

# Ter inleiding

Het Nederlands is gezegend met een geschiedenis van sterke regionale variatie, zoals de dialectologie en historische taalkunde de afgelopen anderhalve eeuw overtuigend hebben laten zien. Tegelijk heeft er de afgelopen vier eeuwen een proces van standaardisatie plaatsgevonden, dat naast of 'boven' de regionale variatie tot eenheid geleid heeft. In woordenboeken als Van Dale en de ANS wordt die eenheidstaal beschreven. Echter, iedereen die meer dan oppervlakkig van deze beschrijvingen gebruikt heeft gemaakt, weet dat zowel Van Dale als de ANS zich bij hun beschrijvingen van woorden en constructies meer dan eens genoodzaakt zagen predikaten als 'regionaal' en 'gewestelijk' te gebruiken. De eenheidstaal is blijkbaar minder één dan de aanduiding 'standaardtaal' suggereert. Het behoeft geen betoog dat daarbij het feit dat het Nederlands in twee landen een nationale status heeft een centraal gegeven is. De kwestie van de nationale identiteiten binnen de eenheid en de eventuele effecten daarvan op die eenheid (de kwestie van convergentie versus divergentie) zal daarom in elk onderzoek naar variatie in de eenheidstaal een prominente plaats innemen.

Zoals de ANS en ook Van Dale zelf toegeven, en zoals ook herhaaldelijk door onderzoekers is opgemerkt (zie de proefschriften van Haeseryn (1990) en Van de Velde (1996)), zijn de predikaten 'regionaal' e.d. grotendeels gebaseerd op globale indrukken en niet op wetenschappelijk onderzoek. De noodzaak van wetenschappelijk onderzoek naar variatie binnen de standaardtaal is door diezelfde onderzoekers meerdere malen benadrukt.

Dat de nationale identiteiten een belangrijke rol lijken te spelen in de variatie binnen de eenheidstaal maakt onderzoek van dit type uitermate geschikt voor subsidiëring door het VNC-fonds. En het VNC-fonds is de afgelopen jaren inderdaad een stimulerend kader gebleken voor het entameren van dergelijk onderzoek.

VNC staat voor 'Vlaams-Nederlands Comité voor Nederlandse taal en cultuur', een commissie die gelden toewijst aan Vlaamse en Nederlandse onderzoekers die in samenwerking met elkaar onderzoek willen doen op de terreinen van de Nederlandse taalkunde, letterkunde en geschiedenis. De gelden zijn afkomstig van FWO-Vlaanderen, NWO, en verder van het Vlaamse Ministerie voor Onderwijs en Ambtenarenzaken en het Nederlandse Ministerie van Onderwijs, Cultuur en Wetenschappen. De looptijd van het VNC-programma is 1993-2002. (Zie de website van NWO, <http://www.nwo.nl> – click op *Internationaal* en ga door naar *FWO*.)

De afgelopen jaren hebben meerdere taalkundige onderzoeksprojecten geprofiteerd van deze subsidiebron. Momenteel lopen er zelfs 4 projecten tegelijk, waarvan er drie al geruime tijd bezig zijn en het vierde betrekkelijk recent, per 1 januari 2000, van start is gegaan. Dit gegeven was voor de redactie van Nederlandse Taalkunde aanleiding om een 'VNC-

nummer' samen te stellen. Medewerkers van de lopende projecten werden uitgenodigd een bijdrage te leveren over hun onderzoek waarin de vraagstelling, aanpak en zo mogelijk eerste resultaten gepresenteerd zouden worden.

De oogst bleek teveel voor één aflevering. Vandaar dat besloten werd om de bijdragen te verdelen over twee afleveringen, het nummer dat nu voor u ligt en het eerste nummer van de volgende jaargang. In het huidige nummer vindt u de volgende bijdragen:

- Grondelaers et al., die verslag doen van hun lexicologisch onderzoek over voorzetsels in het kader van het project "Lexicale variatie in het Standaardnederlands. Convergentie/divergentie en standaardisering/substandaardisering in het Belgische en Nederlandse Nederlands".
- Van den Heede et al., die participeren in hetzelfde project en zich met name richten op verba.
- Cornips en Jongenburger, die de methode van onderzoek uiteenzetten van het op syntactische variatie gerichte project getiteld "Een syntactische atlas van Nederlandse dialecten (SAND)". Dit is het jongste van de vier projecten en ook het enige dat zich juist met variatie in niet-standaard Nederlands bezighoudt.

Hiermee zijn slechts twee van de vier VNC-projecten aan bod gekomen. In nummer 7:1 zult u kennis kunnen nemen van bijdragen vanuit de overige projecten, die gaan over uitspraakvariatie resp. partikelgebruik in Nederland en Vlaanderen.

Zoals gezegd, het Nederlands is gezegend met een rijke taalvariatie, en de standaardtaal blijkt in die rijkdom te delen. Daarbij gaat het niet om een statische variatie maar om een variatie in verandering, een variatie die onderhevig is aan convergerende en divergerende tendenties. Longitudinaal onderzoek is daarom juist op dit gebied interessant en gewenst – ook op het nog niet door een VNC-project afgedekte deel terrein van syntactische variatie in de standaardtaal. Het valt te hopen dat de uitkomsten van de lopende projecten voor het VNC aanleiding zullen vormen om aan de subsidiënten duidelijk te maken dat het zonde zou zijn om de looptijd van het VNC na 2002 niet te verlengen.

De redactie

## ● Verwijzingen

**Haeseryn, W.J.M. (1990).** *Syntactische normen in het Nederlands: een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep*. [S.l.: s.n.] Dissertatie KU Nijmegen.

**Velde, Van de (1996).** *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. [S.l.: s.n.] Dissertatie KU Nijmegen.

# Inhoudswoorden en preposities als standaardiseringsindicatoren

## De diachrone en synchrone status van het Belgische Nederlands

STEFAN GRONDELAERS  
HILDE VAN AKEN  
DIRK SPEELMAN  
DIRK GEERAERTS\*

### Abstract

This paper reports on a series of corpus-based investigations into the diachronic and synchronic status of Belgian Dutch, a variant whose standardisation was delayed in the 16<sup>th</sup> and 17<sup>th</sup> C and is as yet incomplete. Building on a newly developed measure of linguistic overlap and an extensive corpus of synonymous designations of 30 clothing and soccer concepts, a first study (Geeraerts, Grondelaers & Speelman 1999) confirmed the alleged diachronic convergence between Belgian and Netherlandic Dutch from 1950 to 1990, as well as the larger distance traditionally assumed between standard and substandard language in Belgian Dutch (a major indication of the latter's delayed standardisation). Whereas the first investigation relied on shopwindow materials to represent substandard language, the two follow-up studies reported on in this paper make use of a new corpus containing Internet materials as representatives of substandard register – viz. Internet Relay Chat and Usenet. In the first follow-up study, the diachronic and synchronic findings are replicated on the basis of clothing terms. In the second, the basis of comparison between different variants of Dutch is extended from content words to function words, viz. prepositions. Although the lexical as well as the prepositional indicators fully confirm the diachronic convergence between Belgian and Netherlandic Dutch and their different synchronic stratification, the present methodology does not allow us to compare the respective impact of the indicators on the latter.

### I Inleiding

Dit artikel rapporteert over het onderzoek van de Leuvens-Amsterdamse pool van het VNC-project “Convergentie en divergentie in de Nederlandse woordenschat”. Dit project wil een empirisch ondersteund antwoord vinden op de vraag of het Belgische Nederlands

\* Stefan Grondelaers, Dirk Speelman en Dirk Geeraerts zijn verbonden aan het Departement Linguïstiek van de KU Leuven. Hilde Van Aken is OIO bij de Vakgroep Algemene Taalwetenschap van de VU Amsterdam. Stefan Gron-

- convergeert met het Nederlandse Nederlands, of een autonome variant van het Nederlands vertegenwoordigt. Deze descriptieve vraag is gekoppeld aan een controversieel taalpolitiek debat: moet het Belgische Nederlands het Nederlandse Nederlands als norm overnemen, of is er plaats voor een officiële Belgische variant van de Nederlandse standaard? Wij beperken ons in dit artikel uitsluitend tot de descriptieve kwestie.

Wie niet op de hoogte is van de geschiedenis van het Belgische Nederlands, herinnere zich dat in de Renaissance in de meeste Europese landen een linguïstische standaardiseringsbeweging op gang kwam. In Vlaanderen echter werd de ontluikende standaardisering geblokkeerd als gevolg van het feit dat Vlaanderen in de Tachtigjarige Oorlog politiek geïsoleerd raakte van Nederland. In plaats van een eigen standaard te ontwikkelen, nam Vlaanderen in toenemende mate het Frans over voor supraregionale communicatiedoel-einden. Hoewel het Nederlands bij de hereniging van de Nederlanden (1814-1830) opnieuw tot officiële Vlaamse taal gepromoveerd werd, en – vooral vanaf 1930 – zijn standaardisering hervatte, zijn de meeste linguïsten het erover eens dat de standaardisering van het Belgische Nederlands nog niet voltooid is.

Met betrekking tot de status en de ontwikkeling van het Belgische Nederlands zijn in de linguïstische literatuur twee hypothesen te vinden. Om te beginnen wordt een *diachrone convergentie* tussen het Belgische en Nederlandse Nederlands verondersteld, die in hoofdzaak het gevolg is van veranderingen aan Belgische zijde: de standaardisering van het Belgische Nederlands, zo vermoedt men, wordt gekenmerkt door een expliciete normatieve oriëntatie op het Nederlandse Nederlands. Daarnaast manifesteert de standaardiseringsachterstand van het Belgische Nederlands zich in een grotere *synchrone afstand* tussen het Belgische regionale en supraregionale register dan tussen het Nederlandse regionale en supraregionale register; zelfs voor de meest ongetrainde waarnemer is het duidelijk dat de afstand tussen het discours van de modale Belgische wielrenner en het taalgebruik van de Belgische nieuwslezer veel groter is dan de afstand tussen het register van Nederlandse wielrenners en nieuwslezers. In wat volgt verwijzen we naar deze twee hypothesen als resp. de *diachrone convergentiehypothese* en de *synchrone stratificatiehypothese*.

Omdat de synchrone stratificatiehypothese en de diachrone convergentiehypothese haast uitsluitend gebaseerd zijn op anekdotische observatie en m.b.v. enquêtes geëliciteerde data (die niet noodzakelijk overeen hoeven te stemmen met spontaan taalgebruik, zie onder meer Labov 1972, Geeraerts e.a. 1999:24 en Grondelaers 2000:59-63 en 256), lanceerde de onderzoeksgroep rond Dirk Geeraerts in 1994 een grootschalig corpusgebaseerd onderzoek naar de lexicale standaardisering van het Belgische Nederlands (een volledig verslag is te vinden in Geeraerts, Grondelaers & Speelman 1999, verder Geeraerts e.a. 1999). Dat gebeurde op basis van een 40.000 observaties tellende database met synoniemenreeksen voor 15 kleding- en 15 voetbalbegrippen.<sup>1</sup> In de materiaalverzameling

delaers en Hilde Van Aken zijn de Leuvense en Amsterdamse uitvoerder van de Leuven-Amsterdamse pool van het VNC-project “Lexicale variatie in het Standaardnederlands. Convergentie/divergentie en standaardisering/ substandaardisering in het Belgische en Nederlandse Nederlands”. Correspondentie i.v.m. dit artikel wordt bij voorkeur aan hen gericht: stefan.grondelaers@arts.kuleuven.ac.be en hmm.van.aken@let.vu.nl. Het ruimere VNC-project is een samenwerkingsverband tussen de Katholieke Universiteit Leuven, de Vrije Universiteit Amsterdam, de Universiteit Gent, en het I.N.L. Leiden.

1 Appendix 1 en 2 bevatten een overzicht van de gebruikte synoniemenreeksen op het gebied van kleding en voetbal.

werd niet alleen een *geografische* en een *diachrone* dimensie ingebouwd, waardoor Belgisch en Nederlands materiaal uit 1950, 1970 en 1990 vergeleken kon worden; we implementeerden ook een *stilistisch-stratificationele* dimensie, door naast supraregionaal materiaal uit modetijdschriften regionale gegevens uit de etalages van kledingboutiques in twee Belgische en twee Nederlandse steden op te nemen.

In paragraaf 2 van dit artikel worden kort de basisgegevens uit Geeraerts e.a. (1999) samengevat. De rest van het artikel is gewijd aan twee nieuwe studies waarin we m.b.v. een speciaal daarvoor samengesteld tekstcorpus de basisgegevens trachten te repliceren. Paragraaf 3 rapporteert over een lexicale replicatie met de kleding- en voetbaltermen uit Geeraerts e.a. (1999); paragraaf 4 is gewijd aan een replicatiestudie op basis van voorzetselverbindingen. In paragraaf 5, tenslotte, vatten we samen en verkennen we enkele voor de hand liggende parcours voor verder onderzoek.

## 2 Corpusgebaseerd onderzoek naar de standaardisering van het Belgische Nederlands

### 2.1 Materiaal

De empirische basis van Geeraerts e.a. (1999) bestond uit 40.000 observaties van spontaan taalgebruik. We verzamelden de verschillende benamingen (en hun respectievelijke frequenties) die gebruikt worden om 15 kledingbegrippen en 15 voetbalbegrippen aan te duiden. Op basis van de resulterende database kunnen we voor een concept als BUITENSPEL nagaan hoe in Belgische en Nederlandse publicaties de kwantitatieve verhouding is tussen de term *buitenspel* en het leenwoord *offside*; in het geval van het begrip JURK kunnen we vaststellen of er een voorkeur bestaat voor benoeming met *jurk*, *kleed* of *japon*.

Merk op dat de 30 onderzochte kleding- en voetbalbegrippen niet – zoals in het mees- te variatieonderzoek – gepreselecteerd werden op basis van hun variatiegevoeligheid. Door ons onderzoek op een representatieve selectie van begrippen uit twee lexicale velden te baseren, trachten we het risico op beeldvervorming te voorkomen: wanneer de belangstelling vooral gericht is op termen waarvan de variatiegevoeligheid vooraf vermoed wordt, dan heeft men er nog geen idee van wat die termen vertegenwoordigen in het geheel van het taalgebruik. Een uitgesproken Belgische voorkeur voor de purismen *regenscherm*, *duimspijker* en *geldbeugel*, bijvoorbeeld, hoeft nog geen divergentie met het Nederlandse Nederlands te impliceren als die puristische tendens kwantitatief gezien ver- waarloosbaar is in een voor de rest identieke woordenschat.

De kern van de geobserveerde gegevens bestond uit kleding- en voetbalmateriaal uit 1990. Die kern werd in twee richtingen uitgebreid. Om te beginnen werd vergelijkbaar materiaal verzameld voor 1950 en 1970, hetgeen ons in staat stelde een “real-time”- onderzoek naar lexicale convergentie- of divergentieprocessen uit te voeren. Daarnaast werd rekening gehouden met de stilistische stratificatie van het taalgebruik. Tussen stand- daard en dialect bevinden zich namelijk verschillende tussenniveaus waarop registerver- schillen (zoals een toenemende graad van informaliteit) samengaan met een toenemende geografische specialisatie. Voor een onderzoek naar de linguïstische verhouding tussen het Belgische en het Nederlandse Nederlands zijn die tussenstrata tussen standaard en dialect

buitenmate interessant: we kunnen namelijk veronderstellen dat de verschillen tussen Belgisch en Nederlands Nederlands zullen toenemen op het regiolectische niveau.

In Geeraerts e.a. (1999) werd het tussenliggende taalstratum vertegenwoordigd door termen die we hebben aangetroffen op prijskaartjes en labels in etalages van kledingwinkels in de Belgische steden Leuven en Kortrijk en de Nederlandse steden Leiden en Maastricht. Het beoogde publiek van deze vorm van communicatie is beperkter dan het nationale of binationale publiek van de kledingtijdschriften waaruit de kern van de materiaalverzameling geselecteerd werd. Het feit dat we te maken hebben met geschreven semi-formeel taalgebruik belet anderzijds dat we terechtkomen bij het zuiver dialectische eind van het stratificatiele continuüm.

## 2.2 Methode

Hoe kunnen we lexicale convergentie of divergentie tussen twee taalvariëteiten meten? In Geeraerts e.a. (1999) werd een overlappingsmaat gedefinieerd die gebaseerd is op de noties *onomasiologisch profiel* en *uniformiteit*.

Het *onomasiologisch profiel* van een bepaald concept in een bepaalde bron of bronngroep definiëren we als het geheel van de alternatieve benamingen voor dat concept met een specifieke frequentieverhouding in die bron of bronngroep. Tabel 1 bevat de onomasiologische profielen voor *OVERHEMD* in het Belgische en Nederlandse materiaal uit 1990 (resp. B90 en N90 genoemd):

	B90	N90
hemd	31%	17%
overhemd	69%	46%
shirt	0%	37%

Tabel 1: Onomasiologische profielen voor *OVERHEMD* in het Belgische en Nederlandse materiaal uit 1990

*Uniformiteit* is een kwantitatieve maat voor de overeenkomst tussen twee onomasiologische profielen. Onze berekening van uniformiteit vertrekt vanuit het idee dat een gemeenschappelijke linguïstische norm gemeenschappelijk taalgedrag genereert. Maximale uniformiteit bij de benoeming van een begrip komt slechts voor wanneer twee taalvariëteiten dezelfde naam voor dat begrip gebruiken, of verschillende namen die identiek gedistribueerd zijn, i.e. dezelfde relatieve frequentie hebben.

Veel frequenter echter dan zulke voorbeelden van maximale uniformiteit zijn partiële overlappingsen zoals in tabel 1. Laat ons, bij wijze van illustratie, veronderstellen dat de relatieve frequenties in tabel 1 100 reële gebruiksgevallen in elk van beide profielen vertegenwoordigen. De partiële overlapping tussen de profielen in tabel 1 kwantificeren we dan door in beide profielen de gebruiksgevallen te tellen die een tegenhanger hebben in het andere profiel. De ideale scenario's uit de vorige paragraaf – twee taalvariëteiten hebben één term voor hetzelfde concept of meerdere termen maar dan met dezelfde frequentieverhouding – leveren in dit verband een maximale uniformiteit op. In tabel 1, echter, vinden we 14 gevallen van *hemd* in B90 die geen tegenhanger in N90 hebben; 23 Belgische *overhemden* hebben geen Nederlandse tegenhanger, en er zijn geen Belgische tegenhangers voor de 37 Nederlandse *shirts*. Op het totaal van 200 benoemingsinstanties in beide profielen, hebben m.a.w.

slechts  $200 - (14 + 23 + 37) = 126$  gevallen een tegenhanger in het andere profiel, hetgeen een uniformiteit van  $126/2 = 63\%$  oplevert. Ter wille van het kwantitatieve gemak moet opgemerkt worden dat dat percentage altijd gelijk is aan de *som van de kleinste relatieve frequentie voor iedere alternatieve term* (die in tabel 1 telkens schuingedrukt is), i.e.  $17 + 46 + 0 = 63\%$ .

