

Collocaties in online woordenboeken

Ton van der Wouden

Roosevelt Academy/Meertens Instituut/Universiteit Leiden

*je kan nog zulke mooie grammatica's maken,
uiteindelijk zijn de woorden toch nog belangrijker*
(Jan Landsbergen, aangehaald in Van der Beek 2010)

1. Inleiding

Dit opstel gaat over de bruikbaarheid van de elektronische versies van de wetenschappelijke woordenboeken van het Nederlands voor diachroon onderzoek naar collocaties. Om daar iets over te kunnen zeggen, moet ik eerst duidelijk maken wat ik bedoel met de term “collocaties”. Als ik dat gedaan heb, bespreek ik de bruikbaarheid van genoemde woordenboeken aan de hand van een aantal typen collocaties. Ik trek bepaalde conclusies, en ik sluit af met een paar aanbevelingen voor verbetering van de producten.¹

2. Over collocaties

De distributie van lexicaal materiaal is niet willekeurig: veel woorden en uitdrukkingen worden vaak in de buurt van specifieke andere woorden en uitdrukkingen aangetroffen. Soms biedt de grammatica daar een verklaring voor: *postzegel* staat vaker naast *de* dan naast *het*, om de simpele reden dat *postzegel* een *de*-woord is (en *postzegel* is een *de*-woord omdat *zegel*, het rechter lid van de samenstelling, in deze betekenis een *de*-woord is; maar waarom dat zo is, dat weten we niet²). In andere gevallen heeft die onwillekeurige distributie meer te maken met hoe de wereld in elkaar zit, ofte wel onze encyclopedische kennis: om uiteenlopende maar begrijpelijke redenen komt een woord als *postzegel* relatief vaak voor in de omgeving van woorden als *brief*, *frankeren* en *verzamelen*.

Andere distributiepatronen zijn minder goed te begrijpen: om het werkwoord *slapen* te intensiveren gebruikt men bijvoorbeeld het bijwoord *diep*, terwijl men bij *werken* voor precies dezelfde functie *hard* gebruikt, en we weten niet precies waarom.³

¹ De bevindingen uit dit artikel zijn eerder gepresenteerd tijdens de workshop “The construction grammar of Dutch” te Leiden op 25 maart 2011. Ik dank het publiek aldaar en de reviewer(s) voor commentaar en suggesties.

² Het WNT schrijft s.v. *postzegel*: “znw. onz., doch thans gewoonlijk m., daar het woord niet of nauwelijks meer als samenst. wordt gevoeld” maar dat lijkt me niet het hele verhaal: naar mijn gevoel is *zegel* in de betekenis “bedrukt stukje papier dat een zekere waarde vertegenwoordigt” altijd *de*-woord, ook bijvoorbeeld in het geval van de zegeltjes die je bij de kassa krijgt.

³ In de lijn van Lakoff & Johnson (1980) valt te beredeneren dat *diep* als versterkend bijwoord bij *slapen* niet toevallig is – de metafoer dat slaap een put is, is deel van onze cultuur (je *valt* in slaap, enz.). Dit soort verklaringen doet het echter vooral achteraf goed en biedt voor lang niet alle combinaties van werkwoorden en bijwoorden een verklaring: waarom is het (vanuit dit perspectief) *diep nadenken* en *hard werken*?

De voorzetsels bij werkwoorden en bijvoeglijke naamwoorden maken ook een tamelijk willekeurige indruk: het is *hopen op*,⁴ *denken aan*, *verlangen naar*, en *trots op*.

Op zijn minst sinds Jespersen (1917) staat dit fenomeen van samen voorkomende lexicale elementen bekend als **collocatie**. En diezelfde term **collocatie** wordt ook gebruikt voor idiosyncratische combinaties van woorden, bijvoorbeeld *diep slapen* en *blond haar*.

Er zijn tal van definities van het begrip “collocatie” in omloop; vele ervan zijn nodeloos beperkt en ongemotiveerd. Met dit soort beperkte definities blijven allerlei potentieel interessante collocaties ten onrechte buiten beschouwing. Het is dan ook mijn overtuiging dat alleen een pretheoretische, puur kwantitatieve definitie een onbevooroordeelde studie van het fenomeen collocatie in zijn volle breedte mogelijk maakt. Dat neemt overigens niet weg dat een zekere beperking of nadere verfijning voor didactische doeleinden nuttig kan zijn.

Collocaties als *hard werken*, *dol op* en *blond haar* zijn bekend uit de literatuur, evenals bijvoorbeeld groepsnamen (*een roedel herten*, *een school vissen*, *een bende rovers*). Vanuit de aanname dat ook (bepaalde) morfemen en (sommige) grotere uitdrukkingen tot de lexicale elementen gerekend moeten worden (zie bijv. Di Sciullo & Williams 1987), zouden er ook collocaties boven en onder het woordniveau kunnen bestaan. Op verschillende plaatsen (bijv. Van der Wouden 1992) heb ik betoogd dat bijvoorbeeld geïntensiverde adjectieven (augmentatieven) als *spinnijdig* en *beresterk* opgevat kunnen worden als collocaties onder het woordniveau: het linkerdeel *spin* heeft immers alleen in de combinatie met *nijdig* de betekenis “in hoge mate”, en *bere* alleen met *sterk*. *Berenijdig* en *spinsterk* zijn wel mogelijke woorden van het Nederlands, maar ze betekenen niet “erg nijdig” en “erg sterk”.

