

Over lexicale voorkeuren in de alternantie tussen de “skoon bysin” en de “dat-bysin”: Een distinctieve collexeemanalyse*

On lexical preferences in the alternation between the *skoon bysin* “bare complement clause” and the *dat-bysin* “that-complement clause”: a distinctive collexeme analysis

TIMOTHY COLLEMAN*, ILSE FEINAUER[§]

EN CHARLOTTE BRAECKEVELDT

* Universiteit Gent
Vakgroep Taalkunde
België

[§] Dept. Afrikaans en Nederlands
Universiteit Stellenbosch

E-pos: Timothy.Colleman@UGent.be

E-pos: aef@sun.ac.za



Timothy
Coleman



Ilse Feinauer



Charlotte
Braeckveldt

TIMOTHY COLLEMAN is verbonden aan de vakgroep Taalkunde van de Universiteit Gent als hoofddocent Nederlandse Taalkunde. Zijn onderzoek richt zich voornamelijk op de syntaxis en semantiek van grammaticale constructies in het Nederlands en verwante talen, waaronder het Afrikaans. Theoretisch sluit zijn werk aan bij de constructiegrammatica en gerelateerde usage-based benaderingen, methodologisch wordt het gekenmerkt door een corpusgebaseerde aanpak. Hij maakt deel uit van de onderzoeksgroep GLIMS (“Ghent research team on Linguistic Meaning & Structure”) en van het Gents centrum voor het Afrikaans en de studie van Zuid-Afrika.

TIMOTHY COLLEMAN is an Associate Professor in the Dutch section of the Department of Linguistics at Ghent University (Belgium). His research focuses primarily on the syntax and semantics of grammatical constructions in Dutch and related languages, including Afrikaans. This work is informed by key theoretical concepts from construction grammar and related usage-based approaches and is characterized methodologically by the use of qualitative and quantitative techniques for the analysis of corpus data. He is affiliated with the research unit GLIMS (“Ghent research team on Linguistic Meaning & Structure”) and with the Ghent centre for Afrikaans and the study of South Africa.

ILSE FEINAUER is Vicedecaan: Talen in de Faculteit voor Letteren en Sociale Wetenschappen van de Universiteit Stellenbosch en professor in de Afrikaanse taalkunde en vertaalwetenschap in het Departement Afrikaans en Nederlands. In 2014 werd ze benoemd als ereprofessor aan de Taiyuan Universiteit voor Technologie in China. Ze

ILSE FEINAUER is Vice Dean: Languages at the Faculty of Arts and Social Sciences at the University of Stellenbosch. She is also Professor in Afrikaans linguistics and Translation studies in the Dept. of Afrikaans and Dutch. In 2014 she was appointed as Honorary Professor at the Taiyuan University for Technology in China. She co-

* We bedanken Bertus Van Rooy (Noordwes-Universiteit) en twee anonieme beoordelaars voor hun commentaar bij een eerdere versie van dit artikel, en Gilles-Maurice De Schryver (Universiteit Gent) voor de terbeschikkingstelling van het corpusmateriaal uit *Die Burger* 2003.

<p>coördineert en doceert vertaalwetenschappelijke vakken, als onderdeel van de opleidingen vertaling, en daarnaast ook vakken in de Afrikaanse taalkunde, voornamelijk syntaxis. Ze is co-auteur van het <i>Basiswoordeboek vir Afrikaans</i> en redacteur van de verzamelbundel <i>Sintaksis op die voorgrond</i>. Ze heeft gepubliceerd over de syntaxis en lexicografie van het Afrikaans, maar de klemtoon in haar huidige onderzoek ligt vooral op de vertaalwetenschap.</p>	<p>ordinates and teaches translation studies as part of the postgraduate programmes in translation, but she also teaches Afrikaans linguistics, mainly syntax. She is co-author of a Learner's Dictionary of Afrikaans <i>Basiswoordeboek vir Afrikaans</i> as well as co-editor of a volume on Afrikaans syntax. She has published in the field of syntax as well as lexicography, but her research has currently shifted towards translation studies.</p>
<p>CHARLOTTE BRAECKEVELDT studeerde in 2013 af als Master in de Taal- en Letterkunde: Nederlands-Engels aan de Universiteit Gent. In de zomer van 2012 bracht ze een onderzoeksverblijf door aan de Universiteit Stellenbosch, in het kader van haar MA-scriptie over de variatie tussen de bijzin met en zonder <i>dat</i> in het hedendaagse Afrikaans, die werd geschreven onder het promotorschap van Timothy Coleman en Ilse Feinauer. Ze is professioneel actief in het jeugdwerk.</p>	<p>CHARLOTTE BRAECKEVELDT obtained an MA degree in Dutch and English Linguistics and Literature from Ghent University in 2013. She spent a research stay at the University of Stellenbosch in the summer of 2012, in preparation for her MA dissertation on the alternation between the bare complement clause and the <i>dat</i>-complement clause in present-day Afrikaans, which was written under the supervision of Timothy Coleman and Ilse Feinauer. She currently works in youth work.</p>

ABSTRACT

On lexical preferences in the alternation between the skoon bysin “bare complement clause” and the dat-bysin “that-complement clause”: a distinctive collexeme analysis

It is well-known that the alternation in current Afrikaans between (i) the complement clause with dependent word order and an overt complementiser dat (“that”), and the so-called skoon bysin “bare complement clause” without an overt complementiser and with independent word order is determined inter alia by lexical factors. The bare complement clause is said to occur particularly after main clauses with sê (“say”), dink (“think”), glo (“believe”) and similar high-frequency verbs. However, not much empirical research into the preferences of various matrix verbs for the one or the other construction has been done.

This article is aimed at identifying, by means of a distinctive collexeme analysis (Gries & Stefanowitsch 2004), the verbs that show a significant preference for either the dat-complement clause or the bare complement clause in a corpus of general journalistic text. We selected 180 test verbs: 76 verbs were taken from Braeckveldt (2013) and 104 verbs were added to this corpus, mainly from Van Rooy and Kruger (2016). These verbs were all manually analysed and the observed frequencies extrapolated, using all available articles of Die Burger for the period 02/01/2003 to 01/05/2003, in total 6,3 million words. We identified 27 distinctive collexemes for the bare complement clause, i.e. verbs which show a significantly above-average preference for this construction, vs. 59 distinctive collexemes for the dat-complement clause.

When we compare the results of the collexeme analysis with the findings of Van Rooy and Kruger (2016) on the effect of the variable matrix verb, the similarities are greater than the differences. This indicates that the lexical preferences of verbs are quite solid: generally the same verbs are attracted to either the bare complement clause or the dat-clause in different corpora of present-day written Afrikaans. Distinctive collexeme analysis can be seen as an exploratory technique that can yield hypotheses about the semantic variables potentially playing a role in a specific alternation. The results from the quantitative analysis allow for a number of generalisations about the semantic sub-types of verbs attracted towards the bare complement clause or towards the dat-clause. For

instance, causative matrix verbs consistently show a strong preference for the *dat*-clause: five of the ten most distinctive *dat*-collexemes belong to this type (‘provide’), veroorsaak (‘cause’), toelaat (‘allow’), meebring (‘cause/involve’), keer (‘prevent’)). The distinctive collexemes of the bare complement clause, by contrast, almost all belong to the semantic classes of verbs of communication (e.g. sê (‘say’), vertel (‘tell’), laat weet (‘let know’)) and, especially, verbs of cognition (e.g. dink (‘think’), glo (‘believe’), meen (‘think’, ‘mean’), besef (‘realise’), weet (‘know’)), including the closely related sub-class of verbs of perception (sien (‘see’), hoor (‘hear’), voel (‘feel’)).

With this article, we wanted to present empirical data on the lexical preferences of a large and heterogeneous set of matrix verbs for either the bare complement clause or the *dat*-clause, with a view to refining existing views on the exact kinds of verbs attracted to both constructions, and, hence, on the semantic variables that can be included in multifactorial work on the alternation in question. In future research, it would also be interesting to compare the semantic variables co-determining the alternation in Afrikaans to those that potentially play a role in other (Germanic) languages with an alternation between ‘ordinary’ complement clauses and complement clauses displaying main clause characteristics such as different word order and/or the absence of a complementiser (e.g. English, German, Danish).

