabele TYPE SUBSTANTIEF. Bij de meervoudige substantieven is te verwachten dat *zo'n* vaker gebruikt wordt in informele registers dan in formele registers, terwijl er bij de niet-telbare enkelvoudige substantieven geen verschil tussen de registers zou zijn.

Om deze hypotheses te testen, worden de variabelen BELGISCHHEID FRASE, REGISTER en TYPE SUBSTANTIEF in het model opgenomen, alsook een interactie tussen REGISTER en TYPE SUBSTANTIEF. Er is nog geprobeerd de variabele FRASE aan het model toe te voegen als *random effect* met *random*

intercepts, maar dit leidde tot een *singular fit*, zodat deze variabele terug verwijderd diende te worden (Barr e.a. 2013; Bates e.a. 2018). Het Belgische model is dus niet gemengd, maar bevat enkel *fixed effects*. Toch heeft het nog een acceptabele voorspellende kwaliteit, met een C-index van 0,763 (Hosmer & Lemeshow 2000: 162). Alle variantie-inflatiefactoren zijn ook lager dan 5 (Levshina 2015: 180).⁷ Tabel 4 toont de specificaties van het model, terwijl Figuur 3 de effectplots weergeeft.

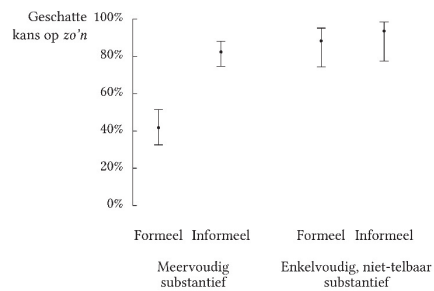
Tabel 4: Specificaties van het Belgische model

AIC: 309,21 Aantal voorkomens van *zulk*: 88
C-index: 0,763 Aantal voorkomens van *zo'n* (succeswaarde): 209

Fixed effects	Waarde	Schatting coëfficiënt	Standaard- fout	Z- waarde	P- waarde
	intercept	1,54	0,25	6,13	< 0,0001
BELGISCHHEID FRASE		0,42	0,47	0,90	0,3663
REGISTER	formeel (vs. informeel)	-0,63	0,23	-2,72	0,0064
TYPE SUBSTANTIEF	meervoudig (vs. enkelvoudig niet-telbaar)	-0,87	0,24	-3,71	0,0002
Interactie REGISTER en TYPE SUBSTANTIEF	formeel & meervoudig	-0,30	0,23	-1,31	0,1910



a. BELGISCHHEID FRASE



b. Interactie REGISTER en TYPE SUBSTANTIEF

Figuur 3: Effectplots van het Belgische model

⁷ Opnieuw zijn deze variantie-inflatiefactoren berekend op hetzelfde model, zonder de interactie.

5 Interpretatie

Dit onderzoek had tot doel drie hypothesen te testen. In de volgende subsecties komen deze hypothesen opnieuw aan bod.

5.1 Hypothese 1: lectale contaminatie

Hypothese 1: Frasen die vaker door Belgen gebruikt worden zullen vaker voorkomen met *zo'n* dan frasen die vaker door Nederlanders gebruikt worden, zowel in het taalgebruik van Belgen als in het taalgebruik van Nederlanders.

Deze hypothese is niet bevestigd. De variabele BELGISCHHEID FRASE heeft in het Nederlandse model noch in het Belgische model een significant positief effect op het gebruik van *zo'n*. In principe kan dat drie dingen betekenen.