Als we de uniformiteit tussen twee taalvariëteiten op meer dan één begrip willen baseren, definiëren we uniformiteitsindex  $U$  als het gemiddelde van de uniformiteitsindexen per individueel begrip. In die berekeningswijze wordt aan een hoogfrequent concept zoals ROK in principe hetzelfde gewicht toegekend als aan een laagfrequent concept zoals, bijvoorbeeld, OVERSCHOEN. Vanuit een pragmatisch perspectief op standaardisering, echter, zou men kunnen stellen dat een gebrek aan uniformiteit minder vaak tot communicatieve misverstanden zal leiden bij laag- dan bij hoogfrequente concepten. Om bij uniformiteitsberekeningen rekening te houden met de frequentie van de onderzochte begrippen, definiëren we  $U'$  als een *gewogen* gemiddelde waarin de relatieve frequentie van elk begrip verrekend wordt.

We illustreren beide berekeningswijzen met een concreet voorbeeld (bijkomende illustraties zijn te vinden in Geeraerts e.a. 1999:41 e.v.). De onderstaande tabel 2 bevat in resp. de tweede en de vierde kolom de onomasiologische profielen (o.p.) van de begrippen JACK, TRUI en JURK in het Nederlandse en het Belgische 1970-corpus. De vijfde kolom bevat de minima per onomasiologisch alternatief, op basis waarvan in de zesde kolom de individuele uniformiteitswaarden 37.3, 55 en 76.8 berekend worden. De globale index  $U = 56.4$  onderaan in tabel 2 is het gemiddelde van de drie individuele uniformiteitswaarden.

	N70	<i>o.p.</i>	B70	<i>o.p.</i>	<i>min</i>	U	G	U'
<b>jack (m/v)</b>								
bloes	0	0,0	0	0,0	0,0	37,3	0,03	1,0
blouson(netje)	0	0,0	31	59,6	0,0			
jack/jek	36	97,3	18	34,6	34,6			
jacket	0	0,0	0	0,0	0,0			
jak(je)	0	0,0	0	0,0	0,0			
windbloes(je)	0	0,0	0	0,0	0,0			
windjack	1	2,7	3	5,8	2,7			
windjacket	0	0,0	0	0,0	0,0			
<i>totaal</i>	37		52					
<b>trui (m/v)</b>								
pull(etje)	3	3,6	105	48,6	3,6	55,0	0,09	4,9
pullover	24	28,6	29	13,4	13,4			
trui(tje)	57	67,9	82	38,0	38,0			
<i>totaal</i>	84		216					
<b>jurk (v)</b>								
japon(netje)	351	25,7	40	2,5	2,5	76,8	0,88	67,9
jurk(je)	1015	74,3	1493	94,7	74,3			
kleed(je)	0	0,0	43	2,7	0,0			
<i>totaal</i>	1366	1576						
						$U = 56,4$	$U' = 73,8$	

Tabel 2: Gewogen en ongewogen uniformiteit tussen Belgische en Nederlandse benamingen uit 1970 voor de begrippen JACK, TRUI en JURK

In deze berekeningswijze van  $U$  wordt aan elk van de drie concepten hetzelfde gewicht toegekend, hetgeen gezien de lage individuele uniformiteitswaarden voor JACK en TRUI tot de lage globale uniformiteit  $U=56.4$  leidt. Als we er (fictief) van uitgaan dat de gehele Belgische en Nederlandse materiaalverzameling voor 1970 in dit voorbeeld uit benamingen voor JACK, JURK en TRUI bestaat, dan vertegenwoordigen de twee begrippen die voor een lage uniformiteit zorgen nauwelijks 12% van de materiaalverzameling, terwijl het veel uniformer benoemde JURK ( $U=76.8$ ) maar liefst 88% van het materiaal representeert. Als we die asymmetrie in onze berekening tot uiting willen brengen – omdat we geloven dat gebrek aan uniformiteit vooral bij hoogfrequente begrippen tot communicatiestoornissen zal leiden –, dan vermenigvuldigen we de individuele uniformiteitswaarden met de wegingsfactor  $G$  uit de voorlaatste kolom, die de relatieve frequentie van de individuele begrippen in de materiaalverzameling reflecteert. Aangezien het begrip JACK in de Belgische en Nederlandse materiaalverzamelingen samen  $37+52=89$  keer voorkomt, en de Belgische en Nederlandse materiaalverzamelingen samen 3331 observaties bevatten, is  $G$  voor JACK gelijk aan  $89/3331=0.03$ . Vermenigvuldigd met de ongewogen uniformiteit voor JACK, levert dat een gewogen uniformiteit  $U'=0.9971$  op (in de tabel afgerond tot 1.0). Dezelfde berekeningswijze levert voor JURK een wegingsfactor  $G$  van 0.8832, en een gewogen uniformiteit van maar liefst 67.8688 op! De gewogen uniformiteit  $U'$  voor de hele materiaalverzameling krijgen we vervolgens door de gewogen waarden voor de drie begrippen bij elkaar op tellen:  $0.9971+4.9499+67.8688=73.82$ ; dat laatste getal reflecteert veel adequater dan de ongewogen uniformiteit het feit dat het meest frequente concept in ons fictieve voorbeeld het meest uniform benoemd wordt. In wat volgt doen we dan ook uitsluitend uitspraken op basis van  $U'$ .

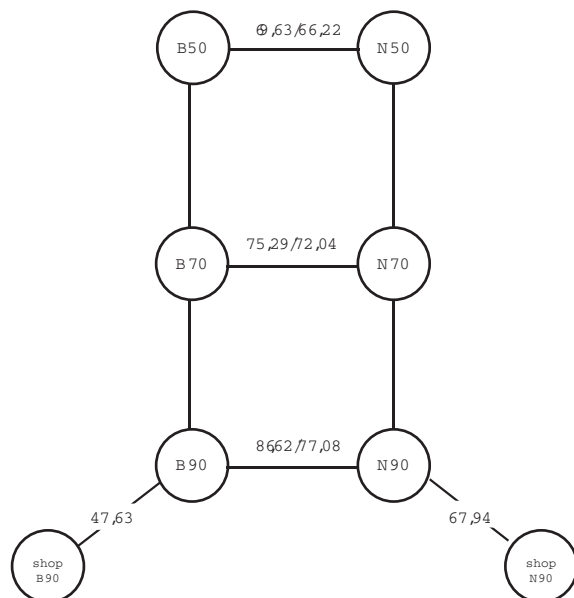
### 2.3. Resultaten

Laten we, alvorens de resultaten te bespreken, de diachrone convergentiehypothese en de synchrone stratificatiehypothese operationaliseren in termen van uniformiteit. Diachroom gezien kunnen we convergentie of divergentie tussen variëteiten kwantificeren als resp. stijgende of dalende uniformiteit tussen die variëteiten. De grotere synchrone afstand tussen Belgisch regionaal en supraregionaal taalgebruik die we veronderstellen zal zich – indien onze hypothese correct is – manifesteren in een lagere uniformiteit tussen het etalage- en het tijdschriftenmateriaal in België dan in Nederland.

Figuur 1 bevat de relevante synchrone en diachrone gegevens. De linkeras in de figuur vertegenwoordigt het Belgische (B) materiaal, terwijl de rechteras het Nederlandse (N) materiaal representeert (“50”, “70”, “90”, en “shop” staan voor resp. het 1950-, het 1970-, het 1990-, en het etalagemateriaal). De waarde op de lijnen tussen de cirkels vertegenwoordigen de gewogen uniformiteit  $U'$  tussen de variëteiten in de cirkels. Bij de vergelijking tussen de meetpunten B50/N50, B70/N70 en B90/N90 geven de cijfers voor de schuine streep de kledinggegevens, en de cijfers na de schuine streep de voetbalgegevens weer.<sup>2</sup>

<sup>2</sup> In verband met de materiaalverzameling waarop figuur 1 gebaseerd is moeten nog twee opmerkingen gemaakt worden. Om te beginnen moet opgemerkt worden dat de uniformiteitswaarden in figuur 1 berekend werden op basis van 12 kledingconcepten i.p.v. de 15 uit de oorspronkelijke studie (hetgeen de afwijkingen tussen de uniformiteitswaarden in figuur 1 en de cijfers in figuur 5.2 van Geeraerts e.a. 1999:69 verklaart). In tegenstelling tot Geeraerts e.a. (1999) is de replicatie die in sectie 3 aan bod komt namelijk gebaseerd op een tekstcorpus zonder afbeeldingen.





Figuur 1: Gewogen uniformiteit bij kleding- en voetbalnamen in Belgisch en Nederlands tijdschriften- en etalagemateriaal uit 1950, 1970 en 1990

De gegevens in figuur 1 vormen een duidelijke bevestiging van de diachrone convergentiehypothese zowel als de synchrone stratificatiehypothese. De stijgende uniformiteit tussen het Belgische en het Nederlandse Nederlands suggereert om te beginnen een uitgesproken lexicale convergentie tussen beide variëteiten:<sup>3</sup>

- |    |             |   |             |   |             |
|----|-------------|---|-------------|---|-------------|
| 1. | U'(B50,N50) | < | U'(B70,N70) | < | U'(B90,N90) |
|    | KL 69.63    | < | 75.29       | < | 86.62       |
|    | VO 66.22    | < | 72.04       | < | 77.08       |

Synchroon gezien manifesteert de vertraagde standaardisering van het Belgische Nederlands zich inderdaad in een opvallend lagere uniformiteit tussen het Belgische tijdschriften- en etalagemateriaal dan tussen het Nederlandse tijdschriften- en etalagemateriaal:

- |    |              |   |              |
|----|--------------|---|--------------|
| 2. | U'(B90,shop) | < | U'(N90,shop) |
|    | 47.63        | < | 67.94        |

dingen van de kledingstukken waarnaar verwezen wordt, en zonder afbeeldingen is de precieze referentie van de begrippen VEST, COLBERT (M) en BLAZER (V) onvoldoende vaststelbaar. Om de gegevens uit Geeraerts e.a. (1999) veilig met de replicatiegegevens te kunnen vergelijken werden die drie begrippen in dit artikel dan ook uit beide analyses verwijderd. Het ontbreken van synchrone gegevens voor de voetbalbegrippen is daarnaast het logische gevolg van het feit dat voetbalbegrippen zoals BUTENSPEL en HOEKSCHOP (uiteraard) niet in winkelatalages voorkomen.

<sup>3</sup> In Geeraerts e.a. (1999:67) wordt van een interpreteerbaar verschil tussen uniformiteitswaarden gesproken wanneer die waarden tenminste 5% verschillen. Verschillen van minder dan 5% vallen binnen een statistische foutmarge.

### ● 3 Naar een lexicale replicatie van de gegevens uit Geeraerts e.a. (1999)

#### 3.1 Materiaal

Hoewel de data in figuur 1 een solide bevestiging van de synchrone stratificatiehypothese en de diachrone convergentiehypothese vormen, is onze zaak voorlopig nog niet waterdicht. Nog afgezien van Van de Velde's (1996) observatie dat er op het vlak van de uitspraak een opvallende diachrone *divergentie* is tussen het Belgische en Nederlandse Nederlands, zijn er in ons eigen onderzoek tekortkomingen die betrekking hebben op de afhankelijke en de onafhankelijke onderzoeksvariabele. Om te beginnen is het empirische fundament van onze bevindingen te smal: twee lexicale velden vormen een te eenzijdige vergelijkingsbasis om veilige uitspraken te kunnen doen over de ontwikkeling en de interne structuur van het Belgische Nederlands. Daarnaast doet de huidige implementering van de stratificatieparameter geen recht aan de synchrone complexiteit van het Belgische Nederlands. Tot nu toe hebben we slechts één tussentaalstratum in het onderzoek betrokken – geïnstantieerd door het etalagemateriaal –, terwijl in de literatuur twee (Geerts 1975), drie (Baetens Beardsmore & Van de Craen 1979), en zelfs vijf (Willemys 1987) stilistisch verschillende variëteiten van het Belgische Nederlands geïdentificeerd werden.

In de rest van dit artikel richten we onze aandacht dan ook op twee bijkomende onderzoeksvragen: (i) kunnen we de bevindingen uit Geeraerts e.a. (1999) met behulp van een ander corpus, en – meer in het bijzonder – een ander type informeel Nederlands bevestigen, en (ii) kunnen we de diachrone convergentiehypothese en synchrone stratificatiehypothese repliceren als we de vergelijkingsbasis uitbreiden van inhoudswoorden zoals kleding- en voetbaltermen naar relationele woorden zoals preposities?

Om die vragen te beantwoorden werd een omvangrijk tekstcorpus van ongeveer 45.000.000 woorden samengesteld. Net zoals het corpus in Geeraerts e.a. (1999), is het CONDIV-corpus gestructureerd op basis van drie types variatie.<sup>4</sup> Het geografische verschil tussen Belgisch en Nederlands Nederlands accommoderen we door in het corpus een Nederlandse tekstenbank, en een Belgische tekstenbank te includeren.

Bij de aanleg van de materiaalverzameling werd eveneens rekening gehouden met de stilistische stratificatie van het Nederlands. Omdat het onmogelijk is a priori vast te stellen hoeveel tussenstrata precies onderscheiden moeten worden (cf. supra), gaan we niet in de eerste plaats op zoek naar taalgebruik dat representatief is voor een bepaald stratum, maar vertrekken we van verschillende types *taalsituaties* als onafhankelijke variabele, en beschouwen we het daarin gebruikte *taalgebruik* als afhankelijke variabele. Stratificatievariëteit accommoderen we dan door taalmateriaal te vergelijken uit vijf stilistisch verschillende taalsituaties. Om te beginnen bevat het CONDIV-corpus drie soorten krantenmateriaal: we onderscheiden *kwaliteitskranten* zoals *De Standaard* en het *NRC Handelsblad* en *massakranten* zoals *Het Laatste Nieuws* en *De Telegraaf*. Er zijn *nationale* massakranten – zoals *Het Laatste Nieuws* en *De Telegraaf* –, maar ook *regionale* massakranten zoals *Het Belang van Limburg*, *De Gazet van Antwerpen* of *De Limburger*.

<sup>4</sup> Het CONDIV-corpus is genoemd naar het VNC-project “Convergentie en divergentie in de Nederlandse woordenschat” waarvoor het werd samengesteld. Een uitgebreide toelichting van het corpus is te vinden in Grondelaers, Deygers, Van Aken, Van Den Heede & Speelman (2000).

Daarnaast werden aan het CONDIV-corpus twee types Internettaal toegevoegd als vertegenwoordigers van het informele Nederlands. IRC – Internet Relay Chat – is een Internetmodule waarin *online* en *synchroon* gecommuniceerd kan worden in babbelkanalen zoals #Leuven en #Vlaanderen. Omdat IRC-ers zoveel mogelijk een conversationeel spreek/schrijf-tempo willen benaderen, en aan dat tempo hun spelling-, morfologie-, en grammaticanormen aanpassen, kan IRC beschouwd worden als “geschreven spreektaal”; IRC-ees manifesteert dan ook colloquialismen en dialectismen die in gewone geschreven teksten ontbreken (voor een overzicht van de linguïstische en sociologische kenmerken van IRC, zie onder meer Herring 1996, Ko 1996, Hentschel 1998, Bays 1998, en Paolillo 1999). Voorbeeld (3) bevat een kort extract uit het babbelkanaal #Leuven:

3. <boink> thebuzz, komde gij ier gewoon reclaam maken of wa?  
<SarAHken> loooooooooooooooooooooooooooooootje !  
<King\_John> jaja  
<atttila> daar kunde ook al een goeie pc voor kopen zenne  
\* DeepStar (((shakes))) lotje  
<Gonzy> lotjeeeeeeeeeeeeeuuuuuuuuuuuuuuuuuh  
<thebuzz> neen, gewoon eens vragen, maar ben nu al weer weg  
<SarAHken> lotje : how are you !? long time no see  
<boink> thebuzz, ge moe daarvoor nie frutten ;)  
<King\_John> zijde weg buzz?  
<Grooveke> atttila : voor 92000 ballen ebde nog gene goeie pc  
<King\_John> allé jong  
<thebuzz> ja, kzijn weg, nog een beke gaan werken

Aangezien Nederlandse en Vlaamse IRC-ers eigen kanalen hebben waarin users uit het andere land niet welkom zijn, en gewoonlijk zelfs automatisch uitgesloten worden (zie onder meer Grondelaers 2000:65 e.v. en Grondelaers, Deygers, Van Den Heede, Van Aken & Spielman 2000:358 e.v.), beschikken we over aparte corpora met zeer informeel Belgisch en zeer informeel Nederlands Nederlands.

Naast IRC werd in het CONDIV-corpus ook Usenetmateriaal geïntegreerd. Usenet is een internetmodule waarin *offline* en *asynchroon* in nieuwsgroepen wordt gediscussieerd d.m.v. e-mailboodschappen.<sup>5</sup> Aangezien e-mail een veel indirecter communicatiemedium is dan IRC, en academische internetoperatoren uitsluitend “nette” nieuwsgroepen op hun server aanbieden, ligt het voor de hand dat het Nederlands op Usenet formeler zal zijn dan dat op IRC. De linguïstische context waarin IRC ontstaat is in elk geval informeler dan die waarin “Usenetese” tot stand komt: terwijl IRC “immediate” en “online” is (Bays 1998), kunnen Usenetters hun boodschappen dankzij het offline karakter van nieuwsgroepen nakijken alvorens ze te posten (cf. onder meer Harrison 1998 voor een gedetailleerder overzicht van de linguïstische kenmerken van Usenet).

<sup>5</sup> IRC is *online* en *synchroon* omdat users onmiddellijk kunnen reageren op een boodschap die in een IRC-kanaal op hun scherm verschijnt. Usenet daarentegen is *offline* en *asynchroon* omdat er “propagation delay” (Paolillo 1999) is tussen het opstellen, nalezen, verzenden, ontvangen, lezen en beantwoorden van een e-mailboodschap.