KEY WORDS: Afrikaans, syntax, word order, complement clause, subordinate clause word order, main clause word order, *dat*-deletion, matrix verb, lexical preference, distinctive collexeme analysis

TREFWOORDEN: Afrikaans, syntaxis, woordvolgorde, complementzin, bijzinsvolgorde, hoofdzinsvolgorde, *dat*-weglating, matrixwerkwoord, lexicale voorkeur, distinctieve collexeemanalyse

SAMENVATTING

Het is bekend dat de variatie in het hedendaagse Afrikaans tussen (i) de complementzin met bijzinsvolgorde en het onderschikkend voegwoord *dat* en (ii) de zgn. ‘skoon bysin’ zonder voegwoord en met hoofdzinsvolgorde onder meer bepaald wordt door lexicale factoren: de skoon bysin zou vooral voorkomen na matrixzinnen met *sê*, *dink*, *glo*, *hoop* of een ander (hoog frequent) werkwoord van communicatie of cognitie (zie bv. Malherbe 1966; Feinauer 1990; Bosch 1999). Empirisch onderzoek naar de lexicale voorkeuren van allerlei matrixwerkwoorden voor deze of gene constructie is echter nog nauwelijks voorhanden. In deze bijdrage wordt aan de hand van een distinctieve collexeemanalyse (Gries & Stefanowitsch 2004) nagegaan welke werkwoorden een significante voorkeur vertonen voor de *dat*-bijzin dan wel de skoon bysin in een corpus hedendaags journalistiek taalgebruik. We vergelijken de resultaten met de bevindingen van Van Rooy & Kruger (2016) en formuleren enkele voorzichtige conclusies over de semantische subtypes van werkwoorden die wel en niet worden aangetrokken tot de constructie met een skoon bysin.

1. INLEIDING

In gebruiksgebaseerde (‘usage-based’) en probabilistische taalkundige theorieën wordt ervan uitgegaan dat er allerlei interacties bestaan tussen *lexicale* en *grammaticale* keuzes: tot de talige kennis van sprekers behoort onder andere gedetailleerde informatie over de neiging van specifieke lexicale items om *samen* voor te komen met specifieke grammaticale constructies – en niet, of relatief minder vaak, met *andere* grammaticale constructies die nochtans in theorie ook mogelijk waren geweest (zie o.m. Goldberg 1995, 2006; Wasow & Arnold 2003; Bresnan 2007; Bybee 2010). Zo is, om een paar voorbeelden te geven uit verschillende talen:

- (i) in het Nederlands bij de transitieve werkwoorden als pakweg *bezorgen* en *gunnen* de kans op het gebruik van de dubbelobjectconstructie met een *nominaal* indirect object veel groter dan die op het gebruik van de constructie met een prepositioneel indirect object, terwijl dat bij werkwoorden als pakweg *afstaan* en *verkopen* precies andersom is (Colleman 2009);
- (ii) in het Zweeds de combinatie van het koppelwerkwoord *vara* (“zijn”) met het perifrastische futurum *ska* (“shall”) veel waarschijnlijker dan de combinatie van *vara* met *ska*’s concurrent *komma att* (“come to”) (Hilpert 2008);
- (iii) in het Engels de kans op een relatiefzin *zonder* het voornaamwoord *that* veel groter na substantieven als pakweg *way of time* dan na substantieven als pakweg *people of stuff* (Wasow, Jaeger & Orr 2011).

De drie voorbeelden komen niet alleen uit verschillende talen, ze betreffen ook heel verschillende fenomenen – resp. de keuze tussen twee argumentstructuur-constructies, de keuze tussen twee perifrastische constructies voor de uitdrukking van toekomstige tijd, en de aan- of afwezigheid van een relativiseerder – zodat we mogen concluderen dat lexicale effecten optreden in uiteenlopende grammaticale domeinen. Een methode die vaak gebruikt wordt om zulke lexicale effecten op te sporen, is *distinctieve collexeemanalyse* (Gries & Stefanowitsch 2004), een techniek die deel uitmaakt van een bredere verzameling van zgn. “collostructionele” methoden die allemaal dienen om de mate van aantrekking na te gaan tussen lexicale items en de constructies waarin ze voorkomen, of tussen verschillende lexicale items die voorkomen in één en dezelfde constructie (zie verder ook nog o.m. Stefanowitsch & Gries 2003, 2005; Gries 2012; Stefanowitsch 2013). In deze bijdrage wordt de methode toegepast op de alternantie in het hedendaagse Afrikaans tussen (1a) de constructie met een complementzin met bijzinsvolgorde (SXV) en ingeleid door de complementeerder *dat* en (1b) de constructie met hoofdzinsvolgorde (SVX) en *zonder dat* (zie daarover onder meer Malherbe 1966; Ponelis 1979: 440-442; Feinauer 1989, 1990; Bosch 1998, 1999; Biberauer 2002; Stell 2011; Kruger & Van Rooy 2015 en ook Van Rooy & Kruger 2016). Het onderzoek is gebaseerd op een uitgebreide en verbeterde versie van de materiaalverzameling van de Gentse masterscriptie van Braeckeveldt (2013), die daarmee ter beschikking komt van de bredere onderzoeksgemeenschap.

- (1) a. Sy het ’n ruk lank uitgegaan en het by haar terugkoms **ontdek** *dat die hut afgebrand het.* (*Die Burger* 14/03/2003)
- b. Hy was net 26 toe Ludwig van Beethoven **ontdek** *hy begin doof word.* (*Die Burger* 19/04/2003)

In wat volgt verwijzen we, in overeenstemming met de gangbare Nederlandse en algemeen-taalwetenschappelijke terminologie, naar de afhankelijke zin met of zonder *dat* als *complementzin* en naar de hoofdzin waarvan die complementzin afhangt als *matrixzin*. Het zinswerkwoord van de hoofdzin is het *matrixwerkwoord*. Naar de complementzin in (1a) verwijzen we in wat volgt als *de volle bijzin* of *dat-bijzin*, de complementzin in (1b) noemen we, met een Afrikaanse term van Ponelis (1979: 440), de *skoon bysin*.¹ In alle aangehaalde voorbeelden staat de complementzin in

¹ In het Nederlands zouden we kunnen spreken van een *kale bijzin*, maar die term is ten eerste niet goed ingeburgerd en dekt ten tweede de lading niet helemaal: “kaal” verwijst naar de afwezigheid van een bindwoord, maar er is natuurlijk nog een tweede belangrijk aspect waarin de complementzin in (1b) afwijkt van die in (1a), nl. in de positie van de persoonsvorm. Daarom gebruiken we in deze bijdrage gewoon de Afrikaanse term.

cursief en het matrixwerkwoord in vetjes. In (informeel) gesproken Afrikaans kom nog een derde constructie voor, nl. een *dat*-bijzin met hoofdzinsvolgorde, zie daarover o.m. Feinauer (1989). In het geschreven Standaardafrikaans is de frequentie van die derde constructie echter verwaarloosbaar klein, zodat we er in dit bestek niet verder op ingaan (zie ook Van Rooy & Kruger 2016).²

De bijdrage is als volgt opgebouwd. Eerst geeft paragraaf 2 een overzicht in vogelvlucht van de bestaande literatuur over de variatie tussen de volle bijzin en de skoon bysin, met bijzondere aandacht voor bestaande hypothesen over de rol van het matrixwerkwoord als factor in die variatie. Daarna worden de werkwijze en de resultaten van het corpusonderzoek gepresenteerd in paragraaf 3 en worden de geobserveerde lexicale voorkeuren nader besproken in paragraaf 4. Paragraaf 5 vat de belangrijkste conclusies van het onderzoek samen.