Ten eerste is het natuurlijk mogelijk dat het falen van de voorspelling van Hypothese 1 eenvoudigweg ligt aan een tekortkoming van het onderzoek. Eén mogelijke tekortkoming is een gebrek aan data. Met name het Belgische model is op erg weinig data gebaseerd; het is niet ondenkbaar dat het positieve effect van BELGISCHHEID FRASE wel significant wordt in België wanneer er een grotere dataset verzameld kan worden. Een andere mogelijke tekortkoming is dat de berekening van BELGISCHHEID FRASE niet de optimale operationalisatie vormt van hoe 'typisch Belgisch' of 'typisch Nederlands' een bepaalde woordcombinatie is. Zoals reeds vermeld in Sectie 4, zijn er tal van andere operationalisaties van dat begrip mogelijk. Mogelijk duiden de belgischheidscores in Tabellen 1-2 eerder op een verschil in de behandelde onderwerpen dan op een verschil tussen beide variëteiten. Dat is echter een fundamentele moeilijkheid bij het berekenen van zulke scores, waar geen eenvoudige oplossing voor bestaat. Het is natuurlijk mogelijk andere manieren om BELGISCHHEID FRASE te berekenen uit te testen op de data, maar daarmee zou ik me statistisch op glad ijs begeven. Wie verschillende operationalisaties op dezelfde data blijft uittesten, vindt er wel uiteindelijk eentje die een significant effect oplevert.

Ten tweede is het mogelijk dat de *zo'n-zulk*-alternantie wel degelijk voldoet aan de vier condities genoemd in Sectie 2, maar dat er nog een ander mechanisme in het spel is dat het ontstaan van lectale contaminatie afblokt. Op dit moment heb ik echter geen weet van een mechanisme dat *zo'n* effect zou kunnen hebben. Ten derde is het mogelijk dat één of meerdere van de vier condities geschonden zijn, of onvoldoende vervuld zijn om een effect te veroorzaken dat duidelijk meetbaar is met de huidige methode.

Conditie 2 en 3 zijn stevig onderbouwd, maar Conditie 1 of 4 zijn mogelijk in twijfel te trekken.

Wat Conditie 1 betreft, kan erop gewezen worden dat *zo'n* wel degelijk vaak voorkomt in Nederland, zeker als ook de voorkomens voor enkelvoudige, telbare substantieven meegeteld worden. Van Olmen & Van der Auwera (2014: 219) stellen zelfs dat het gebruik van *zo'n* in de herkenningfunctie bijzonder frequent is in Nederland. Toch lijkt *zo'n* over het algemeen dominanter in het Belgische Nederlands dan in het Nederlandse, ook als we eenvoudigweg alle gevallen in rekening brengen. In de Belgische WRPPG-component komen *zo'n* en *zo een* in totaal 57.574 maal voor, tegenover *zulk* en *zulke* 5434 maal, wat een verdeling oplevert van 91,4% vs. 8,6%. In de Nederlandse WRPPG-component duiken *zo'n* en *zo een* daarentegen slechts 30.565 maal op, tegenover *zulk* en *zulke* 4288 maal, wat overeenkomt met een verdeling van 87,7% vs. 12,3%.⁸ Bovendien blijkt uit de analyses in Sectie 4 dat *zo'n* in België wel uitgesproken dominanter is dan in Nederland in talige contexten waar beide voornaamwoorden geregeld door elkaar voorkomen, met name voor meervoudige en niet-telbare enkelvoudige substantieven in de identificerings- en intensiveringsfunctie. Er is dus een proportioneel lectaal verschil, en in principe is Conditie 1 vervuld – al is het lectaal verschil misschien te klein om ook een meetbaar effect van lectale contaminatie te veroorzaken.

Als Conditie 4 niet vervuld is, betekent dat dat taalgebruikers frasen met *zo'n* en *zulk* niet als volledige taalvormen in hun geheugen opslaan, of dat alleszins niet voldoende vaak doen om een meetbaar effect van lectale contaminatie teweeg te brengen. Dat is best mogelijk. Zoals vermeld in Sectie 2, is opslag als gehele taalvorm vooral nuttig bij frequente vormen, maar is het niet zo duidelijk vanaf welke frequentie dat type opslag precies nuttig wordt. Misschien zijn nominale constituenten met *zo'n* en *zulk* ze eenvoudigweg niet frequent genoeg, of vertonen ze te veel lexicale verscheidenheid om kant-en-klare opslag op grote schaal nuttig te maken. De tellingen in Tabellen 1-2 tonen immers al dat de genoemde frasen niet bijzonder frequent zijn.

