

# Naar een pluricentrische *Algemene Nederlandse Spraakkunst*

## Het gebruik van productiedata voor de beschrijving van nationale variatie

Arne DHONDT, Timothy COLLEMAN, Johan DE CALUWE en  
Gauthier DELABY

### Abstract

Om van de *Algemene Nederlandse Spraakkunst* (ANS) een pluricentrische referentiegrammatica te maken is er nood aan systematisch onderzoek naar de verschillende grammaticale normen van de drie nationale standaardvariëteiten van het Nederlands, nl. Belgisch Nederlands (BN), Nederlands Nederlands (NN) en Surinaams Nederlands (SN). Op basis van bestaande methodologieën om grammaticale verschillen tussen standaardvariëteiten te beschrijven, zoals die van Taaladvies.net of de Duitse *Variantengrammatik* (cf. Dürscheid & Elspaß 2015), komen we tot een aantal vragen over de methodologie die de herziene ANS moet volgen. We gaan in op het gebruik van productiedata om nationale verschillen te beschrijven, met focus op een vaak ingezette frequentie maat, nl. relatieve tokenfrequentie. Er wordt onderzocht (i) aan de hand van welke criteria kan worden nagegaan of grammaticale varianten al dan niet als equivalenten mogen gelden en (ii) met welke beperkingen rekening moet worden gehouden bij de interpretatie van nationale verschillen in relatieve tokenfrequentie. Verder bespreken we (iii) of er naast productiedata ook evaluatiedata verzameld moeten worden om te bepalen of varianten tot een nationale standaardvariëteit behoren en (iv) welke criteria gehanteerd moeten worden om op basis van productiedata nationale labels toe te kennen aan varianten. Over die vragen reflecteren we aan de hand van literatuuronderzoek en aan de hand van corpusonderzoek voor drie cases van grammaticale variatie tussen BN en NN.

## 1. INLEIDING<sup>1</sup>

Sinds oktober 2015 wordt gewerkt aan een herziening van de tweede editie van de *Algemene Nederlandse Spraakkunst* (ANS2), die dateert uit 1997 en waaraan de E-ANS, die online raadpleegbaar is, inhoudelijk nagenoeg identiek is (alleen een aantal evidente tikfouten en andere formele onvolkomenheden werden verbeterd).<sup>2</sup> De herziene ANS zal enkel online beschikbaar zijn en

<sup>1</sup> We bedanken graag twee anonieme reviewers voor hun opmerkingen en suggesties bij een eerdere versie van dit artikel.

<sup>2</sup> Vanaf hier verwijzen we naar de *Algemene Nederlandse Spraakkunst* met de afkorting ANS. Om specifiek naar de eerste editie (Geerts et al. 1984) of de tweede editie (Haeseryn et al. 1997) te verwijzen, gebruiken we respectievelijk de afkortingen ANS1 en ANS2. Naar de herziene ANS wordt verwezen met de afkorting ANS3. Voor de beschrijving van (de redenen voor) het herzieningsproject hebben we geput uit de aankondiging die op de website van de ANS ([ans.ruhosting.nl](http://ans.ruhosting.nl)) is verschenen over het project.

moet niet alleen toegankelijker en gemakkelijker te gebruiken zijn dan de ANS2, ook de beschrijving van de Nederlandse grammatica moet worden geactualiseerd. Zo'n inhoudelijke update is noodzakelijk om verschillende redenen. Zo is onze kennis over allerlei grammaticale verschijnselen uitgediept sinds 1997, en ook de taal zelf en de normen van de taalgemeenschap zijn sindsdien veranderd. Bovendien is de visie op variatie binnen en buiten de Nederlandse standaardtaal gewijzigd in de afgelopen decennia. Zo heeft de monocentrische visie op het Nederlands bijvoorbeeld plaatsgemaakt voor een pluricentrische visie, waardoor er meer ruimte is gekomen voor diatopische variatie binnen de standaardtaal. In extreme vorm hield de monocentrische visie namelijk in dat het taalgebruik in het noorden van het taalgebied de norm vormde voor het Standaardnederlands en dat alles wat ervan afweek (bv. Zuid-Nederlandse woorden en constructies), niet standaardtalig was (Martin 2001: 709-710; Van Hoof & Jaspers 2012: 102-104). Tegenwoordig gaat de Nederlandse Taalunie echter uit van een pluricentrische visie, waarin er drie nationale standaardvariëteiten van het Nederlands onderscheiden worden, namelijk Belgisch-Nederlands (BN), Nederlands-Nederlands (NN) en Surinaams-Nederlands (SN) (cf. Adviescommissie Taalvariatie 2019: 9-10). Die variëteiten zijn weliswaar voor het grootste deel identiek, maar hebben toch ook eigen normen op het vlak van uitspraak, woordenschat en grammatica.

### 1.1. *Pluricentrische beschrijvingen van het Nederlands*

Wat uitspraak en woordenschat betreft, zijn er al beschrijvingen beschikbaar van de verschillende normen van BN en NN. Van de Velde (1996) heeft bijvoorbeeld de veranderingen in de uitspraaknorm in Vlaanderen en Nederland sinds 1935 onderzocht, en constateerde dat de standaarduitspraak in beide gebieden divergeert. Een pluricentrische beschrijving van de woordenschat vinden we in woordenboeken zoals het *Algemeen Nederlands Woordenboek*<sup>3</sup>, *Prisma Handwoordenboek Nederlands* (Martin & Smedts 2009) en *Van Dale Groot woordenboek van de Nederlandse taal* (Den Boon & Hendrickx 2015), waarin woorden die enkel gebruikelijk zijn in Vlaanderen of Nederland respectievelijk een label BN of NN krijgen (cf. Colman 2009; De Caluwe 2011, 2013).

Voor de grammatica bestaan er nog geen vergelijkbare pluricentrische beschrijvingen. De ANS2 biedt bijvoorbeeld geen pluricentrisch perspectief op de Nederlandse grammatica: de Nederlandse standaardtaal wordt er omschreven als “de taal waarin geen elementen of structuren voorkomen die duidelijk

<sup>3</sup> Het *Algemeen Nederlands Woordenboek* is een digitaal woordenboek dat geraadpleegd kan worden op [anw.inl.nl](http://anw.inl.nl).

opvallen als niet-algemeen” (0.6.2.1), en diatopische variatie wordt beschreven met twee labels<sup>4</sup> (0.6.2.3), namelijk:

- het label ‘geografisch verschillend’, dat wordt toegekend aan “elementen en structuren die wel enigszins streekgebonden zijn, maar voor de meerderheid van de taalgebruikers niet kenmerkend voor de taal in een bepaalde regio”, bijvoorbeeld de verledentijdsvorm *ervoer*, die in het hele taalgebied voorkomt, maar vaker in Nederland dan in België;
- het label ‘regionaal’, dat wordt toegekend aan “[e]lementen en structuren die voor veel taalgebruikers kenmerkend zijn voor een bepaald deel van het taalgebied, of onbekend omdat ze afkomstig zijn uit een andere regio”, bijvoorbeeld de doorbreking van de werkwoordelijke eindgroep door het tweede deel van een voornaamwoordelijk bijwoord in *een onderwerp waar we beter niet kunnen over spreken*, die enkel in België voorkomt.

Geografisch verschillende varianten worden tot de standaardtaal gerekend, terwijl regionale varianten buiten de standaardtaal vallen. De ANS2 gaat dus nog niet uit van nationale standaardvariëteiten met eigen grammaticale normen: elementen en structuren die een beperkte geografische verspreiding hebben, gelden per definitie als niet-standaardtalig.

### 1.2. *De ANS3 als pluricentrische referentiegrammatica*

Doel van de herziene ANS3 is daarentegen om de Nederlandse grammatica vanuit pluricentrisch perspectief te beschrijven. Op basis van empirisch onderzoek moet voor grammaticale varianten beschreven worden of ze verschillend gebruikt worden in de nationale standaardvariëteiten: dat kan inhouden dat bepaalde varianten meer of minder worden gebruikt in een van de landen (= graduele verschillen), maar ook dat varianten slechts in een van de standaardvariëteiten tot de norm behoren (= absolute verschillen). De keuze om van de ANS3 een *pluricentrische* referentiegrammatica te maken impliceert een focus op *nationale* variatie binnen de standaardtaal. Het is natuurlijk niet uitgesloten (en zelfs waarschijnlijk) dat er ook op andere geografische niveaus grammaticale verschillen zijn in de Nederlandse standaardtaal in Europa (bv. *binnen* Nederland en Vlaanderen). De keuze voor een pluricentrisch perspectief en de daarbij horende focus op nationale variatie is echter een taalpolitieke keuze, die vertrekt vanuit de visie dat Vlaanderen en Nederland elk apart een

<sup>4</sup> We spreken hier wel over twee labels, maar in feite kent de ANS2 enkel het label ‘regionaal’ toe aan bepaalde voorbeelden. Bij ‘geografisch verschillende’ varianten wordt in de lopende tekst gesignaleerd dat er diatopische variatie is, maar aan concrete voorbeelden wordt geen *label* ‘geografisch verschillend’ toegekend.

taalgemeenschap vormen. Aangezien bv. bestuur, onderwijs en media (radio, televisie, kranten, magazines) in het Nederlandse taalgebied vooral op nationaal (Vlaams of Nederlands) niveau georganiseerd worden, is dat ook het niveau waarop de meeste gesproken of geschreven standaardtalige teksten (bv. tv- of radioprogramma's, kranten, schoolboeken, beleidsteksten) gericht zijn. Dat impliceert enerzijds dat taalgebruikers vooral behoefte hebben aan een beschrijving van standaardtaalnormen op nationaal niveau (en niet van wat bv. in West-Vlaanderen of Noord-Brabant voorkomt in standaardtalige contexten), en anderzijds dat het vooral op nationaal niveau relevant en haalbaar is om grote corpora met standaardtalige teksten aan te leggen (die de ANS3 dan weer kan gebruiken om grammaticale variatie te onderzoeken).

### 1.3. *Informatie over grammaticale verschillen tussen nationale standaardvariëteiten van het Nederlands*

Om de grammatica van het Nederlands te beschrijven vanuit pluricentrisch perspectief, is er nood aan (empirisch onderbouwde) informatie over grammaticale verschillen tussen de nationale standaardvariëteiten. Zelfs als we het SN buiten beschouwing laten, blijkt die informatie relatief schaars. Zo is het in de eerste plaats al onduidelijk welke verschillen er allemaal zijn. Informatie daarover vinden we bijvoorbeeld bij Haeseryn (1996, 2013); hij stelde een inventaris van grammaticale noord-zuidverschillen op die grotendeels gebaseerd is op de beschrijving van diatopische variatie in ANS1 en ANS2. Daarnaast zijn er casestudy's beschikbaar waarin de variatie tussen België en Nederland voor individuele grammaticale varianten of variantenparen nauwkeuriger in kaart wordt gebracht op basis van corpusonderzoek (o.a. Colleman 2010; De Sutter et al. 2005; Grondelaers et al. 2001; Pijpops & Van de Velde 2018; Speelman & Geeraerts 2009; Tummers et al. 2005). Ook bij Taaladvies.net, de taaladviesite van de Nederlandse Taalunie, heeft een substantieel deel van de grammaticale adviezen betrekking op constructies met nationale variatie. De vraag rijst echter of we op basis van die bronnen een exhaustief overzicht krijgen van de nationale variatie in de grammatica. Zeker aangezien voor (bepaalde types) grammaticale verschillen gesteld wordt dat ze weinig saillant zijn (cf. Van Bree 2000), is het niet onmogelijk dat er verschillen onder de radar gebleven zijn. Om die op te sporen, zijn digitale, computerlinguïstische methodes veelbelovend. Zo trachten Grondelaers et al. (2020) om nog onbekende nationale variatie in de grammatica in kaart te brengen aan de hand van *pivoting*, een computationele techniek waarmee ze in een tweetalig (Nederlands-Engels) corpus van ondertitels zoeken naar equivalente Nederlandse vertalingen van eenzelfde Engels *n*-gram. In een pilootonderzoek vonden ze op die manier een aantal morfosyntactische alternanties die potentieel aan noord-zuidvariatie on-

derhevig zijn; sommige van de gevonden alternanties waren in de literatuur reeds beschreven, andere waren tot nog toe onbekend (bv. de sterkere varianten van diverse types voornaamwoorden lijken vaker gebruikt te worden in Vlaanderen). Van zulk onderzoek mogen in de toekomst wellicht nog interessante resultaten worden verwacht (die dan kunnen worden opgenomen in ANS3), maar het is uiteraard niet de bedoeling van de ANS3-redactie om zelf grootschalig computationeel onderzoek op te zetten om systematisch op zoek te gaan naar nog onbekende grammaticale verschillen.

Hoewel het dus mogelijk is dat nog niet alle noord-zuidverschillen geïdentificeerd zijn, vinden we in de literatuur wel aanknopingspunten voor een beschrijving van nationale variatie in de grammatica. De beschikbare informatie volstaat echter vaak niet als basis voor een pluricentrische referentiegrammatica. In de ANS2 weten we bij als ‘regionaal’ gelabelde varianten bijvoorbeeld niet of ze standaardtaalig zijn in het deel van het taalgebied waarin ze voorkomen: doorbrekingen van de werkwoordelijke eindgroep komen bijvoorbeeld enkel voor in BN, maar zijn ze daar ook standaardtaal? Een ander probleem met de beschrijving in de ANS2 is dat de labels vaak intersubjectief zijn toegekend door de redactie (ANS2 0.6.2.1). Als we echter een accurate beschrijving willen maken van de nationale variatie in het concrete taalgebruik, dan is er nood aan taalgebruiksdata om die beschrijving te onderbouwen. De hierboven genoemde casestudy’s beschrijven wel in detail het *gebruik* van de behandelde constructies in BN en NN (in bepaalde registers), maar ook zij geven geen antwoord op de vraag of en waar de besproken varianten tot de *standaardtaal* behoren. De auteurs hanteren bovendien niet altijd dezelfde methodologie en/of onderzoeken niet dezelfde registers, zodat de gegevens uit die casestudy’s niet optimaal vergelijkbaar zijn. Taaladvies.net biedt dan weer wél informatie over de verschillende grammaticale normen van BN en NN en die adviezen worden bovendien empirisch onderbouwd (cf. infra). Zoals verderop in dit artikel zal blijken, kunnen er echter vragen gesteld worden bij de methodologie die Taaladvies.net hanteert om de normatieve status van varianten in BN en NN te bepalen.

#### 1.4. Overzicht

Kortom, hoewel er al een en ander bekend is over grammaticale verschillen tussen België en Nederland, laat die informatie zich niet zomaar vertalen naar een genuanceerde pluricentrische beschrijving. Als de herziene ANS de Nederlandse grammatica vanuit pluricentrisch perspectief wil beschrijven, zal systematisch onderzocht moeten worden welke verschillen in grammaticale normen er zijn tussen BN, NN én SN. Bij zulk onderzoek worden voor de verschillende varianten op uniforme wijze data verzameld en geïnterpreteerd. Dit

artikel vormt een reflectie over hoe zo'n systematische methode eruit zou kunnen zien. In paragraaf 2 bespreken we twee praktijkvoorbeelden van systematisch onderzoek naar grammaticale verschillen tussen standaardvariëteiten, namelijk de methode van Taaladvies.net en de methode van de *Variantengrammatik*, een beschrijving van de diatopische variatie in de geschreven Duitse standaardtaal (cf. Dürscheid & Elspaß 2015). Op basis daarvan komen we tot een aantal vragen over keuzes die de ANS3 moet maken als ze ook tot een systematische methode wil komen om nationale variatie binnen de grammatica te beschrijven. De vragen hebben betrekking op het gebruik van productiedata om nationale verschillen te beschrijven, met focus op het gebruik van relatieve tokenfrequentie, een frequentie maat die door Taaladvies.net en de Variantengrammatik een prominente rol toebedeeld krijgt. Als illustratie bij die vragen bespreken we in de derde paragraaf de resultaten van verkennend corpusonderzoek naar drie gevallen van grammaticale variatie tussen BN en NN. Op basis van die cases en van aanvullend literatuuronderzoek reflecteren we vervolgens in paragraaf 4 over de vragen die we eerder opwierpen. We focussen bij de casestudy's en de verdere bespreking op verschillen tussen BN en NN: voor die verschillen vinden we in de literatuur voldoende aanknopingspunten, en voor die variëteiten zijn bovendien goede corpora beschikbaar. De bedoeling is uiteraard dat de methodologie uiteindelijk ook kan worden uitgebreid naar SN, zodra ook voor die nationale variëteit omvangrijke en representatieve corpora beschikbaar zijn.

## 2. SYSTEMATISCH ONDERZOEK NAAR GRAMMATICALE VERSCHILLEN TUSSEN STANDAARDVARIËTEITEN

### 2.1. *Variantengrammatik*

Een voorbeeld van systematisch onderzoek naar grammaticale variatie tussen standaardvariëteiten vinden we bij de Variantengrammatik (zie o.a. Dürscheid & Elspaß 2015; Dürscheid et al. 2015; Elspaß 2010; Elspaß et al. 2017; Niehaus 2015). Dat naslagwerk biedt een beschrijving van de diatopische grammaticale variatie binnen de geschreven Duitse *gebruiksstandaard*: alle varianten die regelmatig voorkomen in standaardtalige contexten (bv. krantentaal) in een bepaalde regio worden als standaardtaal binnen die regio beschouwd, ook als die volgens andere naslagwerken niet standaardtalig zijn (Dürscheid & Elspaß 2015: 563). De beschrijving is gebaseerd op frequentiedata uit een digitaal corpus van circa 600 miljoen woorden, met artikelen uit 68 online raadplegbare, regionale kranten uit alle delen van het Duitse taalgebied (Niehaus 2015: 141-144). Het corpus is onderverdeeld in 15 subcorpora, naargelang van het deel van het taalgebied waaruit de artikelen afkomstig zijn, nl.

Noord-West-Duitsland, Noord-Oost-Duitsland, Midden-West-Duitsland, Midden-Oost-Duitsland, Zuid-West-Duitsland, Zuid-Oost-Duitsland, West-Oostenrijk, Midden-Oostenrijk, Zuidoost-Oostenrijk, Oost-Oostenrijk, Zwitserland, Oost-België, Liechtenstein, Luxemburg en Zuid-Tirol.