2. DE SKOON BYSIN VS. DE VOLLE BIJZIN

Het is algemeen aanvaard dat de variatie in (1) onder andere *lexicaal* bepaald is. Op basis van eigen introspectie en/of van frequentietellingen in corpora geven verschillende auteurs een lijstje van matrixwerkwoorden die vaak voorkomen met een skoon bysin. Zo merkt Malherbe (1966:13) op dat de skoon bysin vooral voorkomt bij *belowe, besef, beteken, beweer, bewys, dink, droom, help, hoor, onthou, reken, skat, sweer, toon, uitvind, verbeel, verdra, verneem, verwag, vind, voel* en *weet* – een lijstje dat met instemming wordt geciteerd in Bosch (1999:11). Donaldson (1991:277) zegt dat weglating van *dat* vooral gewoon is na werkwoorden zoals *beweer, dink, glo, hoop, sê, vertel, vertrou, voel, vrees, weet* en *wil hê* en Feinauer (1990:118) noemt *weet, glo, dink, sien, meen, verstaan, hoor, vertel, wens* en *skat*. Hoewel die lijstjes verschillen in de details, tekent er zich een duidelijk algemeen beeld af: de skoon bysin komt vooral voor bij hoogfrequente, semantisch “neutrale” of “algemene” werkwoorden van communicatie en cognitie, terwijl meer formele en laagfrequente werkwoorden met een gespecialiseerdere betekenis de constructie met een *dat*-bijzin verkiezen, al is de skoon bysin ook daarbij niet helemaal uitgesloten. Volgens Feinauer (1990:117) komt in de constructie met een skoon bysin het semantische zwaartepunt op de complementzin te liggen, vandaar dat we die vooral aantreffen na “semanties minder dominante” matrixzinnen: zo is in (2) de skoon bysin veel gewoner in de (a)- dan in de (b)-zin, want “[d]ie hoofsin se semantiese waarde [in 2b] is hoër as dié van [2a], bloot deur die gebruik van die werkwoord *stamel*”; zie ook Kruger en Van Rooy (2015) voor vergelijkbare opmerkingen over de lagere thematische prominentie van de matrixzin in de constructie met een skoon bysin.³

- (2) a. Hy **sê** hy kan nie hoor nie / dat hy nie kan hoor nie.
 b. Hy **stamel** hy kan nie hoor nie / dat hy nie kan hoor nie.

Die hypothese van hogere vs. lagere semantische prominentie is ook bekend uit de uitgebreide literatuur over de weglating van de complementeerder *that* in Engelse complementzinnen, waarin

² In het krantencorpus dat voor dit onderzoek werd gebruikt, troffen we slechts een handvol voorbeelden aan van die derde constructie, waaronder (i) hieronder. Die voorbeelden werden niet meegeteld in de totalen voor de *dat*-bijzin of de skoon bysin.

(i) Hy het gister deur mnr. Milton de la Harpe, sy regsverteenvoerder, ’n beëdigde verklaring aan die hof oorhandig. Daarin **dui** Birgisson **aan** dat hy gaan skuld ontken. (Die Burger 20/03/2003)

³ Overigens kunnen natuurlijk ook andere factoren dan het zinswerkwoord bijdragen tot de lagere semantische dominantie/thematische prominentie van de matrixzin, zoals bv. het gebruik van een pronominaal subject (zie bv. Feinauer 1990:118).

al vaak is gesignaleerd dat *that* het vaakst wegblijft na matrixzinnen met hoogfrequente en semantisch algemene werkwoorden als *think, say, know*, enz. en waarin dat in verband wordt gebracht met de pragmatische functie van zulke zinnen als inleidende epistemische formules veeleer dan als onafhankelijke hoofdzinnen (zie o.m. Bolinger 1972; Thompson & Mulac 1991; Dor 2005; Boye & Harder 2007; Torres Cacoullous & Walker 2009; Wulff *et al.* 2014).

Er is echter weinig of geen *empirisch* onderzoek voorhanden naar de kwestie welke individuele werkwoorden nu de grootste voorkeur voor de skoon bysin of voor de volle bijzin vertonen. In deze bijdrage worden de lexicale voorkeuren van een grote groep van werkwoorden voor de skoon bysin of de *dat*-zin systematisch onderzocht a.d.h.v. een distinctieve collexeemanalyse. De resultaten van de analyse zullen in detail worden vergeleken met de bevindingen van Van Rooy en Kruger (2016), die het matrixwerkwoord meenemen als een van de onafhankelijke variabelen in een multifactoriële studie van de variatie tussen de *dat*-bijzin en de skoon bysin.

3. DISTINCTIEVE COLLEXEEMANALYSE: WERKWIJZE EN RESULTATEN

3.1 Materiaalverzameling

Als corpus gebruikten we de verzamelde artikelen uit de jaargang 2003 van de krant *Die Burger* die digitaal beschikbaar zijn via <http://www.koerantargiewe.media24.com>. *Die Burger* is een groot Afrikaanstalig dagblad met redactiekantoren in de West-Kaap en de Oost-Kaap, dat als kwaliteitskrant een relatief formeel register van geschreven Standaardafrikaans hanteert. De materiaalverzameling voor Braeckevelde (2013) verliep in twee grote fasen. Om er zicht op te krijgen welke matrixwerkwoorden (frequent) voorkomen met een complementzin, werden eerst met behulp van de concordantiemodule van WordSmith Tools (Scott 2012) alle voorkomens van de woordvorm *dat* opgezocht in een sample van ongeveer 6,5 miljoen woorden tekst, nl. alle beschikbare artikelen voor de periode 15/07/2003 tot 31/10/2003. Dat leverde 37409 treffers op, die manueel werden gecontroleerd om alle zinnen te verwijderen waarin *dat* niet als voegwoord werd gebruikt: na die filtering bleven 17192 voorbeelden van complementzinnen met *dat* over, gecombineerd met 332 verschillende matrixwerkwoorden (zie Braeckevelde 2013:87-93 voor de volledige lijst). Daaruit werden alle werkwoorden geselecteerd met een geobserveerde frequentie van minimaal 50 voorkomens met een *dat*-bijzin, 76 werkwoorden in totaal. Die drempel van 50 is uiteraard volkomen arbitrair, de enige bedoeling was om een hanteerbare set samen te stellen van werkwoorden waarvan aangenomen mocht worden dat ze voldoende vaak voorkomen met een complementzin om geschikt te zijn voor de tweede fase van het onderzoek. In die tweede fase werden voor de 76 geselecteerde werkwoorden alle voorkomens van hun verschillende woordvormen opgezocht in een ander corpussample, van vergelijkbare omvang – nl. alle beschikbare artikelen voor de periode 02/01/2003 tot 01/05/2003, samen goed voor 6,3 miljoen woorden tekst. De resultaten van die automatische zoekopdrachten werden manueel doorgenomen om alle voorkomens te identificeren van zowel de constructie met een *dat*-bijzin als de constructie met een skoon bysin, wat uiteindelijk 9046 voorkomens van de skoon bysin opleverde en 5129 voorkomens van de *dat*-bijzin (zie Braeckevelde 2013:21-22).

Voor de huidige studie werd die materiaalverzameling op drie manieren uitgebreid en gecorrigeerd. Ten eerste werden nog 104 matrixwerkwoorden toegevoegd, zodat de uiteindelijke set van testwerkwoorden uit 180 items bestond. Die bijkomende werkwoorden komen vooral uit Van Rooy en Kruger (2016), die eveneens eerst via een zoekopdracht naar *dat* een lijst samenstelden van matrixwerkwoorden die een complementzin kunnen krijgen, maar die (i) daarvoor een ander

corpus gebruikten, nl. het Taalkommissiekorpus⁴ en (ii) uit de resultaten 104 werkwoorden selecteerden “verteenvoerdigend van die volle omvang van werkwoordfrekwensie”, dus opzettelijk inclusief een aantal laagfrequente werkwoorden. Voor alle werkwoorden uit die set van 104 die nog niet in de set van Braeckeveldt (2013) zaten, werden op dezelfde manier als hierboven beschreven alle voorkomens uit het materiaal uit *Die Burger* geëxtraheerd en manueel geanalyseer. Bovendien werden aan de lijst nog een aantal matrixwerkwoorden toegevoegd die genoemd worden bij Malherbe (1966), Feinauer (1990) en Bosch (1999). We beperkten ons daarbij wel tot werkwoorden waarbij een complementzin kan voorkomen als (direct) *object*: werkwoorden als bv. *ontgaan*, *blyk* en *volg* werden niet opgenomen, omdat de complementzin daarbij als subject fungeert – zoals onder meer opgemerkt in Feinauer (1990:117) is de weglating van *dat* in subjectzinnen minder gebruikelijk dan in objectzinnen.⁵