8 In het CGN liggen de verdelingen dicht bij elkaar. Hier komen *zo'n* en *zo een* in het Belgische materiaal 3116 maal voor, tegenover *zulk* en *zulke* 100 maal, oftewel een verdeling van 96,9% vs. 3,1%. In het Nederlandse materiaal komen *zo'n* en *zo een* 8105 maal voor, tegenover *zulk* en *zulke* 416 maal, oftewel een verdeling van 95,1% vs. 4,9%. In het Belgische materiaal is *zo'n* dus nog steeds dominanter dan in het Nederlandse, al is het verschil hier klein. Mogelijks is de sterkere positie van *zo'n* in het Nederlandse CGN ten opzichte van het krantenmateriaal in de Nederlandse WRPPG-component van Sonar te wijten aan de voorkomens van *zo'n* in de herkenningfunctie. Deze functie komt immers vaak voor in Nederland (Van Olmen & Van der Auwera 2014, p. 219), en is wellicht frequenter in gesproken materiaal. Ik dank een anonieme beoordelaar om me hierop te wijzen.

5.2 Hypothese 2: De functies van zo'n en zulk

Hypothese 2: In de Nederlandse data zal er bij de enkelvoudige, niet-telbare substantieven een verschil zijn tussen de functies. In de identificeringsfunctie zal vaker gekozen worden voor *zulk*, terwijl in de intensiveringsfunctie vaker gekozen wordt voor *zo'n*.

Deze hypothese lijkt op het eerste gezicht ook niet bevestigd. De parameter die in het Nederlandse model een onderscheid maakt tussen de intensiverende en identificerende voorkomens vertoont geen significant effect. Het onderscheid tussen katafore en anafore identificering is echter wel significant, en hier zien we inderdaad, zoals voorspeld, een sterk verschil bij de enkelvoudige, niet-telbare substantieven, maar niet bij de meervoudige substantieven.

Wat hier aan de hand is, wordt duidelijker wanneer we rechtstreeks naar de data kijken. Een deel van de katafore voorkomens zijn observaties zoals (25). Op basis van de criteria in Van Olmen (2019: 225–226) zijn deze voorkomens gemarkeerd als katafoor – er bevindt zich in de talige context achter het voornaamwoord *zo'n* of *zulk* immers een referent – maar de grens met intensivering is hier flinterdun. Het is best mogelijk dat de meeste taalgebruikers zulke voorkomens eerder behandelen als intensiverend dan als identificerend, of dat het verschil tussen beide functies in zulke voorkomens eenvoudigweg niet bestaat voor taalgebruikers.

In dat geval gedragen enkel de exofore voorkomens, zoals in (26), zich onverwacht. Ook hier is de grens met de intensiveringsfunctie vaak vaag, zoals in (26a), of met de herkenningfunctie, zoals in (26b). Bovendien komen de exoforen slechts 13 maal voor in de Nederlandse dataset, wat te weinig is om er conclusies op te baseren.

- (25) a. Had in geen jaren iemand ontmoet met **zo'n** boeiende geest als Bozo.
(Sonar-id: WR-P-P-B-000000153.p.1765.s.3)
- a. Boeren protesteren tegen het landbouwbeleid van minister Laurens-Jan Brinkhorst. Wil de veemarkten **zulke** strenge regels opleggen dat sluiting volgens de veeboeren onvermijdelijk is.
(CGN-bestand: fn004335)
- (26) a. 'Ik moet ervandoor.' verroerde zich niet, waardoor ik moeilijk zomaar weg kon lopen. Waarom **zo'n** haast ? 'Geen haast ... je weet wel , ik moet weer eens op huis aan.'
(Sonar-id: WR-P-P-B-000000225.p.179.s.2)

- b. Dat is voor de ramen te zemen hè. 't Is ontzettend nou ja meestal hebben ze **zulke** smalle dingetjes maar dit is vrij breed.
(CGN-bestand: fn000769)

Met die kwalificaties kan Hypothese 2 wel als bevestigd beschouwd worden. Figuur 1e toont immers een groot onderscheid tussen de intensiverende voorkomens en de anafore identificerende voorkomens bij de enkelvoudige, niet-telbare substantieven, terwijl er geen sprake lijkt te zijn van zo'n onderscheid bij de meervoudige substantieven. Onder de enkelvoudige, niet-telbare substantieven blijft *zo'n* echter wel steeds de meest gebruikelijke vorm, ook bij de anaforen. In Nederland lijkt het onderscheid tussen de enkelvoudige en meervoudige substantieven veruit het belangrijkste verschil is tussen *zo'n* en *zulk*, zoals te zien in Figuur 2 en aan de absolute waarde van de coëfficiëntschatting van TYPE SUBSTANTIEF in Tabel 3, die hoog is in vergelijking met de overige categoriale predictoren.