Ook collocaties waarin een van de leden juist groter dan een woord is, zijn gemakkelijk te vinden, bijvoorbeeld bij de intensiverende uitdrukkingen:

- (1) a Het vriest dat het kraakt, het regent dat het giet
 b *Het regent dat het kraakt, *Het vriest dat het giet
 c Je liegt dat je barst, je zuipt als een ketter
 d ?Je liegt als een ketter, *je zuipt dat je barst

Ten slotte kunnen we collocaties onderscheiden waarin geen inhoudswoord voorkomt, bijvoorbeeld het modale partikel *eens* (*'ns*) dat onder meer dikwijls in de buurt van bepaalde andere partikels te vinden is, de partikelcombinatie *zo maar* die tegenwoordig vaak met *kunnen* voorkomt en de combinatie *niet eens* die graag bij de werkwoorden *kunnen* en *weten* staat, en de interjectie *ja* die in spreektaal heel vaak in combinatie met

⁴ Het werkwoord *hopen* komt ook voor met een bijzin (*ik hoop dat het gesmaakt heeft*) en met een voorwerp zonder voorzetsel, maar die laatste mogelijkheid lijkt (tegenwoordig?) beperkt tot (sommige?) voornaamwoorden: *ik hoop het*, *dat hoopt hij*. Dit zou je ook een collocatie-effect kunnen noemen, het is in elk geval een eigenaardigheid van het werkwoord *hopen* die (bij mijn beste weten) niet grammaticaal te verklaren is - en dus in het woordenboek te vinden zou moeten zijn. Een mogelijke tegenwerping is, dat hetzelfde ongeveer ook voor het Engelse *hope* geldt: ook bij dit werkwoord vind je een wel sententieel complement (*We hope that Israel doesn't attack Iran*), maar geen nominaal complement (**We hope a book* is ongrammaticaal). Overigens krijg je in het Engels waarschijnlijk een pro-VP en geen pro-nomen: *I hope so*. Op basis van de overeenkomst tussen het Nederlands en het Engels zou dit gedrag misschien wel volgen uit de semantiek. Maar de vraag is dan wel, hoe?

ja aangetroffen kan worden (VdWouden 2002a,b, 2007) (voorbeelden uit het Corpus Gesproken Nederlands).

- (2) a daar praatten we *dan wel eens* over⁵
b 't zou *zo maar* kunnen.
c en die *weten niet eens* waar Nederland ligt.
d mmm heerlijk. ik *kan niet eens* met vork en mes eten.
e en de hoogleraar in kwestie zei dan *ja ja ja* hè Esprit leuk maar ik weet niets van VBN

Geeraerts schreef vijftientig jaar geleden al “een grondig onderzoek naar collocaties laat nog op zich wachten” (1986:135) en dat is nog steeds waar. Weinig taalkundigen zullen ontkennen dat collocaties bestaan, maar toch zijn ze zelden object van taalkundig onderzoek, misschien omdat de meeste grammaticamodellen er weinig relevants over te zeggen hebben (Van der Wouden 1992). In de computertaalkunde is men vanouds overigens wel geïnteresseerd in meerwoordsuitdrukkingen zoals namen (*Willem Drees*), voorzetseluitdrukkingen (*ten gevolge van*), complexe begrippen (*proces verbaal*) en data (*20 september 1958*): wie echte tekst met de computer wil be- of verwerken, wordt voortdurend geconfronteerd met dit soort “woorden met spaties” (Sag et al. 2001). De theoretische taalwetenschap daarentegen heeft van oudsher belangstelling voor idiomen, meerwoordsuitdrukkingen met een onvoorspelbare semantiek, zoals *de kraaienmars blazen* voor ‘doodgaan’, juist vanwege de mismatch tussen vorm en betekenis (zie bijvoorbeeld Gazdar et al. 1985: 236vv).

Collocaties vormen door hun onvoorspelbaarheid een probleem voor taalleerders, zowel moedertaalsprekers als niet-moedertaalsprekers. In het talenonderwijs en in woordenboeken wordt daarom traditioneel aandacht besteed aan collocaties, in elk geval aan bepaalde typen, zoals vaste voorzetsels (Loonen 2003), min of meer vaste verbindingen tussen werkwoorden en zelfstandige naamwoorden (zie bijv. De Kleijn 2003), groepsnamen enz.⁶

3. Collocaties in de woordenboeken

Na deze inleiding over collocaties kunnen we eindelijk ter zake komen: de bruikbaarheid van de digitale wetenschappelijke woordenboeken van het Nederlands voor onderzoek naar collocaties en collocatieve gedrag. Laat mij vooropstellen dat er in deze bronnen veel relevante informatie te vinden is. Ik illustreer dat in de volgende subparagrafen met vier korte case-studies: eerst behandel ik vaste voorzetsels, en vervolgens intensiverende uitdrukkingen, groepsnamen, en collocaties onder het woordniveau. In de paragraaf

⁵ Getuige de spelling worden sommige partikelcombinaties in de woordenboeken kennelijk opgevat als gelexicaliseerd: het *MNW* bespreekt o.a. *alsmaar* en het *WNT* heeft onder meer *weleens* als ingang. De *Schrijfwijzer* (Renkema 2002) maakt onderscheid tussen *weleens* en *wel eens*, maar het is mij niet duidelijk op welke gronden.