Een tweede wijziging t.o.v. Braeckeveldt (2013) is dat voor de scheidbaar samengestelde werkwoorden ook alle treffers werden opgezocht waarin het partikel en het basiswerkwoord gescheiden voorkomen, aan de hand van automatische zoekopdrachten met WordSmith naar alle voorkomens van het partikel in kwestie in combinatie met een vorm van het basiswerkwoord binnen een maximale afstand van 5 woorden naar links of naar rechts. Ten derde bleek bij een controle op de frequenties in Braeckeveldt (2013) dat voor een vijftiental werkwoorden de gerapporteerde totalen niet klopten, om niet geheel reconstrueerbare redenen (wellicht omdat daarvoor een verkeerde corpusselectie was gebruikt). Voor die werkwoorden werd de analyse overgedaan om tot betrouwbare frequenties te komen. Voor heel frequente werkwoorden werden willekeurige samples van 500 voorkomens (voor werkwoorden met een totale frequentie tussen 1000 en 3000) of 1000 voorkomens (voor werkwoorden met een frequentie van meer dan 3000) manueel geanalyseer en werden de geobserveerde frequenties geëxtrapoleer, zodat we een betrouwbare schatting kregen van de *totale* frequenties van de skoon bysin en de *dat*-bijzin met de werkwoorden in kwestie in het corpus – voor de distinctieve collexeemanalyse is het nl. van belang dat we zicht hebben op de totale frequenties van de twee alternerende constructies in het gebruikte corpus (cf. infra).

Een aantal werkwoorden uit de lijst bleken in het geheel niet voor te komen met een complementzin in het gebruikte corpus: hoofdzakelijk laagfrequente werkwoorden zoals bv. *deklameer*, *snap*, *terugkap* of *wink*, maar ook een aantal werkwoorden die an sich nog relatief frequent zijn maar waarbij het direct object blijkbaar uiterst zelden de vorm van een complementzin aanneemt, zoals *bereik* en *protesteer*. 140 verschillende werkwoorden kwamen wel een of meerdere keren voor met een *dat*-bijzin en/of een skoon bysin. Uiteindelijk werden in totaal 13994 voorkomens van de constructie met een *dat*-bijzin geteld teenover 17868 voorkomens van de skoon bysin: over alle testwerkwoorden heen, komt de skoon bysin dus voor in iets meer dan de helft van de gevallen (56,1 procent). Appendix A geeft een gedetailleer overzicht van de geobserveerde frequenties die dienden als input voor de collexeemanalyse; die tabel vervangt de overzichtstabel in Braeckeveldt (2013:21-22).

⁴ Het Taalkommissiekorpus (Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns 2011) is een groot corpus van hedendaags geschreven Standaardafrikaans dat verschillende registers omvat, waaronder naast journalistiek taalgebruik o.a. ook literaire, academische en administratieve teksten, ten belope van in totaal 57 miljoen woorden lopende tekst.

⁵ Wel opgenomen zijn een aantal werkwoorden waarbij de complementzin strikt genomen niet als direct object, maar als voorzetselsvoorwerp fungeert, al dan niet met weglating van het voorlopige voorzetselvoorwerp, zoals *aandring (daarop)*, *beskuldig (daarvan)*, enz. Bij werkwoorden die zowel een subjectszin als een objectszin kunnen krijgen, zoals *help*, werden enkel de voorkomens met een objectszin meegeteld.

3.2 Resultaten van de distinctieve collexeemanalyse

Distinctieve collexeemanalyse richt zich altijd op een bepaald slot in twee (of meer) alternerende constructies, in dit geval dus het V-slot van de matrixzin waarvan de *dat*-bijzin of skoon bysin afhangt als (direct) object. Op basis van een statistische evaluatie van (i) de frequenties waarmee allerlei lexicale items voorkomen in het betreffende slot van de alternerende constructies in een corpus en (ii) de overkoepelende verhouding tussen die alternerende constructies in dat corpus, wordt berekend welke lexicale items een significante voorkeur hebben voor één van de concurrerende constructies, d.w.z. bij welke items de geobserveerde distributie significant ($p < .05$) verschilt van de distributie die verwacht mocht worden op basis van de totale verhouding tussen de twee constructies in het corpus. Als significantietoets wordt meestal de Fisher-Yates-Exact-test gebruikt, al komen in principe ook andere associatiematen in aanmerking (zie Gries 2012 voor nadere toelichting). De significant aangetrokken lexicale items worden *distinctieve collexemen* genoemd en ze kunnen worden gerangschikt naar de mate van aantrekking tot een van de twee (of meer) alternerende constructies door de p-waarde van de Fisher-Yates-Exact-test als maat te gebruiken: hoe kleiner de p-waarde, hoe sterker het werkwoord in kwestie is aangetrokken tot de betreffende constructie.

De toepassing van een distinctieve collexeemanalyse op de geobserveerde frequenties in Appendix A, met behulp van het R-script voor collostructionele analyse van Gries (2007), levert 12 distinctieve collexemen op voor de skoon bysin, die in (3) worden opgesomd in dalende volgorde van distinctiviteit. Daarnaast blijken maar liefst 88 matrixwerkwoorden significant aangetrokken te zijn tot de *dat*-constructie.

(3) *sê*, *meen*, *dink*, *glo*, *wil hê*, *hoop*, *weet*, *reken*, *voel*, *wens*, *beweer*, *sweer*

Het grote verschil tussen de beide aantallen collexemen valt te verklaren door de bijzondere distributie van de alternerende constructies. Hierboven merkten we op dat de skoon bysin tekent voor 56% van het totale aantal complementzinnen in de materiaalverzameling. Nadere inspectie van de geobserveerde frequenties in Appendix A toont echter dat dat lichte kwantitatieve overwicht in belangrijke mate wordt veroorzaakt door de bijzonder sterke voorkeur voor de skoon bysin van *sê*, *dink* en *meen*, de drie meest frequente werkwoorden in de materiaalverzameling, waarbij de skoon bysin goed is voor resp. 95, 90 en 87 procent van het totale aantal complementzinnen. Geen enkel ander matrixwerkwoord komt bij die scores in de buurt.⁶ De resultaten van het onderzoek leveren dus een duidelijke bevestiging voor het belang van *frequentie* als sturende parameter in de variatie: in overeenstemming met de bevindingen van Van Rooy en Kruger (2016), treffen we de sterkste voorkeur voor de skoon bysin aan bij de werkwoorden die het frequentst worden gecombineerd met een complementzin.

Als we die drie topwerkwoorden niet in rekening brengen, dan zakt de totale frequentie van de skoon bysin tot 6406 en die van de *dat*-bijzin tot 13130, dat is respectievelijk 32.76 en 67.24 procent van de gevallen, dus een veel schevere verhouding. We voerden de distinctieve collexeemanalyse opnieuw uit met die aangepaste totalen. Dat is een enigszins ongebruikelijke ingreep, die echter vanuit gebruiksgebaseerd perspectief goed te motiveren valt. Distinctieve collexeemanalyse houdt rekening met de totale frequenties van de alternerende constructies net omdat ervan wordt uitgegaan dat sprekers probabilistische kennis bezitten over de waarschijnlijkheid waarmee beide constructies voorkomen in het reële taalgebruik: om werkelijk distinctief te

⁶ Met de uitzondering van *proe*, dat echter maar één keer voorkomt met een complementzin in het corpus.