5.3 Hypothese 3: De invloed van register

Hypothese 3: In de Belgische data zal er bij de meervoudige substantieven een verschil zijn tussen de registers. In een formeel register zal *zulk* vaker voorkomen, terwijl in een informeel register *zo'n* vaker zal opduiken.

Deze hypothese is wel duidelijk bevestigd. In het Belgische model zien we vooral een sterk effect van REGISTER bij de meervoudige substantieven, waar *zulk* vaker opduikt in een formeel register. Bij de enkelvoudige, niet-telbare substantieven speelt REGISTER veel minder een rol. In België lijkt *zo'n* dominant te zijn, en houdt *zulk* enkel nog stand in de formele standaardtaal voor meervoudige substantieven.

6 Conclusies

Er zijn in dit onderzoek geen aanwijzingen gevonden dat lectale contaminatie zich voordoet bij de alternantie tussen *zo'n* en *zulk*. Een mogelijke verklaring voor dat resultaat is dat taalgebruikers te weinig, misschien zelfs nooit, gebruik maken van de verwerkingsstrategie om volledige frasen met deze voornaamwoorden in hun geheugen op te slaan. Een denkbare oorzaak daarvan is dat deze frasen niet frequent genoeg zijn of lexicaal te divers. Ik vermoed dat die beperking vaker zal gelden bij gevallen van syntactische

variatie dan bij morfologische variatie, zodat lectale contaminatie zich eerder zal voordoen bij morfologische variatie.

Wat de keuze tussen *zo'n* en *zulk* betreft, lijkt het onderscheid tussen de enkelvoudige en de meervoudige substantieven de belangrijkste tweedeling in de formele Nederlandse standaardtaal, zowel in België als in Nederland. Bij enkelvoudige substantieven is *zo'n* dominant, terwijl bij meervoudige substantieven eerder voor *zulk* wordt gekozen. De oorzaak van deze tweedeling ligt voor de hand: het voornaamwoord *zo'n* heeft zich ontwikkeld uit een samentrekking van het bijwoord *zo* en het enkelvoudige lidwoord *een*, en kan dus moeilijker voor meervoudige substantieven geplaatst worden (Van der Horst 2008: 167, 1102, 1382, 1667). *Zulk* heeft zich daarentegen ontwikkeld uit een adjectief, zodat plaatsing voor meervoudige substantieven geen probleem vormt (Van de Velde 2009: 267-271).

In de competitie tussen *zo'n* en *zulk* zou *zo'n* aan de winnende hand zijn (Ghesquière & Van de Velde 2011: 789). In Nederland houdt *zulk* nog min of meer stand voor enkelvoudige, niet-telbare substantieven bij anafore identificering, al is *zo'n* ook hier de meest gebruikelijke vorm. Mogelijk gaat het slechts om een achterhoedegevecht, en volgt het Nederlandse Nederlands binnenkort de ontwikkeling van het Belgische Nederlands, waar *zulk* nog zwakker staat.

Er zijn nog verschillende open vragen die behandeld kunnen worden in toekomstig onderzoek. Ten eerste zou het nuttig zijn om de verschillende mogelijke predictoren van lectale contaminatie, zoals BELGISCHHEID FRASE in het huidige onderzoek, met elkaar te vergelijken om te onderzoeken welke het meest geschikt is om lectale contaminatie te voorspellen. Dat gebeurt het best aan de hand van een theoretische simulatie. De vraag is immers welke maatstaf vanuit theoretisch oogpunt het meest accuraat de graad aanduidt waarin een lexicaal item geaffecteerd zou zijn door lectale contaminatie, niet welke maatstaf de laagste p-waarde oplevert in deze of gene gevalstudie. Die laatste vraag is immers oninteressant vanuit theoretisch oogpunt. Een nuttige techniek voor zulke simulaties is agent-gebaseerde modellering (Gilbert 2008; Beuls & Steels 2013).