Sommige subcorpora beslaan een staat, andere slechts een deel van een staat. Dat landen nog verder onderverdeeld worden in regio's, komt doordat er volgens Elspaß et al. (2017: 72-73) geen recht wordt gedaan aan de talige werkelijkheid als enkel verschillen tussen landen onderzocht worden. De verspreiding van een variant blijkt namelijk meestal niet samen te vallen met de staatsgrenzen, maar wordt in veel sterkere mate beïnvloed door de grenzen van de oude dialectgebieden; binnen één staat kunnen er dan ook regionale verschillen zijn in het gebruik van varianten, en varianten kunnen ook staatsgrenzen overschrijden: de Duitse regio Beieren sluit op talig vlak bijvoorbeeld nauw aan bij Oostenrijk (cf. Dürscheid & Elspaß 2015: 564). De redacteurs van de Variantengrammatik geven aan dat zij de verspreiding van varianten over *regio's* (in plaats van landen) bestuderen; zo'n perspectief noemen zij een *pluri-areaal* (in plaats van pluricentrisch) perspectief. Hoewel het ook in het Nederlandse taalgebied in Europa niet onmogelijk (en zelfs waarschijnlijk) is dat de verspreiding van varianten niet samenvalt met de staatsgrenzen, kiest de ANS3 wel degelijk voor een pluricentrisch perspectief. Zoals aangegeven in paragraaf 1, wordt er bij die keuze van uitgegaan dat Vlaanderen en Nederland een aparte taalgemeenschap vormen en dat het enkel op nationaal niveau zinvol is om standaardtaalnormen te beschrijven.

Om de distributie van de varianten binnen de subcorpora van de Variantengrammatik te vergelijken, worden twee frequentiematen gebruikt (zie de beschrijving bij Dürscheid & Elspaß 2015; Niehaus 2015). Bij varianten met een of meerdere equivalent(en), bijvoorbeeld de genusvarianten *die/das (E-)Mail*, wordt berekend met welke relatieve tokenfrequentie de variant en zijn equivalent(en) voorkomen in een subcorpus en (aan de hand van een chikwadraattoets) of er significante verschillen in distributie zijn tussen de verschillende subcorpora. Als dat het geval is, dan wordt het distributieverhaal opgenomen in de Variantengrammatik. Om dat verschil vlotter weer te kunnen geven in de lopende tekst, worden aan de waarden voor relatieve tokenfrequentie de verschillende labels uit Figuur 1 (p. 92) gekoppeld.

Daarnaast zijn er ook varianten zonder duidelijk equivalent, bijvoorbeeld de constructie [functiewerkwoord + *am/beim* + infinitief] (zoals in *Sie ist am Schlafen* 'ze is aan het slapen'). Bij die varianten wordt (aan de hand van een chikwadraattoets) berekend of er een significant verschil is in de absolute tokenfrequentie waarmee de variant voorkomt in de verschillende subcorpora. Is

**Figuur 1: labels waarmee de Variantengrammatik in lopende tekst aangeeft hoe frequent een variant is in een bepaalde regio (Dürscheid & Elspaß 2015: 574), met eigen Nederlandse vertaling**

relatieve frequentie	label Variantengrammatik	Nederlands equivalent
< 5%	sporadisch	sporadisch
5-20%	kommt (selten) vor	komt (zelden) voor
21-50%	gebräuchlich	komt geregeld voor
51-94%	üblich/mehrheitlich verwendet	komt in de meerderheid van de gevallen voor
> 95%	gilt (fast) ausnahmslos	komt (zo goed als) uitzonderingsloos voor

er een significant verschil, dan wordt de variant beschreven in de Variantengrammatik, eveneens met de labels uit Figuur 1. De relatieve tokenfrequentie waaraan die labels gekoppeld worden, is in dit geval de verhouding van het aantal tokens van de variant in een subcorpus ten opzichte van het totale aantal tokens van die variant in alle subcorpora samen. Bij alle varianten, met of zonder equivalent(en), is er nog een extra label mogelijk, namelijk ‘unter Schwellenwert’ (‘onder drempelwaarde’): daarmee wordt aangegeven dat het gebruik van een variant of een groep equivalente varianten in een regio niet beschreven kan worden, omdat er minder dan tien tokens zijn voor de variant of de groep equivalente varianten in het subcorpus voor die regio (cf. Niehaus 2015: 144).

## 2.2. *Taaladvies.net*

In tegenstelling tot de Variantengrammatik biedt *Taaladvies.net* geen exhaustieve beschrijving van diatopische grammaticale variatie binnen de Nederlandse standaardtaal: het is een taaladviesite die adviezen publiceert over de standaardtaligheid van allerlei ‘controversiële’ grammaticale, maar ook lexicale en uitspraakvarianten. Hoewel een groot deel van de adviezen betrekking heeft op varianten met nationale variatie, geldt dat zeker niet voor alle adviezen. Om hun adviezen empirisch te onderbouwen, brengen de redacteurs voor de varianten in kwestie verschillende types data samen: ze gaan meer bepaald na welke status er wordt toegekend aan een variant in *naslagwerken* en onderzoeken hoe en hoe vaak hij precies gebruikt wordt in *corpora* (Caluwé & Verreycken 2012: 174-179). Ze maken gebruik van Google, maar ook opnieuw van krantendatabanken, namelijk *Mediargus* (voor Belgische kranten) en *LexisNexis* (voor Nederlandse en Belgische kranten). Volgens Caluwé & Verreycken (2012: 179) zouden krantendatabanken namelijk een goed beeld moeten bieden van wat taalbewuste taalgebruikers als standaardtalig beschouwen, aangezien journalisten en redacteurs “als professionele taalgebruikers geacht worden de standaardtaal te beheersen” en dus “in principe een duidelijker



normbesef [zouden] moeten hebben dan de gemiddelde taalgebruiker.” De frequentie waarmee de variant in die corpora voorkomt ten opzichte van zijn equivalenten, bepaalt of er nog nood is aan extra informatie.<sup>5</sup> Als de variant in Mediargus of LexisNexis in een nationale variëteit voorkomt met een relatieve tokenfrequentie

- tussen 0 en 5%, dan wordt de variant niet als standaardtaal beschouwd in die variëteit;
- hoger dan 50%, dan wordt de variant tot de standaardtaal gerekend in die variëteit;
- tussen 5 en 50%, dan is het onduidelijk of de variant standaardtalig is in die variëteit.

In dat laatste geval wordt een beroep gedaan op een informantpanel, dat bestaat uit een zeventigtal Vlamingen en Nederlanders van verschillende leeftijden en regionale achtergronden, waarvan de meesten professioneel met taal bezig zijn (leerkrachten, journalisten, schrijvers, enzovoort).<sup>6</sup> Zij krijgen een tekst voorgelegd waarin een aantal problematische varianten verwerkt zitten en moeten aangeven welke woorden of wendingen hun onbekend zijn, welke zijzelf niet zouden gebruiken en welke ze in de standaardtaal niet aanvaardbaar vinden. Met die informantpanels gaat Taaladvies.net na hoe er over de variant wordt geoordeeld door de spraakmakende gemeente, de “groep gezaghebbende schrijvers en sprekers die in de publieke, sociaal-culturele sector een belangrijke rol spelen” en van wie het taalgebruik als normbepalend voor de standaardtaal wordt beschouwd.<sup>7</sup> Of die informantpanels representatief genoeg zijn, daarover bestaan er echter verschillende meningen: volgens De Schryver (2012: 153) is het bijvoorbeeld problematisch dat de informantpanels slechts uit een zeventigtal informanten bestaan, en dan nog vooral taal-experten.

<sup>5</sup> Die frequentiecriteria worden niet vermeld bij Caluwé & Verreycken (2012), maar worden wel toegelicht in de interne ‘Handleiding voor het verzamelen en interpreteren van informatie ten behoeve van *Taaladvies(.net)*’. Daar staat ook te lezen dat die criteria gelden “voor kwesties waarbij telkens 2 of 3 varianten worden afgezet tegenover elkaar”, maar dat het cijfermateriaal anders benaderd moet worden als “er meerdere varianten zijn of slechts 1 variant.”

<sup>6</sup> In de eerder vermelde interne handleiding lezen we dat soms toch ook varianten worden voorgelegd die met een relatieve tokenfrequentie van meer dan 50% voorkwamen. Zo werden een aantal varianten voorgelegd om te testen of de selectiecriteria wel streng genoeg waren, en soms worden ook varianten doorgestuurd naar de informanten als de intuïtie van de redacteurs niet overeenstemt met de frequentiedata (aangezien de zoekopdrachten niet altijd voldoende en/of betrouwbare frequentiedata opleverden).

<sup>7</sup> Caluwé & Verreycken (2012) vermelden zelf niet expliciet dat dat de reden is waarom ze die informantpanels inzetten, maar dat wordt wel aangegeven in de beschrijving ‘Werkwijze: welke varianten hebben standaardtaalkarakter (algemeen)’ die online raadpleegbaar is op Taaladvies.net ([https://taaladvies.net/taal/advies/tekst/86/werkwijze\\_welke\\_varianten\\_hebben\\_standaardtaalkarakter\\_algemeen/](https://taaladvies.net/taal/advies/tekst/86/werkwijze_welke_varianten_hebben_standaardtaalkarakter_algemeen/), geraadpleegd op 21 januari 2020). De omschrijving van de spraakmakende gemeente is ontleend aan de tekst ‘Wat is standaardtaal? (algemeen)’ ([https://taaladvies.net/taal/advies/tekst/85/wat\\_is\\_standaardtaal\\_algemeen\\_geraadpleegd\\_op\\_17\\_januari\\_2020](https://taaladvies.net/taal/advies/tekst/85/wat_is_standaardtaal_algemeen_geraadpleegd_op_17_januari_2020/)).

Op grond van de verzamelde informatie krijgt een variant een van de volgende vijf labels: ‘standaardtaal in het hele taalgebied’, ‘standaardtaal in België’, ‘standaardtaal in Nederland’, ‘geen standaardtaal’ of ‘status onduidelijk’ (Caluwé & Verreycken 2012: 181). Het label ‘status onduidelijk’ wordt vooral toegekend aan Belgische varianten die wel veelvuldig voorkomen in de corpora, maar die tegelijkertijd toch niet aanvaard worden door een groot aantal leden van de informantpanels (voorbeelden zijn de lexicale varianten *allergisch aan*, *begoed* en *frigobox*).<sup>8</sup> Taaladvies.net houdt dus rekening met de mogelijke discrepantie tussen het reële taalgebruik en de evaluatie daarvan. Die discrepantie is volgens Caluwé & Verreycken (2012: 181) ontstaan onder invloed van de normatieve traditie in Vlaanderen, die bij bepaalde varianten blijkbaar sterker heeft doorgewerkt dan bij andere. Er is dus in Vlaanderen *linguistic schizophrenia*, een situatie waarin “[t]he proper national norm is heavily practiced but officially depreciated” (Muhr 2012: 39). Zo’n situatie is volgens Muhr (2012: 39) kenmerkend voor niet-dominante variëteiten van pluricentrische talen, zoals BN.

### 2.3. *Vragen bij systematisch onderzoek naar grammaticale verschillen tussen standaardvariëteiten*

Zowel de Variantengrammatik als Taaladvies.net hebben dus een methodologie ontwikkeld om nationale (of regionale) variatie in het gebruik van grammaticale varianten in de standaardtaal te beschrijven aan de hand van (onder meer) frequentiedata uit krantencorpora. Nu roepen de overeenkomsten en verschillen tussen die methodes uiteraard een heel aantal vragen op over de keuzes die gemaakt moeten worden als we tot een systematische methode willen komen om nationale variatie in de grammatica te beschrijven. In de context van dit artikel bespreken we vier vragen met betrekking tot het gebruik van productiedata. Zowel de Variantengrammatik als Taaladvies.net kennen een belangrijke rol toe aan de *relatieve tokenfrequentie* om nationale (of regionale) verschillen te beschrijven. Twee vragen hebben dan ook specifiek betrekking op het gebruik van relatieve tokenfrequentie, nl.

- (i) Taaladvies.net en de Variantengrammatik beschrijven de diatopische variatie op basis van relatieve tokenfrequentie, als het gaat om varianten waarvoor equivalenten bestaan. Wanneer kunnen varianten echter beschouwd worden als *equivalenten*? In geen van beide methodologieën wordt daar aandacht aan besteed.

<sup>8</sup> Wat Caluwé & Verreycken (2012: 181) verstaan onder ‘veelvuldig voorkomen’ is onduidelijk. Volgens de interne handleiding worden enkel varianten met een relatieve tokenfrequentie tussen de 5 en 50% in de krantencorpora (systematisch) doorgestuurd naar het informantpanel.

- (ii) Beide methodologieën lijken ervan uit te gaan dat nationale (of regionale) verschillen in relatieve tokenfrequentie enkel en alleen veroorzaakt worden door de extern-linguïstische factor ‘land/regio’. Bieden verschillen in relatieve tokenfrequentie een goed beeld van de nationale verschillen, of kunnen intern-linguïstische factoren (bv. semantische verschillen tussen varianten) dat beeld vertekenen?

Daarnaast komen ook twee algemenere vragen aan bod, nl.

- (iii) De Variantengrammatik maakt enkel gebruik van frequentiedata uit krantencorpora, terwijl Taaladvies.net daarnaast ook evaluatiedata verzamelt. Volstaat corpusonderzoek voor ANS3, of is er ook nood aan evaluatieonderzoek?
- (iv) Taaladvies.net en de Variantengrammatik hanteren andere labels, en andere criteria om die labels toe te kennen. Welke criteria moet ANS3 gebruiken om labels toe te kennen? Meer specifiek, moet de ANS3 een drempelwaarde definiëren die een variant moet overschrijden binnen een variëteit om onderdeel uit te maken van de standaard van die variëteit?

Hoewel ze naast elkaar gepresenteerd worden, zijn die vragen niet allemaal gelijk van aard. Zo gaat het in (i) om een taalkundig-theoretische vraagstelling, terwijl in (ii) vooral een methodologisch probleem wordt aangekaart. Vraag (iii) is dan weer principieel van aard; ze heeft betrekking op de uitgangspunten van de ANS3, meer bepaald met betrekking tot wat standaardtaal is. Ook de criteria voor labeling in (iv) lijken (minstens gedeeltelijk) afhankelijk van de principiële vraag wat ANS3 beschouwt als standaardtaal. In het vervolg van dit artikel reflecteren we over die vier vragen. Daarvoor gaan we als volgt te werk.

Om een en ander te kunnen illustreren aan de hand van concrete voorbeelden, presenteren we in de volgende paragraaf drie corpusgebaseerde case-study's voor (groepen) grammaticale varianten waarvoor diatopische variatie beschreven wordt in de ANS2, nl.

- A. het gebruik van een infinitief of een voltooid deelwoord bij *komen*;
- B. het gebruik van de aanwijzende voornaamwoorden *zo'n* en *zulke* bij ontelbare en meervoudige substantieven;
- C. het gebruik van het groepsvormende *weten* in de betekenis ‘zich kunnen herinneren’.

Voor elk van die gevallen worden op basis van de literatuur de verschillende varianten (en ook hun normatieve status en geografische verspreiding) beschreven en verzamelen we data in nationale corpora van krantentaal.

Zoals zal blijken uit de bespreking, zijn de cases zo gekozen dat ze ons toelaten om de hierboven opgelijste vragen te illustreren. Zo verschillen de cases in de mate waarin er voor de varianten equivalenten kunnen worden onderscheiden, zodat het niet altijd duidelijk is of relatieve tokenfrequentie kan worden ingezet om nationale verschillen op het spoor te komen (cf. vraag i). Verder is er variatie tussen de cases in de mate waarin naast de extern-linguïstische factor ‘land’ ook intern-linguïstische factoren mogelijk een rol spelen in de nationaal verschillende gebruiksfrequentie van de varianten (cf. vraag ii). Bovendien komen zowel varianten aan bod die in de ANS2 gemarkeerd worden als ‘geografisch verschillend’ (nl. het gebruik van een infinitief of een voltooid deelwoord bij *komen*) als varianten die het label ‘regionaal’ krijgen (nl. het gebruik van *zo 'n* bij ontelbare en meervoudige nomina, en het gebruik van het groepsvormende *weten* ‘zich kunnen herinneren’), zodat we zowel varianten bestuderen die volgens ANS2 deel uitmaken van de standaardtaal als varianten die daarbuiten vallen (wat een invloed kan hebben op hun evaluatie door taalgebruikers, cf. vraag iii). Bij de als ‘regionaal’ gelabelde varianten zijn er verder zowel varianten die volgens de ANS2 beperkt zijn tot een land (nl. *weten* ‘zich kunnen herinneren’: komt voor in België), als varianten die beperkt zijn tot een regio groter dan een land (nl. *zo 'n* bij meervoudige en ontelbare nomina: komt voor in de zuidoostelijke provincies van Vlaanderen én Nederland). We verwachten dus zowel graduele als absolute verschillen te vinden. Op die manier hopen we verschillende types frequentieverschillen vast te stellen, wat ons toelaat om na te denken over de manier waarop die verschillende types moeten worden weergegeven in ANS3 (cf. vraag iv).

Als antwoord op vraag (i) bakenen we bij de bespreking van case A een aantal criteria af waaraan grammaticale varianten moeten voldoen om als equivalenten te kunnen worden beschouwd. Bij de bespreking van cases B en C zullen we vervolgens terugvallen op die criteria. In de vierde paragraaf wordt getracht een antwoord te formuleren op vragen (ii-iv), waarbij geregeld wordt terugverwezen naar de bevindingen uit de casestudy's, en waarbij ook een beroep wordt gedaan op de beschikbare literatuur.

### 3. DRIE CASES VAN GRAMMATICALE VARIATIE TUSSEN BN EN NN

#### 3.1. *Komen met een deelwoord of een infinitief als aanvulling*

##### 3.1.1. Beschrijving van de varianten

De eerste case heeft betrekking op het gebruik van *komen* in combinatie met een werkwoord dat, volgens ANS2, ‘de manier van komen’ aanduidt, zoals in (1).