⁶ Het *Van Dale Groot Woordenboek van Uitdrukkingen* (De Groot et al. 2006) is tamelijk expliciet over de collocaties die wel en niet worden behandeld, de motivatie achter de keuze is evenwel minder duidelijk.

daarna ga ik in op de vraag, of de woordenboeken ook geschikt zijn voor systematischer onderzoek naar collocaties. Ik sluit af met een paar aanbevelingen en conclusies.⁷

3.1 Vaste voorzetsels

We verwachten in de oudste (geboekstaafde) taalfasen minder vaste voorzetsels bij werkwoorden enzovoort dan tegenwoordig: de opkomst van het voorzetselvoorwerp wordt immers gerelateerd aan de deflexie van het Nederlands (Van Haeringen 1956), meer speciaal aan het verdwijnen van het casussysteem. Of in de woorden van Van der Horst (2008: 244-5): “Pas sedert de Middelnederlandse periode kan de voorzetselgroep ook attributieve nabepaling zijn [...], vanaf het late Middelnederlands en de 16de eeuw ook voorzetselvoorwerp [...], en daarbij aansluitend afhankelijk zijn van een substantief of een adjectief”.⁸

De verwachte ontwikkeling zien we niet direct terug bij het werkwoord *hopen*. In het enige relevante voorbeeld in het ONW s.v. regeert het werkwoord voorzetsel *zo*:

- (3) Tho gieng petrus *in* the stat froliche. ande **hoffo do** zo godes riche.⁹
Toen ging Petrus opgewekt de stad in en hoopte op Gods rijk.

Ook in de periode van het VMNW blijkt *hopen* een voorzetselcomplement te krijgen, maar het voorzetsel ligt nog niet vast: we vinden voorbeelden met *in*, *bij*, en *op*.

- (4) a in sinen rike hopic wal.¹⁰
b Cestius hopdte bi desen. Dat van der scare al de scande. Comen soude vp florus hande.¹¹
c Hijs sotter dan en kuekenoet Die hoept op anders mannes doet¹²

In het MNW komt *hopen* (plotseling?) voor met een naamval (7a), maar ook met verschillende voorzetsels (*op*, *in*, *tot*, *van*) (7b-e), en soms twee verschillende voorzetsels binnen een en dezelfde bron (7b-c):

- (5) a Elc sal hopen sijnre ghenaden,¹³
b Here ..., ic hoopte op di, ende du stoeds mi .. bi,¹⁴
c Hoopt in Gode, ende houdt wel sine ghebode, 608.¹⁵

⁷ In het algemeen geef ik de voorbeelden zoals ik die in de woordenboeken vond, en hetzelfde geldt voor de vindplaatsen; afkortingen worden daarbij niet opgelost.

⁸ Overigens valt op p. 252 te lezen: “in het vroeg Middelnederlands zijn er wel al enige min of meer vaste combinaties aan te wijzen van werkwoorden met een ‘vast’ eigen voorzetsel”. Een van de vroegste vaste combinaties zou *geloven aan* zijn, zoals in *nu saget mir einen kuning other greven, the an uren got wille gelouven* (“noem mij nu een koning of graaf, die aan uw god zou wilen geloven”) (Mfr. Reimb.) (id. 253).

⁹ Mfr.Reimb. A, r. 431 Werden, Essen?, Noord-Oost Nederland, 1151-1200.

¹⁰ Aiol p. 315-316, r. 38-1, Nederrijn, 1220-1240.

¹¹ Rijmb. p. 668, r. 9-11, West-Vlaanderen, 1285. Merk op dat ik *bi desen* opvat als voorlopig voorzetselvoorwerp.

¹² En.Cod. p. 479, r. 4-5, Oost-Vlaanderen, Limburg, 1290

¹³ Segh. 11445.

¹⁴ Doct. I, 602.

¹⁵ Doct. I, 608.

- d Wine hebben anders niet een twint, dan wi hopen tote di,¹⁶
- e Die aen di hopen van ghenade, selen van di ooc sijn beraden,¹⁷

In de WNT-periode tenslotte ligt het voorzetsel nog steeds niet vast, maar nominale complementen ontbreken inmiddels weer in de voorbeelden:

- (6) a Die wy eerst in Christo gehoopt hebben,¹⁸
- b Bedrukte ...! hoop op God,¹⁹

Bij het adjectief *dol* zien we in het oudste voorbeeld het voorzetsel *na*, in de latere voorbeelden steeds *op*:

- (7) a Je zult zekerlyk wel weten ..., dat men jonge naar je buurmeisje Agnietje vryd; Hy is 'er als dol na,²⁰
- b (De poes) krijgt van middag ook weêr een aaltje van mij, o daar issie zoo dol op!²¹

Bij *trots*, een relatief late ontlening (1589) die we dan ook alleen in het WNT aantreffen, wordt bij de relevante betekenis 4 “Bewondering en voldoening gevoelend wegens eigen verdienste, bezit of situatie, of die van een ander, waarbij men zich betrokken gevoelt” expliciet gemeld dat het voorzetsel in het noorden *op* is en in het zuiden soms *over*:

- (8) a De Coninghen van dese steden hebben sommighe Mans personen, dat hare Soldaten ende slaven zijn ... Dese zijn seer hoochmoedich ende trots op haer officie, gaende achter Straten, met eenen grooten moedt,²²
- b By elke loftuiging, die den jongen Luitenant werd toegestuurd, vonkelde er geluk en hoogmoed in den blik der maegd; zy was zoo trotsch over haren vriend, dat zy op dit oogenblik onwetend iedereen in haer hart liet lezen,²³

Uit de hierboven gegeven voorbeelden wordt duidelijk dat we aan de hand van het voorbeeldmateriaal van de verschillende woordenboeken gemakkelijk kunnen nagaan of lexicale elementen voorzetsels regeren of regeerden, of dat vaste voorzetsels zijn, of er eventuele regionale verschillen bestaan, en sinds wanneer dat ongeveer het geval is. We moeten er daarbij evenwel van uitgaan dat de steekproef die uit het beschikbare materiaal gemaakt is, in dit opzicht representatief is. Die garantie hebben we natuurlijk niet.