Een tweede mogelijkheid is om verdere gevalstudies van lectale contaminatie onder de loep te nemen. Tot nu toe is enkel variatie onderzocht uit het nominale domein, tussen de Belgische en Nederlandse variëtiëten van het Nederlands. De vier condities voor lectale contaminatie, opgesomd in Sectie 2, zijn echter evengoed toepasbaar op variatie in het verbale domein, bij andere lectale onderscheidingen en in andere talen. Andere mogelijke gevalstudies in het Nederlands zijn te vinden in Haeseryn (1996), en voor

gevalstudies in het Engels en Duits kan men kijken in Algeo (2006), Rohdenburg & Schlüter (2009) en *Variantengrammatik des Standarddeutschen* (2018).

Voorts kan uitgezocht worden of taalgebruikers frasen met *zo'n* en *zulk* ooit als volledige taalvormen in hun geheugen opslaan, en of daarin verschillen tussen taalgebruikers bestaan (Dąbrowska 2012, 2018). Tot slot is de ontwikkeling van *zo'n* en *zulk* wel al apart in kaart gebracht door respectievelijk Van Olmen (2019) en Ghesquière & Van de Velde (2011), maar het is nog niet duidelijk hoe de competitie tussen beide voornaamwoorden precies evolueert in de laatste eeuwen.

Ik dank redacteur Hans Van de Velde en twee anonieme beoordelaars, wier suggesties en opmerkingen dit artikel sterk verbeterd hebben. Voorts wil ik ook mijn dank uiten aan Freek Van de Velde en Karlien Franco voor verscheidene interessante en onderhoudende discussies over de aard van lectale contaminatie, alsook aan Karlien Franco om een eerdere versie van dit artikel na te lezen en van commentaar te voorzien. Tot slot dank ik Benedikt Heller om me te wijzen op het nut van gebruikersgedefinieerde contrastcodering in regressieanalyses.

Bibliografie

- Algeo 2006 – J. Algeo, *British or American English? A handbook of word and grammar patterns*. Cambridge, New York: Cambridge University Press, 2006.
- Altenberg 1994 – B. Altenberg, 'On the functions of such in spoken and written English'. In: N. Oostdijk, P. De Haan & J. Aarts (eds.), *Corpus-based research into language*. Amsterdam: Rodopi, 1994, p. 223-239.
- Arnon & Snider 2010 – I. Arnon & N. Snider, 'More Than Words: Frequency Effects for Multi-Word Phrases'. In: *Journal of Memory and Language* 62 (2010) 1, p. 67-82.
- Barr e.a. 2013 – D.J. Barr e.a., 'Random effects structure for confirmatory hypothesis testing: Keep it maximal'. In: *Journal of Memory and Language* 68 (2013) 3, p. 255-278.
- Bates e.a. 2018 – D. Bates e.a., 'Parsimonious Mixed Models'. In: *arXiv.org*. Ithaca: Cornell University Library, arXiv.org, 2018, 1506.04967v2.
- Beuls & Steels 2013 – K. Beuls & L. Steels, 'Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement'. In: *PLoS ONE* 8 (2013), 3, p. e58960.
- Bod 1995 – L.W.M. Bod, *Enriching linguistics with statistics. Performance models of natural language*. Proefschrift Universiteit van Amsterdam.