De vijf materiaaltypes die in het corpus opgenomen zijn kunnen op een informaliteitsas uitgezet worden die als meest informele pool het IRC-materiaal heeft, en als meest formele pool het materiaal uit de kwaliteitskranten:

4. IRC > Usenet > regionale massakranten > nationale massakranten > kwaliteitskranten<sup>6</sup>

In de derde plaats bevat ook het CONDIV-corpus een belangrijke diachrone dimensie, waarop materiaal uit 1958, 1978 en 1998 gecontrasteerd wordt. Aangezien IRC en Usenet recente mediafenomenen zijn, kunnen ze niet ingevuld worden voor de oudere periodes, en daardoor bevat het diachrone corpus alleen oudere tegenhangers van de kranten. De afwezigheid van elektronisch beschikbare historische krantenedities noopte ons ertoe krantenmateriaal uit 1958 en 1978 in te scannen. Dat het diachrone corpus qua omvang beperkter is dan het synchrone, is een rechtstreeks gevolg van de hoge kostprijs (en de grote foutenmarge) die met het inscannen van sterk verzuurde, poreuze krantenpagina's gepaard gaat.<sup>7</sup>

Aangezien het CONDIV-corpus niet, zoals het corpus uit de eerste studie, een verzameling manueel geselecteerde kleding- en voetbalnamen is, maar een digitaal tekstcorpus, gebeurt de extractie van de taalverschijnselen die we willen bestuderen op basis van elektronische bevraging. Daartoe ontwikkelde de derde auteur van dit artikel een browser op linguïstenmaat: *Abundantia Verborum* (Speelman 1997) bevat een module om digitale bestanden tot bruikbare linguïstische corpora te transformeren, een zoekmachine die de targetdata in die corpora identificeert, een module om de resulterende gegevens te labelen, classificeren en manipuleren, en een component waarin statistische analyses op de gegevens uitgevoerd kunnen worden.<sup>8</sup>

### 3.2 Methode en resultaten

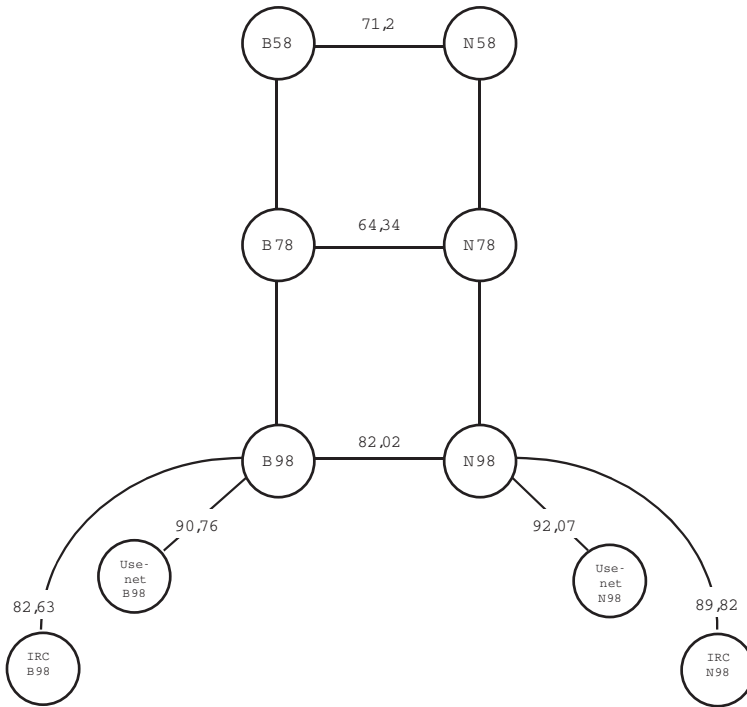
In deze replicatiestudie werd dezelfde kwantitatieve methodologie gehanteerd als in de oorspronkelijke studie. De afhankelijke variabele was uniformiteit in de benaming van 12 kledingconcepten.<sup>9</sup> De onafhankelijke variabelen waren *regio* – België vs. Nederland – en *stratificatie* – IRC vs. Usenet vs. kranten.

<sup>6</sup> Met deze ordening correleren factoren zoals “geïntendeerd bereik” en “mate van redactionele controle” (zie Grondelaers e.a. 2000:358).

<sup>7</sup> Het synchrone corpus bevat 22.984.343 woorden Belgisch Nederlands (IRC = 8.207.007; Usenet = 4.980.780; Het Belang van Limburg = 3.012.330; De Gazet van Antwerpen = 3.068.405; Het Laatste Nieuws = 3.486.911; De Standaard = 3.228.910), en 19.505.008 woorden Nederlands Nederlands (IRC = 6.965.291; Usenet = 7.748.436; De Limburger = 1.680.636; De Telegraaf = 1.590.581; NRC Handelsblad = 1.520.064). Het diachrone corpus telt 3.587.035 woorden Belgisch Nederlands (De Gazet van Antwerpen 1958 = 734.018; De Gazet van Antwerpen 1978 = 699.255; De Standaard 1958 = 1.018.658; De Standaard 1978 = 1.135.104) en 2.079.728 woorden Nederlands Nederlands (Nieuwe Rotterdamsche Courant 1958 = 1.064.001; Nieuwe Rotterdamsche Courant 1978 = 1.015.727).

<sup>8</sup> Meer informatie over *Abundantia Verborum* is te vinden op <http://www.wling.arts.kuleuven.ac.be/genling/abundant>

<sup>9</sup> We beperken de replicatie tot de kledingtermen, waarvoor in Geeraerts e.a. (1999) en in paragraaf 2 etalagemateriaal voorhanden is.



Figuur 2: Gewogen uniformiteit bij kledingnamen uit Belgisch en Nederlands kranten- en Internetmateriaal uit 1958, 1978 en 1998

We brengen nog even in herinnering dat we in dit onderzoek de diachrone convergentie tussen Belgisch en Nederlands Nederlands trachten te repliceren, en de grotere synchrone afstand tussen supraregionaal/formeel en regionaal/informeel Belgisch Nederlands dan tussen supraregionaal/formeel en regionaal/informeel Nederlands Nederlands. Merk op dat de synchrone stratificatiehypothese nog verfijnd kan worden: als de relatief grotere afstand tussen informeel en formeel Belgisch Nederlands in het nieuwe onderzoek inderdaad bevestigd wordt, dan zal het feit dat Usenet minder informeel lijkt dan IRC zich in het Belgische materiaal vertalen in een hogere uniformiteit tussen het Usenet- en het krantenmateriaal, dan tussen het IRC- en het krantenmateriaal.

De relevante gegevens staan in figuur 2, die op dezelfde manier georganiseerd is als figuur 1, met het verschil dat het synchrone deel aan Belgische zowel als Nederlandse kant uit drie vergelijkingspunten is samengesteld. De synchrone gegevens bevestigen trouwens zowel de oorspronkelijke als de verfijnde versie van de synchrone stratificatiehypothese. De grotere Belgische afstand tussen regionale en supraregionale taal manifesteert zich het duidelijkst in IRC:

$$\begin{array}{lcl}
 5. & U'(B98,IRCB98) & < & U'(N98,IRCN98) \\
 & 82.63 & < & 89.82
 \end{array}$$

En de uniformiteit tussen IRC en krantentaal is inderdaad kleiner in het Nederlandse en (vooral) het Belgische Nederlands dan de afstand tussen het Usenet-materiaal en het krantenmateriaal:

$$\begin{array}{lcl}
 6. & U'(B98,IRCB98) & < & U'(B98,UsenetB98) \\
 & 82.63 & < & 90.76 \\
 & & & \\
 & U'(N98,IRCN98) & < & U'(N98,UsenetN98) \\
 & 89.82 & \cong & 92.07
 \end{array}$$

De diachrone convergentiehypothese daarentegen wordt niet rechtstreeks bevestigd door de nieuwe gegevens:

$$\begin{array}{lclcl}
 7. & U'(B58,N58) & < & U'(B78,N78) & < & U'(B98,N98) \\
 & 71.2 & > & 64.34 & < & 82.02
 \end{array}$$

In plaats van een progressieve stijging – zoals in de initiële gegevens – tonen de data in figuur 2 een “geknikt” patroon, met een daling in het materiaal uit 1978. Dat die daling geen probleem vormt voor de diachrone convergentiehypothese tonen de data in de derde kolom van tabel 3. Daarin worden de beide bronnen die B58 en B78 constitueren – *De Gazet van Antwerpen* en *De Standaard* – in de tweede en de derde balk apart met de enige Nederlandse bron – *NRC Handelsblad* – vergeleken; in de eerste balk zijn ter vergelijking nog eens de globale cijfers opgenomen. De gegevens in tabel 3 laten zien dat de initieel vastgestelde convergerende beweging wél opduikt in de uitgesplitste *Standaard*- en *Gazet*-gegevens.

	Abs Freq staal A	Abs Freq staal B	U' globaal	U' n=5
U'(B58,N58)	101	54	71,2	71,67
U'(B78,N78)	109	45	64,34	60,9
U'(B98,N98)	890	391	82,02	82,15
U'(GVA58,NRC58)	36	54	45,27	37,36
U'(GVA78,NRC78)	64	45	46,82	43,36
U'(GVA98,N98)	241	391	81,49	81,49
U'(STA58,NRC58)	65	54	71,38	72,2
U'(STA78,NRC78)	45	45	82,18	78,42
U'(STA98,N98)	103	391	81,37	80,02

Tabel 3: Absolute frequenties en gewogen uniformiteit bij kledingnamen in Belgische en Nederlandse kranten uit 1958, 1978, en 1998.

Het geknikte patroon in de globale gegevens blijkt in belangrijke mate het gevolg van frequentieasymmetriën tussen *De Gazet van Antwerpen* en *De Standaard*: de globale uniformiteitswaarden in 1958 en 1978 correleren namelijk met de individuele uniformiteitswaarde van de Belgische bron die qua frequentie het zwaarst doorweegt op die pun-

ten.<sup>10</sup> Dat de globale Belgische uniformiteit in 1958 ( $U'=71.2$ ) naar de 1958-uniformiteit tussen *Standaard* en *NRC* ( $U'=71.38$ ) neigt is een gevolg van het feit dat de Belgische materiaalverzameling uit 1958 65 observaties uit *De Standaard* en slechts 36 observaties uit *De Gazet* bevat. Het gecombineerde effect van *Gazet*-overwicht in 1978 (64 observaties t.o.v. 45 observaties uit *De Standaard*) en een lage individuele uniformiteit tussen *Gazet* en *NRC* ( $U'=46.82$ ) brengt de globale Belgische uniformiteit in 1978 flink omlaag ( $U'=64.34$ ). In 1998 zijn de afzonderlijke krantenstalen qua omvang evenmin in evenwicht, maar dat is wel het geval voor de uniformiteitscijfers, die resp.  $U'=81.49$  en  $U'=81.37\%$  bedragen, en tot een globale uniformiteit van  $U'=82.02$  leiden.

In verband met de cijfers in tabel 3 moeten nog drie belangrijke opmerkingen gemaakt worden. Om te beginnen kan de materiaalschaarste op de diachrone meetpunten enigszins gecompenseerd worden door in een alternatieve telling alleen begrippen te betrekken die in beide vergeleken stalen tenminste vijf keer voorkomen ( $n=5$ ). De resultaten van deze uitgezuiverde methode – die in de vierde kolom van tabel 3 staan – bevestigen de niet-uitgezuiverde gegevens op elk vlak, zodat we voorlopig (met enige mate van voorzichtigheid en in afwachting van verdere replicatie) kunnen stellen dat we met interpreteerbare tendensen te maken te hebben.

Daarnaast moeten we nog even stilstaan bij de verschillen in diachrone ontwikkeling tussen *De Gazet van Antwerpen* en *De Standaard*. Hoewel beide bronnen globaal gezien een stijgende uniformiteit met het *NRC*-materiaal en een nagenoeg identieke 1998-uniformiteit vertonen ( $U'$  resp. 81.49 en 81.37), verschilt hun specifieke ontwikkeling op twee belangrijke punten. *De Standaard* manifesteert van meet af aan een veel hogere uniformiteit met het *NRC*-materiaal dan *De Gazet* ( $U'=71.38 > 45.27$ ), en kent zijn sterkste stijging op het tweede meetpunt. *De Gazet* is in 1958 zowel als 1978 opvallend laag uniform met *NRC*, en begint pas in 1998 uitgesproken lexicale overlapping met *NRC* te vertonen. Binnen het korte bestek van dit artikel, waarin we ons noodgedwongen tot het externe onderzoeksperspectief uit Geeraerts, Grondelaers & Speelman (1999) beperken, kunnen we niet nagaan welke variabelen verantwoordelijk zijn voor de vertraagde uniformisering in *De Gazet*.<sup>11</sup> Wat de hogere beginuniformiteit van *De Standaard* t.o.v. *De Gazet* wél suggereert, is dat de individuele lexicale evolutie van de Belgische bronnen bepaald lijkt te worden door de stratificatiele status van die bronnen: de hogere aanvangsuniformiteit van *De Standaard* t.o.v. *De Gazet* zou een gevolg kunnen zijn van de hogere positie die de *De Standaard* als kwaliteitskrant op het informaliteitscontinuüm in (4) bekleedt. Wegens het ontbreken van historische gegevens over de lexicale evolutie in de informeelste bronnen kan deze hypothese uitsluitend op basis van replicatie geverifieerd worden. We voegen deze hypothese dan ook toe aan het lijstje van verwachtingen die in de volgende case studie getest moeten worden.

Tenslotte moet nog opgemerkt worden dat de synchrone uniformiteit tussen IRC en krantenmateriaal aan Belgische zowel als Nederlandse zijde onintuïtief hoog ligt (resp.

**10** Tabel 3 bevat in de eerste en tweede kolom de absolute frequenties van de stalen die in de uniformiteitsberekening vergeleken worden. Staal A is daarbij de bron voor de komma en staal B de bron na de komma.

**11** In Geeraerts e.a. (1999:10) wordt een onderscheid gemaakt tussen een *extern* onderzoeksperspectief – waarbij de geanalyseerde taalvariëteiten als één geheel in het onderzoek betrokken worden – en een *interne* benadering, waarbij binnen de onderzochte variëteiten naar specifieke linguïstische variabelen gekeken wordt die met de externe convergentie of divergentie correleren.

82.63 en 89.82).<sup>12</sup> Die opvallende mate van lexicale overlapping tussen IRC en kranten-taal betekent niet alleen dat IRC-ers in weerwil van hun voorliefde voor idioom en jargon relatief conformistisch zijn in hun benoeming van alledaagse begrippen zoals kledingstukken. De hoge uniformiteit tussen IRC en de kranten suggereert tevens dat de globale informaliteit van IRC niet in de eerste plaats door lexicale factoren bepaald wordt.<sup>13</sup> Een correcte inschatting van de linguïstische verhouding tussen formeel en informeel Nederlands is m.a.w. bij uitstek gebaat bij een uitbreiding van de vergelijkingsbasis naar niet-lexicale parameters. In de volgende paragrafen richten we onze aandacht daarom op de rol van voorzetsels als standaardiseringsindicatoren.

#### ● 4 Preposities als standaardiseringsindicatoren

##### 4.1 Selectie van de data

De rest van dit artikel wijden we aan prepositionele variatie in voorzetselverbindingen zoals *zich interesseren aan/voor* of *langs/via het internet*. Hoewel preposities geen typische functiewoorden zijn, verschillen ze van inhoudswoorden doordat ze geen “zelfstandigheid” (ANS 1997:139) aanduiden (zoals de kleding- en voetbaltermen), maar de aard van de relatie tussen twee zelfstandigheden specificeren (idem:507).

Om de digitale extractie van bruikbare prepositionele verbindingen te vergemakkelijken, hebben we de selectie beperkt tot verbindingen die syntagmatisch sterk afgebakend zijn, en daardoor relatief makkelijk opspoorbaar zijn in een elektronisch corpus. Dat zijn in de eerste plaats “vaste verbindingen” (ANS 1997:1169), waarbij een voorzetsel een vaste combinatie vormt met een specifiek werkwoord, adjectief, of substantief. De verbindingen die ons in het licht van dit onderzoek interesseren zijn natuurlijk degene die met meer dan één voorzetsel kunnen voorkomen, zoals *zich interesseren voor/aan* (ibidem). Een tweede groep van voor onze doeleinden interessante voorzetselcombinaties zijn verbindingen waarin de prepositie een beperkte en daardoor makkelijk opspoorbare set van complementen selecteert, zoals “Hij betrad het huis *langs/via het raam/het venster/de deur/de poort*, etc.”

Met het gebruik van preposities als standaardiseringsindicatoren zijn twee belangrijke methodologische problemen verbonden (zie ook Van Aken & Van Den Heede 2001). Om te beginnen ligt de “veldaanpak” die we in 2.1 ter voorkoming van statistische beeldvorming propageerden bij preposities moeilijker. Nog afgezien van het feit dat het minder zinvol is in verband met de onderlinge relaties tussen preposities van een “lexicaal veld”

**12** Het is verleidelijk om deze hoge mate van lexicale overlapping te interpreteren aan de hand van de opvallend lage uniformiteit tussen het etalagemateriaal en het tijdschriftenmateriaal in figuur 1 (B 47.63 en N 67.94). Echter, IRC en etalagemateriaal verschillen linguïstisch gezien zo sterk van elkaar dat ze onmogelijk op een veilige manier met elkaar vergeleken kunnen worden: terwijl etalageverwijzingen naar kledingstukken exclusief lexicaal zijn, manifesteert IRC namelijk een globale informaliteit die door lexicale zowel als niet-lexicale parameters bepaald wordt.

**13** De beperkte impact van lexicale factoren op de globale informaliteit van een bron werd onder meer geobserveerd in Goossens (2000:9 e.v.), waarin beweerd wordt dat de linguïstische eigenheid van het Belgische Nederlands in belangrijker mate door morfo-grammaticale parameters (in het bijzonder de vorm en flexie van enkele determiners) dan door lexicale factoren bepaald wordt.



te spreken,<sup>14</sup> zijn we voor het opstellen van een exhaustieve lijst van variabele preposities in hoofdzaak aangewezen op overzichten uit de taalbeheersingsliteratuur (bijvoorbeeld Penninckx & Buyse 1997 en Paardekoooper 1990), die niet zelden opgesteld zijn om de gebruiker te waarschuwen voor te mijden regionale variatie zoals *zoeken achter* (de “foute” variant die beperkt is tot het Belgische Nederlands) vs. *zoeken naar* (de “juiste” variant). Omdat we in dit stadium van het onderzoek niet kunnen nagaan hoe representatief dergelijke gevallen zijn voor het geheel van het taalgebruik (maar cf. infra), lopen we met het gedwongen gebruik van regionale-variantiegevoelige indicatoren het risico een onrealistisch divergent beeld van de relatie tussen het Belgische en het Nederlandse Nederlands te schetsen. De prepositionele gegevens die verderop besproken worden moeten m.a.w. met de grootst mogelijke omzichtigheid benaderd worden.

Minstens even problematisch is de vaststelling dat het ondanks de rigide syntagmatische afbakening van de gekozen verbindingen moeilijk blijft om volledige synonymie tussen de prepositionele varianten in een onomasiologisch profiel te garanderen. Is denotationele synonymie bij inhoudswoorden relatief gemakkelijk vast te stellen,<sup>15</sup> bij preposities kan betekenisgelijkheid in veel mindere mate gegarandeerd worden. Cuyckens (1991:182 e.v.) geeft een aantal voorbeelden van bijna-synoniemen bij preposities en laat zien dat de keuze voor het ene dan wel het andere voorzetsel heel vaak terug te voeren is op een *verschillende conceptualisering* van de uitgedrukte situatie. De keuze tussen de voorzetsels *in* en *op* in de zin (8), bijvoorbeeld, hangt af van de manier waarop de landmark *veld* wordt waargenomen.<sup>16</sup> Als de landmark gepercipieerd wordt als een onbegrensde tweedimensionele oppervlakte, dan is *op* het aangewezen voorzetsel. Wordt de landmark daarentegen geconcipieerd als een *begrensde* tweedimensionele oppervlakte, dan is *in* geschikter (1991:183):

8. Het elftal staat *in/op* het veld.

Dergelijke conceptuele verschillen zijn veel moeilijker op het spoor te komen wanneer de relatie tussen de verschillende entiteiten figuurlijk is in plaats van ruimtelijk:

9. Hij is leraar *aan/op* de Militaire School.  
10. *Aan/op* onze universiteit gebeurt veel wetenschappelijk onderzoek.

In (9) en (10) (idem:198) is de keuze voor *aan* of *op* wellicht niet semantisch bepaald, en dus lijken beide voorzetsels verwisselbaar (en bruikbaar voor onze onderzoeksdoeleinden).