- (1) Zonder kloppen *kwam* hij de kamer *binnengelopen/binnenlopen*. (ANS2 18.5.3)

De ANS2 (18.5.3) geeft aan dat die werkwoordelijke aanvulling zowel een voltooid deelwoord als een infinitief kan zijn, en onderscheidt drie types werkwoorden die als aanvulling kunnen dienen, namelijk bewegingswerkwoorden die samengesteld zijn met een bijwoord van richting (zoals in 1) of die gecombineerd worden met een bepaling van richting (zoals in 2) en werkwoorden die gecombineerd worden met het bijwoord *aan* (zoals in 3).

- (2) Als de baby maar even hilde, *kwam* vader naar boven *gesneld/snel-len*. (ANS2 18.5.3)
- (3) Van alle kanten *kwamen* mensen *aangelopen/aanlopen*. (ANS2 18.5.3)

De variatie tussen voltooid deelwoord en infinitief wordt in de ANS2 (18.5.3) beschreven als ‘geografisch verschillend’: in het zuiden van het taalgebied (“met name in België, maar in mindere mate ook in het zuiden van Nederland”) zou eerder gekozen worden voor het voltooid deelwoord, terwijl in de noordelijke helft (en dan vooral in het westen) eerder een infinitief wordt gebruikt. Volgens de ANS2 behoren beide varianten dus tot de standaardtaal in het hele taalgebied, en hetzelfde oordeel vinden we bij Taaladvies.net.<sup>9</sup> Taaladvies.net vermeldt niet waarop het zijn oordeel baseert, en in het relevante artikel wordt enkel verwezen naar de beschrijving in de ANS2; vermoedelijk heeft Taaladvies.net dus niet zelf corpus- en evaluatieonderzoek uitgevoerd naar de varianten omdat de ANS2 beide varianten als standaardtalig labelt.

Een belangrijke vraag is nu of de varianten beschouwd kunnen worden als equivalenten. Volgens de ANS2 (18.5.3) zijn ze namelijk niet in alle semantische en syntactische contexten inwisselbaar. Als *komen* zelf als vervangende infinitief optreedt, dus in de voltooide tijden, is er bijvoorbeeld enkel een infinitief mogelijk.<sup>10</sup> Terwijl zin (4) bijvoorbeeld welgevormd is volgens de ANS2, krijgt zin (5) het label “uitgesloten”.

- (4) Ze is hier vanmorgen al twee keer *komen voorbijfietsen*. (ANS2 18.5.3)
- (5) Ze is hier vanmorgen al twee keer *komen voorbijgefietst*. (ANS2 18.5.3)

<sup>9</sup> Over de variatie tussen infinitief en voltooid deelwoord vinden we bij Taaladvies.net informatie in het artikel ‘Komen aanrijden/aangereden’ ([https://taaladvies.net/taal/advies/vraag/458/komen\\_aanrijden\\_aangereden/](https://taaladvies.net/taal/advies/vraag/458/komen_aanrijden_aangereden/), geraadpleegd op 21 januari 2020).

<sup>10</sup> Broekhuis & Corver (2015: 995) signaleren dat sommige sprekers (vooral van zuidelijke variëteiten van het Nederlands) een voltooid deelwoord wel acceptabel vinden in de voltooide tijden als *komen* de vorm heeft van een voltooid deelwoord (m.a.w. als het IPP-effect niet optreedt) en voorafgaat aan het hulpwerkwoord van voltooidheid, bv. in *dat Jan de tuin in gefietst gekomen is* en *dat Jan aangefietst gekomen is*. In de ANS2 (18.5.3) wordt over die contexten niets vermeld.

Ook in bepaalde lexicale contexten is er volgens de ANS2 alleen een infinitief mogelijk, met name in de vaste uitdrukking *komen opdagen* (bv. zin 6) en ook wanneer de combinaties *komen aandragen/aansjouwen/aanslepen/aanzetten* met metaforisch gebruikt worden om uit te drukken dat iemand iets tevoorschijn haalt of te berde brengt zonder dat daar een verplaatsing aan te pas komt (bv. zin 7). Een voltooid deelwoord is in zin (7) niet uitgesloten, maar zou een letterlijke lezing van de zin afdwingen, waarin het subject zijn rijke oom verplaatst.

- (6) Na lang wachten kwam Jan eindelijk *opdagen*. (ANS2 18.5.3)
- (7) Hij kwam weer *aandragen* met zijn rijke oom in Amerika. (ANS2 18.5.3)

Binnen die contexten is er dus geen variatie mogelijk, maar verder is de keuze tussen de varianten volgens de ANS2 vrij: er worden geen semantische verschillen tussen de varianten vermeld. Nochtans claimt Beliën (2016) op basis van een (weliswaar zeer beperkt) corpusonderzoek dat beide varianten niet exact dezelfde semantische lading dekken: ze hebben gemeen dat ze “an unfolding, unidirectional motion event towards a contextually construable vantage point” beschrijven, maar het voltooid deelwoord benadrukt vooral het einde (en dus ook de voltooiing) van die beweging, terwijl dat bij de infinitief niet het geval is (Beliën 2016: 30). Daardoor is er volgens haar in bepaalde semantische contexten een duidelijke voorkeur voor een van beide varianten: als het pad voor de beweging erg kort is, ligt het voor de hand dat de beweging haar eindpunt bereikt en krijgt het voltooid deelwoord dus de voorkeur (zoals in zin 8). De infinitief komt dan weer vaker voor als de beweging de achtergrond vormt voor een simultane gebeurtenis (zoals in zin 9).

- (8) Na lang wachten werden toch de eerste barsten in de eieren zichtbaar en **kwamen** de kuikentjes **uit het ei gekropen**. (Beliën 2016: 28)<sup>11</sup>
- (9) Als er straks dertig skinheads **komen aanrennen**, springen wij ertussen. (Beliën 2016: 26)<sup>12</sup>

Beide varianten komen dus niet in dezelfde lexicale en/of syntactische contexten voor, en in bepaalde semantische contexten is er ook een voorkeur voor een van beide. Leiden die verschillen ertoe dat de infinitief en het voltooid deelwoord bij *komen* niet als equivalenten beschouwd kunnen worden?

<sup>11</sup> Als bron vermeldt Beliën (2016: 28) <http://www.sbo-evenaar.nl/nieuws.html> (geraadpleegd op 29 april 2016).

<sup>12</sup> Als bron vermeldt Beliën (2016: 26) *de Volkskrant* (22/09/2014).

### 3.1.2. Criteria voor equivalentie

Bij de Variantengrammatik noch bij Taaladvies.net is het duidelijk a.d.h.v. welke criteria precies bepaald wordt welke grammaticale varianten ook *equivalenten* zijn. Uit De Smet (2017) blijkt nochtans dat er discussie is over de vraag of volledige grammaticale equivalentie überhaupt bestaat. Hij schetst welke houding de functionalistische en variationistische traditie in hun meest radicale opvatting hebben tegenover situaties waarbij “a language has two or more ways of achieving the same communicative goal” (De Smet 2017: 122), zoals de variatie tussen infinitief en voltooid deelwoord bij *komen*, en laat zien dat ze een tegengestelde visie hebben op het bestaan van equivalenten. In de functionalistische benadering is variatie zelden of nooit semantisch neutraal; Bolinger (1968: 127) stelde bijvoorbeeld dat “a difference in syntactic form always spells a difference in meaning”. Echte grammaticale equivalenten bestaan in die visie dus niet, want varianten zijn niet inwisselbaar zonder (subtiel) betekenisverschil: bij *komen* legt de deelwoordvariant bijvoorbeeld altijd de klemtoon op de voltooiing van de beweging, en heeft hij dus niet exact dezelfde betekenis als de infinitiefvariant. Om semantische verschillen op het spoor te komen, onderzoeken functionalisten welke semantische contextfactoren een invloed hebben op de keuze tussen varianten of gaan ze met collocatieve analyses na met welke lexicale items varianten worden gecombineerd (De Smet 2017: 125). In de variationistische visie kunnen sommige grammaticale en lexicale varianten wel degelijk als equivalenten beschouwd worden, zelfs als er tussen die varianten semantische verschillen zijn: De Smet (2017: 124) verwijst bijvoorbeeld naar Sankoff (1988), die aangeeft dat zulke verschillen in context geneutraliseerd kunnen worden, zodat de varianten dus toch inwisselbaar kunnen zijn. Toegepast op de variatie bij *komen*: het verschil tussen infinitief en voltooid deelwoord kán bijvoorbeeld relevant zijn wanneer de beschreven beweging een achtergrondgebeurtenis vormt of het pad erg kort is, maar in andere contexten wordt abstractie gemaakt van dat verschil. Variationisten willen vooral nagaan welke extern-linguïstische factoren (bv. register of regio) een invloed hebben op de keuze tussen de equivalenten (De Smet 2017: 123).

De Smet (2017: 128) beschrijft een compromis tussen beide posities waarin varianten als equivalenten worden beschouwd als ze *near-synonyms* zijn: het zijn dan wel geen exacte synoniemen, maar in sommige contexten zijn ze wel inwisselbaar. Het is vaak moeilijk om alle contexten af te bakenen waarin de equivalenten inwisselbaar zijn en daarom, zo stelt De Smet (2017: 130), wordt er in onderzoek naar talige variatie vooral gefocust op *choice-dependent variation*, waarbij “a prior syntactic or lexical choice creates a variable context with a limited set of semantically-equivalent options.” Varianten kunnen dus beschouwd worden als equivalenten als het gaat om *near-synonyms* waarvoor

een lexicale en/of grammaticale context af te bakenen valt waarbinnen ze inwisselbaar zijn (al dan niet met subtiele semantische verschillen). Voor de variatie tussen voltooid deelwoord en infinitief bij *komen* kan relatieve tokenfrequentie dus ingezet worden om nationale verschillen in het gebruik van beide varianten op te sporen, zolang we de voltooide tijden en de metaforisch gebruikte combinaties van *komen* met *aandragen*, *aansjouwen*, *aanslepen* en *aanzetten* maar buiten beschouwing laten. Buiten die contexten zijn er weliswaar subtiele semantische verschillen mogelijk, maar daarvan maken we abstractie.

### 3.1.3. Dataverzameling en resultaten

Om zicht te krijgen op de verhouding tussen infinitief en voltooid deelwoord bij *komen* in beide nationale variëteiten, werden data verzameld in OpenSoNaR (Oostdijk et al. 2013).<sup>13</sup> In een eerste fase werd een overzicht opgesteld van werkwoorden die (als infinitief of voltooid deelwoord) voorkomen in de constructie met *komen*. Daartoe werden drie zoekopdrachten uitgevoerd, telkens in drie subcorpora van OpenSoNaR, nl. Chats, Discussion Lists en Newspapers, waarvan de omvang in aantal woorden weergegeven wordt in Tabel 1.

**Tabel 1: corpusgrootte van de collecties Chats, Discussion Lists en Newspapers in OpenSoNaR**

collectie	aantal woorden in BN component	aantal woorden in NN component	totaal aantal woorden
Chats	11 135 773	737 661	11 873 434
Discussion Lists	45 678 562	11 391 992	57 070 554
Newspapers	152 288 524	59 381 224	211 669 748

Er werd gezocht naar het lemma *komen*, onmiddellijk gevolgd of voorafgegaan door een voltooid deelwoord dat niet (pre)nominaal gebruikt is, en naar het lemma *komen*, onmiddellijk gevolgd door een infinitief die niet (pre)nominaal gebruikt is.<sup>14</sup> De zoekopdrachten werden beperkt tot resultaten die zich

<sup>13</sup> De data werden verzameld in het kader van de bachelorproef van Gauthier Delaby (2018).

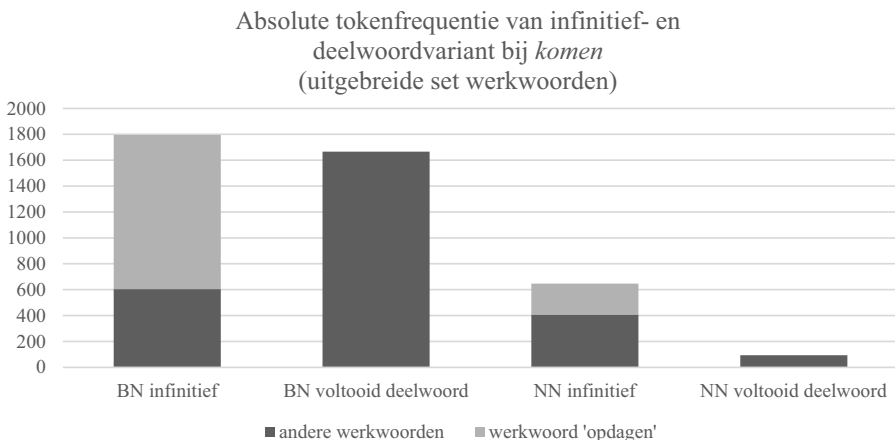
<sup>14</sup> Voor de combinatie van *komen* met een infinitief werd er niet gezocht naar de groene volgorde (waarbij *komen* voorafgegaan wordt door een infinitief). In bijzinnen is die volgorde gemarkeerd (bv. ...*dat mensen aanlopen kwamen*): in tweeledige werkwoordelijke eindgroepen bestaande uit een zelfstandig werkwoord en de persoonsvorm van een groepsvormend werkwoord behoort vooropplaatsing van de infinitief volgens ANS2 (18.5.7.4.i) namelijk tot het formele taalgebruik of tot het regionale taalgebruik (het komt meer bepaald in noordoostelijk Nederland voor) – alleen wanneer het groepsvormend werkwoord een van de modale hulpwerkwoorden *zullen*, *mogen*, *moeten*, *kunnen* of *willen* is, kan die volgorde volgens ANS2 ook standaardtaalig zijn. Ook in hoofdzinnen lijkt die volgorde – waarbij de infinitief dan op de eerste zinsplaats terecht komt – niet voor de hand te liggen (bv. *Aanlopen kwam hij*): vooropplaatsing van een deel van het werkwoordelijk gezegde kan volgens ANS2 (21.3.3) weliswaar in bv. correctiezinnen of om het werkwoordelijk element informatief op de achtergrond te laten treden, maar wijkt wel af van het normale zinspatroon en komt sowieso minder vaak voor bij infinitieven dan bij voltooide deelwoorden.



volledig binnen één zin bevinden. Die zoekopdracht leverde een lijst van 210 werkwoorden op die voorkomen in de constructie.

In een tweede fase werden voor elk van die werkwoorden telkens drie zoekopdrachten uitgevoerd in de Belgische en Nederlandse component van de collecties Newspapers en Discussion Lists. Voor elk van de werkwoorden werd gezocht naar (i) de persoonsvorm van het lemma *komen*, binnen de 25 woorden gevolgd of voorafgegaan door het voltooid deelwoord van het werkwoord, dat niet (pre)nominaal gebruikt is en (ii) de persoonsvorm van het lemma *komen*, binnen de 25 woorden gevolgd door de infinitief van het werkwoord, die niet (pre)nominaal gebruikt wordt. Opnieuw werden de zoekopdrachten beperkt tot resultaten binnen één zin. Er werd dus alleen gezocht naar gebruiken van *komen* als persoonsvorm: op die manier blijven zinnen waarbij *komen* fungeert als vervangende infinitief (in de voltooiden tijden) buiten beschouwing, aangezien in die contexten ook geen variatie mogelijk is. De zoekopdracht leverde in totaal 7806 hits op, waaruit alle irrelevante voorbeelden verwijderd werden.

Figuur 2 toont de verhouding tussen infinitief- en deelwoordvormen in BN en NN, in de collectie Newspapers. We beperken ons tot de resultaten voor de collectie Newspapers – in navolging van Taaladvies.net en de Variantengrammatik (cf. supra) beschouwen we krantentaal namelijk als representatief voor standaardtalige contexten. Bij de resultaten wordt een onderscheid gemaakt tussen voorkomens van *opdagen* en voorkomens van andere werkwoorden (aangezien volgens ANS2 enkel een infinitief mogelijk is bij *opdagen* – zoals overigens ook blijkt uit de resultaten). Tabel 2 biedt een overzicht van de absolute en relatieve tokenfrequenties.



**Figuur 2: absolute tokenfrequentie van infinitief- en deelwoordvormen bij *komen* in de BN- en NN-component van de collectie Newspapers van OpenSonaR**

**Tabel 2: absolute en relatieve tokenfrequentie van infinitief- of deelwoordvormen bij *komen* in de BN- en NN-component van de collectie Newspapers van OpenSoNaR**

	vorm →	infinitief	voltooid deelwoord	TOTAAL
	werkwoord ↓			
BN	<i>opdagen</i>	1191 (99.92%)	1 (0.08%)	<b>1192</b>
	<i>andere</i>	603 (26.59%)	1665 (73.41%)	<b>2268</b>
	<b>TOTAAL</b>	<b>1794 (51.85%)</b>	<b>1666 (48.15%)</b>	<b>3460</b>
NN	<i>opdagen</i>	239 (100.00%)	0 (0.00%)	<b>239</b>
	<i>andere</i>	407 (81.56%)	92 (18.44%)	<b>499</b>
	<b>TOTAAL</b>	<b>646 (87.53%)</b>	<b>92 (12.47%)</b>	<b>738</b>

Voor NN zijn er minder data, aangezien de collectie Newspapers voor NN opvallend kleiner is dan voor BN. Als we de verhouding tussen infinitief en voltooid deelwoord bij alle werkwoorden in BN en NN vergelijken, dan blijkt de verdeling significant verschillend in beide deelgebieden ( $\chi^2=316.75$ ,  $df=1$ ,  $p<0.0001$ , Cramer's  $V=0.2753$ ). Zowel in BN als in NN is er een voorkeur voor de infinitiefvariant, maar die voorkeur is meer uitgesproken in NN (87.53%) dan in BN (51.85%). Als we de resultaten voor *opdagen* buiten beschouwing laten, blijkt de verdeling nog altijd significant verschillend ( $\chi^2=530.97$ ,  $df=1$ ,  $p<0.0001$ , Cramer's  $V=0.439$ ), maar nu is er in BN een uitgesproken voorkeur voor het voltooid deelwoord (73.41%); in NN is de voorkeur voor de infinitief licht gedaald naar 81.56%.