- Cappelle 2006 – B. Cappelle, ‘Particle placement and the case for “allostructions”’. In: *Constructions* 7 (2006), p. 1-28.
- Centraal Bureau voor Statistiek – Beschikbaar op: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83926NED/barv?dl=17256&ts=1584114358740> (Bezocht op 15 september 2021).
- Croft 2003 – W. Croft, ‘Lexical rules vs. constructions. A false dichotomy’. In: H. Cuyckens e.a. (eds.), *Motivation in language: studies in honor of Günter Radden*. Stanford: CSLI Publications, 2003, p. 49-68.
- Dąbrowska 2006 – E. Dąbrowska, ‘Low-level schemas or general rules? The role of diminutives in the acquisition of Polish case inflections’. In: *Language Sciences* 28 (2006) 1, p. 120-135.
- Dąbrowska 2012 – E. Dąbrowska, ‘Different Speakers, Different Grammars: Individual Differences in Native Language Attainment’. In: *Linguistic Approaches to Bilingualism* 2 (2012) 3, p. 219-253.
- Dąbrowska 2014 – E. Dąbrowska, ‘Recycling utterances: A speaker’s guide to sentence processing’. In: *Cognitive Linguistics* 25 (2014) 4, p. 617-653.
- Dąbrowska 2018 – E. Dąbrowska, ‘Experience, aptitude and individual differences in native language ultimate attainment’. In: *Cognition* 178 (2018), p. 222-235.
- Daems, Heylen & Geeraerts 2015 – J. Daems, K. Heylen & D. Geeraerts, ‘Wat dragen we vandaag: een hemd met blazer of een shirt met jasje?’ In: *Taal en Tongval* 67 (2015) 2, p. 307-342.
- Den Boon & Geeraerts 2005 – *Van Dale Groot woordenboek van de Nederlandse taal*. 14th ed. Antwerpen/Utrecht: Van Dale Lexicografie, 2005.
- Diessel 2019 – H. Diessel, *The Grammar Network*. Cambridge: Cambridge University Press, 2019.
- Duihoven 1988 – A. M. Duihoven, *Middel nederlandse syntaxis, synchroon en diachroon. Deel 1: de naamwoordgroep*. Leiden: Nijhoff, 1988.
- Gazzaniga, Ivry & Mangun 2009 – M. Gazzaniga, R. Ivry & G. Mangun, *Cognitive neuroscience. The biology of the mind*. 3rd edn. New York: Norton, 2009.
- Geeraerts & Van de Velde 2013 – D. Geeraerts & H. Van de Velde, ‘Supra-regional characteristics of colloquial Dutch’. In: F. Hinskens & J. Taeldeman (eds.), *Language and space: Dutch*. Berlin: Mouton de Gruyter, 2013, p. 532-556.
- Ghesquière & Van de Velde 2011 – L. Ghesquière & F. Van de Velde, ‘A corpus-based account of the development of English such and Dutch zulk: Identification, intensification and (inter)subjectification’. In: *Cognitive Linguistics* 22 (2011) 4, p. 765-797.
- Gilbert 2008 – N. Gilbert, *Agent-based models*. Los Angeles: Sage, 2008.
- Gries 2013 – S. T. Gries, *Statistics for linguistics with R. A practical introduction*. 2e ed. Berlin: De Gruyter, 2013.

- Grondelaers, Speelman & Geeraerts 2008 – S. Grondelaers, D. Speelman & D. Geeraerts, 'National variation in the use of er "there". Regional and diachronic constraints on cognitive explanations'. In: G. Kristiansen & R. Dirven (eds.), *Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 2008, p. 153-203.
- Haeseryn 1996 – W. Haeseryn, 'Grammaticale verschillen tussen het Nederlands in België en het Nederlands in Nederland: Een poging tot inventarisatie'. In: R. van Hout & J. Kruijssen (red.), *Taalvariëties: Toonzettingen en modulaties op een thema*. Dordrecht: Foris, 1996, p. 109-126.
- Haeseryn e.a. 1997 – W. Haeseryn e.a., *Algemene Nederlandse Spraakkunst*. Groningen: Nijhoff, 1997.
- Heller 2018 – B. Heller, *Stability and fluidity in syntactic variation world-wide*. Proefschrift KU Leuven.
- Hosmer & Lemeshow 2000 – D. Hosmer & S. Lemeshow, *Applied logistic regression*. 2e ed. New York: Wiley, 2000.
- Levshina 2015 – N. Levshina, *How to do linguistics with R*. Amsterdam: John Benjamins, 2015.
- Oostdijk e.a. 2002 – N. Oostdijk e.a., 'Experiences from the Spoken Dutch corpus project'. *Proceedings of the third international conference on language resources and evaluation (LREC)*, 2002, p. 340-347.
- Oostdijk e.a. 2013 – N. Oostdijk e.a., 'The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch'. In: P. Spyns & J. Odiijk (eds.), *Essential Speech and Language Technology for Dutch*. Heidelberg: Springer, 2013, p. 219-247.
- Pijpops 2020 – D. Pijpops, 'What is an alternation? Six answers'. In: *Belgian Journal of Linguistics* 34 (2020), p. 283-294.
- Pijpops te versch. – D. Pijpops, 'Lectal contamination. Evidence from corpora and from agent-based simulation'. In: *International Journal of Corpus Linguistics*, te verschijnen.
- Pijpops, Beuls & Van de Velde 2015 – D. Pijpops, K. Beuls & F. Van de Velde, 'The rise of the verbal weak inflection in Germanic. An agent-based model'. In: *Computational linguistics in the Netherlands Journal* 5 (2015), p. 81-102.
- Pijpops & Van de Velde 2016 – D. Pijpops & F. Van de Velde, 'Constructional contamination: How does it work and how do we measure it?' In: *Folia Linguistica* 50 (2016) 2, p. 543-581.
- Pijpops & Van de Velde 2018 – D. Pijpops & F. Van de Velde, 'A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion'. In: *Corpus Linguistics and Linguistic Theory* 14 (2018) 1, p. 99-131.
- Pijpops e.a. 2018 – D. Pijpops e.a., 'Comparing explanations for the Complexity Principle. Evidence from argument realization'. In: *Language and Cognition* 10 (2018) 3, p. 514-543.