zijn, moet een bepaald lexicaal item niet gewoon significant vaker dan in 50 procent van de gevallen in één van de twee constructies voorkomen, maar significant vaker dan op basis van de overkoepelende verhouding (die zelden fifty-fifty is) verwacht mocht worden. Echter, *sê*, *meen* en *dink* komen *zoveel* vaker voor met een skoon bysin dan andere matrixwerkwoorden dat we ons kunnen afvragen of ze nog veel invloed hebben op de mentale representatie van de relatieve frequenties van de skoon bysin en de *dat*-bijzin in *het algemeen*: in gebruikgebaseerde theorieën wordt ervan uitgegaan dat hoogfrequente vertegenwoordigers van een categorie als afzonderlijke units zitten opgeslagen in het geheugen en in mindere mate bijdragen tot de cognitieve verankering (“entrenchment”) van de categorie als zodanig (zie o.m. Bybee 2010:33-56 en de referenties daar). Als abstractie wordt gemaakt van *sê*, *dink* en *meen*, dan is de kans op een *dat*-bijzin gemiddeld dubbel zo groot als de kans op een skoon bysin: dat betekent dat een werkwoord waarbij de skoon bysin goed is voor pakweg de helft van de gevallen zich daarmee eigenlijk al onderscheidt van het “gemiddelde” alternerende werkwoord. Door in de collexeemanalyse geen rekening te houden met de treffers van *sê*, *meen* en *dink*, zal een dergelijk werkwoord inderdaad worden herkend als een werkwoord dat vaker dan verwacht voorkomt met een skoon bysin. Omgekeerd moet een werkwoord op die manier significant vaker dan in twee derde van de gevallen voorkomen met een *dat*-bijzin om te mogen gelden als werkwoord met een bovengemiddelde voorkeur voor die constructie. De belangrijkste resultaten van de collexeemanalyse met de aangepaste totalen worden gepresenteerd in Tabel 1. Het aantal distinctieve collexemen van de skoon bysin loopt op tot 24, zie de rechterkolom van de tabel; met de toppers *sê*, *meen* en *dink* erbij, hebben we dus in totaal 27 werkwoorden geïdentificeerd die een lexicale voorkeur vertonen voor de skoon bysin. Het aantal collexemen van de *dat*-bijzin loopt terug tot 59: de linkerkolom van Tabel 1 geeft daarvan enkel de 25 meest distinctieve, zie Appendix B voor de volledige lijst. We volgen in Tabel 1 de conventie van later collostructioneel werk om als maat van distinctiviteit niet de p-waarde van de Fisher-Yates-Exact-test zelf maar een negatieve logaritmische transformatie daarvan te rapporteren, zodat die scores makkelijker te interpreteren zijn: hoe groter de waarde, hoe sterker de aantrekkings. In de volgende paragraaf worden de bevindingen van de collexeemanalyse verder besproken.

TABEL 1: Resultaten van de collexeemanalyse (op basis van aangepaste totalen zonder *sê*, *meen* en *dink*)

<i>Dat</i> -bijzin (N=13149)		Skoon bysin (N=6404)	
Collexeem	Distinctiviteit	Collexeem	Distinctiviteit
		<i>sê</i> , <i>meen</i> , <i>dink</i> +	
1. sorg	43.3	1. glo	206.93
2. veroorsaak	34.91	2. weet	77.67
3. beskuldig	32.90	3. hoop	68.2
4. vra	32.71	4. beweer	38.92
5. meebring	30.83	5. wil hê	36.84
6. toelaat	27.92	6. voel	36.62
7. verwag	27.27	7. reken	18.47

TABEL 1: Resultaten van de collexeemanalyse (op basis van aangepaste totalen zonder *sê, meen* en *dink*) (vervolg)

Dat-bijzin (N=13149)		Skoon bysin (N=6404)	
Collexeem	Distinctiviteit	Collexeem	Distinctiviteit
		sê, meen, dink +	
8. ontken	27.01	8. besef	15.31
9. verseker	23.92	9. vermoed	12.2
10. keer	20.96	10. sien	11.86
11. versoek	18.87	11. hoor	11.63
12. toesien	18.33	12. berig	11.58
13. aandrang	17.83	13. wens	10.61
14. aanbeveel	16.67	14. erken	7.74
15. eis	16.63	15. getuig	7.17
16. aanvaar	16.45	16. vertel	5.82
17. voorkom	15.23	17. besluit	4.54
18. beklemtoon	11.73	18. agterkom	4.50
19. voorstel	11.38	19. sweer	3.33
20. voorsien	9.89	20. aanneem	3.01
21. verduidelik	9.52	21. verbeel	2.81
22. bewys	9.06	22. laat weet	2.6
23. bepaal	8.99	23. beslis	2.08
24. noem	8.51	24. droom	1.5
25. bydra	8.47		

4. DISCUSSIE

Er zijn uiteraard een aantal methodologische verschillen tussen onze studie en die van Van Rooy en Kruger (2016). Het huidige onderzoek is niet multifactorieel, zodat we niet kunnen uitsluiten dat sommige van de geobserveerde lexicale effecten geheel of gedeeltelijk voortvloeien uit andere variabelen. Daartegenover staat dat in het huidige onderzoek *meer* matrixwerkwoorden zijn opgenomen en dat de collexeemanalyse twee *gerangschikte* lijsten van collexemen als output geeft, zodat een fijnmaziger beeld kan ontstaan van de lexicale voorkeuren en hun sterkte.

Als we de resultaten van de collexeemanalyse vergelijken met de bevindingen van Van Rooy en Kruger (2016) over het effect van de variabele matrixwerkwoord, dan zijn de overeenkomsten duidelijk groter dan de verschillen. De meeste werkwoorden die Van Rooy en Kruger (2016) opsommen in hun lijstjes (8) en (10) – d.w.z. de groepen van werkwoorden die, op twee verschillende

plaatsen in de classificatieboom, de waarschijnlijkheid van de skoon bysin sterk doen toenemen – komen terug bij de distinctieve collexemen in de rechterkolom van Tabel 1. Bij de werkwoorden met een zo goed als categorische voorkeur voor de *dat*-bijzin in de dataverzameling van Van Rooy en Kruger (2016), opgesomd onder (9) in hun artikel, zitten naast een groot aantal laagfrequente werkwoorden die in ons corpus niet of nauwelijks voorkomen met een complementzin ook allerlei werkwoorden die terugkomen in de linkerkolom van Tabel 1 of verderop in de lijst van distinctieve collexemen van de *dat*-bijzin. Nergens zijn de resultaten echt tegenstrijdig, d.w.z. er is geen enkel werkwoord dat in de data van Van Rooy en Kruger (2016) een voorkeur vertoont voor de *dat*-bijzin maar in onze test geklasseerd wordt bij de significante collexemen van de skoon bysin of andersom. Dat wijst erop dat de lexicale voorkeuren van allerlei werkwoorden behoorlijk robuust zijn: in verschillende corpora van hedendaags geschreven Standaardafrikaans (het Taalkommissiekorpus in het onderzoek van Van Rooy & Kruger 2016, een krantencorpus met materiaal uit Die Burger 2003 voor het huidige onderzoek) zijn het grotendeels dezelfde werkwoorden die worden aangetrokken tot de skoon bysin of de *dat*-bijzin (wat overigens niet wegneemt dat er registerverschillen kunnen optreden, zie hieronder). Onze set van testwerkwoorden bevatte naast de 104 werkwoorden uit het onderzoek van Van Rooy en Kruger (2016) nog 76 andere werkwoorden, waarvan er uiteraard ook verschillende behoren tot de twee reeksen van significant aangetrokken werkwoorden. Een aantal voorbeelden van werkwoorden waarvan de lexicale voorkeur nog niet eerder empirisch was vastgesteld, zijn *sorg*, *veroorsaak*, *beskuldig*, *meebring*, *verwag*, enz. bij de *dat*-collexemen en *meen*, *hoop*, *beweer*, *besef*, *wil hê*, *reken*, *berig*, enz. bij de collexemen van de skoon bysin. Het valt niet uit te sluiten dat sommige van die werkwoorden vooral in het *journalistieke* taalgebruik frequent voorkomen met de *dat*-bijzin of de skoon bysin. Zo laat Bertus van Rooy (persoonlijke communicatie) weten dat *meen*, dat niet voorkwam in de testset van Van Rooy en Kruger (2016), in het Taalkommissiekorpus vooral in de component krantentaal met een skoon bysin voorkomt, terwijl het in andere registers op het eerste gezicht slechts zelden een complementzin als object krijgt. Mogelijk zijn er nog andere werkwoorden waarbij de sterkte van het gemeten effect samenhangt met de keuze voor een corpus krantentaal (zie ook noot 8 verderop).