3.2 Intensiveerders

¹⁶ Alex. VI, 1130.

¹⁷ Boetps. 32, 51.

¹⁸ Statenb., Eph. 1, 12 [ed. 1688].

¹⁹ BILD. 1, 86 [1798].

²⁰ V. EFFEN, Spect. 6, 6 [1733].

²¹ V. LENNEP, K. Zev. 4, 3 [1865].

²² O.-I. e. W.-I. Voyag. 5, 37 d [1602].

²³ CONSC., Bat. 2, 66 [1858].

Ook voor het onderzoeken van vaste versterkende bijwoorden en uitdrukkingen bij werkwoorden lijken de woordenboeken geschikt, maar bij nadere beschouwing vallen de resultaten toch wat tegen. Het werkwoord *slapen* is al heel vroeg geattesteerd, maar in het materiaal van het ONW komt het niet met een versterkend bijwoord voor. In het VMNW vinden we een vergelijking *slapen als een zwijn* die waarschijnlijk intensiverend bedoeld is (9):

- (9) Ende als hi siet dat dat hi beghert. Dranc hi van ioien so den wijn. Dat hi ghinc slapen als een suijn.²⁴

In het MNW vinden we verschillende voorbeelden van de combinatie *vast slapen*:

- (10) a Die moede es van dachvaerde slaept vast op der aerden²⁵
b Nember en slaept ooc so vaste Parthonopeus, hine sal ontspringhen²⁶

Het WNT geeft een forse lijst van versterkende vergelijkingen, waarvan sommige nog steeds bekend zijn:

- (11) a De hulptroepen die op de wallen bivakkeren slapen als ossen en moeten wakker gehouden worden²⁷
b Zij hadden zoo schoon toegetast in de vaderlandsche producten, dat zij alle volmondig of liever volmagig adhaereerden aan het advies van den Burger Succade — zij sliepen als een mof²⁸
c Een gat in den dach slapen. Dormir la grasse matinée²⁹

Voorbeelden met versterkende bijwoorden als *vast* en *diep* treffen we in het lemma over *slapen* echter niet aan.³⁰ Toch is de combinatie *vast slapen* wel degelijk te vinden in het WNT, zij het niet bij *slapen*, maar bij *vast*, waar ze apart besproken wordt (II, 7), en geïllustreerd met tal van voorbeelden:

- (12) a Het is vele te vast ende te scandelijck gheslapen³¹
b Dat daer worden gheschapen zoude De Vrouwe, te wijle dat hy (Adam) vast slapen zoude³²
c Mijn Wijf slaept so vast als een paert³³

²⁴ Rijmb. p. 421, r. 32-34, West-Vlaanderen, 1285

²⁵ Bloeml. 1, 52, 195.

²⁶ Parth. 6537

²⁷ V. REES, T. Poland 2, 33 [1867].

²⁸ DE WAKKER v. z., Steenb. Fam. 4, 34.

²⁹ PLANT. [1573].

³⁰ Wel vinden we in het lemma *slapen* melding van *vast* met een semantisch verwante uitdrukking (*Hij ... was ... door een slaapdrankje ... zoo vast in den dut geraakt, dat hij nog niet regt wakker is, en ik mag dan om den dood niet graag met die luije slaapzakken te doen hebben*, FOKKE, B.R. 3, 67 [1803]) en *diep* met het nomen *slaap* (*Slaap-kruid, dat diepen slaap verwekt, en haar d'ooogen soo vast toe nijpt, of sij Lethes water gedronken hadden*, VALENTIJN, Ovid. 1, 222.)

³¹ in FREDERICQ, Pamfl. 20 [1568].

³² HOUWAERT, Lusth. d. M. 2, 526 [1582].

³³ BREDERO 1, 281 [1613].

Als *vast slapen* te vinden is onder *vast*, dan zou *diep slapen* misschien wel te vinden moeten zijn onder *diep*. Dat blijkt evenwel niet het geval, al vinden we wel vermelding van het adjectief *diep* dat het nomen *slaap* versterkt:

- (13) a Hoewel nu so vele vermaningen tot noch toe gedaen uwen diepen slaep
 weinich hebben cunnen keeren³⁴
 b Geen slaep, hoe diep en vast, Geen' stille nachtrust my van sorregen
 ontlast³⁵

Dat betekent niet dat *diep slapen* een heel moderne combinatie is: via de Digitale Bibliotheek voor de Nederlandse Letteren (www.dbnl.org) en Google vinden we gemakkelijk voorbeelden vanaf het begin van de negentiende eeuw:

- (14) a Althans ik sliep diep, zeer diep in³⁶
 b De heer sliep nog diep, ter wyl de paerden te drinken kregen, en de
 koetsier zyne leden uitstrekke³⁷
 c Ik sliep lang en verkwikkend. Ik droomde helder, terwijl ik diep sliep.³⁸

Ook *werken* is een werkwoord dat al in de oudste fasen van het Nederlands geattesteerd is. De combinatie *hard werken* vinden we andermaal niet bij het werkwoord, maar gelukkig wel bij het bijwoord:³⁹

- (15) a Hy moet hard werken,⁴⁰
 b Hetgeen ik met hard werken verdiend heb,⁴¹

Uit deze voorbeelden blijkt dat er veel informatie over versterkende bijwoorden en – uitdrukkingen in de woordenboeken staat. Door de niet al te systematische behandeling van dit soort informatie is het evenwel niet altijd gemakkelijk, die informatie ook boven water te krijgen – maar misschien toch gemakkelijker dan uit een papieren woordenboek.