- Plevoets 2008 – K. Plevoets, *Tussen spreek- en standaardtaal. Een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfosyntactische verschijnselen uit het gesproken Belgisch-Nederlands*. Proefschrift KU Leuven.
- 'Referentiebestand Nederlands – RBN' 2014. Beschikbaar aan het Instituut voor de Nederlandse taal: <http://hdl.handle.net/10032/tm-a2-n2>, 2014.
- Rohdenburg & Schlüter 2009 – G. Rohdenburg & J. Schlüter, *One language, two grammars? Differences between British and American English*. Cambridge, New York: Cambridge University Press, 2009.
- Ruette 2012 – T. Ruette, *Aggregating Lexical Variation: towards large-scale lexical lectometry*. Proefschrift KU Leuven.
- De Smet & Van de Velde 2019 – I. De Smet & F. Van de Velde, 'Reassessing the evolution of West Germanic preterite inflection'. In: *Diachronica* 36 (2019) 2, p. 139-180.
- Speelman 2014 – D. Speelman, 'Logistic regression: A confirmatory technique for comparisons in corpus linguistics'. In: D. Glynn & J. A. Robinson (eds.), *Corpus Methods for Semantics*. Amsterdam: John Benjamins, 2014, p. 487-533.
- Speelman e.a. 2020 – D. Speelman e.a., 'Schaalvergroting in het syntactische alternantieonderzoek. Een nieuwe analyse van het presentatieve er met automatisch gegenereerde predictoren'. In: *Nederlandse Taalkunde* 25 (2020) 1, p. 101-123.
- Speelman & Geeraerts 2009 – D. Speelman & D. Geeraerts, 'Causes for causatives: the case of Dutch "doen" and "laten"'. In: T. Sanders & E. Sweetser (eds.), *Causal Categories in Discourse and Cognition*. Berlin: Mouton de Gruyter, 2009, p. 173-204.
- Speelman, Grondelaers & Geeraerts 2008 – D. Speelman, S. Grondelaers & D. Geeraerts, 'Variation in the choice of adjectives in the two main national varieties of Dutch'. In: G. Kristiansen & R. Dirven (eds.), *Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 2008, p. 205-233.
- Speelman, Heylen & Geeraerts 2018 – D. Speelman, K. Heylen & D. Geeraerts, 'Introduction'. In: D. Speelman, K. Heylen & D. Geeraerts (eds.), *Mixed-Effects Regression Models in Linguistics*. Cham: Springer, 2018, p. 1-10.
- Stefanowitsch & Flach 2016 – A. Stefanowitsch & S. Flach, 'The corpus-based perspective on entrenchment'. In: H.-J. Schmid (eds.), *Entrenchment and the Psychology of Language Learning*. Berlin: Mouton de Gruyter, 2016, p. 101-127.
- Stefanowitsch & Gries 2003 – A. Stefanowitsch & S.T. Gries, 'Collostructions: Investigating the interaction of words and constructions'. In: *International Journal of Corpus Linguistics* 8 (2003) 2, p. 209-244.
- Tremblay e.a. 2011 – A. Tremblay e.a., 'Processing Advantages of Lexical Bundles: Evidence from Self-Paced Reading and Sentence Recall Tasks'. In: *Language Learning* 61 (2011) 2, p. 569-613.