**14** Volgens Van Aken & Van Den Heede 2001 is het in principe mogelijk om spatiële voorzetsels die doel, herkomst, pad e.d. uitdrukken als een lexicaal veld te zien. Dergelijke aanpak is bij voorzetsels die louter op basis van variabiliteit en opzoekbaarheid geselecteerd werden natuurlijk veel minder evident.

**15** Omdat Geeraerts, Grondelaers & Bakema (1994) gebaseerd is op een semasiologisch verrijkt corpus (waarin de referenten van de bestudeerde kledingnamen afzonderlijk toegankelijk zijn), konden we het onderzoek in Geeraerts, Grondelaers & Speelman (1999) beperken tot begrippen waarvoor namen bestaan die tenminste denotatief synoniem zijn.

**16** Het begrippenpaar “trajector-landmark” is uit de Cognitieve Linguïstiek afkomstig (zie onder meer Langacker 1987:217). In de prepositionele relatie die uitgedrukt wordt door een zin als *Het glas staat op tafel*, is *het glas* de “trajector”, i.e. het gelokaliseerde object, terwijl *de tafel* de “landmark” is die fungeert als referentiepunt voor de lokalisering van de trajector.

den). Voorzichtigheid blijft hoe dan ook geboden, want specifieke semantische of syntactische contexten kunnen die verwisselbaarheid blokkeren. Dat illustreert Taylor (1988:312-313) met de Engelse preposities *over* en *above*:

11. He lives *over* me.
12. He lives *above* me.

Hoewel informanten nauwelijks verschil merken tussen (11) en (12), bewijst Taylor dat er wel degelijk een semantisch onderscheid is, dat zichtbaar wordt wanneer preciezere informatie over de afstand tussen trajector en landmark aan de zin wordt toegevoegd:

13. He lives ten floors *above* me.
14. \*He lives ten floors *over* me.

Volgens Taylor is *over* onmogelijk in (14) omdat de afstand tussen trajector en landmark te groot is voor invloed van de trajector op de landmark. Dat laatste blijkt nochtans een voorwaarde voor het gebruik van *over*: waar het voorzetsel *above* uitsluitend signaleert dat de trajector zich op een hoger niveau bevindt dan de landmark, bezit *over* de extra betekenisnuance dat de trajector de landmark op de een of andere wijze moet kunnen beïnvloeden.

In de analyse werden alleen verbindingen betrokken waarvan de prepositionele varianten daadwerkelijk onderling verwisselbaar bleken. Na uitzuivering van ongewenste semasiologische variatie werden 16 bruikbare verbindingen geselecteerd die met 2 of 3 prepositionele varianten per profiel in totaal 7937 keer in het totale materiaal voorkwamen.<sup>17</sup>

#### 4.2 Resultaten

Op basis van die gegevens werden drie reeksen diachrone uniformiteitscijfers berekend, die in tabel 4 gepresenteerd worden.

	B-N	gva-nrc	sta-nrc
U' 1958	58.66	49.46	66.02
U' 1978	61.97	61.02	63.36
U' 1998	67.14	73.62	73.26

Tabel 4: Gewogen uniformiteit bij voorzetsels uit Belgisch en Nederlands krantenmateriaal uit 1958, 1978 en 1998

In de linkerkolom bevinden zich de globale gegevens die gebaseerd zijn op alle beschikbare bronnen per meetpunt. Aangezien de globale ontwikkeling in de vorige case studie sterk contrasteerde met de individuele ontwikkeling in de constituerende bronnen, worden de gegevens in de tweede en derde kolom uitgesplitst voor, respectievelijk, *De Gazet van Antwerpen* en *De Standaard*. Alvorens de resultaten te bespreken brengen we nog even

<sup>17</sup> De 16 prepositionele profielen worden in Appendix 3 met een voorbeeld toegelicht.

in herinnering dat we op basis van de gegevens in tabel 4 twee diachrone convergentie-hypothesen over de status van het Belgische Nederlands trachten te bevestigen: we zoeken niet alleen bijkomende confirmering voor de verwachting dat het Belgische en Nederlandse Nederlands progressief convergeren, maar we trachten ook de correlatie tussen de aard van de convergerende beweging in een bepaalde bron en de stratificatiele status van die bron te repliceren.

Aangezien het prepositiemateriaal op de verschillende meetpunten geen vertekende absolute-frequentieverschillen laat zien<sup>18</sup> – zoals bij de inhoudswoorden – zijn de globale prepositiedata in de linkerkolom interpreteerbaar. Getuige die gegevens bevestigen de prepositionele data i.t.t. de lexicale replicatiedata ook globaal de convergerende ontwikkeling tussen het Belgische en het Nederlandse Nederlands die in Geeraerts e.a. (1999) gevonden werd. Afgezien van de lichte terugval in *De Standaard* 1978 (die weliswaar binnen de vijf percent-marge valt), bevestigen de individuele ontwikkelingen per bron eveneens de convergerende tendens.

De interessantste observatie die in verband met tabel 4 gemaakt kan worden is de verschillende aard van de convergerende ontwikkeling in *De Gazet* en *De Standaard*, die in belangrijke mate het patroon bij de inhoudswoorden bevestigt. Net zoals bij de kledingtermen heeft *De Gazet* in 1958 een markante uniformiseringsachterstand op *De Standaard*, maar eindigen beide bronnen in 1998 op een nagenoeg identieke uniformiteitshoogte van  $U' =$  resp. 73.62 en 73.26! Hoewel de interpretatie van absolute cijfers in deze case studie wegens de huidige grofkorreligheid van het meetinstrument gevaarlijk is, en hoewel de gegevens van slechts twee kranten in de berekeningen betrokken werden, suggereren de dubbel gerepliceerde individuele ontwikkelingskenmerken van *De Gazet* en *De Standaard* dat diachrone ontwikkeling en stratificatiele-stilistische status wel degelijk correleren in de Belgische bronnen: de bulk van de convergerende ontwikkeling tussen het Belgische en het Nederlandse Nederlands komt niet op rekening van de Belgische kwaliteitskranten (die van meet af aan al vrij uniform zijn met de Nederlandse kranten en slechts beperkt verder convergeren), maar van de Belgische massakranten, die aanvankelijk opvallend afwijken van de Nederlandse kranten, maar in 40 jaar tijd hun uniformiteitsachterstand op de kwaliteitskranten inhalen.

Tabel 5 bevat de synchrone uniformiteitsgegevens (“Use”, “IRC”, “Inf” en “Kr” staan voor het Usenet-materiaal, het IRC-materiaal, het totale informele materiaal (Usenet + IRC) en het krantenmateriaal):

	B	N
U'(Use,Kr)	81.57	96.58
U'(IRC,Kr)	75.21	93.85
U'(Inf,Kr)	80.02	96.74

Tabel 5: Gewogen uniformiteit bij voorzetsels uit Belgisch en Nederlands kranten- en Internetmateriaal uit 1998

**18** Behalve dan op het meetpunt '78, waar het *Standaard*-materiaal 263 observaties bevat, terwijl het *Gazet*-materiaal slechts 164 observaties telt. Omdat de uniformiteit tussen *Gazet* en *NRC en Standaard* en *NRC* in 1978 resp. 61.02 en 63.36 bedraagt, levert dat frequentieverschil geen afwijkende globale uniformiteit tussen het beschikbare Belgische en Nederlandse materiaal op.

Met deze gegevens trachten we de twee synchrone bevindingen uit de vorige casus studie te bevestigen. Dat de afstand tussen formeel en informeel Belgisch Nederlands groter is dan de afstand tussen formeel en informeel Nederlands Nederlands blijkt om te beginnen eveneens uit het prepositiemateriaal, en zelfs in veel hogere mate dan bij de inhoudswoorden:

$$15. \quad \begin{array}{lcl} U'(\text{Inf B}, \text{Kr B}) & < & U'(\text{Inf N}, \text{Kr N}) \\ 80.02 & < & 96.74^{19} \end{array}$$

En ook de verfijnde synchrone stratificatiehypothese over de hogere uniformiteit in het Belgische materiaal tussen Usenet en kranten dan tussen IRC en kranten wordt door de prepositiegegevens bevestigd; in het Nederlandse materiaal overlappen Usenet en IRC in vergelijkbare mate met het krantenmateriaal:

$$16. \quad \begin{array}{lcl} U'(\text{Use B}, \text{Kr B}) & > & U'(\text{IRC B}, \text{Kr B}) \\ 81.57 & > & 75.21 \\ \\ U'(\text{Use N}, \text{Kr N}) & > & U'(\text{IRC N}, \text{Kr N}) \\ 96.58 & \equiv & 93.85 \end{array}$$

Als uit het globale overzicht van lexicale en prepositionele resultaten één observatie omtrent houden moet worden, dan is het wel de vaststelling dat er tussen de geobserveerde ontwikkelingspatronen bij inhoudswoorden en preposities uitgesproken parallellieën bestaan. In afwachting van veiligere meetinstrumenten (cf. infra) rechtvaardigen die parallellieën voorlopig de grofkorreligheid van de prepositionele tellingen.

## ● 5 Conclusies en consequenties

We vatten samen. Aan de hand van 12 kledingbegrippen, 15 voetbalbegrippen en 16 voorzetselverbindingen hebben we in dit artikel met behulp van de kwantitatieve methodologie uit Geeraerts, Grondelaers & Speelman (1999) onderzocht of er in de periode van 1958 tot 1998 sprake is (i) van een reële convergentie of divergentie tussen het Belgische en het Nederlandse Nederlands, en (ii) van een grotere afstand tussen het formele en informele Belgische Nederlands dan tussen het formele en informele Nederlandse Nederlands. Anders dan in het meeste traditionele variatieonderzoek, is onze benadering niet gebaseerd op geëliciteerde gegevens, maar op spontaan tot stand gekomen taalgebruik uit

<sup>19</sup> Aangezien bij 11 van de 16 prepositionele verbindingen daadwerkelijk variatiegevoeligheid werd vastgesteld (omdat één of meer varianten per profiel wel in België maar niet in Nederland voorkomt), is enige voorzichtigheid geboden bij de interpretatie van de uitzonderlijk hoge synchrone uniformiteitscijfers in de Nederlandse kolom van tabel 5: uniformiteitscijfers die in 11 van de 16 gevallen gebaseerd zijn op profielen met Belgische varianten zullen logischerwijze hoger liggen voor het Nederlandse materiaal (waarin die Belgische varianten ontbreken) dan cijfers die op basis van een kleiner aantal niet-variatiegevoelige verbindingen tot stand kwamen. Dat de gegevens in tabel 5 niettemin betrouwbaar zijn toont een herrekening die beperkt is tot de 5 niet-variatiegevoelige verbindingen in het materiaal (met prepositionele alternatieven die zowel in België als in Nederland voorkomen): noch voor Nederland noch voor België levert die reanalyse namelijk uitgesproken lagere cijfers op:  $U'(\text{IRC N}, \text{Kr N}) = 92.43$ ;  $U'(\text{Use N}, \text{Kr N}) = 93.26$ ;  $U'(\text{Inf N}, \text{Kr N}) = 94.51$  en  $U'(\text{IRC B}, \text{Kr B}) = 73.88$ ;  $U'(\text{Use B}, \text{Kr B}) = 80.80$ ;  $U'(\text{Inf B}, \text{Kr B}) = 78.33$ .

een omvangrijk corpus dat niet alleen regionaal en historisch, maar ook stilistisch-stratificationeel gecontroleerd werd: in het corpus zijn naast formele taal uit massa- en kwaliteitskranten namelijk ook twee soorten Internettaal – IRC en Usenet – opgenomen als vertegenwoordigers van het informele Nederlands. Met die nadruk op feitelijk taalgebruik hebben we vertekeningen trachten te voorkomen die het gevolg zouden kunnen zijn van de kloof tussen wat taalgebruikers doen met taal en wat ze denken te doen met taal.

Onze analyse levert de volgende empirische bevindingen op:

- Er zijn onweerlegbare kwantitatieve aanwijzingen dat het Belgische en het Nederlandse Nederlands in de periode 1958-1998 in belangrijke mate naar elkaar toegroeien op het vlak van de inhoudswoorden zowel als dat van de preposities.
- Die globale convergentie moet echter enigszins genuanceerd worden. Zowel de analyse op basis van de inhoudswoorden als die op basis van de preposities suggereert namelijk dat de precieze diachrone convergentietoename per Belgische bron afhankelijk is van de stilistisch-stratificationele status van die bron. Individuele convergentiemetingen tussen *De Gazet van Antwerpen* en *NRC* en *De Standaard* en *NRC* tonen m.n. aan dat het taalgebruik van Belgische kwaliteitskranten in 1958 in hoge mate aansluit bij de Nederlandse krantentaal, terwijl de taal van Belgische regionale massakranten aanvankelijk flink afwijkt van Nederlandse krantentaal. Het is in dit licht opvallend dat kwaliteits- zowel als massakranten op lexicaal zowel als prepositioneel vlak naar precies hetzelfde convergentieniveau evolueren.
- I.v.m. de synchrone stratificatie van het Belgische Nederlands bevestigen inhoudswoorden zowel als preposities dat de afstand tussen supraregionaal/formeel en regionaal/informeel taalgebruik in België beduidend groter is dan in Nederland. De aanwezigheid in het corpus van twee soorten Internetmateriaal met een verschillende stratificationele status stelt ons in staat de net genoemde synchrone conclusie te preciseren: de intermediaire positie tussen IRC en krantentaal die “Usenetese” stilistisch gezien lijkt in te nemen blijkt uit een grotere afstand tussen IRC en krantentaal dan tussen Usenet en krantentaal.

Met deze meervoudig gerepliceerde bevindingen is de zaak van het Belgische Nederlands echter nog niet rond. Op korte termijn zijn m.n. twee types vervolgonderzoek nodig als we de uitgangsvragen van dit artikel over de status en de structuur van het Belgische Nederlands met enige mate van vertrouwen willen beantwoorden.

Om te beginnen moeten een aantal praktische en methodologische problemen aangepakt worden die met het beschreven onderzoek verbonden zijn. Zo is het diachrone corpus dat we momenteel hanteren te klein voor onderzoek naar inhoudswoorden; dat de globale diachrone uniformiteitscijfers in tabel 3 oninterpreteerbaar zijn heeft in de eerste plaats met extreme materiaalschaarste te maken, en dus moet gezocht worden naar haalbare alternatieven voor de dure en momenteel nog weinig efficiënte methode van het krantenscannen.

Een ernstiger methodologisch defect is de basis waarop de prepositionele indicatoren gepreselecteerd werden. Dat de uit taalbeheersingsoverzichten geselecteerde verbindingen haast per definitie variatiegevoelig zijn is vooral problematisch omdat in dit stadium van het onderzoek niet nagegaan kan worden hoe representatief de gekozen indicatoren voor het geheel van het taalgebruik zijn. In een volgende fase van het onderzoek pakken we

dit probleem aan door *Abundantia Verborum* op basis van een zo exhaustief mogelijke adjectief- en prepositielijst (uit bijv. CELEX) op zoek te laten gaan naar *alle* adjectief-prepositieverbindingen waarvan het collocatiegehalte – berekend op basis van formele collocatiematen – een bepaalde drempel overschrijdt.<sup>20</sup> Het spreekt vanzelf dat deze methode de kwestie van de variatiegevoeligheid én het probleem van de representativiteit meteen oplost.<sup>21</sup> Aangezien dergelijke “veilige” data momenteel nog niet voorhanden zijn, moeten de bestaande prepositionele gegevens met de grootst mogelijke voorzichtigheid benaderd worden. Dat het methodologisch gevaarlijk zou kunnen zijn individuele prepositionele uniformiteitsgegevens in absolute zin te interpreteren, neemt echter niet weg dat er tussen de geobserveerde patronen bij inhoudswoorden en preposities een opvallende en geruststellende parallelie bestaat, die in zekere zin als een a posteriori rechtvaardiging van de verzamelde prepositionele gegevens beschouwd kan worden.

De hoge uniformiteit tussen IRC-ees en krantentaal die inhoudswoorden en preposities tonen, suggereert tenslotte dat de globale informaliteit van IRC wellicht niet in de eerste plaats door specifieke woord- of prepositiekeuzes bepaald wordt. En dus dient onze vergelijkingsbasis niet alleen van lexicale naar niet-lexicale standaardiseringsindicatoren uitgebreid te worden, maar moeten we ook kwantitatieve methodes vinden om de respectievelijke impact van lexicale en niet-lexicale factoren op de globale informaliteit van IRC en andere bronnen te berekenen. Dat laatste is een noodzakelijke voorwaarde voor inzicht in de complexe stratificationele structuur van het Belgische Nederlands; dat inzicht is op zijn beurt een essentiële voorwaarde voor een empirisch gefundeerd antwoord op de uitgangsvraag waarmee we dit artikel begonnen. Hoe overtuigend ook, de hier gepresenteerde gegevens beantwoorden die vraag (nog) niet.

## ● Bibliografie

- Aken, H. Van & V. Van Den Heede (2001).** Measuring lexical variation: a corpus-based study of prepositions and verbs in Standard Dutch. Verschijnt in *Language variation in Europe. Proceedings of the First International Conference on Language Variation in Europe*.
- Baetens Beardsmore, H. & P. Van de Craen (1979)** . The development of triglossia in Flemish Belgium. In: M. van de Velde & W. Vandeweghe (red.), *Sprachstruktur, Individuum und Gesellschaft. Akten des 13. Linguistischen Kolloquiums 1*. Tübingen: Niemeyer, 191-200.
- Bays, H. (1998).** Framing and face in internet exchanges: a socio-cognitive approach. *Linguistik Online* 98/1. Online publicatie op <http://viadrina.euv-frankfurt-o.de/~wjournal/bays.html>
- Cuyckens, H. (1991).** *The semantics of spatial prepositions in Dutch: A cognitive-linguistic exercise*. Doctorale dissertatie Universiteit Antwerpen.

<sup>20</sup> Zie Stubbs (1996) voor een overzicht van collocatiematen.

<sup>21</sup> Merk op dat de automatische methode die we hier voorstellen bij voorkeur met adjectief-voorzetselcombinaties zoals *gek op/van* uitgevoerd wordt, als we tenminste de verwarring tussen bepalingen en voorzetselvoorwerpen willen vermijden die frequent optreedt bij werkwoord-prepositieverbindingen. De computer maakt namelijk geen onderscheid tussen het voor onze doeleinden interessante Belgische voorzetselvoorwerppatroon *Wij zoeken achter onze portefeuille* en het oninteressante bepalingspatroon *wij zoeken achter de kast*.