3.3 Groepsnamen

Groepsnamen of collectieven vormen een interessante mengeling van gewone taal, vakjargon en sjibbolet: voor een stadsbewoner zijn vrijwel alle wilde dieren onbekend en dus gevaarlijk en eng; het heeft weinig zin om verschillende namen te onderscheiden voor verschillende soorten groepen (enge) beesten (ik beperk me hier tot dieren, maar

³⁴ in FREDERICQ, Pamfl. 24 [1568].

³⁵ VONDEL 2, 708 [1629].

³⁶ Adriaan Loosjes, *Zedelijke verhalen* (1814) 67, via Google books.

³⁷ Hans Christian Andersen, Jacques François Jean Heremans *Prentenboek zonder Prenten* (1857), via Google books.

³⁸ Frederik van Eeden, *Dromenboek* 5-6 februari 1906. Naar de editie Dick Schlüter (Bert Bakker, Amsterdam 1979) via DBNL.

³⁹ De combinatie *hard werken* is ook te vinden via de optie “zoeken binnen citaten”.

⁴⁰ wolff en deken, Leev. 3, 32 [1784].

⁴¹ v. lennep, K. Zev. 4, 323 [1865].

hetzelfde geldt voor andere groepen: *een bos hout, een riem papier, ...*). Voor de boer daarentegen zijn de verschillen in gedrag tussen een groep schapen en een groep koeien en een groep wolven waarschijnlijk groter en interessanter dan de overeenkomsten, en dus gebruikt men verschillende namen voor de verschillende groepen.⁴² Voor de (jagende) elite tenslotte is het wellicht nuttig onderscheid te maken tussen *roedels* en *kudden* en *pakken* en *troepen* omdat die zich verschillend gedragen en verschillend benaderd en bejaagd worden; bovendien is die verfijnde terminologie een leuk gezelschapsspel en een middel om zich te onderscheiden van het vulgus (cf. Lipton 1991, Pauw van Wieldrecht 1985).⁴³

De woordenboeken bieden vanzelfsprekend ook informatie over dit soort groepsnamen. Zo leren we uit het WNT s.v. *trop* dat er naast elkaar twee woorden *trop* en *troep* bestaan, waarvan het eerste de oudste rechten heeft, maar inmiddels alleen nog gewestelijk voorkomt, “met name in het O. en Z. van ons taalgebied”. Het VMNW laat zien dat *trop* in de *Rijmbijbel* gebruikt kon worden voor groepen van mensen en dieren:

- (16) a Daernaer sultu ghemoeten saen. Een trop van propheten gaen.⁴⁴
 b Doe baden si (t.w. de boze geesten die uit een bezetene worden gedreven) dat si moesten varen. In eenen trop suijn die darbi waren. Ende dar in die weide ghinghen.⁴⁵

Het MNW geeft voorbeelden waarbij *trop* mensen, dieren, vogels of vruchten verenigt:

- (17) a Een trop van misseliken (allerlei) lieden⁴⁶
 b Hi sach eenen trop van herten ..., die de coninc hadde ... in een parc besloten⁴⁷
 c Daer heefti inder weiden vonden een groot trop scapen⁴⁸
 d Dese appele sijn harde soete ende men vinter wel C (100) tenen troppe op die bome⁴⁹

Het WNT schrijft s.v. *kudde*: “Waarschijnlijk is de bet. ‘troep vee’ de oorspronkelijke, en is de toepassing van 't woord op een troep van andere wezens, die men in allerlei talen aantreft, eene uitbreiding van dat oorspronkelijk gebruik.” Dit vermoeden kan juist zijn, maar het wordt niet bevestigd door voorbeelden uit de oudere woordenboeken. Het VMNW stelt bijvoorbeeld expliciet: “kudde, troep, van zowel viervoeters als gevleugelde dieren”, en de voorbeelden gaan over varkens en bijen:

⁴² Een andere mogelijke verklaring voor de verschillende namen voor verschillende groepen dieren zou kunnen zijn dat de gewoonte om schapen te houden, met bijbehorende terminologie, uit een andere cultuur overgenomen zou kunnen zijn dan de gewoonte om koeien te houden.

⁴³ Redacteur Frans Heyvaert informeert me dat groepsnamen een aparte categorie zullen krijgen in het *Algemeen Nederlands Woordenboek (ANW)*, het volgende wetenschappelijke woordenboek van het Nederlands.

⁴⁴ Rijmb. p. 210, r. 35-36, West-Vlaanderen, 1285.

⁴⁵ Rijmb. p. 546, r. 31-33, West-Vlaanderen, 1285.

⁴⁶ Alex. II, 341.