- Tremblay & Baayen 2010 – A. Tremblay & R.H. Baayen, 'Holistic Processing of Regular Four-word Sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall'. In: D. Wood (eds.), *Perspectives on formulaic language: acquisition and communication*. London: Continuum, 2010, p. 151-173.
- Tummers, Speelman & Geeraerts 2005 – J. Tummers, D. Speelman & D. Geeraerts, 'Inflectional variation in Belgian and Netherlandic Dutch: a usage-based account of the adjectival inflection'. In: N. Delbecque, J. Van der Auwera & D. Geeraerts (eds.), *Perspectives on variation: sociolinguistic, historical, comparative*. Berlin: Mouton de Gruyter, 2005, p. 93-110.
- Van Agtmaal-Wobma e.a. 2007 – E. van Agtmaal-Wobma e.a., *Belgen in Nederland en Nederlanders in België*. Centraal Bureau voor de Statistiek (CBS).
- Van de Velde 2009 – F. Van de Velde, *De nominale constituent. Structuur en geschiedenis*. Leuven: Leuven University Press, 2009.
- Van der Horst 2008 – J. van der Horst, *Geschiedenis van de Nederlandse syntaxis*. Leuven: Universitaire Pers Leuven, 2008.
- Van der Vliet 2007 – H. van der Vliet, 'The Referentiebestand Nederlands as a Multi-Purpose Lexical Database'. In: *International Journal of Lexicography* 20 (2007) 3, p. 239-257.
- Van Eerten 2007 – L. van Eerten, 'Over het Corpus Gesproken Nederlands'. In: *Nederlandse Taalkunde* 12 (2007) 3, p. 194-215.
- Van Olmen 2019 – D. Van Olmen, 'A diachronic corpus study of prenominal zo'n "so a" in Dutch: Pathways and (inter)subjectification'. In: *Functions of Language* 26 (2019) 2, p. 216-247.
- Van Olmen & Van der Auwera 2014 – D. Van Olmen & J. van der Auwera, 'Over zo'n en zo meer'. In: F. Van de Velde e.a. (red.), *Patroon en argument. Een dubbelveestbundel bij het emeritaat van William Van Belle en Joop van der Horst*. Leuven: Universitaire Pers Leuven, 2014, p. 215-228.
- Variantengrammatik des Standarddeutschen* 2018 – *Ein Online-Nachschlagewerk. Verfasst von einem Autorenteam unter der Leitung von Christa Dürscheid, Stephan Elspaß und Arne Ziegler*. 2018. Beschikbaar op http://mediawiki.ids-mannheim.de/VarGra/index.php/Substantive_auf_ation/_ung.
- Wallis 2012 – S. Wallis, 'That vexed problem of choice'. *Presentatie op ICAME33*, KU Leuven, 30 mei – 3 juni 2012. Beschikbaar op: www.ucl.ac.uk/english-usage/statspapers/vexedchoice.pdf.
- Wolk e.a. 2013 – C. Wolk e.a., 'Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change'. In: *Diachronica* 30 (2013) 3, p. 382-419.

Zehentner & Traugott 2020 – E. Zehentner & E.C. Traugott, 'Constructional networks and the development of benefactive ditransitives in English'. In: L. Sommerer & E. Smirnova (eds.), *Nodes and Networks in Diachronic Construction Grammar*. Amsterdam: John Benjamins, 2020, p. 168-211.

Over de auteur

Dirk Pijpops, Bâtiment A2, Departement de Langues modernes: linguistique, littérature et traduction, Place Cockerill 3-5, 4000 Liège, België
dirk.pijpops@uliege.be