Hierboven werd aangegeven dat er volgens de ANS2 nog een andere context is waarin er geen variatie mogelijk is, namelijk bij *komen aandragen/aansjouwen/aanslepen/aanzetten met*, wanneer die geen verplaatsing uitdrukken, maar metaforisch aangeven dat iets tevoorschijn wordt gehaald of te berde wordt gebracht. Van die vier combinaties treffen we enkel *komen aanzetten* aan tussen de resultaten.

**Tabel 3: absolute en relatieve tokenfrequentie van infinitief- en deelwoordvormen van het werkwoord *aanzetten* bij *komen* in de BN- en NN-component van de collectie Newspapers van OpenSoNaR**

	vorm →	infinitief	voltooid deelwoord	TOTAAL
	werkwoord ↓			
BN	<i>aanzetten</i>	83 (97.65%)	2 (2.35%)	<b>85</b>
NN	<i>aanzetten</i>	58 (96.67%)	2 (3.33%)	<b>60</b>

Zoals Tabel 3 aangeeft, komt dat werkwoord in de meerderheid van de gevallen in de infinitiefvorm voor (bijvoorbeeld in zinnen 10-12), maar toch ook vier keer in de deelwoordvorm, zoals in de zinnen (13-14). In de deelwoordvorm blijkt in de zin inderdaad steeds een ruimtelijke interpretatie mogelijk, in de infinitief-

vorm vinden we zowel ruimtelijke als metaforische interpretaties.

- (10) In mijn dagelijks leven erger ik mij enorm aan winkelwagentjesagressie. **Komen** mensen van achter in de rij **aanzetten** en planten ze zo'n wagentje gewoon voor je. ('Binnenhof over middelvingers en winkelwagentjes', *Algemeen Dagblad* (11/12/2003))
- (11) Er zijn zaken waar ik absoluut niet tussen kom. Daar ben ik duidelijk over. Mensen die in rechtszaken verwikkeld zijn of met een verkeersboete **komen aanzetten**, daar kan en wil ik niets mee te maken hebben. ('Een les in nederigheid', *Trouw* (16/04/2003))
- (12) In De Vlieger **komen** de kinderen dikwijls zelf met eigen thema's **aanzetten**. ('Leerlingen leefschool De Vlieger 'trouwen' in stadhuis', *Het Laatste Nieuws* (21/02/2006))
- (13) Toen stond hij daar in het landschap als steenrots, trots. Tot daar een jonge steenhouwer **kwam aangezet** met een groot houweel. Hij begon te houwen en te hakken in de rots en sloeg er grote brokken uit. ('Korte berichten', *Het Laatste Nieuws* (22/07/2006))
- (14) De rits die het zwempak van achteren hoog en strak moest sluiten, was eruit gerukt. Assistent Hans Schwarte **kwam** met een nieuw zwartrood hansop **aangezet** en dat werd onder water razendsnel aangetrokken door Quint. ('Onderwateroorlog eindigt in 8-8', *de Volkskrant* (18/07/2003))

Die voorkeur voor de infinitief bij metaforische betekenissen treffen we echter niet alleen aan bij *komen aanzetten*. Uit de resultaten van een distinctieve collexeemanalyse blijkt namelijk dat de werkwoorden die aangetrokken worden tot de infinitiefvariant vooral werkwoorden zijn die in combinatie met *komen* vaak een metaforische beweging uitdrukken. Het gaat daarbij om werkwoorden zoals *aandrijven*, *bovendrijven*, *opdraven*, *opduiken* en *bijspringen*, die metaforisch kunnen uitdrukken dat iets tevoorschijn komt of gebracht wordt, zoals in (15-19).

- (15) De tabloid News of the World **kwam aandrijven** met het verhaal van zijn minnares, Rebecca Loos, en een week later meldde ook model Sarah Marbeck dat ze twee keer het bed deelde met de blonde god. ('De toernooziekte van David Beckham', *De Morgen* (01/07/2006))
- (16) Alles verliep immers naar wens zolang het resultaat goed was, maar eens we in een wedstrijd, en al te vaak voor eigen publiek, op achterstand kwamen, **kwamen** de kleine strubbelingen **bovendrijven** en bakten we er niets meer van. ('Iedereen knokt voor elkaar', *Het Laatste Nieuws* (22/09/2006))
- (17) Ook Brugge **kwam** niet met het verwachte elftal **opdraven**. ('Tchité orgelpunt van feestweek', *Het Nieuwsblad* (28/08/2006))

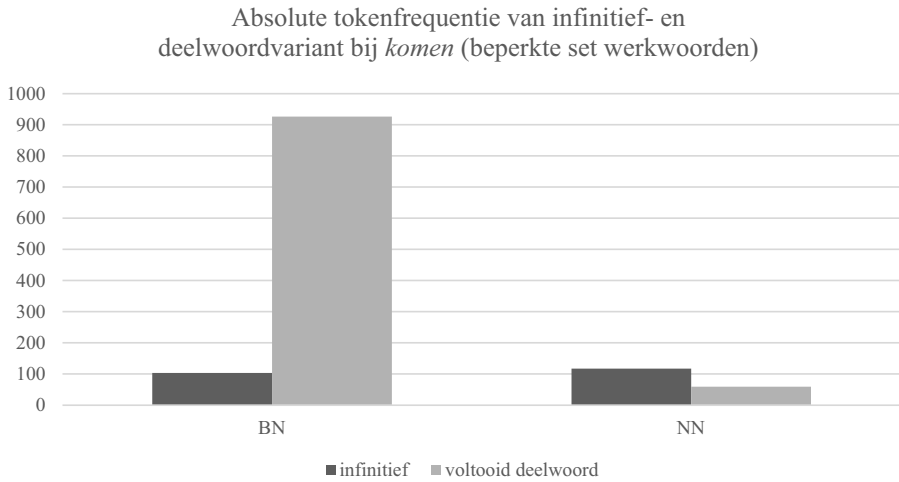
- (18) Op zich zonder enig belang, maar na de veldloop in Genk **kwamen** ontstekingsverschijnselen **opduiken**. ('Jos Maes, eeuwig veldloper', *Het Nieuwsblad* (21/12/2006))
- (19) In het verleden **kwam** Zelzate in noodgevallen **bijspringen**, maar soms duurt het achttien minuten voor ze het centrum bereiken. ('Brandweer kan alle hoogtes aan', *Het Nieuwsblad* (17/10/2006))

Nu is het niet duidelijk hoe sterk de voorkeur is voor een infinitief bij zulke metaforische betekenissen: is er bij metaforische betekenissen ook een mogelijkheid om voor een voltooid deelwoord te kiezen?<sup>15</sup> Volgens de ANS2 (18.5.3) is dat niet het geval bij *komen aandragen/aansjouwen/aanslepen/aanzetten met*, maar het is niet duidelijk of dat als een exhaustieve lijst bedoeld is. Om de invloed van de factor 'ruimtelijke/metaforische betekenis' te kunnen inschatten, zou voor alle resultaten in de dataset moeten worden aangegeven of het werkwoord een ruimtelijke verplaatsing of een metaforisch tevoorschijn komen uitdrukt. Dat is geen sinecure, want het onderscheid tussen die twee is eerder gradueel dan absoluut: in zin (11) is het bijvoorbeeld al onduidelijk of daar een verplaatsing plaatsvindt of niet, en zelfs als er geen verplaatsing plaatsvindt, blijft de verplaatsingsbetekenis vaak toch op de achtergrond aanwezig (bijvoorbeeld in 17). Tegelijkertijd zijn onze resultaten onbetrouwbaar als we metaforisch gebruikte werkwoorden in de dataset houden, zonder dat we weten welke invloed de factor 'ruimtelijke/metaforische betekenis' heeft op de keuze tussen infinitief- en deelwoordvorm. Om dat probleem te omzeilen, werden een aantal testwerkwoorden geselecteerd, nl. *fietsen, lopen, rennen, rijden, vliegen, wandelen* (daaronder verstaan we zowel de enkelvoudige werkwoorden als de scheidbaar samengestelde werkwoorden die gevormd worden door die werkwoorden te combineren met *aan-, af-, binnen-, buiten-, in-, langs-, op-, over-, tegemoet-, toe-, uit-, voor- en voorbij-*). Die werkwoorden vertegenwoordigen een behoorlijk groot deel van de resultaten in BN en NN en worden in de dataset uitsluitend<sup>16</sup> in ruimtelijke betekenis gebruikt, zodat metaforische betekenissen de resultaten niet kunnen vertekenen. Figuur 3 toont de verdeling tussen infinitief- en deelwoordvariant bij die testwerkwoor-

<sup>15</sup> Een reviewer merkt terecht op dat het onderscheid tussen een metaforische en een letterlijke betekenis niet volledig onafhankelijk is van het onderscheid tussen een beweging waarbij het eindpunt bereikt wordt en een waarbij dat niet het geval is. Aangezien er bij de metaforische betekenissen (bv. in 15-19) geen sprake is van een echte beweging, kan er ook geen echt eindpunt bereikt worden (wat mogelijk ook verklaart waarom er in die metaforische gebruiken een voorkeur is voor de infinitief). Toch valt het onderscheid tussen een metaforische en letterlijke betekenis niet samen met het onderscheid tussen een beweging met en zonder eindpunt: bij letterlijke gebruiken is er nl. variatie in de mate waarin een eindpunt/voltooiing van de beweging gesuggereerd wordt. In onderstaande zin ligt bv. een lezing voor de hand waarin de beweging niet voltooid is. *Zijn motor sloeg af en toen kwamen ze aanrennen, de boeren, met stokken en knuppels uit de weilanden.* ('Na die gijzeling door boeren heeft hij dagenlang gehuild', krant en publicatiedatum niet gespecificeerd in SoNaR, maar NN)

<sup>16</sup> Bij twee voorkomens van *fietsen* een één voorkomen van *aanvliegen* werd geen verplaatsing uitgedrukt. Die hits werden buiten beschouwing gelaten.

den, Tabel 4 biedt een overzicht van de absolute en relatieve tokenfrequentie van die varianten bij de verschillende werkwoorden.



**Figuur 3: absolute tokenfrequentie voor de infinitief- en deelwoordvorm bij *komen* in de BN- en NN-component van de collectie Newspapers van OpenSoNaR, voor de werkwoorden *fietsen, lopen, rennen, rijden, vliegen, wandelen* (en hun samengestelde vormen)**

**Tabel 4: absolute en relatieve tokenfrequentie voor de infinitief- en deelwoordvorm bij *komen* in de BN- en NN-component van de collectie Newspapers van OpenSoNaR, voor de werkwoorden *fietsen, lopen, rennen, rijden, vliegen, wandelen* (en hun samengestelde vormen)**

	vorm →	infinitief	voltooid deelwoord	TOTAAL
	werkwoord (enkelvoudige en samengestelde vormen ↓)			
BN	<i>fietsen</i>	2 (12.50%)	14 (87.50%)	<b>16</b>
	<i>lopen</i>	7 (3.52%)	192 (96.48%)	<b>199</b>
	<i>rennen</i>	3 (23.08%)	10 (76.92%)	<b>13</b>
	<i>rijden</i>	76 (11.36%)	593 (88.64%)	<b>669</b>
	<i>vliegen</i>	8 (11.76%)	60 (88.24%)	<b>68</b>
	<i>wandelen</i>	6 (9.68%)	56 (90.32%)	<b>62</b>
	<b>TOTAAL</b>		<b>102 (9.93%)</b>	<b>925 (90.07%)</b>
NN	<i>fietsen</i>	3 (33.33%)	6 (66.67%)	<b>9</b>
	<i>lopen</i>	42 (60.00%)	28 (40.00%)	<b>70</b>
	<i>rennen</i>	16 (76.19%)	5 (23.81%)	<b>21</b>
	<i>rijden</i>	47 (85.45%)	8 (14.55%)	<b>55</b>
	<i>vliegen</i>	3 (33.33%)	6 (66.67%)	<b>9</b>
	<i>wandelen</i>	5 (50.00%)	5 (50.00%)	<b>10</b>
	<b>TOTAAL</b>		<b>116 (66.67%)</b>	<b>58 (33.33%)</b>

Als we de verhouding tussen infinitief en deelwoord bij die testwerkwoorden in beide variëteiten vergelijken, dan blijkt die significant te verschillen ( $\chi^2=318.56$ ,  $df=1$ ,  $p<0.0001$ , Cramer's  $V=0.5181$ ). In NN is er een lichte voorkeur (66.67%) voor de infinitiefvariant, terwijl er in BN een uitgesproken voorkeur (90.07%) is voor de deelwoordvorm. Bekijken we de individuele werkwoorden, dan zien we in BN bij alle testwerkwoorden een sterke voorkeur voor het voltooid deelwoord; in NN is er bij *lopen*, *rennen* en *rijden* een uitgesproken voorkeur voor de infinitiefvorm. In beide variëteiten is het zo dat het aandeel infinitieven lager is bij deze testwerkwoorden met exclusief ruimtelijke betekenissen dan wanneer alle werkwoorden behalve *opdagen* in rekening gebracht worden (zoals in Tabel 2). In BN wordt bij deze werkwoorden zelfs maar in 9.93% van de gevallen een infinitief gebruikt. Zowel in Tabel 2 als Tabel 4 vinden we beide varianten evenwel terug in het hele taalgebied, zij het in verschillende mate; het gaat hier dus om een gradueel verschil tussen BN en NN. In de vierde paragraaf zullen we nagaan hoe goed deze resultaten ons inzicht bieden in de nationale variatie in het gebruik van beide varianten, en hoe de resultaten kunnen worden vertaald naar nationale labels in ANS3.

### 3.2. Zo'n of zulk(e) bij ontelbare en meervoudige nomina

#### 3.2.1. Beschrijving van de varianten

Het tweede geval van diatopische variatie dat we hier onderzoeken betreft de combinatiemogelijkheden van de aanwijzende voornaamwoorden *zo'n* en *zulk(e)*. In standaardtaal zijn de combinatiemogelijkheden van die twee voornaamwoorden complementair volgens de ANS2 (5.6.6): “[v]oor substantieven in het enkelvoud die met het onbepaald lidwoord *een* gecombineerd kunnen worden, wordt *zo'n* gebruikt, in andere gevallen *zulk(e)*.” Hoewel het niet expliciet op die manier verwoord wordt, lijkt dat erop neer te komen dat *zulk(e)* combineert met ontelbare en meervoudige substantieven (*zulk hout*, *zulk vee*, *zulke mannen*, *zulke huizen*, *zulke dingen*) en *zo'n* met telbare, enkelvoudige nomina (*zo'n man*, *zo'n huis*, *zo'n ding*).

Nu duikt *zo'n* volgens de ANS2 (5.6.6) in “een groot zuidoostelijk deel van het Nederlandse taalgebied” ook op in contexten waar in standaardtaal alleen *zulk(e)* mogelijk is. Als voorbeelden worden de zinnen (20-22) gegeven, waarin *zo'n* gecombineerd wordt met meervoudige en ontelbare nomina. Die gebruiken van *zo'n* krijgen het label ‘regionaal’ taalgebruik.

- (20) *Zo'n* dingen zeggen ze hier niet. (ANS2 5.6.6)
- (21) *Zo'n* vee zie je bij ons niet. (ANS2 5.6.6)
- (22) Vind je *zo'n* donker hout mooi? (ANS2 5.6.6)

Wat “een groot zuidoostelijk deel van het Nederlandse taalgebied” precies is, wordt hier niet gespecificeerd, waardoor het niet duidelijk is of de variant beperkt is tot Vlaanderen, of dat hij ook in Nederland voorkomt. Waarschijnlijk bedoelt de ANS2 echter het zuidoosten van Vlaanderen én Nederland, want de beschrijving van de geografische verspreiding werd vermoedelijk ontleend aan De Rooij (1989). Hij deed op basis van dialectonderzoek naar het gebruik van *zo'n/zulk(e)* namelijk de volgende suggestie voor de herziening van ANS1:

Als in een herziene druk van de ANS [ANS1, AD] t.a.v. het gebruik van *zo'n* voor standaardtalig *zulk(e)* een nadere geografische specificatie wordt gegeven van de kwalifikatie “regionaal”, kan die b.v. zijn: “een groot zuidoostelijk deel van het Nederlandse taalgebied”, of wat specifieker: “Vlaams-Brabant, Antwerpen, Oost-Noord-Brabant en de beide Limburgen, met uitlopers naar de aangrenzende gebieden”. (De Rooij 1989: 198)

Net zoals bij de variatie tussen infinitief en voltooid deelwoord bij *komen* gaat het hier dus niet om een absoluut nationaal verschil: het gebruik van *zo'n* bij meervoudige en ontelbare nomina stopt niet aan de rijksgrens. Terwijl de beide varianten bij *komen* echter in het hele taalgebied voorkwamen, komt de variant met *zo'n* dit keer enkel in het zuidoosten van het taalgebied voor. De vraag rijst hoe dat verschil weerspiegeld wordt in nationale corpora.

Bij de diatopische variatie tussen *zo'n* en *zulk(e)* is er echter nog een complicerende factor: de ANS2 (5.6.6) laat zien dat *zo'n* in standaardtaal wel degelijk gecombineerd kan worden met ontelbare nomina, maar alleen als het een intensiverende betekenis heeft, zoals in (23), en geen identificerende betekenis, zoals in (20-22).

(23) Hij had *zo'n* pijn dat hij er niet van kon slapen. (ANS2 5.6.6)

De verschillende betekenissen van *zo'n* en *zulk(e)* worden onder meer bij Van Olmen & Van der Auwera (2014: 215-216) beschreven. Zij geven aan dat *zo'n* en *zulk(e)* beide zowel een identificerende als een intensiverende functie kunnen hebben. In hun identificerende functie drukken ze “een gelijkenis uit van een onbepaald token met een bepaald type waarvoor een bepaald token als voorbeeld dient” (Van Olmen & Van der Auwera 2014: 215). In (24) bijvoorbeeld verwijst *zo'n commentaar op de islam* naar commentaar van het type waarvoor het commentaar van de journalist uit Istanbul als voorbeeld dient. Bij *zulk* identificerend gebruik is er dus altijd een (anaforische of cataforische) relatie met een referent in de context (Ghesquière & Van de Velde 2011: 772). Zin (25) bevat een intensiverend *zo'n*. Intensivering omschrijven Ghesquière & Van de Velde (2011: 782) als “the measuring or heightening of the degree

of a certain evaluative or scalar notion inherent in the modified element.” In (25) geeft *zo 'n* inderdaad aan dat de spreker een hoge mate van haat voelt.