⁴⁷ Vl. Rijmk. 2698, Vlaanderen, 1380-1425.

⁴⁸ Franc. 4365, Vlaanderen, 1301-1350 (noot van de MNW-redactie: “het woord schijnt hier en elders onz.”).

⁴⁹ Mandev.2 40, 30, Holland, 1462.

- (18) a Aldaerbi so stont en kudde suijn ende weidde.⁵⁰
 b Aldos so sach die maget goet Of daer si sat. of daer si stoet Of daer si lach in haren beden Die euele gheeste aldar si leden Hare *ende* ghins beneuen hare Gelijcdat het en kudde ware Van bien (*bijen*). die uligen teene swarme⁵¹

In het MNW vinden we vergelijkbare informatie: “*Kudde, troep*, ruimer van opvatting dan ndl. *kudde*, daar het niet alleen van vee gezegd wordt, zoals nog heden, maar ook van vogels en zelfs nu en dan van mensen.”

- (19) a Der gheiten cudde⁵²
 b (Doe) vloghen ... diversche groote cudden van raven over beede die leghers⁵³
 c Tfolc quam met cudden groot, die dat wonder Gods anesaghen⁵⁴

Uit dit kleine onderzoekje kunnen we als hypothese afleiden dat de veronderstelling van het WNT misschien niet correct is: *kudde* blijkt al in de oudste Nederlandse voorbeelden zowel met vee als met andere dieren gebruikt te kunnen worden. We kunnen ons overigens voorstellen dat de uitbreiding al heel erg oud, en de laatste tijd wellicht tegenwerkt wordt door normatieve tendensen, bijvoorbeeld in het onderwijs. Nader onderzoek, ook in verwante talen, zou daar meer licht kunnen doen schijnen. Maar in elk geval is duidelijk dat we de elektronische woordenboeken regelmatig goed kunnen gebruiken om de gebruiksmogelijkheden van groepsnamen te onderzoeken.

3.4 Onder het woordniveau

Bleek het nog vrij gemakkelijk om informatie te vinden over vaste voorzetsels, versterkende bijwoorden en groepsnamen, voor collocaties onder het woordniveau is dat lastiger. Het valt niet mee om systematisch naar versterkende samenstellingen te zoeken. Een woord als *splinternieuw* vinden we alleen in het WNT, met attestaties vanaf 1724:

- (20) a Aan welk heel gebouw wy ook niet konden bemerken, dat een eenig stukje oud hout was, maar alles splinter nieuw⁵⁵
 b In dat ... geval, zal ik het (zeker oud verhaal) zo ten agtersten voor en ten ondersten boven weten te zetten, dat het by de liefhebbers zelfs ... voor splinter nieuw zal kunnen doorgaan⁵⁶
 c Ik bewaarde voor elk kind, Een splinternieuw zeeuwsch Quartje⁵⁷

⁵⁰ Diat. p. 66, r. 3-4, Brabant-West, 1291-1300.

⁵¹ Lutg.K p. 130, r. 24-30, Brabant-West, 1265-1270.

⁵² Hs. Hoogl., 1376-1400.

⁵³ Despars 3, 71, Vlaanderen/Brabant, 1562-1592.

⁵⁴ Brab. Y. II, 582.

⁵⁵ VALENTIJN, O.-I. II, 1, 270 b [1724].

⁵⁶ V. EFFEN, Spect. 9, 146 [1734].

⁵⁷ WOLFF en DEKEN, Econ. L. 1, 145 [1781].

Ook vele andere woorden van dit type vinden we alleen in het WNT, zoals *spinnijdig*, *poedelnaakt*, *ijskoud*, *ijzersterk*, *pikdonker*, *stikdonker*, *bloedheet*, en *pisnijdig*. Op grond daarvan zouden we misschien willen concluderen dat dat woordtype nog niet zo oud is, maar dat blijkt toch ook weer niet te kloppen: *stokoud* blijkt namelijk al in het MNW voor te komen, niet alleen dankzij vermelding door Kiliaan (21a), maar ook in een levensecht voorbeeld (21b):

- (21) a *stockoud*, senex decrepitus, fessus aetate senex; senex tripes, baculo gradum adjuvans (Kil.)
 b Sullen oock ... die schepenen zeer neerstige upsicht hebben up den stockolden personen, die haer goet niet regieren en konnen⁵⁸

Via *stokII* in het WNT (waar ook melding gemaakt wordt van *stokdoof*, *stokarm* (in Gron.), *stokblind* (Kil.), *stokdom* (eenmaal aangetroffen), *stokdonker* en *stokduister* (verouderd), en *stokdood* (in het Zuiden)) komen we er bovendien achter dat de verbinding *stokstil* mogelijk nog aanzienlijk ouder is: het VMNW geeft verschillende voorbeelden, waarbij blijkt dat het woord zowel op beweging (22a) als op geluid (22b) betrekking kon hebben. In het WNT heeft het woord nog steeds beide gebruiksmogelijkheden (22c-f):

- (22) a Mar doe har seluen uant so cranc Die maget ende so ongedegen Dat si ne conste nit volplegen Dis bedens noch na haren wille Bleef si noch liggende al stocstille Ende alineen met groter cracht Met beden worstelde ende vacht Jegen den hogsten van den trone⁵⁹
 b Doen hiefse haer ogen op mettyen ende sprac. en machs mi nv nyet gescien so moet gescien sijn heilege wille jc wil dan swigen al stocstille⁶⁰
 c De Veerman staat stok-stil by zynen Helschen schuyt, En Cerberus hoort toe na 't lieffelyk geluyt⁶¹
 d Ze zag hem staan ..., ginder, stokstil midden in 't groen⁶²
 e Mijn kleyn hartje! waarom Zydy zoo sprakeloos, zo stock-stil en zoo stom?⁶³
 f Maar Iesus zwijgt stok-stil⁶⁴