- Geeraerts, D., S. Grondelaers & P. Bakema (1994).** *The structure of lexical variation. Meaning, naming, and context.* (Cognitive Linguistics Research 5). Berlijn: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & D. Speelman (1999).** *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen.* Amsterdam: Meertensinstituut.
- Geerts, G. (1975).** *Voorlopers en varianten van het Nederlands. Een gedocumenteerd dia- en synchroon overzicht.* Leuven: Acco.
- Goossens, J. (2000).** De toekomst van het Nederlands in Vlaanderen. *Ons Erfdeel* 43, 3-13.
- Grondelaers, S. (2000).** *De distributie van niet-anaforisch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal.* Doctorale dissertatie kU. Leuven.
- Grondelaers, S., K. Deygers, H. Van Aken, V. Van Den Heede & D. Speelman (2000).** Het CONDIV-corpus geschreven Nederlands. *Nederlandse Taalkunde* 5/4, 356-363.
- Harrison, S. (1998).** E-mail discussions as conversation: moves and acts in a sample from a listserv discussion. *Linguistik Online* 98/1. Online publicatie op <http://viadrina.euv-frankfurt-o.de/~wjournal/harrison.htm>
- Hentschel, E. (1998).** Communication on IRC. *Linguistik Online* 98/1. Online publicatie op <http://viadrina.euv-frankfurt-o.de/~wjournal/irc.html>
- Herring, S. (1996).** *Computer-mediated communication: linguistic, social and cross-cultural perspectives.* Amsterdam: Benjamins.
- Ko, Kwang-Kyu (1996).** Structural characteristics of computer-mediated language: a comparative analysis of Interchange discourse. *Electronic Journal of Communication* 6/3. Online publicatie op <http://ublib.buffalo.edu/libraries/eresources/ejournals/records/ejc.html>
- Labov, W. (1972).** Some principles of linguistic methodology. *Language in Society* 1, 97-120.
- Langacker, R.W. (1987).** *Foundations of cognitive grammar. Volume I: Theoretical prerequisites.* Stanford, CA: Stanford University Press.
- Paardekooper, P.C. (1990).** *ABN gids: taalnormboek.* Antwerpen: Standaard.
- Paolillo, J. (1999).** The virtual speech community: social network and language variation on IRC. *Journal of Computer-Mediated Communication* 4/4. Online publicatie op <http://jcmc.huji.ac.il/paolillo.htm>.
- Penninckx, W. & P. Buyse (1997).** *Correct taalgebruik.* Kortrijk-Heule: UGA
- Speelman, D. (1997).** *Abundantia Verborum. A computer tool for carrying out corpus-based linguistic case studies.* Doctorale dissertatie kU. Leuven.
- Stubbs, M. (1996).** *Text and corpus linguistics.* Oxford: Blackwell.
- Taylor, J.R. (1988).** Contrasting prepositional categories: English and Italian. In: B. Rudzka-Ostyn (red.), *Topics in Cognitive Linguistics.* Amsterdam: Benjamins, 299-326.
- Velde, H. van de (1996).** *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993).* Doctorale dissertatie K.U. Nijmegen.
- Willemyns, R. (1987).** Norm en variatie in Vlaanderen. In: J. de Rooij (red.), *Variatie en norm in de standaardtaal.* Amsterdam: Meertensinstituut, 142-164.

**Appendix 1: onomasiologische profielen voor kledingconcepten****overhemd (m)**

hemd  
overhemd  
shirt

**bloes (v)**

bloes(je)  
bloeze/bl(oe/ou)se

**t-shirt (m/v)**

T-shirt  
shirt

**trui (m/v)**

pull(etje)  
pullover  
trui(tje)

**broek (m/v)**

broek  
pantalon

**jeans (m/v)**

jeans  
spijkerbroek

**legging (v)**

caleçon  
legging  
leggings

**rok (v)**

rok(je)

**jurk (v)**

japon(netje)  
jurk(je)  
kleed(je)

**jack (m/v)**

bloes  
blouson(netje)  
jack/jek  
jacket  
jak(je)  
windbloes(je)  
windjack  
windjacket

**kostuum (m)**

colbertcostuum  
costuum  
pak

**mantelpak (v)**

complet  
costuum(pje)  
deux-pièces  
ensemble  
geheel  
mantelcostuum(pje)  
mantelpak(je)  
pak(je) tailleur(tje)  
tailleurcostuum  
tweestuks



## Appendix 2: onomasiologische profielen voor voetbalconcepten

### **aanvaller**

aanvaller  
forward  
puntspeler  
speerpunt  
spits  
spitsspeler  
voorhoedespeler  
voorman  
voerspeler  
voorwaarts

### **aftrap**

aftrap  
kick-off

### **buitenspel**

buitenspel  
off-side

### **doelman**

doelkeeper  
doelman  
doelverdediger  
doelwachter  
goal-keeper  
goalie  
keeper

### **doelpunt**

doel  
doelpunt  
goal  
treffer

### **handspel**

handbal  
handfout  
hands  
handspel  
handsspel

### **hoekschop**

corner  
hoekschop  
hoekschot  
hoekshot

### **overtreding**

foul  
fout  
overtreding

### **scheidsrechter**

arbiter  
kampleider  
match-leader  
ref(eree)  
scheids  
scheidsrechter  
spelleider  
voetballeider  
wedstrijdleider

### **strafschop**

elfmeter  
elfmetertrap  
penalty  
strafschop

### **tegenaanval**

counter  
tegenaanval  
tegenactie  
tegenoffensief  
tegenstoot

### **verdediger**

achterhoedespeler  
achterman  
achterspeler  
back  
back-speler  
verdediger

### **voorzet**

center  
centerschot  
voorzet

### **vrije trap**

free kick  
vrij schot  
vrije schop  
vrije trap  
vrijschop  
vrijschot

### **wedstrijd**

confrontatie  
duel  
kamp  
match  
ontmoeting  
partij  
treffen  
wedstrijd

### Appendix 3: onomasiologische profielen voor voorzetselverbindingen

**interesse hebben** [in/voor]

‘Een Zuid-Koreaanse club zou interesse hebben [in/voor] Calo’

[op/aan] **het einde**

‘Pas [op/aan] het einde van het seizoen waren die groeipijnen voorbij’

**zeggen** [aan/tegen/tot]

‘Ze kan dan rechtstreeks [aan/tegen/tot] kardinaal Danneels zeggen dat de paus zich wat moderner moet kleden’

**kant** [op/uit] **gaan**

‘Om maar te zeggen dat ‘t met Canal+ inderdaad niet de goeie kant [op/uit] gaat’

[aan/in/tegen] een **tempo**

‘Bij het Toneelhuis volgen de voorstellingen elkaar [aan/in/tegen] een razend tempo op’

[langs/via] (figuurlijke) **weg**

‘Misschien kan ik je [langs/via] deze weg helpen’

[langs/via] (de) **radio/telefoon/internet/...**

‘[langs/via] de telefoon kan zij dan onmiddellijk antwoorden en zien wat zij voor de patiënt kan doen’

[langs/via] **raam/venster/deur/poort**

‘Telkens drongen ze de woningen binnen [langs/via] de achterdeur’

[aan/van/langs] **alle/beide/geen/sommige/twee... kanten/zijden**

‘Djeezus mja, da piept hier [aan/van/langs] alle kanten’

[aan/langs] **weerszijden/weerskanten**

‘In de Zeelaan stonden de wagens [aan/langs] weerszijden gestationeerd’

[aan/tegen/voor] ...**prijs/prijzen**

‘Dan zijn er in de meeste studentenkroegen consumpties [aan/tegen/voor] scherpe prijzen te bekomen’

**kopen/verkopen...**[aan/tegen/voor] + **geldbedrag**

‘Eind vorig jaar hadden we in voorverkoop kaarten [aan/tegen/voor] 450 frank gekocht’

[aan/tegen/voor] + **geldbedrag per stuk/eenheid**

‘Toen sommige multinationals begonnen, kregen ze grond [aan/tegen/voor] 50fr. per vierkante meter’

**allergisch/[over][on]gevoelig** [aan/voor]

‘Ik ben allergisch [aan/voor] snorren’

**zich interesseren** [aan/voor]

‘Het kongres staat open voor alle dames en heren die zich [aan/voor] deze problemen interesseren’

[langs/aan/van] **de ene kant/de andere kant**

‘De kranten zijn onafhankelijker geworden. Maar [langs/aan/van] de andere kant is er de toenemende invloed van de commercie’

# Lexicale Noord/Zuid-variatie in het Standaardnederlands

Een synchrone en diachrone studie van werkwoorden op basis van corpora.

VICKY VAN DEN HEEDE\*

KATRIEN DEYGERS\*\*

JOHAN DE CALUWE\*

JOHAN TAEDEMAN\*

## Abstract

In this paper we explore the basic methodological assumptions that are characteristic of the corpus-based study of verbs in lexical variation research. This part of CONDIV2 (= the name of the research project) is based on a computer corpus of texts and elaborates on similar research of Geeraerts, Grondelaers & Speelman (1999) (GGs) on nouns in a corpus of manually collected instances. GGS study lexical variation from an onomasiological point of view.

It is made clear that corpus-based research that is to be conducted automatically, and not by way of manual screening of primary data, brings about some technical problems caused by the homonymous and polysemous nature of many lexical items. These problems are hard to overcome unless corpora are enriched with conceptual information. We also point out that the study of relational content words, viz. verbs, should not be limited to the study of onomasiological variation, but must also take into account other kinds of variation, on the level of valency, selectional restrictions and collocations. This kind of variation research requires a very detailed analysis of concrete instances of verb-usage, which entails a syntactic as well as a semantic subcategorisation. Unsolved fundamental issues in lexical semantics and syntax make this analysis a very challenging task.

## 1 Inleiding

Dit artikel rapporteert over het onderzoek van de Gents-Leidse pool van het VNC-project "*Lexicale variatie in het Standaardnederlands. Convergentie / divergentie en standaardisering / substandaardisering in Vlaanderen en Nederland*" (hierna CONDIV2). Elders in dit nummer brengen Grondelaers, Van Aken, Speelman en Geeraerts verslag uit over de Leuven-

\* (\*) Universiteit Gent, Vakgroep Nederlandse Taalkunde; (\*\*) Instituut voor Nederlandse Lexicologie (Leiden). Alle auteurs zijn bereikbaar via e-mail op [vicky.vandenheede@rug.ac.be](mailto:vicky.vandenheede@rug.ac.be). Met dank aan Timothy Coleman en een anonieme lezer voor hun kritische opmerkingen bij een eerdere versie van dit artikel.

Amsterdamse pool van dit project. In die bijdrage wordt uitgebreid aandacht besteed aan de historische context en de voorgeschiedenis van CONDIV2. Wij beperken ons hieronder tot een korte recapitulatie van de belangrijkste punten.

CONDIV2 is een VNC-project<sup>1</sup> van vier universitaire instellingen in Vlaanderen en Nederland en ging van start in 1997-1998. De projectmedewerkers zijn Katrien Deygers (INL Leiden), Stefan Grondelaers (KU Leuven), Vicky Van Den Heede (Universiteit Gent) en Hilde Van Aken (V.U. Amsterdam). De promotoren c.q. onderzoeksleders zijn Piet Van Sterkenburg, Dirk Geeraerts, Johan Taeldeman, Johan De Caluwe en Jaap Van Marle. Bij de aanvang kreeg het project drie belangrijke methodologische opdrachten mee: (1) het ontwikkelen van technieken om lexicale Noord/Zuid-variatie te detecteren en te beschrijven; (2) het ontwikkelen van technieken om de evolutie van de lexicale variatie te bepalen; (3) het ontwikkelen van technieken om – met behulp van labeling – lexicale variatie lexicologisch in kaart te brengen. Daarbij vormden de belangrijkste methodologische uitgangspunten en concepten (zoals *onomasiologisch profiel* en *externe uniformiteit*) zoals beschreven in Geeraerts, Grondelaers en Speelman (1999: 36-41) (hierna GGS) het vertrekpunt.<sup>2</sup>

Als materiaalbasis voor het project werd een eigen corpus samengesteld. Het CONDIV-corpus geschreven Nederlands is een computercorpus van ruwe, d.w.z. niet-geannoteerde, teksten en bevat Belgische en Nederlandse data voor drie meetpunten in de twintigste eeuw: de jaren vijftig, zeventig en negentig. Het hedendaagse luik van het corpus (ruim 42 miljoen woorden groot) bevat enerzijds internetmateriaal, afkomstig uit ‘chatgroups’ en ‘newsgroups’; anderzijds bevat het krantentaal uit drie soorten kranten: regionale massakranten, nationale massakranten en nationale kwaliteitskranten. Met deze samenstelling werd gepoogd om – stilistisch gezien – een zo breed mogelijk gamma geschreven Nederlands bijeen te brengen. Het historische luik van het corpus (ruim 5,5 miljoen woorden groot) bestaat, om praktische redenen, enkel uit krantentaal, die speciaal voor CONDIV2 werd ingescand<sup>3</sup>.

Grosso modo bestaat de opdracht van de Gents / Leidse pool erin na te gaan hoe we op basis van een corpus<sup>4</sup> van teksten (in tegenstelling tot het corpus van manueel verzamelde attestaties van GGS) de Noord/Zuid-variatie die we aantreffen bij werkwoorden (en niet bij substantieven, zoals in GGS) in kaart kunnen brengen. Daarbij moeten we niet alleen aandacht besteden aan woorden maar ook aan gebruikswijzen van woorden die typisch zijn voor het hedendaagse Belgische c.q. Nederlandse Nederlands. Beide aspecten, met name het onderzoek van een andere lexicale categorie dan substantieven enerzijds en het werken met een computercorpus van teksten anderzijds, hebben tot gevolg dat de methode GGS op een aantal punten moet worden aangepast. Deze aspecten betreffen het methodologische gedeelte van CONDIV2 en staan centraal in paragraaf 2.

In paragraaf 3 ten slotte, evalueren wij in welke mate de methode GGS bruikbaar is voor grootschalig onderzoek naar werkwoordelijke variatie in het Nederlands.

1 VNC: Vlaams-Nederlands Comité voor Nederlandse Taal en Cultuur.

2 Voor een definitie van die methodologische concepten verwijzen we naar GGS (1999) en naar de bijdrage van Grondelaers, Van Aken, Speelman en Geeraerts elders in dit nummer.

3 Voor meer bijzonderheden over het corpus verwijzen we naar Grondelaers, Deygers, Van Aken, Van Den Heede en Speelman (2000) en naar de bijdrage van Grondelaers, Van Aken, Speelman en Geeraerts elders in dit nummer.

4 Voor een definitie van een corpus zie o.m. Mc Enery & Wilson (1998: 24) en Johansson (1998: 3).

## ● 2 Condiv-onderzoek op basis van werkwoorden: methodologische beschouwingen

Logischerwijze dienen voor een corpusgebaseerd onderzoek naar lexicale variatie de volgende stappen te worden gevolgd: (1) selectie van de variabelen; (2) corpusanalyse; (3) kwantitatieve en kwalitatieve interpretatie van de resultaten. Hieronder bespreken we voor elk van die stappen de problematiek die specifiek met de aard van de onderzochte variabelen en het CONDIV-corpus samenhangt.

### 2.1 Selectie van de variabelen

#### *Woordveldgebaseerde aanpak*

De keuze voor een woordveldgebaseerde aanpak lag in zekere zin voor de hand, omdat ook GGS (1999) op die manier te werk zijn gegaan. Toch is eerst over een alternatieve aanpak nagedacht. CONDIV2 blijft immers een methodologisch project en moet dus een onderzoeksmethode ontwikkelen waarmee in principe de Nederlandse woordenschat in zijn geheel kan worden onderzocht. Los van de vraag of het ooit mogelijk zal zijn de hele woordenschat te categoriseren met behulp van woordvelden<sup>5</sup>, moet erover worden gewaakt dat op zijn minst elk type werkwoord na afloop van het project in principe kan worden geanalyseerd. Daarbij kan niet a priori worden aangenomen dat een hulpwerkwoord als *hebben* eenzelfde aanpak vraagt als bijvoorbeeld het zelfstandig werkwoord *huiten*. Idealiter zou voor elk semantisch en syntactisch type werkwoord van het Nederlands verkennend onderzoek moeten worden gedaan.

Uit praktische overwegingen was het niet haalbaar volgens dit typologische principe te werken. Wat de syntactische typologie betreft zou de – vrij rudimentaire – indeling van de ANS (1997:46) als leidraad kunnen dienen. De ANS geeft vier mogelijke indelingen aan: (1) zelfstandige werkwoorden, koppelwerkwoorden en hulpwerkwoorden; (2) overgangelijke versus onovergangelijke werkwoorden; (3) wederkerende versus niet-wederkerende werkwoorden; (4) persoonlijke versus onpersoonlijke werkwoorden. Een vergelijkbare semantische typologie is echter niet voorhanden<sup>6</sup>, en binnen CONDIV2 is er geen ruimte om zo'n typologie uit te werken. Daarom heeft een andere overweging, met name het streven naar maximale parallelie met het onderzoek van GGS (1999) de bovenhand gekregen<sup>7</sup>. Dat maakt dat ook in CONDIV2 de woordveldstrategie zal worden toegepast. Op die manier wordt vergelijking van de resultaten mogelijk over de verschillende lexicaal categorieën heen.

De keuze van een woordveld en de bijhorende representatieve begrippen is in principe vrij. In de praktijk is de onderzoeker natuurlijk wel gebonden aan het corpus dat hij/zij gebruikt. Voor CONDIV2, dat niet alleen in synchrone, maar ook in diachrone werkwoordvariatie is geïnteresseerd, houdt dat onder meer in dat zeer recente begrippen niet

<sup>5</sup> Van der Wouden (2000) plaatst zowel vraagtekens bij de haalbaarheid van de woordveldgebaseerde als van de onomasiologische aanpak.

<sup>6</sup> De meeste pogingen tot nu toe, zoals bijvoorbeeld Levin (1993) en Oppentocht (1999), zijn vaak ofwel te algemeen ofwel voor maar een bepaald type werkwoorden bruikbaar.

<sup>7</sup> Dat neemt niet weg dat binnen deze aanpak bewust verschillende types werkwoorden behandeld zullen worden. Zie bijvoorbeeld Van Den Heede (2000) voor een Noord/Zuid-vergelijking van causaal *doen* en *laten*.

in aanmerking komen voor onderzoek. Een concept als *E-MAILEN* bijvoorbeeld, zal vanzelfsprekend niet voorkomen in oudere Nederlandse teksten en leent zich dan ook niet voor CONDIV-onderzoek. Een bijkomende beperking is dat vooral frequentere concepten en de bijhorende benamingen geschikte CONDIV-variabelen zijn<sup>8</sup>. Aangezien het de bedoeling is de kwantitatieve gegevens ook m.b.v. statistiek te interpreteren, kan met begrippen/ termen die niet of nauwelijks in de CONDIV-corpora voorkomen immers weinig worden aangevangen.