- (24) Een journalist uit Istanbul kreeg er in 1998 een jaar gevangenisstraf, omdat hij had geschreven dat de islam sociale onrechtvaardigheid legitimeert door ze als “Gods wil” voor te stellen. Volgens het Europees Hof moet **zo'n commentaar op de islam** worden getolereerd. (‘Aanval op godsdienst moet kunnen’, *De Standaard* (07/06/2006))
- (25) Aan het Astridplein in Antwerpen overheersten echter radeloze woede en onversneden wrok. “Ik heb **zo'n haat** gekregen, ik ben echt een racist geworden.” (‘De afbreker breekt af’, *De Morgen* (05/07/2006))

Volgens Van Olmen & Van der Auwera (2014) kent *zo 'n* bovendien nog twee functies die *zulk(e)* niet kent, namelijk de benaderingsfunctie in (26) en de herkenningfunctie in (27). In (26) verschijnt *zo 'n* met de betekenis ‘ongeveer’ bij een telwoord. In (27) verwijst *zo 'n* *puber* naar een prototypische puber, een puber zoals iedereen er volgens de spreker wel een kent, zonder dat er in de context een voorbeeld gegeven wordt van dat type puber (Van Olmen & Van der Auwera 2014: 219).

- (26) Het gaat goed met de koolmees. Er vliegen er tegenwoordig **zo'n miljoen** in Nederland rond, een verdubbeling sinds 1970. (‘Het stille creperen in een opgewarmd Nederland’, *Trouw* (13/11/2004))
- (27) Maar op de HAVO werd ik echt een ramp. Word ik echt wel uh was ik echt **zo'n puber** weet je wel. Echt best wel lastige lastige meid was ik. (Van Olmen & Van der Auwera 2014: 219)<sup>17</sup>

Er is dus diatopische variatie tussen *zo 'n/zulk(e)* bij meervoudige nomina en ook bij ontelbare nomina, maar bij die laatste categorie alleen als het voorwoord identificerend wordt gebruikt. De case verschilt op drie vlakken van de variatie tussen infinitief en deelwoord bij *komen*. Ten eerste zijn er hier geen aanwijzingen dat er semantische verschillen zijn tussen de varianten, waardoor er geen twijfel lijkt te bestaan dat de varianten als equivalenten mogen worden beschouwd. Verder krijgen de beide varianten (*zo 'n* en *zulke*) in deze case niet dezelfde status toegekend: het gebruik van *zulke* bij meervoudige en ontelbare nomina is in standaardtalige contexten onproblematisch, maar het gebruik van *zo 'n* bij meervoudige en ontelbare nomina wordt door ANS2 als ‘regionaal’ gelabeld, wat inhoudt dat die variant buiten de standaardtaal valt. Ook op Taaladvies.net krijgen *zo 'n* en *zulk(e)* niet dezelfde status: de combinatie van *zo 'n* met ontelbare nomina is volgens de taaladviesite enkel standaardtaal in België, en voor de combinatie van *zo 'n* met een meer-

<sup>17</sup> Van Olmen & Van der Auwera (2014: 219) ontlene hun voorbeeld aan de component ‘spontane conversaties’ van het Corpus Gesproken Nederlands.



voudig substantief staat er te lezen dat het onduidelijk is of die tot de standaardtaal in België behoort: hoewel bij meervoudige nomina veelvuldig *zo 'n* gebruikt wordt, “ook door veel standaardtaalsprekers”, is er toch “een niet te verwaarlozen groep taalgebruikers die dergelijke combinaties afkeurt.”<sup>18</sup> Taaladvies.net vermeldt niet expliciet waarop het die uitspraken baseert, maar uit de formulering van het advies valt af te leiden dat hier een beroep werd gedaan op corpus- en evaluatiedata. Een derde verschil is dat beide varianten dit keer niet voorkomen in het hele taalgebied, althans volgens ANS2; *zo 'n* is beperkt tot de zuidoostelijke provincies.

### 3.2.2. Dataverzameling en resultaten

Om na te gaan hoe frequent de combinaties zijn, verzamelden we data in de BN- en NN-component van de collectie Newspapers van OpenSoNaR. In beide componenten werd gezocht naar combinaties van *zo 'n* en *zulk(e)*, onmiddellijk gevolgd door een substantief. Dat leverde 29 595 resultaten op voor BN en 16 881 voor NN, te veel om manueel te verwerken. Daarom werd voor beide variëteiten een random sample van 2000 hits samengesteld. Alle treffers werden handmatig gecontroleerd; hits waarin *zo 'n* in de benaderingsfunctie (zoals in 26) of herkenningfunctie (zoals 28) wordt gebruikt, of hits met de vaste verbinding *zo 'n beetje* (zoals in zin 29) werden niet meegenomen in de verdere annotatie.

- (28) Zalmroze vind ik nu niet meteen de mooiste kleur voor een koerstrui, zegt Roy Sentjens, de nieuwe aanwinst van het team. Ze gaan er in het peloton eens goed mee lachen, vrees ik. Zeker als het echt **zo 'n baby-roze** wordt, zegt Johan Vansummeren. ('Renners zijn er niet gerust op', *Het Laatste Nieuws* (08/11/2006))
- (29) Hij is als schrijver **zo 'n beetje** aan het jongleren, meer niet. ('Vloed tegen kruis en keel', *Trouw* (04/10/2003))

Vervolgens werd iedere hit geannoteerd: er werd aangegeven welk voornaamwoord (*zo 'n* of *zulk(e)*) hij bevatte en met welk type nomen het voornaamwoord werd gecombineerd: een enkelvoudig telbaar nomen, een meervoudig telbaar nomen of een ontelbaar nomen.

Bij de annotatie was er soms twijfel of een nomen telbaar of ontelbaar was. Dat is op zich niet verwonderlijk: zoals Joosten (2003: 164-168) aangeeft, is de telbaarheid van individuele nomina namelijk geen binaire, maar een graduele kwestie. Om nomina toch dichotomisch als telbaar of ontelbaar te kunnen

<sup>18</sup> Over de variatie tussen *zo 'n* en *zulk(e)* bij meervoudige en ontelbare nomina vinden we bij Taaladvies.net informatie in het artikel 'Zo 'n / zulke mensen; zo 'n / zulk bier' ([https://taaladvies.net/taal/advies/vraag/944/zo\\_zulke\\_mensen\\_zo\\_n\\_zulk\\_bier/](https://taaladvies.net/taal/advies/vraag/944/zo_zulke_mensen_zo_n_zulk_bier/), geraadpleegd op 21 januari 2020).

labelen, werd besloten om één criterium te hanteren: als een nomen (op basis van de eigen intuïtie van de eerste auteur) gecombineerd kan worden met het onbepaald lidwoord *een*, wordt het als telbaar beschouwd (merk op dat de combinatiemogelijkheid met *een* ook door de ANS2 (5.6.6) genoemd wordt als criterium om uit te maken of *zo 'n* mogelijk is). Aangezien de context vaak een invloed heeft op de mogelijkheid om *een* toe te voegen, werd die ook telkens in acht genomen. In (30) verwijst *zo 'n nalatigheid* bijvoorbeeld niet naar één specifiek voorval, waardoor het moeilijker is om *een* toe te voegen, zodat die hit dus het label ‘ontelbaar’ krijgt. In (31) daarentegen verwijst *zo 'n schaalvergroting* wel naar één concreet geval van schaalvergroting, en krijgt de hit dus het label ‘telbaar’.

- (30) Controle op het bestuur van Lierse was onmogelijk, want tussen 2002 en november 2005 is geen enkele bestuurswissel officieel bekendgemaakt bij de rechtbank van koophandel. Voor **zo'n nalatigheid** moet de voetbalbond volgens het reglement een licentie aan een club weigeren. ('Ik wil een tweede kans', *Het Laatste Nieuws* (28/10/2006))
- (31) In het verlengde van de verhoopte samenwerking tussen België, Nederland en Luxemburg ontstaat opnieuw het idee om tot één gezamenlijke voetbalcompetitie te komen. "De voordelen van **zo'n schaalvergroting** zijn duidelijk." ('Beneluxpolitici willen WK 2018 binnenhalen', *De Morgen* (04/12/2006))

Enkel bij de ontelbare nomina werd nog een extra annotatielaag toegevoegd waarin werd aangegeven of het voornaamwoord een identificerende of intensiverende functie had. Net zoals Van Olmen (2019) kenden we het label 'identificerend' toe als er een referent in de context genoemd werd waarnaar de constituent met *zo 'n/zulke* terug- of vooruitverwijst, zoals in (30-31). Resultaten kregen het label 'intensiverend' als er sprake was van "the existence of some gradable notion in the noun phrase introduced by *zo 'n* and the presence of other indications of subjectivity in the context" (Van Olmen 2019: 227), zonder dat daarbij vooruit- of terugverwezen werd naar een referent in de context, zoals in (25). Bij een combinatie van intensiverende en identificerende functie, zoals in (32), werd het label 'ambigu' gebruikt.

- (32) [...] hij zei dat de burgerlijke partijen hun hoop nu moesten vestigen op een eventuele misstap van Annick De Gucht. De enige manier om haar op te pakken is na het plegen van strafbare feiten. Moet de maatschappij nu al **zulke onzin** hopen? ('Advocaat verrast door commotie rond vrijlating moordende moeder', *Het Nieuwsblad* (02/05/2006))

Tabel 5 geeft een overzicht van de resultaten van het corpusonderzoek. De ontelbare nomina worden in Tabel 6 nog verder onderverdeeld naargelang van

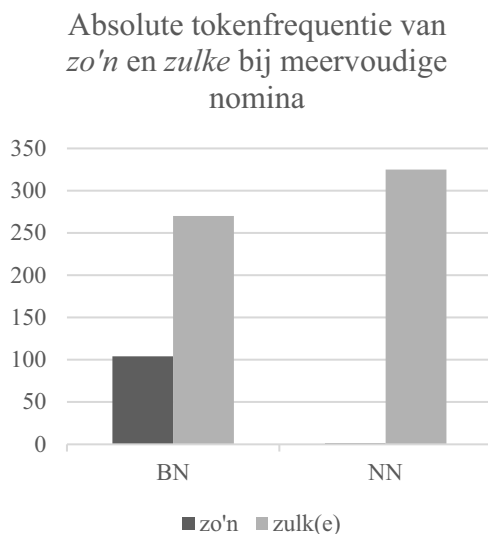
de semantische functie van het voornaamwoord. In de contexten waarin we geïnteresseerd zijn, de meervoudige en ontelbare nomina met identificerend voornaamwoord, is er inderdaad een verschillende verdeling tussen *zo'n* en *zulk(e)* in de beide nationale variëteiten, zoals we ook kunnen aflezen uit Figuren 4 en 5.

**Tabel 5: absolute en relatieve tokenfrequentie voor *zo'n* en *zulk(e)* in verschillende syntactische contexten in een sample van 2000 hits voor de zoekopdracht '*zo'n/zulk(e)* + nomen' in de BN- en NN-component van de collectie Newspapers van OpenSoNaR**

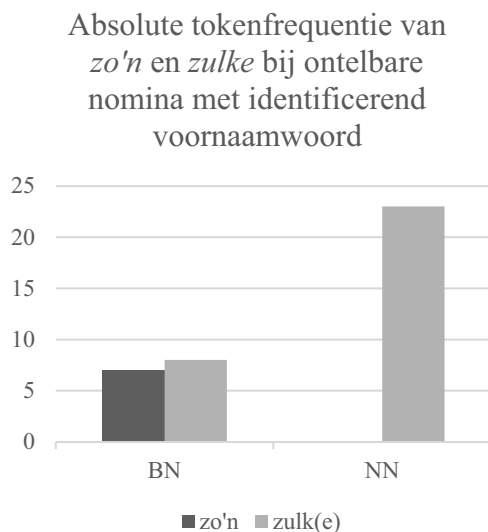
	voornaamwoord →	zo'n	zulk(e)	TOTAAL
	syntactische context ↓			
BN	enkelvoudig, telbaar nomen	1497 (99.47%)	8 (0.53%)	1505
	meervoudig nomen	104 (27.81%)	270 (72.19%)	374
	ontelbaar nomen	38 (74.51%)	13 (25.49%)	51
	TOTAAL	1639	291	1930
NN	enkelvoudig, telbaar nomen	1554 (99.81%)	3 (0.19%)	1557
	meervoudig nomen	1 (0.30%)	325 (99.70%)	326
	ontelbaar nomen	24 (46.15%)	28 (53.85%)	52
	TOTAAL	1579	356	1935

**Tabel 6: absolute en relatieve tokenfrequentie voor *zo'n* en *zulk(e)* bij ontelbare nomina (ingedeeld naargelang van de semantische functie van het voornaamwoord), in een sample van 2000 hits voor de zoekopdracht '*zo'n/zulk(e)* + nomen' in de BN- en NN-component van de collectie Newspapers van OpenSoNaR**

	voornaamwoord →	zo'n	zulk(e)	TOTAAL
	semantische functie ↓			
BN	identificerend	7 (46.67%)	8 (53.33%)	15
	ambigu	12 (75.00%)	4 (25.00%)	16
	intensiverend	19 (95.00%)	1 (5.00%)	20
	TOTAAL	38	13	51
NN	identificerend	0 (0.00%)	23 (100.00%)	23
	ambigu	2 (28.57%)	5 (71.43%)	7
	intensiverend	22 (100.00%)	0 (0.00%)	22
	TOTAAL	24	28	52



**Figuur 4:** absolute tokenfrequentie voor *zo'n* en *zulk(e)* bij meervoudige nomina in een sample van 2000 hits voor de zoekopdracht '*zo'n/zulk(e)* + nomen' in de BN- en NN-component van de collectie Newspapers van OpenSoNaR



**Figuur 5:** absolute tokenfrequentie voor *zo'n* en *zulk(e)* bij ontelbare nomina met identificerend voornaamwoord in een sample van 2000 hits voor de zoekopdracht '*zo'n/zulk(e)* + nomen' in de BN- en NN-component van de collectie Newspapers van OpenSoNaR

Bij een meervoudig nomen komt in NN slechts een keer *zo 'n* voor (in 33), terwijl dat in BN in 27.81% van de gevallen gebruikt wordt in plaats van *zulke*.<sup>19</sup> Dat verschil blijkt statistisch significant ( $\chi^2=101.17$ ,  $df=1$ ,  $p<0.0001$ , Cramer's  $V=0.3842$ ).

- (33) Het gebrek aan kennis over wetten en regels rond reïntegratie speelt het MKB vaak parten. 'Grotebedrijven (sic) hebben personeelsafdelingen die **zo'n zaken** bijhouden. MKB-werkgevers niet', weet Brouwer. ('Het MKB zucht onder reïntegratie van zieken', *Algemeen Dagblad* (15/10/2003))

Bij ontelbare nomina met identificerend voornaamwoord wordt in NN nooit *zo 'n* gebruikt, terwijl dat in BN bij 46.67% van de relevante hits het geval is. Omdat we echter maar een beperkt aantal hits hadden voor ontelbare nomina met identificerend voornaamwoord, hebben we voor die combinatie getracht om extra data te verzamelen. Daarvoor voerden we in de BN- en NN-component van de collectie Newspapers van OpenSoNaR een nieuwe zoekopdracht uit, dit keer naar *zo 'n/zulk(e)*, onmiddellijk gevolgd door een van de nomina in (34):

- (34) \*werk, antiek, artillerie, **bemoeienis**, bezoek, brandhout, **contact**, fruit, gebak, gebladerte, gereedschap, \*gerei, geschut, gevogelte, \*goed, grut, \***hulp**, ijzer, **informatie**, keramiek, kleding, lectuur, lingerie, materiaal, **muziek**, **nieuws**, ongedierte, patisserie, porselein, post, **reclame**, vee, **veiligheid**, volk, \*waar, wild, zilver

Die nomina werden geselecteerd op basis van de set nomina die in de oorspronkelijke dataset als ontelbaar gelabeld werden, en op basis van de set nomina die in het corpusonderzoek van Joosten (2003: 169-183) uitsluitend in ontelbare contexten bleken voor te komen. Nomina waarbij de kans groot leek dat ze met een intensiverend voornaamwoord zouden worden gecombineerd, werden buiten beschouwing gelaten. De vetgedrukte nomina zijn afkomstig uit onze eigen dataset, de andere uit Joosten (2003). Bij nomina met een asterisk is zowel gezocht naar het nomen zelf als naar samenstellingen waarin het nomen als rechterlid fungeert. Na filtering en annotatie kwamen we tot de resultaten in Tabel 7.

<sup>19</sup> In het artikel zijn er geen aanwijzingen dat de man die geciteerd wordt in (33) *geen* spreker van NN zou zijn.

**Tabel 7: absolute en relatieve tokenfrequentie voor *zo'n* en *zulk(e)* in verschillende semantische functies bij de ontelbare nomina in de resultaten voor de zoekopdracht naar *zo'n/zulk(e)* onmiddellijk gevolgd door de nomina in (34) in de BN- en NN-component van de collectie Newspapers van OpenSoNaR**

	voornaamwoord →	<i>zo'n</i>	<i>zulk(e)</i>	TOTAAL
	semantische functie ↓			
BN	identificerend	20 (50.00%)	20 (50.00%)	40
	ambigu	2 (100.00%)	0 (0.00%)	2
	intensiverend	3 (100.00%)	0 (0.00%)	3
	TOTAAL	25 (55.56%)	20 (44.44%)	45
NN	identificerend	4 (8.51%)	43 (91.49%)	47
	ambigu	0 (0.00%)	0 (0.00%)	0
	intensiverend	1 (100.00%)	0 (0.00%)	1
	TOTAAL	5 (10.42%)	43 (89.58%)	48

De verhoudingen tussen identificerend *zo'n/zulk(e)* bij ontelbare nomina in Tabel 7 lijken overeen te stemmen met die uit de beperktere dataset: in BN wordt in 50% van de gevallen *zo'n* gebruikt, terwijl dat in NN slechts in 8.51% van de relevante contexten het geval is. Dat verschil is statistisch significant ( $\chi^2=16.6$ ,  $df=1$ ,  $p<0.0001$ , Cramer's  $V=0.4626$ ).