Het is uiteraard ook interessant om het niveau van de individuele werkwoorden te verlaten en op zoek te gaan naar generalisaties. Collexeemanalyse wordt het best gezien als een exploratieve techniek, die hypothesen kan opleveren over de *semantische* variabelen die potentieel een rol spelen in de alternantie – variabelen die dan in toekomstig onderzoek kunnen worden meegenomen in een multifactoriële test. Van Rooy en Kruger (2016) namen ook een onafhankelijke variabele “semantische categorie van het matrixwerkwoord” op in hun model, waarvoor vijf klassen werden onderscheiden, nl. “kommunikasieproses”, “mentale proses”, “wyse van kommunikasieproses”, “waarborg” en “andere”, hoofdzakelijk voortbouwend op een voorstel van Dor (2005) voor het Engels (cf. infra). Die variabele bleek echter een statistisch onbeduidende rol te spelen. De vraag is dan ook of er op basis van de output van de distinctieve collexeemanalyse misschien andere semantische generalisaties mogelijk zijn: vallen er groepjes van werkwoorden aan te wijzen met een gemeenschappelijk betekenismoment die een consistente voorkeur aan de dag leggen voor de skoon bysin dan wel de *dat*-bijzin?

Een eerste vaststelling in dat verband is dat *causatieve* werkwoorden een consistent sterke voorkeur vertonen voor de *dat*-bijzin: vijf van de tien meest distinctieve *dat*-collexemen zijn van dat type (*sorg*, *veroorsaak*, *toelaat*, *meebring* en *keer*), verderop in de lijst volgen onder meer nog *toesien*, *voorkom*, *bydra* en *verhoed*. Verhagen (2005:141-149) wijst er in zijn constructiegrammaticale analyse van (Nederlandse) complementzinnen al op dat causatieve werkwoorden verschillen van andere types van matrixwerkwoorden doordat ze moeilijk als “mental-space builders” kunnen worden beschouwd. Terwijl matrixzinnen met allerlei werkwoorden van communicatie en cognitie zoals *zeggen*, *beweren*, *denken*, *geloven*, enz. in de meerderheid van de gevallen fungeren als een

instructie aan de hoorder om de inhoud van de complementzin – die het zwaartepunt van de mededeling vormt – vanuit een bepaald perspectief te zien (d.w.z., in plaats van als een verslag van een situatie waarin iemand iets zegt of denkt), benoemen matrixzinnen met werkwoorden als *zorgen, maken, voorkomen*, enz. wel degelijk causatieve *gebeurtenissen*. In het Nederlands correspondeert het semantisch-pragmatische verschil tussen de twee niet met formele verschillen zoals de aan- of afwezigheid van het voegwoord *dat* of hoofdzins- vs. bijzinsvolgorde, maar de parallelie met de bestaande hypothesen over *that*-deletie in het Engels en de skoon bysin in het Afrikaans is duidelijk: het ligt niet voor de hand om een complementzin die afhangt van een causatief zinswerkwoord als semantisch dominant voor te stellen, vandaar de sterke voorkeur voor de *dat*-bijzin bij werkwoorden als *sorg, veroorsaak*, enz. voor de *dat*-bijzin. Toch moet worden opgemerkt dat de skoon bysin bij causatieve werkwoorden niet *volkomen* uitgesloten is: verschillende van de bovengenoemde werkwoorden komen wel degelijk ook een of twee keer voor met de skoon bysin in het corpus. (4) is een geobserveerd voorbeeld met *sorg*.

- (4) Daarom is dit verblydend dat die baanopsigter, AD Carter, alles in sy vermoë doen om te **sorg** *dinge verloop Dinsdag vlot*. (Die Burger 13/03/2003)

Dor (2005) formuleert een algemene semantische hypothese over *that*-deletie in het Engels: dat verschijnsel kan enkel optreden bij “epistemische” werkwoorden, d.w.z. werkwoorden “which entail that a cognitive agent ... has made an epistemic claim concerning the truth of the proposition denoted by the embedded clause” (p. 347). In de praktijk zijn dat werkwoorden die een speech act benoemen zoals *say, tell, claim*, enz., werkwoorden van cognitie zoals *believe, think, conclude*, enz. en uitzonderlijk ook “manner-of-speaking”-werkwoorden zoals *scream, yell* en *whisper*. Bij allerlei andere werkwoorden is *that* verplicht.

We laten in het midden of er in het Engels inderdaad zo’n categorische semantische beperking is, maar in het Afrikaans is dat in elk geval *niet* zo. Hierboven zagen we al dat zelfs bij causatieve werkwoorden uitzonderlijk een skoon bysin kan voorkomen, onder (5) geven we nog een paar voorbeelden met andere matrixwerkwoorden die niet impliceren dat iemand beweert (of ontkent, gelooft, ontdekt, enz.) dat een propositie waar is. Echter, in zulke gevallen is de skoon bysin wel relatief *zeldzaam*: de werkwoorden in (5) behoren tot de distinctieve collexemen van de *dat*-bijzin – al volgen ze wel wat verder in de ranglijst dan de hierboven genoemde causatieve werkwoorden. Nog andere *dat*-collexemen die niet behoren tot de werkwoorden van communicatie of cognitie zijn *ooreenkom, waarborg* en *behels*. Blijkbaar speelt het verschil tussen “epistemische” en andere werkwoorden dus wel degelijk een rol in de alternantie, al is er dan geen categorische beperking.

- (5) a. Die hofbevel **bepaal** *Metrorail moet alle stappe doen wat redelikerwys moontlik is om “behoorlike en voldoende veiligheidsdienste” te voorsien*. (Die Burger 07/02/2003)
- b. Die Kiwi’s se sege oor die Windies Donderdag in Port Elizabeth was nie vir Suid-Afrika ’n goeie uitslag nie. Dit **beteken** *Nieu-Seeland is nou terug in die prentjie*. (Die Burger 15/02/2003)

Merk op dat sommige van die werkwoorden – waaronder in elk geval *bepaal* en *beteken*, en ook de meeste causatieve werkwoorden – relatief vaak een *onbezield* subject krijgen, zoals in (5). Ook bij werkwoorden als *bewys, verklaar* en *bevestig*, die eveneens behoren tot de *dat*-collexemen, is het subject niet noodzakelijk bezield. Van Rooy en Kruger (2016) nemen in hun statistisch model wel de vorm (voornaamwoord of lexicale NP) en de persoon van het subject op als onafhankelijke variabelen, maar geen semantische variabelen zoals de bezieldheid van de subjectreferent: het zou de moeite waard zijn om in toekomstig onderzoek ook die variabele te testen.

De distinctiewe collexemen van de skoon bysin behoren wel (zo goed als) allemaal tot de werkwoorde van communicatie of van cognitie, insluitief de daarby aansluitende werkwoorde van waarneming *sien, hoor* en *voel*.⁷ In dat verband is het nog interessant om even nader in te gaan op een opmerking van Bolinger (1972) over Engels *decide* die word aangehaald in Dor (2005:359-360). *Decide* kan gebruik word as een epistemiesch werkwoord waarvan die betekenis aanleunt by *realize* en *understand* en dan kan *that* wegblywen (bv. *Why didn't you take the job? – I decided I was too old*). Echter, als het gebruik word om een situasie te benoemen waarin “the deciding agent has the power or authority to change reality simply by making the decision” (Dor 2005:360), dan is *that* verpligt aanwesig (bv. *Yesterday the Supreme Court decided *(that) wiretaps must stop*). In Tabel 1 staan drie verskillende Afrikaanse werkwoorde die kunnen diene as vertaalequivalent van *decide*: daarvan behoren *besluit* en *beslis* allebei tot die distinctiewe collexemen van die skoon bysin, en inderdaad word die werkwoorde vaak gebruik om situasies te benoemen waarin iemand op grond van bepaalde oewegingen tot een conclusie komt, zoals in (6) hieronder. *Bepaal* echter, dat enkel die “objectiewe” betekenis heeft – beter geglost as “stipuleren, vastleggen” dan as “tot die conclusie komen” – behoort tot die distinctiewe collexemen van die *dat*-bijzin.