De bewijsplaatsen waarbij *stokstil* over beweging gaat, lopen evenwel tot laat in de negentiende eeuw (22d), terwijl het jongste voorbeeld van *stokstil* dat op geluid slaat, uit de zeventiende eeuw dateert. Overigens verbaast het me niet dat we geen jongere voorbeelden van *stokstil* in deze betekenis aantreffen: Van Dale vermeldt deze betekenis niet.

4. Systematische collocatiekunde

⁵⁸Stadr. v. Steenw. 47, 20.

⁵⁹Lutg.K p. 433, r. 4-11, Brabant-West, 1265-1270.

⁶⁰Lutg.K p. 433, r. 22-25, Brabant-West, 1265-1270.

⁶¹SCRIVERIUS, Ged. 61.

⁶²LOVELING, D.E. 73 [1891].

⁶³BREDERO 2, 307 [161.].

⁶⁴OUDAAN, Aand. Treurigh. 20 [1660].

In de voorgaande paragrafen heb ik laten zien dat de gedigitaliseerde grote woordenboeken van het Instituut voor Nederlandse lexicologie van grote waarde kunnen zijn voor casuïstisch onderzoek naar (sommige soorten) collocaties: zo kan men een stukje reconstrueren van de historische ontwikkeling van de versterkende bijwoorden *hard* en *zeer*, wordt duidelijk dat groepsnamen als *kudde* niet zo strikt beperkt zijn tot bepaalde diersoorten als we misschien zouden denken, en vinden we bevestiging voor de opkomst van verschillende vaste voorzetsels bij werkwoorden en bijvoeglijke naamwoorden. Met enig geluk kwamen we ook een ontwikkeling in de betekenis van het versterkte bijvoeglijk naamwoord *stokstil* op het spoor.

Als de onderzoeker evenwel minder geïnteresseerd is in een lexicografische, en meer in een algemenere benadering van collocaties, dan zijn de woordenboeken in hun huidige digitale vorm even ongeschikt als hun papieren bronnen. Met de standaard interface op de woordenboeken is het voor zover ik dat kan overzien onmogelijk om een antwoord te krijgen op vragen als: welke bijvoeglijke naamwoorden selecteren het voorzetsel *op*, welke dieren komen voor in *kuddes*, welke werkwoorden komen voor met welke intensiverende uitdrukkingen, welke adjectieven kunnen worden versterkt met *kei* (*keihard*, *keitof*?) of *bloed* (*bloedheet*, *bloedhard*?), en welke partikels en interjecties vaak in elkaars buurt staan.⁶⁵

Toch is die informatie wel aanwezig. In principe hebben we immers met (het citatenmateriaal van) de digitale woordenboeken een prachtig diachroon corpus in handen dat vele eeuwen geschiedenis van het Nederlands overspant. Voor het soort onderzoek dat ik hier beschreven heb zou de toegankelijkheid van dat corpus echter aanzienlijk verbeterd moeten worden. En het kan niet heel moeilijk zijn de relevante informatie toegankelijk te maken.

Stel eens dat de zoek-interface van de woordenboeken uitgebreid zou worden met een module die het mogelijk maakt, alle voorbeeldzinnen (of een selectie op grond van periode, genre, locatie – metadata doen het werk) te onderzoeken met collocatie-software, die regelmatig terugkerende combinaties aan het licht brengt. Dus: we zoeken bijvoorbeeld naar *hopen* en we krijgen bijvoorbeeld een lijst van alle woorden die vaker dan vijf keer onmiddellijk rechts van het gezochte woord werden aangetroffen – niet alleen binnen het lemma *hopen*, maar ook in alle andere lemmata. Zonder twijfel zou het voorzetsel *op* hoog in die lijst voorkomen, en het voegwoord *dat*, maar waarschijnlijk ook wel andere voorzetsels, en misschien ook wel woorden die we niet verwachtten – maar zo gaat dat altijd met corpusgebaseerd onderzoek, er zit ook altijd nog een flinke portie handwerk in.

Voor zover ik kan overzien is het woordenboekenmateriaal in de digitale vorm al voldoende netjes gestructureerd om gelijk aan de slag te kunnen. De algoritmen voor dergelijke zoekoperaties zijn niet bijzonder moeilijk, dus deze verrijking van de zoekmachin zou niet bijzonder veel tijd en inspanning hoeven te kosten. De programmeurs zouden bovendien inspiratie op kunnen doen bij commerciële programma's als *monoconc* (<http://www.monoconc.com/>) en *Wordsmith tools* (<http://www.lexically.net/wordsmith/>), de zoekmachines van het Institut für Deutsche Sprache (<https://cosmas2.ids-mannheim.de/cosmas2-web>) of het Digitale Wörterbuch der deutsche Sprache (<http://www.dwds.de/>), bij een webapplicatie als *webcorp*

⁶⁵ Ook de optie **uitgebreid zoeken** kan ons hier niet helpen.