Bij meervoudige en – in identificerende functie – ontelbare nomina wordt in BN dus significant vaker *zo'n* gebruikt dan in NN. Bij de ontelbare nomina wordt het wel vaker gebruikt – in ongeveer de helft van de gevallen – dan bij meervoudige nomina, waar ook in BN in driekwart van de gevallen voor *zulk(e)* gekozen wordt. In NN wordt *zo'n* zo goed als nooit gebruikt bij meervoudige nomina, en slechts in een kleine 10% van de gevallen bij ontelbare nomina. Hoewel *zo'n* in die contexten volgens de ANS2 zowel in het zuidoosten van België als in de zuidoostelijke provincies van Nederland voorkomt, dringt het gebruik in België in veel sterkere mate door in standaardtalige contexten dan in Nederland: bij meervoudige nomina vertaalt zich dat in een (zo goed als) absoluut verschil tussen BN en NN, bij ontelbare nomina is het verschil eerder gradueel. In de vierde paragraaf zullen we nagaan hoe goed die resultaten ons inzicht bieden in de nationale variatie in het gebruik van beide varianten, en hoe de resultaten kunnen worden vertaald naar nationale labels in ANS3.

### 3.3. Gebruik van weten met betekenis ‘zich kunnen herinneren’

#### 3.3.1. Beschrijving van de varianten

Het laatste geval van diatopische variatie dat hier aan bod komt, heeft betrekking op het groepsvormende gebruik van het werkwoord *weten*. De ANS2 (18.5.4.12) stelt dat het groepsvormende *weten* drie betekenissen kan uitdrukken: zo heeft *weten* in (35) de betekenis ‘in staat zijn’ en in (36) ‘weten waar’. De betekenis van zin (37) wordt geparafraseerd als ‘Ik heb nog meegemaakt dat hij in een orkest speelde’ of ‘Ik kan me nog herinneren dat hij in een orkest speelde’, maar die parafrases lijken de semantische waarde niet exact te vatten. De zin betekent zoveel als ‘ik was ervan op de hoogte dat hij in een orkest speelde in de periode waarin hij dat deed (maar nu is die periode voorbij)’, of – als het een eenmalige gebeurtenis betreft – ‘ik was ervan op de hoogte toen hij eenmalig in een orkest speelde’. In beide gevallen blijft het in het midden hoe de spreker dat wist, bv. doordat hij hem toen al kende of doordat hij hem zelf aan het werk heeft gezien in het orkest.

(35) Hij *heeft* zich altijd heel goed *weten te redden*. (ANS2 18.5.4.12)

(36) Ik heb jaren met haar samengewerkt, maar ik *heb* haar nooit *weten te wonen*. (ANS2 18.5.4.12)

(37) Ik *heb* hem nog in een orkest *weten spelen*. (ANS2 18.5.4.12)

In de drie betekenissen treedt in de voltooide tijden de vervangende infinitief van *weten* op. In de betekenis ‘zich kunnen herinneren’ wordt *weten* volgens de ANS2 zelfs alleen gebruikt in de voltooide tijden, dus als vervangende infinitief. Zowel bij ‘weten waar’ als bij ‘zich kunnen herinneren’ is er diatopische variatie. In de betekenis ‘weten waar’ wordt vooral in Belgisch Nederlands vaak een infinitief zonder *te* gebruikt, maar dat wordt door de ANS2 als regionaal, niet-standaardtalig taalgebruik gelabeld. Het gebruik van *weten* ‘zich kunnen herinneren’ is beperkt tot Belgisch Nederlands en krijgt dus eveneens het label ‘regionaal’.

In wat volgt focussen we op de diatopische variatie in het gebruik van de variant *weten* ‘zich kunnen herinneren’. Die case verschilt op verschillende vlakken van de vorige twee. Zo suggereert de verwoording in de ANS2 dit keer dat de variant beperkt is tot België, terwijl dat voor de varianten uit de vorige twee cases niet het geval was: terwijl we bij de vorige cases dus verwachten om de varianten telkens in beide nationale variëteiten aan te treffen – zij het in verschillende mate – is dat nu niet het geval. Net zoals bij de vorige case is er ook hier een variant in het spel die niet-standaardtalig is – althans volgens ANS2, want op [Taaladvies.net](http://Taaladvies.net) of in Van Dale (Den Boon & Hendrickx 2015) blijft de kwestie onvermeld. Het verschil is echter dat er in dit geval geen *near*

*synonym* lijkt te bestaan waarmee *weten* ‘zich kunnen herinneren’ inwisselbaar is in een welbepaalde lexicale en/of grammaticale context. De ANS2 (18.5.4.12) suggereert weliswaar een aantal alternatieve formuleringen met *zich kunnen herinneren* en *meemaken*, en in dat rijtje zouden we ook *weet hebben van* kunnen toevoegen (bv. *Ik had er weet van dat hij in een orkest speelde*). Dat blijven echter parafrases die niet exact dezelfde semantische waarde hebben en die bovendien een andere syntactische context vereisen: het gaat telkens om zelfstandige werkwoorden die gecombineerd worden met een afhankelijke zin ingeleid door het grammaticaal voegwoord *dat*, waarin gespecificeerd wordt wat het onderwerp van de matrixzin zich precies herinnert, wat hij heeft meegemaakt of waar hij weet van heeft. In die alternatieven wordt de betekenis dus op een ‘lexicalere’ manier overgebracht dan in de constructie met het groepsvormende *weten*. Dit lijkt dus een geval te zijn waarin we geen relatieve tokenfrequentie kunnen inzetten om nationale verschillen op het spoor te komen, maar waarbij we moeten vertrouwen op verschillen in absolute en genormaliseerde tokenfrequentie.

### 3.3.2. Dataverzameling en resultaten

Ook voor deze case werden data verzameld in de BN en NN krantencomponent van OpenSoNaR. In beide componenten werd gezocht naar alle voorkomens van het woord *weten*, onmiddellijk gevolgd door een infinitief of een *te*-infinitief. Naar andere vormen van *weten* werd niet gezocht, noch naar zinnen waarin *weten* niet onmiddellijk voorafgaat aan de infinitief, aangezien de ANS2 (18.5.4.12) aangeeft dat *weten* ‘zich kunnen herinneren’ enkel voorkomt in de voltooiden tijden, als vervangende infinitief. De *te*-infinitieven werden meegenomen in de zoekopdracht om na te gaan of *weten* ‘zich kunnen herinneren’ ook met *te* voorkomt, aangezien de twee andere betekenissen van het groepsvormende *weten* telkens met een *te*-infinitief voorkomen.

Voor BN leverde de zoekopdracht 1628 hits op, voor NN 1445 hits. In die hits werden alle voorkomens van het groepsvormende *weten* gelabeld naargelang van de betekenis die *weten* uitdrukt. Een bijzondere vermelding verdienen de zinnen in (38-41). Het gebruik van *weten* in die zinnen lijkt duidelijk verwant aan het gebruik van *weten* in (37): ook hier wordt telkens beschreven dat het onderwerp ergens weet van heeft of iets meemaakt. Het verschil is echter dat het hier niet gaat om zinnen in de voltooiden tijd: wat het onderwerp weet of wat hij meemaakt, is dan ook niet voorbij. De zinnen kunnen dan ook niet geparafraseerd worden met *zich kunnen herinneren*, maar bijvoorbeeld wel met *meemaken* (bv. zin 38 als *er is niks ergers dan meemaken dat een kind sterft*). In tegenstelling tot wat de ANS2 (18.5.4.12) aangeeft, lijkt dat gebruik van *weten* dus ook voor te komen buiten de voltooiden tijd. Zinnen (38-39) la-



ten bovendien zien dat *weten* ook in deze betekenis gecombineerd kan worden met een *te*-infinitief. Het betreft hier echter wel een minderheid van de hits: zinnen (38-39) zijn de enige hits met een *te*-infinitief, en naast zinnen (38-41) waren er slechts twee andere hits waarin *weten* buiten de voltooide tijd de betekenis ‘meemaken’ heeft.

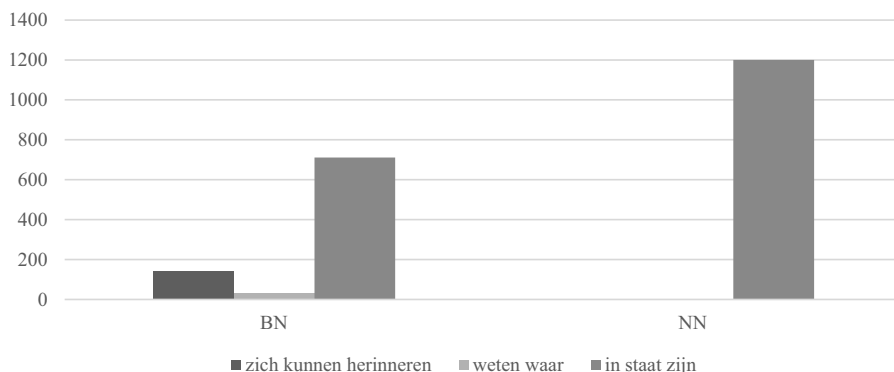
- (38) Ik weet nu, ook door zelf vader te zijn: er is niks ergers dan een kind **weten te sterven**. (‘Goed dat ik hem heb gestopt, maar het is geen heldendaad’, *Het Laatste Nieuws* (15/05/2006))
- (39) Ik en de vrienden beleven er nog steeds lol aan om hem te zien staan draaien en keren en hem sms’jes **weten te sturen** à la: ‘Stop met lachen en red mij hieruit, rotzakskes.’ (Artikel zonder titel, *Het Laatste Nieuws* (18/07/2006))
- (40) Misschien slapen we vooral samen om elkaar weer wakker te **weten worden**. (Artikel zonder titel, *De Morgen* (23/09/2006))
- (41) “Wij kennen ook perfect onze sportieve mogelijkheden, je zal ons dan ook niet **weten zweven**.” (“Ik moet nog groeien”, *Het Nieuwsblad* (12/10/2006))

Tabel 8 biedt een overzicht van de resultaten: telkens wordt de absolute tokenfrequentie gegeven waarmee een betekenis van groepsvormend *weten* voorkomt; om de cijfers optimaal te kunnen vergelijken wordt ook de genormaliseerde tokenfrequentie per miljoen woorden gegeven, aangezien de corpora voor BN en NN niet even groot zijn. In Figuur 6 worden de verschillen in absolute tokenfrequentie gevisualiseerd. Focussen we op de betekenis ‘zich kunnen herinneren’, dan zien we meteen dat die betekenis enkel aangetroffen wordt in de dataset voor BN. Als we het totaal aantal voorkomens voor die betekenis in BN vergelijken met het aantal voorkomens in NN, dan is dat – rekening houdende met de verschillende corpusgrootte (cf. Tabel 1) – een statistisch significant verschil (Goodness of fit:  $\chi^2=54.32$ ,  $df=1$ ,  $p<0.0001$ ). Wat de geografische verspreiding betreft, blijkt de beschrijving in de ANS2 dus correct: *weten* ‘zich kunnen herinneren’ is (in ieder geval in standaardtalige contexten) beperkt tot BN. Het gaat hier dus om een absoluut verschil tussen BN en NN. In paragraaf 4 zullen we nagaan hoe dat verschil vertaald kan worden naar labels in een referentiegrammatica.

**Tabel 8:** absolute en genormaliseerde tokenfrequentie voor de verschillende betekenissen van het groepsvormende *weten* in de resultaten voor de zoekopdracht ‘*weten* + (te) + infinitief’ in de BN- en NN-component van de collectie Newspapers van OpenSoNaR. Bij de absolute tokenfrequentie wordt tussen haakjes het aantal *te*-infinitieven gegeven.

	tokenfrequentie →	absolute tokenfrequentie	genormaliseerde tokenfrequentie (/miljoen woorden)
	betekenis van <i>weten</i> ↓		
BN	‘zich kunnen herinneren’	143 (2)	0.939
	‘weten waar’	32 (0)	0.210
	‘in staat zijn’	711 (694)	4.67
	<b>TOTAAL</b>	<b>886 (696)</b>	<b>5.82</b>
NN	‘zich kunnen herinneren’	0 (0)	0
	‘weten waar’	0 (0)	0
	‘in staat zijn’	1200 (1196)	20.2
	<b>TOTAAL</b>	<b>1200 (1196)</b>	<b>20.2</b>

Absolute tokenfrequentie voor de verschillende betekenissen van groepsvormend *weten*



**Figuur 6:** absolute tokenfrequentie voor de verschillende betekenissen van het groepsvormende *weten* in de resultaten voor de zoekopdracht ‘*weten* + (te) + infinitief’ in de BN- en NN-component van de collectie Newspapers van OpenSoNaR

#### 4. NAAR EEN SYSTEMATISCHE METHODE VOOR DE BESCHRIJVING VAN GRAMMATICALE VERSCHILLEN TUSSEN NATIONALE STANDAARDVARIËTEITEN

Als samenvatting van de casestudy's wordt in Tabellen 9 en 10 een beknopt overzicht gegeven van de resultaten. Tabel 9 bevat de twee cases waarbij we relatieve tokenfrequentie gebruikten om nationale verschillen op te spo-

ren.<sup>20</sup> Die cases zouden we kunnen beschrijven als grammaticale verschillen vanuit *onomasiologisch* perspectief: er zijn verschillende varianten om eenzelfde stand van zaken te beschrijven, en de nationale variëteiten maken niet in dezelfde mate gebruik van die varianten. Tabel 10 bevat de absolute en genormaliseerde tokenfrequentie voor de variant *weten* ‘zich kunnen herinneren’. Hier gaat het om een grammaticaal verschil vanuit *semasiologisch* perspectief: eenzelfde vorm wordt in verschillende delen van het taalgebied op een andere manier gebruikt. In dit geval wordt de vorm met een andere frequentie gebruikt, maar het is ook denkbaar dat eenzelfde vorm bv. in andere syntactische of semantische contexten voorkomt in de verschillende nationale variëteiten: Coleman (2010) laat bijvoorbeeld zien dat het nominaal belanghebbend voorwerp in België met een groter aantal types werkwoorden kan worden gecombineerd dan in Nederland. Om zo’n verschil in kaart te brengen, zouden we de typefrequentie (het aantal types werkwoorden waarmee het belanghebbend voorwerp voorkomt) in beide landen moeten vergelijken.

Met die casestudy’s in het achterhoofd keren we in deze paragraaf terug op de vragen uit paragraaf 2.3, over de keuzes die ANS3 moet maken om nationale variatie op een systematische manier te beschrijven. De vragen worden hieronder beknopt herhaald. Vraag (i) is reeds aan bod gekomen in paragraaf 3, de overige vragen worden besproken in paragrafen 4.1 t.e.m. 4.3.

- (i) Wanneer kunnen grammaticale varianten beschouwd worden als *equivalenten*?
- (ii) Bieden verschillen in relatieve tokenfrequentie een goed beeld van de nationale verschillen, of kunnen intern-linguïstische factoren (bv. semantische verschillen tussen varianten) dat beeld vertekenen?
- (iii) Volstaat corpusonderzoek voor ANS3, of is er ook nood aan evaluatieonderzoek?
- (iv) Welke criteria moet ANS3 gebruiken om labels toe te kennen? Meer specifiek, moet de ANS3 een drempelwaarde definiëren die een variant moet overschrijden binnen een variëteit om onderdeel uit te maken van de standaard van die variëteit?

<sup>20</sup> Voor de casestudy naar de combinatie van *komen* met een infinitief of een voltooid deelwoord worden zowel de cijfers gegeven voor de uitgebreide set werkwoorden (alle werkwoorden behalve *opdagen*) als voor de beperkte set (*fietsen, lopen, rennen, rijden, vliegen, wandelen* en hun samengestelde vormen). Voor de casestudy over de combinatie van identificerend *zo’n* en *zulk(e)* met een ontelbaar nomen worden enkel de cijfers gegeven voor de extra zoekopdracht met een set van nomina die vaak ontelbaar worden gebruikt (overzicht in 34).