Voor CONDIV2 bleek vooral de relatief kleine omvang van het historische luik van het corpus een beperkende factor bij de selectie van variabelen. Overigens vertoont dit deel van het corpus, behalve zijn beperkte omvang, nog een bijkomend manco: in het materiaal komen nogal wat scanfouten voor<sup>9</sup>, waardoor het moeilijk wordt om – met behulp van woordvormgebaseerde zoekopdrachten – alle reële voorkomens van een variabele uit het corpus te extraheren. Dat laatste kon echter gedeeltelijk worden opgevangen door een studie te maken van voorspelbare scanfouten en die fouten ook in onze zoekopdrachten te integreren. Verkennende onderzoeken met werkwoorden die blijken een kleine manuele steekproef in een aantal krantenvolumes, relatief vaak in krantenartikelen blijken voor te komen, leverden toch vrij teleurstellende resultaten op. In tabel 1 geven we een overzicht van de genoteerde frequenties voor *gooien* (betekenis: 'letterlijk: een voorwerp gooien'), *kopen*, *opereren* (betekenis: 'medische ingreep uitvoeren') en *telefoneren*. Op het eerste gezicht lijken die frequenties werkbaar; als men echter, bijvoorbeeld bij *gooien*, ook het onderscheid tussen *gooien*, *gooien met* en *gooien naar* wil kwantificeren blijken de genoteerde aantallen volstrekt ontoereikend. Enkel voor *kopen* waren er voldoende attestaties om ook de afzonderlijke gebruikswijzen in kaart te brengen.

	BN 58	NN58	BN78	NN78
<i>gooien</i>	16	13	67	29
<i>opereren</i>	9	0	10	2
<i>telefoneren</i>	9	3	15	2
<i>kopen</i>	88	62	172	92

Tabel 1: frequenties van *gooien*, *kopen*, *opereren* en *telefoneren* in het historische corpus

In het Leidse luik van CONDIV2 speelt de beperkte omvang van het historische corpus geen rol: de aandacht van de Leidse medewerker gaat namelijk uit naar de synchrone variatie in het Nederlands, waarbij het uitwerken van een geschikte manier om Noord/Zuid-variantie te labelen centraal staat. Het Leidse onderzoek is gevoerd op basis van werkwoorden in het woordveld *EMOTIES TONEN* (bijvoorbeeld de concepten *HUILEN*, *SCHELDEN* en *BELEDIGEN*). Het Gentse luik houdt zich vooral bezig met het in kaart brengen van de evolutie van het Belgische en het Nederlandse Nederlands, en moet dus noodgedwongen rekening houden met de beperkte omvang van het historische corpus bij de selectie

<sup>8</sup> Sinclair (1991: 18) wijst erop dat vooral voor onderzoek waarin ook de gebruikswijzen van woorden centraal staan, zoals het onze dus, de grootte van een corpus vaak een probleem is; men heeft nu eenmaal voldoende attestaties nodig van de onderzochte variabelen om er statistisch zinvolle uitspraken over te doen.

<sup>9</sup> Ook andere auteurs rapporteren over dat probleem. Barnbrook (1996: 32) zegt dat manuele correctie van ingescand materiaal in feite noodzakelijk is; binnen CONDIV2 was hier echter geen ruimte voor. Het historische luik van het corpus kwam immers pas ter beschikking in de loop van 2000.

van een woordveld. De keuze is hier dan ook gevallen op een woordveld met een wijd gamma van hoogfrequente concepten, met name VERBALE COMMUNICATIE (met concepten als VRAGEN, OVERLEGGEN EN WAARSCHUWEN).

*Constructie van het onomasiologische profiel*

De parallelle met GGS wordt verder doorgetrokken in de manier waarop het onderzoek wordt gevoerd. Zo zal CONDIV2 in de eerste plaats op zoek gaan naar *formele onomasiologische variatie*. De hoofdvraag voor elk onderzocht begrip luidt: welke van de mogelijke benamingen voor dit begrip is het frequentst in het Belgische en het Nederlandse Nederlands? Daartoe moet voor elk onderzocht begrip eerst het *onomasiologische profiel* worden opgesteld, of het geheel van mogelijke benamingen voor dat begrip met hun relatieve frequenties (GGS 1999: 5). Die frequenties moeten door middel van zoekopdrachten in de verschillende deelcorpora worden opgespoord (cf. 2.2). Daarna kan, onder meer door toepassing van de *externe uniformiteitsmaat* uit GGS (1999)<sup>10</sup> worden nagegaan hoe groot de overlapping tussen de verschillende deelverzamelingen is (cf. 2.3).

Twee punten maken dat CONDIV2 de werkwijze van GGS met betrekking tot het onomasiologische profiel enigszins moet aanpassen: (1) het feit dat er in CONDIV2 met een anderssoortige materiaalbasis wordt gewerkt en (2) het feit dat het onderzoek wordt gevoerd op basis van werkwoorden in plaats van substantieven. Zoals reeds gezegd wordt in CONDIV2 met een tekstcorpus gewerkt dat met de computer moet worden bevraagd, en niet met een corpus van manueel bijeengebrachte attestaties zoals bij GGS. De werkwijze van GGS had als gevolg dat het onomasiologische profiel van de onderzochte concepten geleidelijk werd geconstrueerd, als resultaat van het manueel nalezen van een groot aantal bronnen. Gaandeweg verzamelden de onderzoekers, met behulp van de afbeeldingen die in de gescreende bronnen werden aangetroffen, alle mogelijke benamingen voor concepten als JURK, BLOES etc. Niet alleen konden zij, door de grote hoeveelheid geëxcerpeerde bronnen, vrijwel zeker zijn van de volledigheid van de onomasiologische profielen; bovendien kregen ze door hun werkwijze op geen enkele manier af te rekenen met problemen van homonymie en polysemie, die bij het werken met computercorpora helaas schering en inslag zijn (cf. 2.2).

Aangezien CONDIV2 op termijn grootschalig onderzoek naar Noord/Zuid-variatie mogelijk moet maken, behoorde de manuele werkwijze van GGS, wegens te tijdrovend en te adhoc, niet tot de mogelijkheden. Automatische constructie van de onomasiologische profielen in het computercorpus was echter evenmin een optie: zoals reeds gezegd bevat het CONDIV-corpus geen enkele vorm van linguïstische annotatie, waardoor het enkel op woordvormen, en niet op concepten kan worden bevraagd.<sup>11</sup> Dat maakt dat in CONDIV2 het verzamelen van alle mogelijke benamingen voor een begrip geen resultaat is van het corpusonderzoek zoals bij GGS, maar eraan voorafgaat. Een belangrijke conse-

<sup>10</sup> Voor technische informatie over de berekening van deze maat verwijzen we naar Geeraerts e.a. 1999: 41 e.v.

<sup>11</sup> De aanwezigheid van dergelijke conceptuele informatie in corpora is veeleer uitzondering dan regel. Bouma en Schuurman (1998:23) wijzen er in een rapport over 'De positie van het Nederlands in Taal- en Spraaktechnologie' op dat er voor het Nederlands geen corpora bestaan die voorzien zijn van constituentestructuur of semantische annotatie. Dat geldt dus ook voor de INL-corpora, waarin weliswaar op LEMMA kan worden gezocht, waardoor alle verschijningsvormen van een woord kunnen worden gegenereerd met een eenvoudige query, maar waarbij geen onderscheid wordt gemaakt tussen de verschillende betekenissen van een woord.

quentie hiervan is dat het dankzij het CONDIV-corpus weliswaar mogelijk is om formele onomasiologische variatie van vooraf bepaalde synoniemen te kwantificeren, maar dat het corpus niet toelaat om ‘nieuwe’ woorden aan het licht te brengen.<sup>12</sup>

Concreet houdt de constructie van een onomasiologisch profiel in CONDIV2 in dat de onderzoeker d.m.v. de eigen intuïties, het raadplegen van lexicologische naslagwerken (waaronder, specifiek voor het Belgische Nederlands, ook taalzuiveringswerken), en eventueel door navraag bij native speakers van het Belgische en het Nederlandse Nederlands, op zoek gaat naar alle mogelijke benamingen voor het geselecteerde begrip. Pas daarna kunnen, door het bevragen van het corpus, de relatieve frequenties van die benamingen worden verzameld. Het is van belang dat het onomasiologische profiel zo volledig mogelijk is: benamingen die over het hoofd worden gezien, zullen immers niet in het corpus opgezocht en gekwantificeerd worden (cf. supra). Een onvolledig onomasiologisch profiel kan dan ook een verkeerde uitkomst opleveren.

Op het eerste gezicht lijkt het niet zo problematisch om alle mogelijke synoniemen voor een bepaald begrip bijeen te zoeken. Zoals gezegd liggen de zaken voor werkwoorden echter niet altijd even gemakkelijk.<sup>13</sup> Werkwoorden kunnen semantisch gekarakteriseerd worden als woorden die ‘een werking’ uitdrukken (ANS 1997: 45). Uiteraard wordt ‘werking’ daarbij ruimer opgevat dan ‘handeling’; de ANS vermeldt ook ‘bewegen’, ‘gebeuren’, ‘veranderen’, ‘in een toestand verkeren’ en ‘een eigenschap hebben’. In de praktijk blijkt dat de uitdrukking van de ‘werking’ waarnaar werkwoorden verwijzen, semantisch complexer is dan de uitdrukking van een ‘zelfstandigheid’, die karakteristiek is voor substantieven (zie ook de ANS 1997: 139). Problemen worden veroorzaakt door het criterium van de onderlinge verwisselbaarheid enerzijds en door het veelvuldig voorkomen van collocaties als synoniem voor een werkwoord anderzijds. In principe moeten alle woorden in een onomasiologisch profiel *onderling verwisselbaar* zijn, ten minste in de betekenis waarin ze in het onomasiologische profiel zijn opgenomen (cf. Baldinger 1980, Miller ea 1990<sup>14</sup>). Door het feit dat werkwoorden relatiewoorden zijn, en daardoor een groot deel van hun betekenis ontleen aan hun onmiddellijke context (zie Fellbaum 1990: 40), blijkt die onderlinge verwisselbaarheid van ‘synonieme werkwoorden’ bij momenten echter zeer relatief. Zo lijken de werkwoorden *bellen/opbellen/telefoneren* op het eerste gezicht goede synoniemen voor het begrip TELEFONEREN. In de zin *gisteren heeft hij naar me gebeld/gisteren heeft hij me opgebeld/gisteren heeft hij naar me getelefoneerd* zijn de drie in elk geval onderling verwisselbaar. *Opbellen* is echter niet verwisselbaar met *bellen* en *telefoneren* in de zin *ik bel met/ telefoneer met mijn moeder*.

Voor een algemeen begrip als bijvoorbeeld *MAKEN* is het probleem nog groter. Afhanke-

**12** Dit probleem heeft zich uiteraard niet gesteld in de replicatie van het onderzoek GGS in het CONDIV-corpus, waarvan de resultaten worden gepresenteerd in Grondelaers, Van Aken, Speelman en Geeraerts elders in dit nummer. In die zin is de replicatie weliswaar een verrijkende uitbreiding voor het onderzoek GGS, maar is het niet representatief voor de methodologische doelstellingen van CONDIV2.

**13** We verwijzen hier nogmaals naar Van der Wouden (2000: 95) die zich afvraagt bij welk concept het werkwoord *zich inschrijven in* (bijvoorbeeld: een bepaalde onderzoekstraditie) moet worden ondergebracht.

**14** Baldinger (1980: 216) definieert synonimie als volgt: “This [namelijk: of twee woorden synoniem zijn] can be verified by substitution: in each position I can replace any synonym within the group with any other without modifying the conceptual substance.” In Miller ea (1990: 6) vinden we: “A weakened version of this definition would make synonymy relative to a context: two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value.”



lijk van het voorwerp van het maken worden bepaalde werkwoorden gebruikt: ETEN MAKEN heeft een heel ander onomasiologisch profiel dan LAWAAI MAKEN. Fellbaum (1990:41) stelt dat werkwoorden, door hun relationele karakter, semantisch gezien de meest complexe, en moeilijkst te bestuderen woordsoort vormen. Ze wijst daarbij op het feit dat door onderzoek gebleken is dat werkwoorden veel polysemer van aard zijn dan substantieven. Hieruit concludeert ze, met Gentner en France (1988):

*(...) verb meanings are more flexible than noun meanings. Verbs can change their meanings depending on the kinds of noun arguments with which they co-occur, whereas the meanings of nouns tend to be more stable in the presence of different verbs. Gentner and France (1988) have demonstrated what they call the high mutability of verbs.*

Die ‘mutability’ van werkwoorden is een factor die het construeren van het onomasiologische profiel behoorlijk kan compliceren: de impact van de lexicale context op de betekenis van een term moet heel goed overzien worden, wat op basis van eigen intuïtie en informatie in lexicografische naslagwerken niet altijd mogelijk is. In de praktijk komt het erop neer dat het onomasiologische profiel van een begrip in de loop van het eigenlijke corpusonderzoek wordt bijgestuurd en verfijnd.

De ‘werking’ waarnaar werkwoorden verwijzen, kan in veel gevallen behalve door individuele lexicale items, ook worden uitgedrukt door een woordgroep die in meerdere of mindere mate gelexicaliseerd is. Dat heeft uiteraard gevolgen voor het onomasiologische profiel van een concept. Bij het begrip TELEFONEREN bijvoorbeeld, zijn de woordgroepen *een telefoontje plegen* of *een telefoontje doen* zeker in bepaalde contexten volwaardige synoniemen voor *bellen/opbellen/telefoneren*. In bepaalde gevallen zijn zelfs woordgroepen als *na dat telefoontje met zijn moeder*, waarin niet eens een werkwoord voorkomt, semantisch equivalent met uitdrukkingen als *nadat hij met zijn moeder had getelefoneerd*. Woordgroepen die slechts bij toeval voorkomen zullen wellicht weinig verschil uitmaken voor de resultaten van een variatieonderzoek; *collocaties* daarentegen mogen in het onomasiologische profiel van een begrip zeker niet ontbreken. Het probleem hier is dat collocaties niet altijd even gemakkelijk traceerbaar zijn als individuele lexicale items. De eigen intuïties laten de onderzoeker hier al sneller in de steek, en woordenboeken zijn lang niet consequent in het opnemen van collocaties. Toch wordt er in CONDIV2 naar gestreefd het onomasiologische profiel ook op dit punt zo volledig mogelijk te maken.

## ● 2.2 Corpusonderzoek

### *Bevraging van het corpus*

In het CONDIV-corpus zijn, door de afwezigheid van elke vorm van linguïstische annotatie, zoals gezegd enkel woordvormgebaseerde zoekopdrachten mogelijk. Dat brengt een paar heel technische problemen mee die het bevragen van het corpus bijzonder tijdrovend maken. Ten eerste genereert een woordvormgebaseerde query alle homonieme vormen van een werkwoord. Een zoekopdracht naar het werkwoord *wenen* bijvoorbeeld wordt gedefinieerd door alle mogelijke verschijningsvormen van het werkwoord op te geven : [*ween, weent, wenen, geweend, weende, weenden, wenend*]. Hiermee oogst de onderzoeker ook

alle attestaties *Wenen*, waarbij het uiteraard niet om de werkwoordvorm, maar om de hoofdstad van Oostenrijk gaat<sup>15</sup>. Een tweede probleem is dat de onderzoeker, behalve met homonieme vormen, bij polyseme woorden ook nog moet afrekenen met andere dan de onderzochte betekenissen<sup>16</sup>. Een zoekopdracht naar causaal *doen* bijvoorbeeld, genereerde duizenden treffers, waarvan na zifting slechts een dertigtal attestaties overbleven (zie Van Den Heede 2000). Het komt er dus op neer dat een omslachtige schiftingsprocedure noodzakelijk is vooraleer de studie van onomasiologische variatie mogelijk is.

#### *Analyse van werkwoorden*

In principe houden we in dit stadium van het onderzoek enkel die attestaties over die passen binnen het onomasiologisch profiel van het bestudeerde concept, en kan de methode GGS, met het berekenen van de externe uniformiteit, verder worden toegepast. Het relationele karakter van werkwoorden maakt echter dat ook de specifieke syntactische / semantische relaties die een werkwoord aangaat in de zin voorwerp van onderzoek worden. Op die manier wordt het relevant om de verschillende gebruikswijzen binnen één betekenis van een werkwoord verder onder te verdelen en te kwantificeren, en met andere woorden een *gebruiksprofiel* op te stellen. Het is immers gebleken dat bij bepaalde werkwoorden eventuele Noord / Zuid-verschillen enkel optreden in een of meerdere subcategorieën die in het profiel onderscheiden zijn, terwijl op het niveau van de algemene betekenis geen contrasten blijken<sup>17</sup>. Om te beginnen houdt zo'n gebruiksprofiel een verdere semantische categorisering van het onderzochte werkwoord in. Betekenissen zoals we ze in woordenboeken – en in veel gevallen ook in onze eigen intuïtie – hebben opgeslagen, blijken immers vaak veel te ruim gedefinieerd. *Krassen* bijvoorbeeld is toch weer net anders in *je naam in een boom krassen* dan in *een lottobiljetje krassen*. Bovendien blijkt het werkwoord in de eerste context zowel in Vlaanderen als in Nederland voor te komen, terwijl het in de tweede context enkel in Nederlands-Nederlandse bronnen werd aangetroffen. Bij deze semantische subcategorisering spelen het domein van de woordenschat en eventuele semantische selectierestricties t.a.v. de argumenten van het werkwoord een rol. Daarnaast moet de onderzoeker op dit niveau oog hebben voor eventuele collocaties, die – net als de andere subcategorieën – voor alle deelcorpora moeten worden gekwantificeerd.

Naast deze semantische categorisering is echter ook een analyse nodig van de types constructies waarin een werkwoord optreedt. In eerste instantie is het nodig de valentie van het werkwoord in kaart te brengen. Werkwoorden die divalent-transitief zijn in de ene variëteit van het Nederlands, blijken soms divalent-intransitief te zijn in de andere variëteit. In België kun je *je moeder telefoneren* of *naar je moeder telefoneren*, met of zonder voorzetsel dus; in de Nederlands-Nederlandse corpora werd alleen de constructie *telefoneren*

<sup>15</sup> Dit probleem kan niet opgelost worden door de query hoofdlettergevoelig te maken en enkel *wenen* zonder hoofdletter te vragen; de woordvorm kan namelijk op de eerste zinsplaats voorkomen, bijvoorbeeld in inversie. Overigens doen gelijkaardige problemen zich voor bij het bevragen van andere taalkundige corpora. Het probleem met 'Wenen' bijvoorbeeld, blijkt zich ook in het INL-corpus voor te doen, ondanks de codering.

<sup>16</sup> Ook dit probleem geldt evenzeer bij het bevragen van bijvoorbeeld het INL-corpus.

<sup>17</sup> Dit is ongetwijfeld ook de reden waarom werkwoorden in contrastief onderzoek zowat de meest bestudeerde lexicale categorie vormen. We verwijzen hier bijvoorbeeld naar de Gentse CONTRAGRAM-groep, die rond deze problematiek al heel wat denkwerk heeft verricht (zie o.m. Devos 1996).

naar gevonden. In België kan je probleemloos *naar iets peilen*; in Nederland zegt men *iets peilen*. Verder is er de vraag met welke voorzetselconstructies een werkwoord kan voorkomen, ook als het om een voorzetselconstituent gaat die niet tot het valentiepatroon van het werkwoord behoort. Naarmate een voorzetsel minder inherent is aan de betekenis van het werkwoord, komt die vraag echter meer en meer op het terrein van de Leuvense en de Amsterdamse CONDIV-onderzoekers te liggen, die zich o.a. bezig houden met onderzoek naar functiewoorden, m.n. de voorzetsels. In België lezen mensen hun krant *op de trein*; in Nederland zegt men *in de trein*<sup>18</sup>, maar in dit geval is het duidelijk dat de voorzetselvariatie niet werkwoordgebonden kan worden genoemd. Ten slotte is het in theorie zelfs niet ondenkbaar dat Noord / Zuid-verschillen optreden in het negatief/gepassiveerd/vragend gebruik van een bepaald werkwoord. Hier moet gezegd dat het onderzoek nog geen voorbeelden van zulke verschillen aan het licht heeft gebracht.