- (6) Williams het gesê met die telling op 1-4 het sy **besluit** sy wil nie meer balle mis nie en graag die wedstryd wen. (Die Burger 18/01/2003)

Die epistemiese werkwoorde vorme op zich natuurlik een heterogene kategorie. Kruger en Van Rooy (2015) suggereren dat werkwoorde van communicatie relatief vaker een voorkeur voor die *dat*-bijzin hebben dan werkwoorde van cognitie: “Verbs that do not allow omission, or generally prefer the full form, tend to be more formal, low-frequency verbs, with a higher incidence of communication verbs”. Dat lijkt te kloppen: sterker selfs, het geldt niet enkel voor laagfrekwente werkwoorde, want by die distinctiewe *dat*-collexemen zitten ook allerlei werkwoorde van communicatie die wel degelik relatief frekwent zijn, zoals *beskuldig, vra, ontken, verseker, aandrang, versoek, eis, aanbeveel, beklemtoon, voorstel*, enz. Die enige werkwoorde van cognitie in die top-25 in die linkerkolom van Tabel 1 zijn *verwag, aanvaar* en *voorsien* (in sommige van zijn gebruikswijze, *voorsien* kan ook as causatief werkwoord word gebruik, verwant aan *sorg*). Andersom zijn by die distinctiewe collexemen van die skoon bysin die werkwoorde van cognitie duidelik in die meerderheid. Die relevante generalisasie lijkt te zijn dat vooral werkwoorde die een *neutrale mededeling* benoemen een voorkeur hebben voor die skoon bysin: *sê, beweer, berig, getuig, vertel, laat weet*.⁸ Het enige niet-neutrale kommunikatiewerkwoord in die rechterkolom is

⁷ Die enige moegelijke uitzondering is het volitionele komplekse werkwoord *wil hê*; echter, ook al is dat strikt genomen geen werkwoord van communicatie of cognitie, het is er via werkwoorde as *hoop, wens* en *droom* – die ook allemaal behoren tot die distinctiewe collexemen van die skoon bysin – wel nauw aan verwant.

⁸ Een anonieme beoordelaar is het niet eens met onze typering van *beweer* as een semantiesch neutraal kommunikatiewerkwoord en vermoedt dat die sterke voorkeur van dat werkwoord voor die skoon bysin veeleer een *register effect* is, dat te maken heeft met die hoge frekwentie van *beweer* in het journalistieke taalgebruik. We laten het oewer aan toekomstig onderzoek om na te gaan of *beweer* ook in andere registers vaker dan verwacht een skoon bysin krijgt. Merk echter op dat we met onze semantiese typering vooral een onderscheid hebben willen maken tussen kommunikatiewerkwoorde die een gewone mededeling benoemen en kommunikatiewerkwoorde die andere types speech acts benoemen. Die keuze van sprekers voor *beweer* as matrixwerkwoord i.p.v. bijvoorbeeld *sê* of *vertel* dient vaak om een zekere twifel te suggereren oewer die waarheid van die gerapporteerde mededeling – in die zin is het werkwoord inderdaad “niet neutraal” (zie Verhagen 2005: 115-117 voor discussie van vergeljikbare werkwoorde). Echter, dat neemt niet weg dat *beweer* in tegenstelling tot bv. *vra, eis, aanbeveel, verseker*, enz. wel degelik een oewerdracht van informasie benoemt waarbij niet spesiaal iets van die toegesprokene verwacht word. In die zin is het werkwoord dan weer wel “neutraal”.

sweer, maar de voorkeur van dat werkwoord voor de skoon bysin lijkt vooral op het conto te komen van de min of meer vaste formules *Ek sweer* en *Jy/n Mens sou sweer*. De communicatiewerkwoorden onder de *dat*-collexemen daarentegen, benoemen vooral (zij het niet uitsluitend) andere types van communicatieve gebeurtenissen: vragen/verzoeken (*vra, eis, versoek, aandring*), ontkenningen (*ontken*), bevelen/suggesties (*beveel, aanbeveel, voorstel*) en diverse minder neutrale mededelingen (*beskuldig, beklemtoon, kla*). Bij de communicatiewerkwoorden lijkt het voor de alternantie tussen de *dat*-bijzin en de skoon bysin dus een rol te spelen *welk soort* speech act benoemd wordt.

5. SLOT

De belangrijkste doelstelling van dit artikel was om empirische data aan te leveren over de lexicale voorkeuren voor de skoon bysin dan wel de *dat*-bijzin van een grote en diverse groep matrixwerkwoorden. De generalisaties in de vorige paragraaf blijven uiteraard relatief aan de oppervlakte, al hebben ze wel een aantal hypothesen opgeleverd over semantische variabelen die kunnen worden meegenomen in toekomstig onderzoek: zo is het aan te bevelen om de causatieve werkwoorden als een afzonderlijke klasse te beschouwen en bij de werkwoorden van communicatie een onderscheid te maken tussen verschillende soorten speech acts, en ook zou het de moeite waard zijn om het effect van de variabele [+/- bezieldeheid van het matrixsubject] systematisch te testen. We hopen dat deze bijdrage zal aanzetten tot verder empirisch onderzoek naar de lexicale en semantische variabelen die de alternantie tussen de skoon bysin en de *dat*-bijzin mee bepalen en naar de overeenkomsten en verschillen met de variabelen die vergelijkbare alternanties beregelen in andere (Germaanse) talen, zoals *that*-deletion in het Engels maar bv. ook het voorkomen van complementzinnen met hoofdzinsvolgorde in het Duits (Auer 1998) en in het Deens (Jensen & Christensen 2013).

BIBLIOGRAFIE

- Auer, P. 1998. Zwischen Parataxe und Hypotaxe: "Abhängige Hauptsätze" im gesprochenen und geschriebenen Deutsch. *Zeitschrift für Germanistische Linguistik*, 26(3):284-307.
- Biberauer, T. 2002. Verb second in Afrikaans: Is this a unitary phenomenon? *Stellenbosch Papers in Linguistics*, 34:19-69.
- Bolinger, D. 1972. *That's that*. Den Haag: Mouton.
- Boye, K. & Harder, P. 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in Language*, 31(3):569-606.
- Bosch, B. 1998. Die onderskikker *dat*: 'n Korpus-gebaseerde bespreking (Deel 1). *South African Journal of Linguistics*, 16(4):120-126.
- Bosch, B. 1999. Die onderskikker *dat*: 'n Korpus-gebaseerde bespreking (Deel 2). *South African Journal of Linguistics*, 17(1):2-15.
- Braeckeveldt, C. 2013. *De skoon bysin in Afrikaanse krantentaal: Voorkomen en parameters*. Ongepubliceerde MA-scriptie. Gent: Universiteit Gent.
- Bresnan, J. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Featherston & Sternefeld (eds). *Roots: Linguistics in search of its evidential base*. Berlin: Mouton De Gruyter, pp. 77-96.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Colleman, T. 2009. Verb disposition in argument structure alternations: A corpus study of the Dutch dative alternation. *Language Sciences*, 31: 593-611.
- Donaldson, B.C. 1991. *The influence of English on Afrikaans. Second edition*. Pretoria: Academica.
- Dor, D. 2005. Toward a semantic account of *that*-deletion in English. *Linguistics*, 43(2):345-382.
- Feinauer, A.E. 1989. Plasing in Afrikaanse afhanklike sinne. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 7(1):30-37.