Tabel 9: variantenparen waarvoor relatieve tokenfrequentie werd ingezet om nationale verschillen te beschrijven

	variant	regionaal volgens ANS2?	relatieve tokenfrequentie in standaardtalige contexten in BN	relatieve tokenfrequentie in standaardtalige contexten in NN	beschrijving volgens methode Variantengrammatik
Case A(1)	<i>komen</i> + INF (uitgebreide set)	nee	26.59%	81.56%	BN: komt geregeld voor NN: komt in de meerderheid van de gevallen voor
	<i>komen</i> + VD (uitgebreide set)	nee	73.41%	18.44%	BN: komt in de meerderheid van de gevallen voor NN: komt (zelden) voor
Case A(2)	<i>komen</i> + INF (beperkte set)	nee	9.93%	66.67%	BN: komt (zelden) voor NN: komt in de meerderheid van de gevallen voor
	<i>komen</i> + VD (beperkte set)	nee	90.07%	33.33%	BN: komt in de meerderheid van de gevallen voor NN: komt geregeld voor
Case B(1)	<i>zo 'n</i> + meervoudig nomen	ja	27.81%	0.30%	BN: komt geregeld voor NN: komt sporadisch voor
	<i>zulk(e)</i> + meervoudig nomen	nee	72.19%	99.70%	BN: komt in de meerderheid van de gevallen voor NN: komt (zo goed als) uitzonderingsloos voor
Case B(2)	identificerend <i>zo 'n</i> + ontelbaar nomen	ja	50%	8.51%	BN: komt geregeld voor NN: komt (zelden) voor
	identificerend <i>zulk(e)</i> + ontelbaar nomen	nee	50%	91.49%	BN: komt geregeld voor NN: komt in de meerderheid van de gevallen voor

**Tabel 10: varianten waarvoor absolute tokenfrequentie werd ingezet om nationale verschillen te beschrijven**

	variant	regionaal volgens ANS2?	frequentie in standaardtalige contexten in BN	frequentie in standaardtalige contexten in NN	beschrijving volgens methode
Case C	weten 'zich kunnen herinneren'	ja	absolute tokenfrequentie	absolute tokenfrequentie	Variantengrammatik BN: komt (zo goed als) uitzonderingsloos voor NN: komt (vrijwel) niet voor
			genormaliseerde tokenfrequentie per miljoen woorden	genormaliseerde tokenfrequentie per miljoen woorden	
			relatieve tokenfrequentie t.o.v. NN	relatieve tokenfrequentie t.o.v. BN	
			143	0	
			0.939	0	
			100.000%	0.00%	

#### 4.1. *De verhouding tussen extern- en intern-linguïstische factoren*

In Tabel 9 wordt duidelijk dat de verhouding tussen deelwoord- en infinitief-variant bij *komen* verschillend is in beide nationale corpora. Hetzelfde geldt voor de verhouding tussen de voornaamwoorden *zo 'n* en *zulk(e)* bij meervoudige en ontelbare nomina. De vraag rijst nu echter hoe goed die verschillen in relatieve tokenfrequentie ook de nationaal verschillende voorkeur voor een bepaalde variant weerspiegelen. Mogelijk worden de frequentieverschillen namelijk niet alleen beïnvloed door de extern-linguïstische factor 'land', maar spelen ook intern-linguïstische factoren (bv. de voorkeur van een van beide varianten voor een bepaalde semantische of syntactische context) een rol in het frequentieverschil. In hun artikel over de intern- en extern-linguïstische factoren die de keuze tussen causatief *doen* en *laten* beregelen, onderscheiden Levshina et al. (2013: 35) drie mogelijke scenario's voor de verhouding tussen intern- en extern-linguïstische factoren bij de beregeling van alternaties. Die scenario's illustreren we hieronder aan de hand van de variatie tussen deelwoord en infinitief bij *komen*.

In een eerste scenario (*independent lectal and conceptual variables*) hebben extern- en intern-linguïstische factoren onafhankelijk van elkaar invloed op de keuze voor een bepaalde variant: voor de variatie bij *komen* zou dat betekenen dat de semantische context en de nationale variëteit los van elkaar een invloed hebben op de keuze tussen infinitief en voltooid deelwoord. In dat geval vormen de verschillen in relatieve tokenfrequentie dus een goede afspiegeling van de nationale verschillen in het gebruik van de varianten. Een tweede scenario wordt door Levshina et al. (2013: 35) aangeduid met de term *socioconceptual variation* en is eigenlijk een specifieke vorm van het eerste scenario. Intern- en extern-linguïstische factoren zijn opnieuw onafhankelijk van elkaar, maar bepaalde waarden van de intern-linguïstische factoren komen wel vaker voor in (de corpora voor) bepaalde variëteiten dan in andere. Daardoor krijgen we een vertekend beeld van de invloed van de variëteit. Dat de deelwoordvariant met een hogere frequentie voorkomt in BN, zou bijvoorbeeld als oorzaak kunnen hebben dat er in de Belgische krantencomponent van OpenSoNaR gewoon meer contexten beschreven worden waarin het pad erg kort is en waarin het deelwoord dus de voorkeur krijgt. Zo kan de perceptie ontstaan dat BN een voorkeur heeft voor voltooide deelwoorden bij *komen*, terwijl dat eigenlijk niet het geval is. In een derde scenario interageren extern- en intern-linguïstische factoren met elkaar: bij zulke *lectally moderated semantic variation* verschilt de impact van bepaalde intern-linguïstische factoren naargelang van extern-linguïstische factoren. Toegepast op de variatie bij *komen* zou dat betekenen dat de invloed die semantische contextfactoren hebben op de keuze tussen infinitief en voltooid deelwoord verschilt in BN en NN. In dat geval zijn ver-

schillen in relatieve tokenfrequentie weinig informatief: het fundamentele verschil tussen de variëteiten is dan namelijk niet dat beide varianten met een verschillende frequentie worden gebruikt, maar wel dat de keuze op een andere manier beregeld wordt.

Alleen wanneer er sprake is van *independent lectal and conceptual variables* zijn verschillen in relatieve tokenfrequentie tussen nationale variëteiten dus voldoende informatief om uitspraken te doen over nationale verschillen in het gebruik van bepaalde varianten. Met welk van de drie scenario's we te maken hebben, valt echter enkel uit te maken aan de hand van multifactorieel onderzoek (bv. regressieonderzoek). Het lijkt evenwel weinig haalbaar voor een grootschalig project zoals de ANS3 om voor alle gevallen van grammaticale variatie multifactorieel onderzoek op te zetten, aangezien de daarvoor noodzakelijke (manuele) controle en annotatie van de observaties erg tijdrovend is. Als we ons beperken tot een vergelijking van de relatieve tokenfrequentie waarmee varianten in de nationale variëteiten voorkomen, dan moeten we echter in het achterhoofd houden dat er mogelijk sprake is van *lectally moderated semantic variation* of van *socioconceptual variation*.

Die laatste twee situaties lijken evenwel enkel mogelijk als er tussen de varianten ook daadwerkelijk intern-linguïstische (bv. semantische of syntactische) verschillen zijn. Dat is het geval bij de infinitief- en deelwoordvariant bij *komen*, maar voor de variatie tussen *zo'n* en *zulke* bij meervoudige en ontelbare nomina hebben we daar op basis van de literatuur geen aanwijzingen voor. Hoewel nader onderzoek nodig zou zijn om dat te bevestigen, lijkt hier geen *lectally moderated semantic variation* of *socioconceptual variation* mogelijk en bieden de verschillen in relatieve tokenfrequentie tussen de corpora dus een goed beeld van de nationale variatie.

#### 4.2. *Productiedata en/of evaluatiedata?*

Een andere vraag bij de resultaten in Tabellen 9 en 10 is of die frequentiedata volstaan om uit te maken wat tot de standaard behoort in BN en NN. Zoals beschreven in paragraaf 2 maakt de Variantengrammatik énkél gebruik van frequentiedata: zij gaan uit van een gebruiksstandaard per regionale variëteit waartoe alles behoort wat met een zekere frequentie voorkomt in het kranten corpus voor die variëteit. Taaladvies.net daarentegen verzamelt in bepaalde gevallen ook nog evaluatiedata, omdat het uitgaat van *linguistic schizophrenia*, van een discrepantie tussen wat *gebruikt* wordt in standaardtalige contexten en wat *aanvaard* wordt als standaardtaal. Bij onze casestudy's bleek dat varianten die volgens de ANS2 'regionaal' zijn en dus buiten de standaardtaal vallen, wel degelijk voorkomen in krantencorpora (soms zelfs behoorlijk frequent, bv. het gebruik van *zo'n* bij meervoudige nomina in BN). Volstaat dat

voor ANS3 om die varianten (in een van beide variëteiten) als standaardtaal te labelen, of moet ANS3 ook nagaan hoe taalgebruikers die varianten evalueren? Bij die vraag kunnen we twee bedenkingen plaatsen.

De verschillen tussen de Variantengrammatik en Taaladvies.net met betrekking tot evaluatiedata vallen vermoedelijk ten minste voor een deel te verklaren doordat beide bronnen een verschillend doel hebben. Waar de Variantengrammatik wil beschrijven welke diatopische verschillen er zijn in het gebruik van grammaticale varianten in standaardtalige contexten, is het primaire doel van Taaladvies.net om taaladvies te verstrekken. Terwijl de Variantengrammatik een descriptief doel heeft, heeft Taaladvies.net een prescriptief doel. Het is dan ook niet onlogisch dat die laatste evaluatiedata gebruikt om standaardtaalstatus toe te kennen: wie op zoek is naar taaladvies, wil graag varianten vermijden die door een deel van de taalgebruikers worden afgekeurd. Of ook ANS3 gebruik moet maken van evaluatiedata, is dus verbonden met de vraag welk doel de grammatica heeft: taalgebruik beschrijven of voorschrijven? Het doel van ANS3 is op dat vlak ongewijzigd ten opzichte van ANS2, die “principiële beschrijvend te werk [gaat], dat wil zeggen dat ze niet voorschrijft hoe iets moet, maar dat ze de vormen en constructies die in het taalgebruik voorkomen, registreert en bespreekt”<sup>21</sup> (ANS2 0.2). Vanuit die optiek zou het dus moeten volstaan om productiedata te verzamelen.

Nog los daarvan is het ook onduidelijk of er wel zulke grote verschillen zijn tussen de gebruiksstandaard en datgene wat taalgebruikers als standaardtaal aanvaarden. Dat Taaladvies.net vaststelt dat er varianten zijn waarbij corpus- en evaluatiegegevens elkaar tegenspreken, lijkt te bevestigen dat die schizofrenie echt bestaat en dat kan rechtvaardigen dat evaluatiegegevens betrokken worden. Echter, volgens De Schryver (2012: 153) zijn de evaluatiedata van Taaladvies.net weinig representatief, doordat de informantpanels niet groot genoeg zijn, maar vooral ook doordat ze grotendeels bestaan uit taalexperthen. Die taalexperthen zullen immers nog in sterkere mate beïnvloed zijn door de vroegere monocentrische taalpolitiek, waarbij Belgisch-Nederlandse items afgekeurd werden, en zo wordt het onderzoek voor De Schryver (2012) een self-fulfilling prophecy die de oude monocentrische norm bevestigt. De vraag is dus of we die discrepantie ook vaststellen als de spraakmakende gemeente wat ruimer opgevat wordt, en als bijvoorbeeld niet alleen taalexperthen, maar ook andere hoogopgeleiden bevraagd worden. Overigens speelt de discrepantie tussen corpus- en evaluatiedata mogelijk sowieso al een minder grote rol bij grammatica. Van Bree (2000) stelt bijvoorbeeld dat de domeinen van syntaxis en flexie weinig bewuste taaldomeinen zijn (cf. Auer 2014): het is dus goed

<sup>21</sup> Daar wordt nog de bedenking aan toegevoegd “dat wie daar behoefte aan heeft, wel degelijk een normatief gebruik kan maken van de ANS” (ANS2 0.2).



mogelijk dat taalgebruikers allerlei grammaticale varianten met diatopische variatie niet eens opmerken.

Om de vraag te beantwoorden of evaluatiedata een zinvolle aanvulling vormen op corpusdata om de normatieve status van grammaticale items te bepalen, is er dus onderzoek nodig naar de evaluatie van allerlei grammaticale varianten waarvoor er diatopische verschillen zijn. Hoewel er voor Nederlandstalig België al heel wat onderzoek (zie overzicht in Grondelaers 2013) is verricht naar de *attitudes* van taalgebruikers tegenover variëteiten en – in veel mindere mate, bv. bij Ghyselen (2009) en Grondelaers & Speelman (2013) – tegenover specifieke varianten, is onderzoek naar de evaluatie van varianten als standaardtaal voorlopig schaars. Een uitzondering vormt het onderzoek van de krant *De Standaard* uit 2014 waarin 40 zinnen met typisch Vlaamse items en 10 zinnen met typische Nederlandse items ter beoordeling werden voorgelegd aan 3226 taalprofessionals (zie De Schryver 2015 voor methodologie en resultaten), en waaruit bleek dat veel van die typisch Vlaamse items door de meerderheid van de respondenten aanvaard werden.

#### 4.3. Een drempelwaarde voor standaardtaligheid?

Als de ANS zich zou beperken tot productiedata, dan is een bijkomende vraag hoe frequentiedata vertaald moeten worden naar nationale labels. We weten op basis van de gegevens in Tabellen 9 en 10 wel hoe frequent een variant voorkomt in BN en NN in standaardtalige contexten (nl. in krantentaal), maar we kunnen ons afvragen of een variant die voorkomt in BN of NN ook altijd tot de standaard van die variëteit behoort, of dat we een drempelwaarde moeten definiëren die varianten binnen een variëteit moeten overschrijden om onderdeel uit te maken van de standaard van die variëteit (i.e. om een label BN of NN te krijgen).

Volgens Taaladvies.net is zo'n drempelwaarde noodzakelijk: zij hanteren een relatieve tokenfrequentie van 5% als ondergrens om in aanmerking te komen voor standaardtaalstatus: varianten met een relatieve tokenfrequentie hoger dan 50% zijn sowieso standaardtaal<sup>22</sup>, maar varianten met een relatieve tokenfrequentie tussen 5% en 50% hebben een onduidelijke status en moeten worden voorgelegd aan een informantenpanel. Die werkwijze roept verschillende bedenkingen op. Ten eerste lijken de drempelwaarden van 5% en 50% voor discussie vatbaar. Zo is een drempelwaarde van 50% voor standaardtaligheid aan de hoge kant: het is niet duidelijk of die drempelwaarde geldt voor *alle* varianten of alleen voor varianten die om een of andere reden problematisch zijn. Als de drempelwaarde geldt voor *alle* varianten, dan zou dat bete-

<sup>22</sup> Maar zie voetnoot 6 voor uitzonderingen op die regel.

kenen dat van bijna alle varianten in Tabel 9 de normatieve status onduidelijk is: behalve voor *zo 'n* met meervoudig nomen geldt namelijk voor elke variant dat die in een of beide nationale variëteiten met een relatieve tokenfrequentie van 50% of minder voorkomt. Het lijkt waarschijnlijker dat de drempelwaarde van Taaladvies.net dus enkel geldt voor ‘problematische’ varianten, maar daarbij rijst dan weer de vraag welke varianten problematisch zijn. Een mogelijk criterium voor *zo 'n* ‘problematische status’ zou de status van de variant in naslagwerken zijn, maar dat lijkt een weinig geschikt criterium: volgens de ANS2 vallen bijvoorbeeld alle regionaal beperkte varianten buiten de standaardtaal. Als we zulke varianten labelen als problematisch – en ze dus aan strengere criteria voor standaardtaalstatus onderwerpen – lijkt dat een bias in te bouwen die moeilijk te rijmen valt met het pluricentrische perspectief. Ook de ondergrens van 5% lijkt streng: als een variant *zo* weinig voorkomt in verhouding tot zijn equivalenten, is een *mogelijke* verklaring dat die variant niet tot de standaardvariëteit behoort, maar er kunnen ook andere verklaringen zijn voor de lage frequentie van de variant. Mogelijk wordt de variant bijvoorbeeld enkel in bepaalde (vak)contexten gebruikt. Verder is het ook onduidelijk *op welk niveau* de drempelwaarde gehanteerd moet worden, een vraag die bij grammaticale varianten (die vaak op verschillende niveaus van abstractie bestudeerd kunnen worden) prominenter is dan bij lexicale varianten. Voor de casestudy over *komen* zien we bijvoorbeeld dat de cijfers erg verschillen tussen de uitgebreide en de beperkte set werkwoorden: op welk van beide sets werkwoorden passen we die drempelwaarde dan toe?

De Variantengrammatik hanteert dan weer geen drempelwaarden om te bepalen of een variant onderdeel uitmaakt van de standaard van een variëteit: als een variant voldoende frequent voorkomt in het corpus voor een standaardvariëteit, dan wordt de variant als onderdeel beschouwd van de gebruiksstandaard van die variëteit.<sup>23</sup> Dat die werkwijze eveneens bedenkingen oproept, wordt duidelijk als we Tabellen 9-10 opnieuw bekijken. In de laatste kolom hebben we daar telkens aangegeven welk label van de Variantengrammatik de gebruiks-frequentie van een variant in BN en NN zou beschrijven. Bijna alle varianten komen voor in beide variëteiten, met uitzondering van *weten* ‘zich kunnen herinneren’. Die variant treffen we alleen aan in standaardtalige contexten in BN, en behoort dus volgens de Variantengrammatik enkel tot de gebruiksstandaard in BN (en niet in NN). De combinatie van *zo 'n* met een meervoudig nomen komt met een relatieve tokenfrequentie van 0.30% voor in NN en wordt dus volgens de Variantengrammatik ‘sporadisch’ gebruikt in NN. De

<sup>23</sup> Zoals aangegeven in 2.1 krijgt een (groep) variant(en) het label ‘unter Schwellenwert’ in een regio als er in het corpus voor die regio minder dan tien tokens zijn voor die variant(en). Dat label geeft enkel aan dat het gebruik van die variant(en) in die regio niet beschreven kan worden doordat er te weinig data zijn, en zegt dus niets over de standaardtaligheid van de variant (cf. Niehaus 2015: 144). Het gaat hier dus om een ander soort drempelwaarde dan de drempelwaardes voor standaardtaligheid die Taaladvies.net hanteert.

vraag is echter of een frequentie van 0.30% hoog genoeg is om de variant als onderdeel van de gebruiksstandaard in NN te beschouwen, zeker als we weten dat er voor de combinatie maar één attestatie was in de Nederlandse krantencomponent. Het ging daarbij dan wel om één attestatie *binnen een sample*, maar de kwestie laat niettemin zien dat de grens tussen een gradueel en absoluut verschil niet altijd scherp te trekken is: het is goed mogelijk dat allerlei varianten uit de Belgisch Nederlandse gebruiksstandaard ook met een zekere frequentie voorkomen in standaardtalige contexten in NN, zonder dat dat hoeft te betekenen dat die daarom per se deel uitmaken van de gebruiksstandaard in NN. Dezelfde vraag kunnen we ons overigens ook stellen bij *weten* ‘zich kunnen herinneren’: aangezien ook die variant niet erg frequent is in BN, is er mogelijk twijfel of die daar wel tot de standaard behoort.

Zelfs als de ANS zich beperkt tot een beschrijving van de variatie in standaardtalige contexten, dan nog lijkt er dus nood aan een drempelwaarde voor standaardtaligheid. Het lijkt evident dat die drempelwaarde moet gelden voor *alle* varianten, en dat ze lager moet zijn dan de 50% die Taaladvies.net hanteert, maar hoe hoog ze dan wel moet zijn, is moeilijk te bepalen op basis van de casestudy's in paragraaf 3. Er zouden voor veel meer varianten frequentie-data verzameld moeten worden, zodat we kunnen bekijken welke varianten uit de boot vallen bij verschillende mogelijke drempelwaardes. Zowel varianten met als varianten zonder equivalenten moeten bestudeerd worden, zodat we een drempelwaarde voor relatieve tokenfrequentie en voor absolute tokenfrequentie (of eventueel: genormaliseerde tokenfrequentie) kunnen afbakenen.