De constructie van het gebruiksprofiel brengt heel wat problemen mee. Op basis waarvan wordt bepaald wanneer een bepaalde gebruikswijze, semantisch of syntactisch, een aparte categorie moet vormen? Pasklare classificatiesystemen die bovendien gedetailleerd en concreet genoeg zijn om ze in de praktijk van het corpusonderzoek toe te passen, zijn nog lang niet voorhanden<sup>19</sup>, en de traditionele woordenboeken zijn op dit vlak hoogstens als controle achteraf bruikbaar. Voor sommige werkwoorden is binnen een bepaald paradigma een semantische analyse uitgedokterd. Vooral binnen de cognitieve semantiek is op dit vlak al wat werk verzet, bijvoorbeeld voor speech act verbs (Dirven e.a. 1982, Wierzbicka 1987) of voor de causale hulpwerkwoorden *doen* en *laten* (Verhagen & Kemmer 1992). Maar hier blijft de vraag of een analyse binnen een ander paradigma niet tot een heel andere subcategorisering, en dus tot andere kwantitatieve resultaten, zou hebben geleid.

In de onderzoekspraktijk van CONDIV2 wordt voor de constructie van het gebruiksprofiel meestal een 'bottom up'-strategie gevolgd: bij de analyse van de verschillende attestaties ontstaat geleidelijk aan een categorisatie op basis van eigenschappen die de onderzoeker in de semantische en syntactische context van het onderzochte werkwoord aantreft. In tweede instantie worden woordenboeken en naslagwerken als de ANS gebruikt om de voorlopige categorisering bij te sturen. Daarbij is een zekere subjectiviteit van de kant van de onderzoeker nooit uit te sluiten.

### ● 2.3 Interpretatie van de data

Zodra alle attestaties van een bepaald concept gecategoriseerd en gekwantificeerd zijn, kunnen de data kwantitatief en kwalitatief worden geïnterpreteerd. Voor de kwantitatieve interpretatie is vooral de *externe uniformiteitsmaat* (GGS 1999: 41 e.v.) van belang. Met de externe uniformiteit kan zowel diachrone als synchrone variatie worden gemeten. Deze maat, die nagaat hoe groot de overlapping tussen twee taalverzamelingen is ten aanzien van een bepaald fenomeen, kan zowel worden toegepast op het gebruiksprofiel van

18 Zie elders in dit nummer de bijdrage van Grondelaers, Van Aken, Spielman en Geeraerts.

19 We verwijzen hier opnieuw naar Levin 1993 en Oppentocht 1999.

een term in een bepaalde betekenis als op het onomasiologische profiel van een concept<sup>20</sup>. Om na te gaan of verschillen in de genoteerde relatieve frequenties significant zijn wordt gebruik gemaakt van statistische tests zoals chi kwadraat en goodness of fit. Voor het begrip TELEFONEREN (zie tabel 2) vinden we bijvoorbeeld een – vrij hoge – uniformiteit van 80% voor het hedendaagse Belgische en Nederlandse Nederlands. Dat neemt niet weg dat er toch significante verschillen zijn genoteerd: in beide corpora is *bellen* de dominante term, maar in het Belgische Nederlands heeft *bellen* meer te duchten van de termen *telefoneren* en *opbellen* dan in het Nederlandse Nederlands.

	BN(98)	NN(98)
<i>bellen</i>	73%	93%
<i>telefoneren</i>	10%	2%
<i>opbellen</i>	17%	5%

Tabel 2: relatieve frequenties voor het begrip TELEFONEREN in het hedendaagse corpus

In GGS (1999) worden verder formules voorgesteld om ook een kwalitatieve interpretatie van de data mogelijk te maken: zo kan onder meer de impact van factoren als Franse invloed, Engelse invloed en invloed van taalpropaganda worden gekwantificeerd.

### 3 Conclusie

Op basis van de beschouwingen in paragraaf 2 is een voorlopige evaluatie van de methode GGS mogelijk. Voor het meten van synchrone en diachrone variatie blijkt de woordveldgebaseerde, onomasiologische methode van GGS (1999) een goed vertrekpunt. Niettemin hebben we vastgesteld dat het werken met een niet geannoteerd computercorpus van teksten het verzamelen van data toch wel bemoeilijkt. Enerzijds is de onderzoeker gebonden door de grenzen van zijn corpus (en dan vooral door de grenzen van het historische luik ervan): weinig frequente begrippen/benamingen komen niet in aanmerking voor dit onderzoek. Anderzijds is het bevragen van het corpus met behulp van woordvormgebaseerde zoekopdrachten dikwijls erg tijdrovend vanwege de vele ongewenste hits in het geval van homoniemen en polysemen. Dat maakt dat aan de vrij rechttoe-rechtaan toepassing van de uniformiteitsmaat van GGS voor sommige variabelen een vrij omslachtige schifting van attestaties voorafgaat.

Naar analogie van het onderzoek van GGS is in CONDIV2 voor een woordveld gebaseerde aanpak gekozen. Het opstellen van een onomasiologisch profiel is minder eenvoudig gebleken voor werkwoorden dan voor zelfstandige naamwoorden. We hebben vastgesteld dat het erg belangrijk is dat de onderzoeker inzicht heeft in de precieze contexten waarin de verschillende termen/ woordgroepen van een profiel onderling verwisselbaar zijn. Verder is het van belang dat het onomasiologische profiel zo volledig mogelijk is, wat inhoudt dat ook eventuele collocaties moeten worden opgenomen.

<sup>20</sup> Aangezien tot nog toe enkel alleenstaande cases zijn onderzocht, achten wij het niet zinvol in dit stuk al resultaten mee te geven; voor een voorbeeld van een afgeronde case study verwijzen wij naar de replicatie van de GGS-studie in het CONDIV-corpus, in het stuk van Grondelaers, Van Aken, Spielman en Geeraerts elders in dit nummer.

Ten slotte is gebleken dat het bij een onderzoek naar werkwoorden wenselijk is om ook eventuele Noord / Zuid-verschillen op het vlak van o.a. selectierestricties, valentie en collocaties aan het licht te brengen. Bij het opstellen van een gebruiksprofiel wordt de CONDIV-onderzoeker echter geregeld met vragen geconfronteerd waarop geen eenduidig antwoord voor handen is. Een bottom up-strategie waarbij de lexicale en syntactische context van de variabele als uitgangspunt wordt genomen lijkt tot nog toe de beste aanpak.

Echt grootschalig lexicaal variatieonderzoek lijkt pas realiseerbaar zodra er ‘slimmere’ corpora ter beschikking komen, waarin liefst ook op betekenis, maar op zijn minst in elk geval op woordsoort kan worden gezocht. Voor diepgaand onderzoek naar het Nederlands uit het pre-computertijdperk is in elk geval behoefte aan een corpus dat op zijn minst even groot is als het hedendaagse luik uit het CONDIV-corpus. In het algemeen moet trouwens wellicht naar steeds groter wordende corpora worden gestreefd, zodat ook relatief zeldzame fenomenen kunnen worden bestudeerd (zie ook Sinclair 1991: 18). Ongetwijfeld kan een samengaan van linguïstische, statistische en technologische know-how leiden tot het ontwikkelen van efficiënte automatische procedures, waardoor de taak van de linguïst grotendeels wordt herleid tot het interpreteren van data waarin de computer relevante patronen heeft ontdekt<sup>21</sup>. Zodra dit mogelijk wordt is een woordveldgebaseerde aanpak niet langer nodig, of zelfs zinvol: met behulp van de computer kan de volledige woordenschat uit een ‘slim’ corpus van a tot z worden onderzocht. Vooraleer dergelijke ‘slimme’ corpora, zeker voor een relatief kleine taal als het Nederlands, tot de mogelijkheden behoren, zullen echter nog heel wat fundamenteel-theoretische semantische en syntactische knopen moeten worden doorgenhakt m.b.t. de analyse van lexicale items. Tot die tijd is kleinschaliger onderzoek, op basis van zorgvuldig geselecteerde woordvelden – waarbij praktische overwegingen zoals frequentie een belangrijke rol spelen – wellicht het beste alternatief.

## ● Bibliografie

- Algemene Nederlandse Spraakkunst*. **W. Haeseryn e.a. (1997)**. Groningen: Wolters-Noordhoff.
- Aken, H. Van & V. Van Den Heede (2001)**. Measuring lexical variation: a corpus-based study of prepositions and verbs in Standard Dutch. Te verschijnen in *Proceedings ICLaVE1* (First International Conference on Language Variation in Europe, Barcelona, 2000).
- Baldinger, K. (1980)**. *Semantic Theory*. Oxford: Blackwell.
- Barnbrook, G. (1996)**. *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh: University Press.
- Bouma, G. & I. Schuurman (1998)**. *De positie van het Nederlands in Taal- en Spraaktechnologie*. Rapport in opdracht van de Taalunie.
- Devos, F. (1996)**. Contrastive verb valency: overview, criteria, methodology and applications. *Studia Germanica Gandensia* 40, 15-71.

21 Hierbij komen we op het domein van *data mining*, specifiek toegepast op teksten ook wel *text mining* genoemd.

- Dirven, R., Goossens, L. & Putseys, Y. (1982).** *The scene of linguistic action and its perspectivization by speak, talk, say and tell.* Amsterdam: Benjamins.
- Fellbaum, C. (1990).** English verbs as a semantic net. *International Journal of Lexicography* 3 (4), 278 – 301. <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>
- Geeraerts, D., Grondelaers, S. & Speelman, D. (1999).** *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen.* Amsterdam: Meertens Instituut.
- Gentner, D. & I. France (1988).** The Verb Mutability Effect: Studies of the Combinatorial Semantics of Nouns and Verbs. In: S. Small, G. Cottrell & M. Tanenhaus (eds.): *Lexical Ambiguity Resolution.* Los Altos, Calif.: Morgan Kaufmann.
- Grondelaers, S., Deygers, K., Van Aken, H., Van Den Heede, V. & Speelman, D. (2000).** Het CONDIV-corpus geschreven Nederlands. *Nederlandse Taalkunde* 5/4, 356-363.
- Heede, V. Van Den (2000).** Corpusonderzoek: doen of laten? In: V. De Tier, M. Devos & J. Van Keymeulen (eds.), *Nochtans was scherp van zin. Huldealbum Hugo Ryckeboer.* Vakgroep Nederlandse Taalkunde: 429-438.
- Johansson, S. (1998).** On the role of corpora in cross-linguistic research. In: Johansson, Stig and Signe Oksefjell (eds.), *Corpora and Crosslinguistic Research: Theory, Method, and Case Studies.* Amsterdam: Rodopi, 1-24.
- Levin (1993).** *English verb classes and alternations: a preliminary investigation.* Chicago: University of Chicago Press.
- McEnery, T. & A. Wilson (1996).** *Corpus Linguistics.* Edinburgh: University Press.
- Miller, G. A., R. Beckwith, C.Fellbaum, D.Gross and K.J. Miller. (1990).** Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography* 3 (4), 235 – 244. <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>
- Oppendoort, A.L. (1999).** *Lexical Semantic Classification of Dutch verbs. Towards constructing NLP and human-friendly definitions.* Utrecht: Led.
- Sinclair, J. (1991).** *Corpus, Concordance, Collocation.* Oxford: University Press.
- Verhagen, A. & Kemmer, S. (1992).** Interactie en oorzakelijkheid. Een cognitieve benadering van causatief-constructies in het Nederlands. *Gramma / TTT, tijdschrift voor taalkunde* & (1), 1-20.
- Wierzbicka (1987).** *English speech act verbs: a semantic dictionary.* London: Academic Press.
- Wouden, T. van der (2000).** Boekbespreking: D. Geeraerts, S. Grondelaers & D. Speelman: *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen.* In: *Tabu* (30), 1-2, 94-95.

# Het design en de methodologie van het SAND-project

LEONIE CORNIPS  
WILLY JONGENBURGER\*

## Abstract

In this paper we present the design and the methodology of the SAND-project. Dialectological, sociolinguistic and generative syntactic researchers in the Netherlands and in Dutch-speaking Belgium join forces in order to make a syntactic atlas that, unlike the traditional atlases, is based on theoretical generative insights. The design of the data collection of syntactic variation in the Dutch dialects is focussed on the following research topics: left and right periphery of the clause, pronominal reference and negation and quantification.

The SAND-project is demanding with respect to the methodology. In order to be able to investigate the distribution of syntactic variables in the Dutch dialects in a systematical way, several elicitation techniques, such as written questionnaires and oral elicitation tasks, are used in the different phases of fieldwork.

## • I Inleiding

SAND is het acroniem van de Syntactische Atlas van de Nederlandse Dialecten. Het is een project van het Vlaams Nederlands Comité en het is begin 2000 gestart.<sup>1</sup> Het belangrijkste doel van dit project is het in kaart brengen en analyseren van geografisch bepaalde syntactische variatie in het Nederlandstalig gebied. De output van dit project is een syntactische atlas van de Nederlandse dialecten aan de hand van vier inhoudelijke thema's en wel: woordvolgordevariatie in de rechterperiferie van de zin, woordvolgordevariatie in de linkerperiferie van de zin, variatie in pronominale verwijzing en variatie in negatie en kwantificatie.<sup>2</sup> De thema's zijn geselecteerd omdat we op basis van reeds verricht dialect-

\* Met dank aan de SAND-promovendi en/of veldwerkers: Jeroen van Craenenbroeck, Gunther De Vogelaer, Tamar Israel, Susanne van der Kleij, Marjo van Koppen, Annemie Neuckermans, Henk Wolf en Hedde Zeijlstra. Leonie Cornips en Willy Jongenburger zijn beiden als wetenschappelijk onderzoeker verbonden aan het Meertens Instituut, Joan Muyskenweg 25, 1096 CJ Amsterdam. Email: leonie.cornips@meertens.knaw.nl en willy.jongenburger@meertens.knaw.nl

1 Aan het SAND-project werken de volgende instellingen mee: Universitaire Instelling Antwerpen, Rijksuniversiteit Gent, Katholieke Universiteit Brussel, Universiteit van Amsterdam, Universiteit Leiden en Meertens Instituut (zie ook <http://www.meertens.knaw.nl/projecten/sand/sand.html>)

2 Tevens zullen er over de thema's individuele proefschriften verschijnen.

onderzoek en al bestaande literatuur verwachten dat zij een ruime mate van variatie zullen laten zien met een duidelijke geografische spreiding.<sup>3</sup> Verder is de theoretische relevantie van de thema's een belangrijk criterium evenals het feit dat zij een representatief beeld van de syntactische variatie binnen het Nederlandstalig gebied geven (zie Bennis en Van der Auwera 1999).

Het SAND-project vormt een ideale kruisbestuiving tussen de traditionele dialectologie en de theoretische syntaxis. Het SAND-project zet de traditie van grotere en kleinere dialectologische atlasprojecten zoals respectievelijk de MAND (Morfologische Atlas van de Nederlandse Dialecten) (zie Goeman & Taeldeman 1996) en de FAND (Fonologische Atlas van de Nederlandse Dialecten) (zie Taeldeman & Verleyen 1999) en de Atlas van de Algemene Dialectsyntaxis (AND) voort (zie Gerritsen 1991). Het is overigens belangrijk om te vermelden dat syntactische variatie in atlasprojecten te midden van andere linguïstische variatie altijd een zeer kleine plaats inneemt. Volgens Gerritsen (1991:9) betreft van het totaal aantal gepubliceerde kaarten in een atlas nooit meer dan vijf procent syntactische variatie.

Daarnaast concentreert het SAND-project zich op de theoretische vraag welke bouwprincipes alle dialecten gemeen hebben, op welke punten zij verschillen en wat de grenzen van de variatie zijn. Het SAND-project gaat dus niet uitsluitend op zoek naar de geografische spreiding van een syntactische variant zoals in meer traditioneel dialectologisch onderzoek het geval is maar het onderzoekt die altijd in samenhang met andere syntactische varianten die volgens de theorie van belang kunnen zijn. Bijvoorbeeld congruentiemorfologie op het voegwoord is volgens een generatieve hypothese identiek aan de verbale congruentiemorfologie die aangetroffen wordt in inversiecontexten (zie Vanden Wyngaerd 2000). Het maken van een op generatieve theorie gebaseerde syntactische dialectatlas is nieuw voor het Nederlandstalig gebied maar niet voor het buitenland. Als voorbeeld van een soortgelijk atlasproject noemen we het ASIS-project (Syntactische Atlas van de Noorditaliaanse Dialecten) in Padua<sup>4</sup>.

In dit artikel bespreken we het design en de methodologie van het SAND-project. Er zijn vele factoren die bij een dergelijk grootscheeps atlasproject van wezenlijk belang zijn en waar in het design rekening mee gehouden moet worden. Belangrijke vragen voor het design zijn bijvoorbeeld: (i) op welke wijze verzamelen we de data voor de atlas en hoe valide zijn die data? (ii) op welke wijze visualiseren we straks de data? (iii) welke meetpunten in Nederland en België selecteren we en hoe ziet de spreiding van die meetpunten eruit? (iv) hoeveel en welke informanten ondervragen we per meetpunt? (v) op welke wijze ondervragen we de informanten ofwel welke elicitatietechnieken zetten we in? en tenslotte (vi) wat is de meest efficiënte veldwerkcyclus om zo optimale gegevens te verzamelen?

Met betrekking tot de vraag op welke wijze we straks de data visualiseren, kunnen we al gebruik maken van ervaringen opgedaan in het ASIS-project. In dit project is men uiteindelijk afgeweken van het oorspronkelijke doel, het maken van een atlas. Visualisering

<sup>3</sup> Zie voor de dialectbibliografie die door de SAND-uitvoerders is aangelegd: <http://www.meertens.knaw.nl/projecten/sand/sand.html>

<sup>4</sup> Zie <http://asis-cnr.unipd.it/>



van de ruimtelijke spreiding van varianten in de vorm van kaarten is ondoenlijk gebleken omdat bij sommige thematische onderwerpen zoveel verschijnselen met elkaar samenhangen, dat het vanwege het grote aantal symbolen niet meer zinvol is om deze op één kaart te visualiseren. Ook voor de Nederlandse syntactische atlas zal de complexe relaties tussen variabelen die nauw met elkaar samenhangen niet op overzichtelijke wijze op kaarten gevisualiseerd kunnen worden. Ter illustratie: voor de variatie in de rechterperiferie zouden variatie in de volgorde van werkwoordclusters, plaatsing van *te* bij infinitief, het IPP-effect en de aanwezigheid van *ge-*, en de opbreekbaarheid van werkwoordclusters op één kaart gevisualiseerd moeten worden omdat al deze verschijnselen theoretisch gezien met elkaar samenhangen. Wanneer elke variant van de genoemde verschijnselen wordt gevisualiseerd door middel van een uniek symbool, dan zou het resultaat een kaart zijn met een wirwar van symbolen, die niet voor interpretatie vatbaar is. In eerste instantie heeft men bij ASIS gedacht aan het werken met doorzichtige bladen met diverse symbolen voor een verschijnsel, die over elkaar heen gelegd kunnen worden. Het gebruik van een dergelijke losbladige atlas laat uiteraard te wensen over. Een adequate oplossing voor dit probleem is een algoritme gekoppeld aan de syntactische databank waarmee door een zoekleutel in te tikken, een kaart gevisualiseerd en geprint kan worden met de spreiding van de gewenste syntactische verschijnselen. Ook voor het SAND-project lijkt dit de aangewezen techniek. Op het Meertens Instituut is ruime ervaring en expertise met cartografische technieken aanwezig. Het is de overweging waard om, in navolging van het ASIS-project, het eindproduct in de vorm van een databank op internet te zetten. Op deze wijze zouden onderzoekers van elders (na autorisatie) gebruik kunnen maken van de waardevolle dialectdata die in de loop van het onderzoek verzameld zijn.