(<http://www.webcorp.org.uk/>), en bij de gratis toolbox NSP (<http://www.d.umn.edu/~tpederse/nsp.html>).

5. Besluit

Het lijkt erop dat collocaties in de grote woordenboeken niet systematisch behandeld worden. Natuurlijk zijn deze woordenboeken geen productiewoordenboeken, dus misschien mogen we die eis van systematische behandeling van collocaties niet stellen. Aan de andere kant: dit soort informatie is wel relevant, onvoorspelbaar, en in andere naslagwerken niet te vinden. Omdat de informatie primair lexicaal van aard is, zouden we ze in het woordenboek verwachten.

Voor casuïstisch onderzoek naar (sommige soorten) collocaties bleken de gedigitaliseerde wetenschappelijke woordenboeken van het Nederlands van grote waarde: zo kan men de historische ontwikkeling reconstrueren van de versterkende bijwoorden *hard* en *zeer*, van groepsnamen als *kudde*, en van verschillende vaste voorzetsels.

Voor systematisch onderzoek naar collocaties bleken de woordenboeken in huidige vorm echter veel minder geschikt: de schatten in deze schatkamer van de Nederlandse taal zijn wel erg goed verstopt. In de laatste paragraaf van dit opstel heb ik een suggestie gedaan, hoe ze beter toegankelijk gemaakt zouden kunnen worden.

Bibliografie

- Beek, L. van der (2010).** *Van rekenmachine tot taalautomaat.*
- Gazdar, G., E. Klein, G. Pullum & I. Sag (1985).** *Generalized phrase Structure Grammar.* Cambridge, Mass., Harvard University Press.
- Geeraerts, D. (1986).** *Woordbetekenis: een overzicht van de lexicale semantiek.* Leuven: ACCO.
- Groot, H. de, M. de Smit, T. Hacquebord & R. Reinsma (2006).** *Van Dale Groot uitdrukkingenwoordenboek: verklaring en herkomst van moderne uitdrukkingen.* Utrecht [etc.] : Van Dale Lexicografie, Amsterdam [etc.]: Reader's Digest.
- Haeringen, C.B. van (1956).** *Nederlands tussen Duits en Engels*
- Horst, J. van der (2008).** *Geschiedenis van de Nederlandse syntaxis.* 2 dln. Leuven, Universitaire Pers Leuven.
- Jespersen, O. (1917).** *Negation in English and other languages.* Copenhagen: Host.
- Kleijn, P. de (2003).** *Combinatiewoordenboek van Nederlandse substantieven (zelfstandige naamwoorden) met hun vaste verba (werkwoorden).* Rozenberg, Amsterdam.
- Lakoff, G. & M. Johnson (1980).** *Metaphors we live by.* Chicago [etc.]: University of Chicago Press
- Lipton, J. (1991).** *An exaltation of larks, The Ultimate Edition.* Viking.
- Loonen, N. (2003).** *Stante pede gaande van dichtbij langs AF bestemming @ 2003,* diss. Universiteit Utrecht.
- Pauw van Wioldrecht, A. (1985).** *Het dialect van de adel.* Utrecht, HES Uitgevers.
- Renkema, J. (2002).** *Schrijfwijzer.* Vierde druk, Den Haag, Sdu Uitgevers, 2002.

- Royen, P. Gerlach (1947-1954).** *Buigingsverschijnselen in het Nederlands* 4 dln, Amsterdam Noord-Hollandse Uitgevers Maatschappij (Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Afd. letterkunde; nieuwe reeks dl.50 no. 1, dl.52, dl.58 no. 1 en 2, dl.59).
- Sag, I., T. Baldwin, F. Bond, A. Copestake & D. Flickinger (2001).** Multiword Expressions: A Pain in the neck for NLP, Ms Stanford 2001 (LinGO Working Paper No. 2001-03 (CSLI Linguistic Grammars Online (LinGO) Lab, Stanford University); also in *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, Mexico City, Mexico, pp. 1-15.
- Sciullo, A.-M. di & E. Williams (1987).** *On the definition of word*. Cambridge, Mass. MIT Press 1987
- Wouden, T. van der (1992).** Beperkingen op het optreden van lexicale elementen. *De Nieuwe Taalgids* 85, 513-38.
- Wouden, T. van der (2001).** Collocational behaviour in non content words. In Béatrice Daille & Geoffrey Williams: COLLOCATION: Computational Extraction, Analysis and Exploitation. Proceedings of a Workshop during the 39th Annual Meeting of the Association for Computational Linguistics and the 10th Conference of the European Chapter, Toulouse, France, July 7th. Toulouse, France: CNRS - Institut de Recherche en Informatique de Toulouse, and Université de Sciences Sociales, 16-23.
- Wouden, T. van der (2002a).** Particle research meets corpus linguistics: on the collocational behavior of particles. In T. van der Wouden, A. Foolen, & P. Van de Craen (red.), *Particles*. (*Belgian Journal of Linguistics* 16). Amsterdam: John Benjamins, 151-174.
- Wouden, T. van der (2002b).** Partikels: naar een partikelwoordenboek voor het Nederlands. *Nederlandse Taalkunde* 7, 20-43.
- Wouden, T. van der (2007).** On the phraseology of stop words. *Leiden Papers in Linguistics* 4, 56-67. <http://www.hum.leiden.edu/lucl>.
- Wouden, T. van der (2011).** *Over de Nederlandse spreektaal*. Lezing IVG, Warschau, zomer 2010, te verschijnen in de proceedings