- Feinauer, A.E. 1990. Skoon afhanklike sinne in Afrikaanse spreektaal. *Suid-Afrikaanse Tydskrif vir Taalkunde*, 8(3):116-120.
- Goldberg, A.E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A.E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gries, S.Th. 2007. Coll.analysis 3.2a. A script for R to compute collostructional analyses.
- Gries, S.Th. 2012. Frequencies, probabilities, association measures in usage/exemplar-based linguistics: some necessary clarifications. *Studies in Language*, 36:477-510.
- Gries, S.Th. & Stefanowitsch, A. 2004. Extending Collostructional Analysis: A corpus-based perspective on “alternations”. *International Journal of Corpus Linguistics*, 9:97-129.
- Hilpert, M. 2008. *Germanic Future Constructions. A usage-based approach to language change*. Amsterdam: John Benjamins.
- Jensen, T.J. & Christensen, T.K. 2013. Promoting the demoted: The distribution and semantics of “main clause word order” in spoken Danish complement clauses *Lingua* 137:38-58.
- Kruger, H. & Van Rooy, B. 2015. Finite declarative complement clauses: Construction forms. *Taalportaal*. <http://www.taalportaal.org/taalportaal/topic/pid/topic-14295917758711433> (geraadpleeg op 21/12/2015).
- Malherbe, G.H. 1966. *Die gebruik van “dat” as verbindingswoord in Afrikaans*. Ongepubliseerde MA verhandeling. Stellenbosch: Stellenbosch Universiteit.
- Ponelis, F.A. 1979. *Afrikaanse sintaksis*. Pretoria: JL van Schaik.
- Scott, M. 2012. Wordsmith Tools version 6. Stroud: Lexical Analysis Software.
- Stefanowitsch, A. & Gries, S.Th. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8:209-243.
- Stefanowitsch, A. & Gries, S.Th. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1:1-43.
- Stefanowitsch, A. 2013. Collostructional analysis. In Hoffmann & Trousdale (eds). *The Oxford handbook of construction grammar*. Oxford: Oxford University Press, pp. 290-306.
- Stell, G. 2011. *Ethnicity and language variation: Grammar and code-switching in the Afrikaans speech community*. Frankfurt: Peter Lang.
- Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2011. Taalkommissiekorpus 1.1. Noordwes-Universiteit: CText.
- Thompson, S.A. & Mulac, A. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, 15(3):237-251.
- Torres Cacoullos, R. & Walker, J.A. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics*, 47(1):1-43.
- Van Rooy, B. & Kruger, H. 2016. Faktore wat die weglating van die Afrikaanse onderskiker *dat* bepaal. *Tydskrif vir Geesteswetenskappe*, 56(1):102-116.
- Verhagen, A. 2005. *Constructions of intersubjectivity: Discourse, syntax and cognition*. Oxford: Oxford University Press.
- Wasow, Th. & Arnold, J. 2003. Post-verbal constituent ordering in English. In Rohdenburg & Mondorf (eds). *Determinants of grammatical variation in English*. Berlin: Mouton De Gruyter, pp. 119-154.
- Wasow, Th., Jaeger, T.F. & Orr, D. 2011. Lexical variation in relativizer frequency. In Simon & Wiese (eds). *Expecting the unexpected: Exceptions in grammar*. Berlin: Mouton De Gruyter, pp. 175-195.
- Wulff, S., Lester, N. & Martinez-Garcia, M.T. 2014. *That*-variation in German and Spanish L2 English. *Language and Cognition*, 6(2):271-299.

APPENDIX A: Overzicht van de geobserveerde frekwensies voor de 180 testwerkwoorde, met het aandeel van die *dat*-bijzin in het totaal

werkwoord	#dat	#sko	%dat	werkwoord	#dat	#sko	%dat	werkwoord	#dat	#sko	%dat
aanbeveel	132	5	96	herbeklemtoon	0	0	-	spekuleer	0	0	-
aandring	103	0	100	herbevestig	1	0	100	stamel	0	0	-
aandui	202	51	80	herhaal	10	5	67	stateer	0	0	-
aankondig	335	89	79	herinner	11	0	100	stel	46	15	75
aanneem	18	24	43	hoop	207	414	33	stipuleer	3	0	100
aansê	3	1	75	hoor	117	137	46	sweer	2	10	17
aanstip	0	0	-	inhou	3	0	100	te kenne gee	74	19	80
aanvaar	176	14	93	inlig	79	10	89	teekap	0	0	-
afkondig	0	0	-	insien	9	1	90	teoretiseer	0	0	-
aframmel	0	0	-	insinueer	3	0	100	terugkap	0	0	-
afspreek	0	0	-	keer	121	0	100	terugskryf	0	0	-
agiteer	0	0	-	kerm	0	0	-	toelaat	161	0	100
agterkom	52	56	48	kla	64	13	83	toesien	115	1	99
antisipeer	0	0	-	konkludeer	0	0	-	toesnou	3	0	100
argumenteer	15	2	88	konstateer	2	0	100	toon	294	70	81
baklei	1	0	100	laat weet	46	42	52	uitbasuin	1	0	100
beaam	8	0	100	leer	46	10	82	uitblaker	0	0	-
bedoel	13	4	76	lees	23	6	79	uitspreek	3	0	100
begryp	6	2	75	meebring	188	1	99	uitvind	14	12	54
behels	30	0	100	meedeel	70	12	85	verbeel	4	11	27
beken	9	2	82	meen	255	1685	13	verdra	1	0	100
bekendmaak	141	45	76	meld	36	3	92	verduidelik	125	14	90
beklemtoon	150	16	90	mompel	0	0	-	vergeet	31	6	84
beloof/belowe	41	25	62	mor	1	0	100	verhoed	49	1	98
bemerk	2	0	100	neul	0	0	-	verifieer	0	0	-
benadruk	101	26	80	noem	63	2	97	verklaar	60	6	91
bepaal	136	18	88	ontdek	17	14	55	verkondig	9	0	100
bepleit	1	0	100	ontgaan	0	0	-	verlang	2	0	100
bereik	0	0	-	onthou	121	52	70	vermaan	0	0	-
berig	164	171	49	onthul	23	8	74	vermoed	65	98	40
besef	323	300	52	ontken	296	24	93	verneem	117	69	63
beskuldig	200	1	100	oordeel	4	0	100	veroorzaak	201	0	100
beslis	72	55	57	ooreenkom	27	3	90	verordineer	0	0	-
besluit	139	114	55	oortuig	163	63	72	verseker	439	67	87
bespiegel	41	4	91	oppas	3	0	100	versoek	109	0	100
beteken	708	237	75	paai	0	0	-	verstaan	56	19	75
betreur	4	0	100	postuleer	0	0	-	verswyg	3	0	100
beveel	29	1	97	prewel	1	0	100	vertel	187	153	55
bevestig	412	131	76	proe	0	1	0	vertrou	30	13	70
bevind	235	82	74	profeteer	0	0	-	verwag	510	79	87
beweer	220	328	40	proklameer	0	0	-	verwittig	6	1	86
bewys	334	81	80	propageer	2	0	100	verwyty	4	0	100

APPENDIX A: Overzicht van de geobserveerde frekwenties voor de 180 testwerkwoorden, met het aandeel van de *dat*-bijzin in het totaal (*vervolg*)

werkwoord	#dat	#sko	%dat	werkwoord	#dat	#sko	%dat	werkwoord	#dat	#sko	%dat
bieg	8	4	67	protesteer	0	0	-	vind	89	18	83
bluf	1	0	100	raai	7	5	58	voel	150	259	37
bydra	49	0	100	raaklees	0	0	-	voorgee	25	14	64
byvoeg	204	73	74	reken	29	83	26	voorkom	88	0	100
deklameer	0	0	-	rondvertel	0	0	-	voorsien	65	1	98
deursein	0	0	-	saamstem	103	16	87	voorspel	94	20	82
dikteer	0	0	-	sê	476	8564	5	voorstel	135	13	91
dink	133	1213	10	sien	117	138	46	vra	301	17	95
droom	5	8	38	sinspeel	28	0	100	vrees	76	25	75
eis	105	1	99	skat	12	12	50	waarborg	9	0	100
erken	263	214	55	skel	0	0	-	waarsku	330	114	74
fantaseer	0	0	-	skerts	7	0	100	wed	6	1	86
fluister	4	3	57	skimp	3	1	75	weet	305	535	36
getuig	91	97	48	skryf	41	23	64	wegneem	4	0	100
gewaar	0	0	-	smeek	2	2	50	wens	15	45	25
gis	10	1	91	snap	0	0	-	wil hê	32	135	19
glo	338	934	27	soebat	0	0	-	wink	0	0	-
help	28	0	100	sorg	294	5	98	wys	620	219	74

APPENDIX B: Volledige lijst van alle distinctiewe ($p < .05$) *dat*-collexemen

sorg, veroorsaak, beskuldig, vra, meebring, toelaat, verwag, ontken, verseker, keer, versoek, toesien, aandrang, aanbeveel, eis, aanvaar, voorkom, beklemtoon, voorstel, voorsien, verduidelik, bewys, bepaal, noem, bydra, toon, aankondig, verhoed, beteken, saamstem, inlig, bevestig, aandui, verklaar, behels, wys, sinspeel, help, beveel, meedeel, vind, bespiegel, voorspel, meld, waarsku, kla, benadruk, ooreenkom, bevind, te kenne gee, bekendmaak, leer, byvoeg, herinner, vergeet, waarborg, verkondig, beaam, argumenteer