Belangrijk is tot slot dat de drempelwaarde die we voor ogen hebben geen onderscheid maakt tussen wat standaardtalig is en wat niet. Zoals aangegeven kunnen ook varianten met een zeer lage frequentie namelijk perfect bruikbaar zijn in standaardtalige contexten; de lage frequentie kan immers ook andere oorzaken hebben (de variant kan bijvoorbeeld beperkt zijn tot bepaalde (vak)contexten). De drempelwaarde moet dus gelden als een soort ‘alarmdrempel’: als een variant die drempel niet overschrijdt, moet uitgezocht worden wat de oorzaak daarvan is. Er zou onderzocht moeten worden welke middelen in zulke gevallen ingezet kunnen worden om die oorzaak aan het licht te brengen. Als de variant niet geschikt is voor geschreven standaardtalige contexten, kan dat bijvoorbeeld blijken uit het feit dat de variant wél frequent voorkomt in gesproken en/of niet-standaardtalige contexten, of dat de variant ongepast bevonden wordt door taalgebruikers voor standaardtalige contexten. Als de variant wél geschikt is voor geschreven standaardtalige contexten, maar slechts in zeer beperkte (vak)contexten voorkomt, kan dat bijvoorbeeld blijken uit de dispersie van de variant in het corpus, i.e. de mate waarin de variant verspreid is over de verschillende onderdelen waaruit het corpus bestaat (cf. Gries

2008, 2010); voor varianten die beperkt zijn tot bepaalde contexten, verwachten we dan een lage mate van dispersie.

## 5. BESLUIT

Zoals aangegeven aan het begin van dit artikel is er nood aan systematisch onderzoek naar de verschillende grammaticale normen van BN, NN en SN als we de Nederlandse grammatica vanuit pluricentrisch perspectief willen beschrijven. Op basis van de methodologie die de Variantengrammatik en Taaladvies.net hanteren om grammaticale variatie tussen nationale of regionale standaardvariëteiten te beschrijven op basis van productiedata in (kranten)corpora (eventueel aangevuld met evaluatiedata), wierpen we een aantal vragen op over het gebruik van productiedata (en in het bijzonder: het gebruik van relatieve tokenfrequentie) om nationaal verschillende grammaticale normen te beschrijven. Die vragen hebben betrekking op keuzes die ANS3 moet maken als ze een systematische methode wil ontwikkelen voor de beschrijving van nationale variatie in de grammatica. Aan de hand van literatuuronderzoek en aan de hand van corpusonderzoek voor drie gevallen van diatopische grammaticale variatie tussen BN en NN trachtten we tot een antwoord op die vragen te komen.

Ten eerste werd nagegaan bij welke varianten we de nationale variatie kunnen beschrijven aan de hand van de relatieve tokenfrequentie van de variant en zijn equivalent(en) in nationale (kranten)corpora; daartoe werden een aantal criteria afgebakend waaraan varianten moeten voldoen om als equivalenten te worden beschouwd. We stipten ook aan dat het gebruik van relatieve tokenfrequentie voor de beschrijving van nationale variatie een aantal beperkingen vertoont, nl. als er tussen de varianten (syntactische of semantische) verschillen zijn die leiden tot een situatie van *socioconceptual variation* of *lectally moderated semantic variation*. In die situaties is relatieve tokenfrequentie onvoldoende informatief om nationale verschillen te beschrijven. Verder bogen we ons over de vraag of ANS3 zich kan beperken tot het verzamelen van productiedata of dat ook evaluatiedata noodzakelijk zijn. We gaven aan dat dat samenhangt met de vraag of ANS3 een descriptief of prescriptief doel heeft, maar daarnaast ook dat het onduidelijk is of er in BN wel zo'n sterke discrepantie is tussen productie- en evaluatiedata als weleens gesuggereerd wordt. Tot slot wierpen we ook de vraag op welke frequentiecriteria ANS3 moet hanteren om varianten te labelen, meer bepaald of er een drempelwaarde moet worden gedefinieerd die een variant moet overschrijden in een variëteit om tot de standaard van die variëteit te behoren. Zo'n drempelwaarde lijkt nuttig, maar is niet onproblematisch, aangezien een lage frequentie niet noodzakelijk impliceert dat een item niet tot de standaard behoort. We suggereerden daarom

dat de drempelwaarde geen criterium voor standaardtaligheid mag zijn, maar eerder moet functioneren als een ‘alarmdrempel’: bij varianten die onder de alarmdrempel blijven, is verder onderzoek nodig om te bepalen wat de oorzaak is van hun lage frequentie. Hoe hoog die drempelwaarde moet zijn, blijft wel onduidelijk.

Wat de prioriteiten voor het vervolg van ons onderzoek betreft, zijn er twee thema’s waarin we beter inzicht willen krijgen. Ten eerste willen we proberen een inschatting te maken van de verhouding tussen productie- en evaluatiedata. Daartoe testen we voor een set Belgisch Nederlandse grammaticale items hoe die geëvalueerd worden door een groep taalgebruikers die niet alleen taalexperten, maar ook andere hoogopgeleiden bevat. We gaan daarbij na of er verschillen zijn in de evaluatie naargelang van het profiel van de sprekers, zoals hun taalexpertise, taalideologie, leeftijd en regionale achtergrond. Ten tweede willen we beter inzicht krijgen in de frequentiecriteria voor standaardtaligheid. Daarvoor verzamelen we in een volgende stap in nationale kranten corpora frequentiedata voor een groter aantal grammaticale varianten, zowel varianten mét equivalenten als zonder. Bijzondere aandacht moet uitgaan naar varianten waarvan we verwachten dat hun frequentie in (een van) de nationale variëteiten om een of andere reden beperkt zal zijn, bv. omdat ze in naslagwerken afgekeurd worden in geschreven standaardtaal, omdat ze een beperkte geografische verspreiding kennen, of omdat ze enkel in bepaalde (vak)contexten worden gebruikt. Voor laagfrequente items zou vervolgens onderzocht moeten worden aan de hand van welke middelen we kunnen bepalen *waarom* ze laagfrequent zijn, bv. door na te gaan wat hun dispersie is in standaardtalige registers, met welke frequentie ze voorkomen in niet-standaardtalige registers of hoe ze geëvalueerd worden door taalgebruikers. Op basis van de frequentiedata in nationale corpora kunnen we ook verschillende mogelijke drempelwaarden voor relatieve en absolute (of genormaliseerde) tokenfrequentie evalueren, door na te gaan welke varianten al dan niet de drempelwaarde overschrijden. Op die manier moet duidelijk worden hoe werkbaar de door ons voorgestelde ‘alarmdrempel’ voor standaardtaligheid is.

## REFERENTIELIJST

- Adviescommissie Taalvariatie. (2019). Visie op taalvariatie en taalvariatiebeleid. Visietekst van de Adviescommissie Taalvariatie in opdracht van het Algemeen Secretariaat van de Taalunie. Retrieved from <http://taalunieversum.org/sites/tuv/files/downloads/Visietekst%20taalvariatie%20-%20februari%202019.pdf>
- Auer, P. (2014). Anmerkungen zum Salienzbegriff in der Soziolinguistik. *Linguistik online*, 66(4), 7-20.

- Beliën, M. (2016). Exploring semantic differences in syntactic variation: Dutch komen ('come') with a past participle or an infinitive. In: A. Bannink & W. Honseelaar (red.), *From variation to iconicity: Festschrift for Olga Fischer on the occasion of her 65th birthday*, 17-32. Amsterdam: Pegasus.
- Bolinger, D. (1968). Entailment and the meaning of structures. *Glossa*, 2, 119-127.
- Broekhuis, H., & N. Corver. (2015). *Syntax of Dutch: Verbs and verb phrases. Volume 2*. Amsterdam: Amsterdam University Press.
- Caluwé, D., & V. Verreycken. (2012). Standaardtaal en tussentaal in de taaladviesing. In: K. Absillis, J. Jaspers, & S. Van Hoof (red.), *De manke usurpator: over verkavelingsvlaams*, 167-185. Gent: Academia Press.
- Colleman, T. (2010). Lectal variation in Constructional Semantics: Benefactive ditransitives in Dutch. In: D. Geeraerts, G. Kristiansen, & Y. Peirsman (red.), *Advances in Cognitive Sociolinguistics*, 191-221. Berlijn/New York: Mouton de Gruyter.
- Colman, L. (2009). De verscheidenheid in de eenheid: het Belgische en Nederlandse Nederlands in het Algemeen Nederlands Woordenboek (ANW). In: E. Beijk, L. Colman, M. Göbel, F. Heyvaert, T. Schoonheim, R. Tempelaars, & V. Waszink (red.), *Fons verborum. Feestbundel voor prof. dr. A.F.M.J. (Fons) Moerdijk, aangeboden door vrienden en collega's bij zijn afscheid van het Instituut voor Nederlandse Lexicologie*, 131-142. Leiden/Amsterdam: Instituut voor Nederlandse Lexicologie/Gopher BV.
- De Caluwe, J. (2011). Dutch as a bicentric language: a lexicographic (r)evolution. In: R. Muhr (red.), *Non-dominant varieties of pluricentric languages. Getting the picture. In memory of Michael Clyne*, 141-152. Berlijn: Peter Lang Verlag.
- De Caluwe, J. (2013). Nederland en Vlaanderen: (a)symmetrisch pluricentrisme in taal en cultuur. *Internationale Neerlandistiek*, 51(1), 45-59.
- De Rooij, J. (1989). Zo'n dingen zeggen ze hier (niet). Regionale verschillen in het gebruik van zo'n en zulk(e). In: S. Theissen & J. Vromans (red.), *Album Moors. Een bundel opstellen aangeboden aan Joseph Moors ter gelegenheid van zijn 75ste verjaardag*, 181-201. Liège: CIPL.
- De Schryver, J. (2012). Het einde van de tussentaal en de Vlaamse standaardtaaldiscussie. In: K. Absillis, J. Jaspers, & S. Van Hoof (red.), *De manke usurpator: over verkavelingsvlaams*, 141-165. Gent: Academia Press.
- De Schryver, J. (2015). Hoe Vlaams is het Standaardnederlands van taalprofessionelen. *Over Taal*, 54(1), 6-9.
- De Smet, H. (2017). The motivated unmotivated: Variation, function and context. *Leuven Working Papers in Linguistics* 37, 6, 122-142.
- De Sutter, G., D. Speelman, & D. Geeraerts. (2005). Regionale en stilistische effecten op de woordvolgorde in werkwoordelijke eindgroepen. *Nederlandse Taalkunde*, 10(2), 97-128.
- Delaby, G. (2018). *Komen met een voltooid deelwoord of infinitief. Een constructieel en kwantitatief corpusonderzoek*. Bachelorproef. Universiteit Gent. U.

- Den Boon, T., & R. Hendrickx. (2015). *Van Dale Groot woordenboek van de Nederlandse taal*. Utrecht/Antwerpen: Van Dale Uitgevers.
- Dürscheid, C., & S. Elspaß. (2015). Variantengrammatik des Standarddeutschen. In: R. Kehrein, A. Lameli, & S. Rabanus (red.), *Regionale Variation des Deutschen. Projekte und Perspektiven*, 563-584. Berlin/Boston: de Gruyter.
- Dürscheid, C., S. Elspaß, & A. Ziegler. (2015). Variantengrammatik des Standarddeutschen. Konzeption, methodische Fragen, Fallanalysen. In: A. N. Lenz & M. M. Glauninger (red.), *Standarddeutsch im 21. Jahrhundert: theoretische und empirische Ansätze mit einem Fokus auf Österreich*, 205-233. Göttingen: V&R Unipress.
- Elspaß, S. (2010). Regional standard variation in and out of grammarians' focus. In: A. N. Lenz & A. Plewnia (red.), *Grammar between norm and variation*, 127-144. Frankfurt am Main: Peter Lang.
- Elspaß, S., C. Dürscheid, & A. Ziegler. (2017). Zur grammatischen Pluriarealität der deutschen Gebrauchsstandards – oder: Über die Grenzen des Plurizentritätsbegriffs. *Zeitschrift für deutsche Philologie*, 136, 69-91.
- Geerts, G., W. Haeseryn, J. de Rooij, & M. C. van den Toorn. (1984). *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters-Noordhoff/Wolters.
- Ghesquière, L., & F. Van de Velde. (2011). A corpus-based account of the development of English *such* and Dutch *zulk*: identification, intensification and (inter)subjectification. *Cognitive Linguistics*, 22(4), 765-797.
- Ghyselen, A.-S. (2009). Ne zelfzekere leraar of gewoon nen enthousiaste mens? Een matched-guise onderzoek naar de attitude tegenover tussentaal bij West-Vlamingen. *Taal en Tongval*, 61, 83-113.
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437.
- Gries, S. T. (2010). Dispersions and adjusted frequencies in corpora: further explorations. In: S. T. Gries, S. Wulff, & M. Davies (red.), *Corpus-linguistic applications. Current studies, new directions*, 197-212. Amsterdam/New York: Brill/Rodopi.
- Grondelaers, S. (2013). Attitude measurements in the Low Countries. In: F. Hinskens & J. Taeldeman (red.), *Language in time and space: Dutch*. Berlin: Mouton de Gruyter.
- Grondelaers, S., R. De Troij, D. Speelman, & A. Van den Bosch. (2020). Vissen naar variatie. Digitaal op zoek naar onbekende Noord/Zuid-verschillen in de grammatica van het Nederlands. *Nederlandse Taalkunde*, 25(1), 73-99. doi:10.5117/NED-TAA2020.1.003.GRON
- Grondelaers, S., & D. Speelman. (2013). Can speaker evaluation return private attitudes towards stigmatised varieties? Evidence from emergent standardisation in Belgian Dutch. In: T. Kristiansen & S. Grondelaers (red.), *Language (de)standardisation in Late Modern Europe: experimental studies*, 171-191. Oslo: Novus.
- Grondelaers, S., D. Speelman, & A. Carbonez. (2001). Regionale variatie in de postverbale distributie van presentatief er. *Neerlandistiek.nl*, 01(04).
- Haeseryn, W. (1996). Grammaticale verschillen tussen het Nederlands in België en het Nederlands in Nederland: een poging tot inventarisatie. In: R. Van Hout & J.

- Kruijssen (red.), *Taalvariëties. Toonzettingen en modulaties op een thema*, 109-129. Dordrecht: Foris Publications.
- Haeseryn, W. (2013). Belgian Dutch. In: F. Hinskens & J. Taeldeman (red.), *Dutch*, 700-720. Berlijn: De Gruyter Mouton.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij, & M. C. van den Toorn. (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Joosten, F. (2003). *Collectiva en aggregaatsnamen in het Nederlands: begripsbepaling en typologie*. (Doctoraatsproefschrift), Katholieke Universiteit Leuven.
- Levshina, N., D. Geeraerts, & D. Speelman. (2013). Towards a 3D-grammar: Interaction of linguistic and extralinguistic factors in the use of Dutch causative constructions. *Journal of Pragmatics*, 52, 34-48.
- Martin, W. (2001). Natiolectismen in het Nederlands en hun lexicografische beschrijving. *Revue belge de philologie et d'histoire*, 79(3), 709-736.
- Martin, W., & W. Smedts. (2009). *Prisma Handwoordenboek Nederlands, met onderscheid tussen Nederlands-Nederlands en Belgisch-Nederlands. Derde, herziene druk*. Houten: Het Spectrum.
- Muhr, R. (2012). Linguistic dominance and non-dominance in pluricentric languages: a typology. In: R. Muhr (red.), *Non-dominant varieties of pluricentric languages. Getting the picture. In memory of Michael Clyne*, 23-48. Wenen: Peter Lang Verlag.
- Niehaus, K. (2015). Areale Variation in der Syntax des Standarddeutschen. Ergebnisse zum Sprachgebrauch und zur Frage Plurizentrik vs. Pluriarealität. *Zeitschrift für Dialektologie und Linguistik*, 82(2), 133-168.
- Oostdijk, N., M. Reynaert, V. Hoste, & I. Schuurman. (2013). The construction of a 500 million word reference corpus of contemporary written Dutch. In: P. Spijns & J. Odijk (red.), *Essential Speech and Language Technology for Dutch: Results by the STEVIN-project*, 219-247. Berlijn: Springer.
- Pijpops, D., & F. Van de Velde. (2018). A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion. *Corpus Linguistics and Linguistic Theory*, 14(1), 99-131.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In: F. Newmeyer, J. (red.), *Linguistics: The Cambridge Survey, Vol. 4 Language: The socio-cultural context*. Cambridge: Cambridge University Press.
- Speelman, D., & D. Geeraerts. (2009). Causes for causatives: the case of Dutch *doen* and *laten*. In: T. Sanders & E. Sweetser (red.), *Causal categories in discourse and cognition*, 173-204. Berlijn: Walter de Gruyter.
- Tummers, J., D. Speelman, & D. Geeraerts. (2005). Inflectional variation in Belgian and Netherlandic Dutch: a usage-based account of the adjectival inflection. In: N. Delbecque, J. Van der Auwera, & D. Geeraerts (red.), *Perspectives on variation: sociolinguistic, historical, comparative*. Berlijn: Walter de Gruyter.
- Van Bree, C. (2000). Taalbewustzijn, taalverandering en regionale taalvariatie. Een benadering. *Taal en Tongval*, 52, 22-46.



- Van de Velde, H. (1996). *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. Katholieke Universiteit Nijmegen.
- Van Hoof, S., & J. Jaspers. (2012). Hyperstandaardisering. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 128(1), 97-125.
- Van Olmen, D. (2019). A diachronic corpus study of prenominal *zo 'n* 'so a' in Dutch. Pathways and (inter)subjectification. *Functions of Language*, 26(2), 217-248.
- Van Olmen, D., & J. Van der Auwera. (2014). Over *zo 'n* en *zo meer*. In: F. Van de Velde, H. Smessaert, F. Van Eynde, & S. Verbrugge (red.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 215-228. Leuven: Universitaire Pers.