De opbouw van dit artikel ziet er als volgt uit. In het tweede deel gaan we in op de relatie tussen dialect en standaardtaal en de validiteit van de te verzamelen data. In het derde deel bespreken we het aantal en de spreiding van de meetpunten én het aantal en het profiel van de informanten en tenslotte het design van de mondelinge interviews. In het vierde deel passeren de (in)directe schriftelijke en mondelinge elicitatietechnieken de revue. In het vijfde deel bespreken we de opzet van het veldwerk in het SAND-project.

## ● 2 De relatie standaardtaal en dialect

In dit project staan de Nederlandse dialecten in Nederland en in België centraal. Het te onderzoeken taalgebied is niet homogeen. Immers, de relatie tussen de standaardtaal en het dialect in Nederland is anders dan in België. Omdat veel van de westelijke dialecten in Nederland in taalkundig opzicht erg dicht bij de standaardtaal liggen, is de scheidslijn tussen dialect en standaardtaal daar niet altijd duidelijk. Zo zullen dialectsprekers in de Randstad niet snel aangeven dat ze dialect spreken, en kunnen we potentiële informanten dus niet eenvoudigweg vragen welk dialect ze spreken. Bij de selectie van informanten gaan we er daarom vanuit dat alle informanten geboren en getogen moeten zijn in hun huidige woonplaats, evenals hun ouders (zie § 3.4). In de periferie van Nederland is het structurele verschil tussen de standaardtaal en de dialecten aanzienlijk groter en is het dialect minder sociaal gemarkeerd dan in het westen. Het standaardnederlands in Neder-

land kan beschouwd worden als het sociolect van de midden- en hoge klasse. In België spreekt iedereen dialect, ongeacht de sociale status. Het gebruiksdomen van het standaardnederlands in België is volgens Van de Velde (1996:31) voornamelijk beperkt tot formele situaties: het is een tamelijk vastliggende norm, die men zo dicht mogelijk probeert te benaderen. De meeste sprekers echter houden het bij het Belgisch Beschaafd als standaardtaal, een soort van regionaal gekleurd Nederlands, waarin grotere interferentie van het dialect is dan gebruikelijk bij een standaardtaal. De dialectsituatie in België is te vergelijken met de periferie van Nederland, bijvoorbeeld Limburg, waar iedereen zich ervan bewust is wanneer iemand dialect en wanneer iemand standaardtaal spreekt.

### 2.1 De validiteit van de data

Een belangrijke vraag in het SAND-project is de wijze waarop de dialectdata verzameld gaan worden. Onderzoek naar de geografische spreiding van syntactische verschijnselen in niet-standaardvariëteiten zoals dialecten is problematisch. Een onderzoeker kan niet beschikken over kant en klare grammatica's zoals dat wel het geval is bij standaardtalen (zie bijvoorbeeld de ANS voor de Nederlandse standaardtaal). Ten tweede kan er nauwelijks direct onderzoek van de (informele) spreektaal plaatsvinden hoewel deze code de meest valide data levert. Dit is praktisch niet haalbaar omdat voor een dialectatlas vele dialecten ofwel een groot geografisch gebied onderzocht moeten worden. Bovendien komen in de spreektaal vele syntactische varianten slechts sporadisch voor. Een goed voorbeeld daarvan zijn drieledige werkwoordsclusters met twee modale werkwoorden (zie Barbiers 2000). Er is daardoor weinig inzicht in het voorkomen van meerdere volgordes binnen een dergelijk cluster als *ik weet dat Jan hard moet kunnen werken* of *ik weet dat Jan hard werken kunnen moet*. Bovendien is bij onderzoek naar spreektaal de samenhang tussen diverse variabelen afhankelijk van toevallige voorkomens en is er geen inzage in de begrenzing van de variatie (negatieve evidentie), met andere woorden men kan nooit achterhalen of een bepaalde constructie ongrammaticaal is zoals bijvoorbeeld de volgorde *ik weet dat Jan hard kunnen moet werken* in het drieledige werkwoordscluster. Om bovenstaande redenen heeft een experimentele setting de voorkeur.

Zodra er een soort experimentele setting ontstaat, hoe subtiel ook, zoals in een mondeling 'spontaan' gesprek tussen interviewer en informant, spreken we van geobserveerd in plaats van feitelijk gedrag (Labov 1972:209). Het observeren van gedrag heeft als gevolg dat alleen al de aanwezigheid van de onderzoeker/interviewer het taalgedrag van de informant kan beïnvloeden: mogelijk past de informant zijn taalgebruik aan de gesprekspartner aan (zie Giles & Powesland 1975). Ook is er een experimentele setting mogelijk waarin gerapporteerd gedrag van de informant verzameld wordt doordat de interviewer vragen aan de informant stelt over zijn taalgebruik. De term gerapporteerd gedrag geeft aan dat het gaat om subjectieve antwoorden: datgene wat de informant zelf aangeeft te doen, hoeft niet per se hetzelfde te zijn als wat hij in werkelijkheid doet (feitelijk gedrag). Het is uit de sociolinguïstische literatuur bekend dat antwoorden van sprekers van niet-standaardvariëteiten die een lager sociaal prestige bezitten meestal de standaardtaalnorm zullen benaderen (Labov 1972:21). Ook zijn sprekers nauwelijks in staat om aan te geven welke vormen zij al dan niet zelf gebruiken of (on)aanvaardbaar vinden. Uit Labov (1972:177) en vele studies daarna is het duidelijk geworden dat slechts weinig sprekers zich realiseren dat zij

gestigmatiseerde vormen waarover zij zelf een negatief oordeel hebben, ofwel een negatieve attitude hebben, toch zelf gebruiken. Zij vinden van zichzelf dat zij prestigevarianten spreken (Labov 1972:213). Uit het bovenstaande blijkt dat we voor de SAND-atlas moeilijk rechtstreeks aan informanten kunnen vragen of zij bepaalde constructies al dan niet grammaticaal vinden. Op zich is er niets tegen het gebruik van deze methode, zolang de onderzoeker zich realiseert dat data die op deze manier verzameld zijn, betrekking hebben op gerapporteerd gedrag en dus subjectief zijn. De interpretatie van gerapporteerd gedrag dient dan ook met grote voorzichtigheid te gebeuren. Waar mogelijk zou de onderzoeker ernaar moeten streven het beeld dat ontstaat op basis van gerapporteerd gedrag te completeren en te valideren aan de hand van onderzoek naar geobserveerd gedrag (zie Jongenburger & Aarssen te verschijnen; Cornips & Jongenburger 1999).

In de traditionele dialectologie maakt men voor de grotere atlasprojecten veelvuldig gebruik van schriftelijke vragenlijsten. Deze methode zullen we ook in het SAND-project toepassen als voorbereiding op de mondelinge interviews. De uiteindelijke atlas wordt echter uitsluitend gebaseerd op data verkregen uit de *mondelinge* veldwerkronden.

### 3 SAND-atlas

Het primaire doel van het SAND-project is het maken van een op generatieve theorie gebaseerde syntactische dialectatlas waarin variatie die de syntactische variabelen vertonen, uitsluitend toegeschreven mag worden aan de geografische spreiding van die variabelen. Het is daarom van wezenlijk belang om bij voorbaat variatie die veroorzaakt kan worden door sociale variabelen zoals leeftijd, opleiding, beroep en etniciteit van de informant evenals effecten van taalcontact, zoveel mogelijk uit te sluiten. Het design van het SAND-project houdt hiermee rekening door deze sociale variabelen constant te houden.

#### 3.1 Spreiding van de meetpunten

In het verleden hebben verschillende onderzoekers gewerkt aan een classificatie van de Nederlandse dialecten in Nederland en België. Op basis van een dergelijke dialectindeling van het Nederlandstalig gebied zou een eerste selectie gemaakt kunnen worden van voor de SAND mogelijk interessante gebieden. Om tot een dialectindeling te komen zijn verschillende methoden gebruikt. Nerbonne & Heeringa (1998) behandelen beknopt Goossens' overzicht van methoden (zie Goossens 1977). De besproken methoden zijn: 1) de vlakkenmethode, waarbij op basis van bekende dialectgeografische verschillen en op grond van intuïties een kaart verdeeld wordt in vlakken; 2) de isoglossenmethode, waarbij de beslissing welke samenvallende isoglossen al dan niet beschouwd moeten worden als dialectgrens, wordt genomen op basis van subjectieve gronden; 3) de pijltjesmethode, waarbij dialecten die volgens de sprekers ervan (vrijwel) gelijk zijn, met elkaar worden verbonden door pijltjes.

Nerbonne en Heeringa concluderen dat de traditionele methoden voor dialectclassificatie subjectief van aard zijn met een atomistische aanpak waardoor het niet mogelijk is observaties te aggregeren naar een hoger niveau en uitspraken te doen over streektalen. Bovendien zijn de genoemde methoden dichotoom: dialecten zijn gelijk of ongelijk, er

kan niet aangegeven worden in welke mate dialecten van elkaar verschillen. Nerbonne en Heeringa presenteren een nieuwe, computationele methode voor de analyse van dialectologische gegevens die gebaseerd is op de Levenshtein-afstand, die voor het eerst in de dialectologie werd gebruikt door Kessler (1995). Bij deze methode worden dialecten met elkaar vergeleken door een woord uit het ene dialect te vergelijken met het corresponderende woord in het andere dialect. Het vergelijken gebeurt door te bepalen hoe het woord zo eenvoudig mogelijk kan worden veranderd in het andere woord door klanken toe te voegen, te verwijderen of te vervangen (Nerbonne & Heeringa 1998:169). Met behulp van multidimensional scaling brengen de auteurs de 104 dimensies (of dialecten) terug tot drie dimensies, ofwel drie hoofdgroepen: de Friese, de Saksische en de Frankische dialecten. Deze indeling komt overeen met resultaten uit de traditionele dialectologie<sup>5</sup>.

De meest eenvoudige oplossing voor de spreiding van de meetpunten ten behoeve van de SAND-atlas is een constante geografische spreiding voor het hele Nederlandstalige gebied. Volgens deze methode ligt de fysieke afstand tussen de te onderzoeken plaatsen vast, door bijvoorbeeld over het taalgebied een denkbeeldig raster te leggen met hokjes met constante afmetingen en binnen elk hokje een plaats te selecteren. Nadeel van een dergelijke aanpak is dat er op geen enkele wijze gebruik wordt gemaakt van de dialectologische kennis die voorhanden is. Een methode waarbij het grootste deel van het totaal aantal meetpunten wordt geselecteerd op basis van een gelijkmatige geografische spreiding en het resterende aantal meetpunten wordt aangewezen op basis van overgangsgebieden zoals die door Weijnen (1966) en Daan en Blok (1977) beschreven zijn, lijkt optimaal. Met deze methode is een gelijkmatige geografische spreiding van de meetpunten over het gebied gewaarborgd en maken we efficiënt gebruik van dialectologische kennis (zie § 3.3 voor details over de uiteindelijke selectie van de meetpunten).

### 3.2 *Typologie van de meetpunten*

Om te voorkomen dat er een traditionele, agrarische atlas wordt gemaakt, zijn een aantal grote steden in Nederland en België als meetpunt opgenomen. In Nederland is o.a. gekozen voor Amsterdam, Utrecht, Rotterdam, Nijmegen en Heerlen. De representatie van grote steden in de atlas heeft een aantal consequenties. Allereerst bestaan er vaak verschillende variëteiten in een grote stad naast elkaar. Informanten uit de grote steden kunnen daarom niet beschouwd worden als representatief voor het Amsterdams of het Utrechts, maar als representatief voor de wijk/straat waar ze wonen.

Voor de overige meetpunten stellen we als uitgangspunt dat de plaatsen een al wat oudere ontstaansgeschiedenis moeten hebben, en geen recente, grote demografische veranderingen hebben doorgemaakt. Op basis van deze criteria worden groeikernen zoals Zoetermeer, Almere en Lelystad uitgesloten.

<sup>5</sup> Vergelijking van de klassieke dialectindeling op basis van fonologische eigenschappen met de binnen de syntactische atlas te verzamelen dialectgegevens zal nieuw licht werpen op de verhouding tussen fonologische en syntactische variatie binnen de Nederlandse dialecten.

### 3.3 Uiteindelijke selectie meetpunten

Praktische overwegingen zoals menskracht en bewaking van de fasering in de tijd spelen een belangrijke rol bij het aantal meetpunten dat we kunnen verwerken voor de atlas. Uitgaande van zeven veldwerkers (vijf in Nederland en twee in België) lijkt een aantal van 250 meetpunten voor de *mondellinge* interviews voor België en Nederland samen realistisch. In Nederland zijn 150 meetpunten geselecteerd; tweederde (98 meetpunten) daarvan is op basis van een gelijkmatige geografische spreiding gekozen (zie §3.1). In tweede instantie zijn de resterende 52 meetpunten gekozen op basis van dialectologische criteria. Volgens Weijnen (1966) en Daan en Blok (1977) kent Nederland vele overgangsgebieden waar te verwachten is dat de taalkundige variabelen een grote variatie zullen laten zien. De resterende 40 meetpunten zijn vooral over deze overgangsgebieden, namelijk Limburg en het verdere oosten van Nederland gespreid. Ook in België zijn 100 meetpunten geselecteerd op basis van geografische spreiding waarbij de FAND-atlas als uitgangspunt is genomen. Een aantal meetpunten zijn – net als in Nederland – achter de hand gehouden voor de overgangsgebieden.

Bij het schriftelijk opvragen van gegevens wordt een fijnmaziger netwerk gebruikt (zie §4.1).

### 3.4 Profiel van de informanten

Voor elk van de 250 meetpunten in het Nederlandstalig gebied wordt één informant gekozen. De volgende selectiecriteria worden gehanteerd voor geschikte informanten:

- persoon spreekt het dialect van de plaats waarin hij woont;
- persoon is geboren en getogen in de plaats waarin hij woont, evenals zijn ouders. De informanten moeten in elk geval tot minimaal hun 18e jaar in de plaats opgegroeid zijn en niet langer dan 7 jaar ergens anders hebben gewoond (bijvoorbeeld voor hun studie);
- persoon spreekt buiten het huiselijk domein in minstens één publiek domein dialect. De informanten zullen over het algemeen tweetalig zijn wat ook wenselijk is met het oog op de elicitatietoetsen die we zullen afnemen (vertalen vanuit standaardnederlands zie §4.1 & 4.2). De mate van tweetaligheid zullen we bevragen aan de hand van een domeinvragenlijst, waarbij voor een aantal communicatieve situaties wordt gevraagd of er dialect wordt gesproken, bijvoorbeeld met de partner, de burens en op het werk.

Overige richtlijnen bij de selectie zijn:

- persoon komt bij voorkeur uit lage of lage middenklasse;
- persoon is bij voorkeur tussen de 55-70 jaar oud.

Uit een aantal onderzoeken blijkt dat vrouwen in Nederland in grote steden frequenter vormen uit de standaardtaal gebruiken dan mannen (zie Brouwer 1989; Schatz 1986) en dat het patroon omgekeerd is op het platteland waar in het algemeen de vrouwen meer dialectvarianten gebruiken dan mannen (zie van Hout 1989:76, Münstermann 1986). Boves & Gerritsen (1995:212) merken hierover op dat:

[...] door de toenemende verstedelijking en de opkomst van de moderne vervoer- en communicatiemiddelen het verschil tussen stad en het platteland aan het afnemen is en het daarom te verwachten is dat in de toekomst de seksespecifieke taalpatronen op het platteland hetzelfde zullen worden als in de stad.

Recent onderzoek naar de mate van dialectgebruik onder een representatieve steekproef van 5000 informanten in Nederland (Jongenburger & Vousten, 2000) heeft aangetoond dat er geen interactie (meer) bestaat tussen de mate van stedelijkheid en geslacht. Bovendien is het verschil in de mate van dialectgebruik tussen mannen en vrouwen slechts acht procent (53% van de mannen geeft aan dialect te spreken versus 45% van de vrouwen). Op basis van deze onderzoeksresultaten is besloten zowel mannelijke als vrouwelijke informanten te werven voor de atlas.

### *3.5 Design mondelinge interviews met hulpinterviewer*

Omdat geen van de veldwerkers (c.q. uitvoerders) die Nederland voor hun rekening nemen, dialect spreekt, worden dialectsprekende hulpinterviewers ingeschakeld om effecten van accommodatie tussen informant en interviewer zoveel mogelijk te voorkomen, dat wil zeggen dat de standaardtaal interfereert in het lokale dialect (zie § 2.2). Dit betekent dat voor elk meetpunt c.q. plaats één informant en één hulpinterviewer geworven worden en voor deze hulpinterviewer gelden dezelfde selectiecriteria als voor de informant. De dialectsprekende hulpinterviewer die wordt ingeschakeld, heeft tijdens de gehele procedure een actieve rol. Elke hulpinterviewer gaat allereerst zelfstandig op zoek naar een geschikte informant (iemand die niet bij hem/haar in huis woont of heeft gewoond), en neemt in de eerste ronde van het interview samen met een veldwerker een gestructureerde vragenlijst door. De hulpinterviewer spreekt de vragen in het lokale dialect in op een DAT-/minidiscorder. Bovendien wordt de hulpinterviewer in deze ronde naar metalinguïstisch commentaar gevraagd. De informant is niet aanwezig tijdens dit gesprek. In de tweede ronde speelt de veldwerker in aanwezigheid van de hulpinterviewer de opgenomen zinnen met instructie af voor de informant. De hulpinterviewer herhaalt desgewenst 'live' de opgenomen zinnen voor de informant en verduidelijkt eventueel de opgenomen instructies. De veldwerker is dus wel aanwezig in de tweede ronde van het interview, maar het initiatief tijdens het interview ligt duidelijk bij de hulpinterviewer. Dit hele gesprek vindt plaats in het dialect en de standaardnederlands-sprekende veldwerker houdt zich zoveel mogelijk op de achtergrond om accommodatie te voorkomen. Teneinde zuiver dialect te ontlokken en een ongedwongen sfeer te creëren, opent de hulpinterviewer het gesprek met de informant in het dialect en spreekt eventueel over bekende of plaatselijke onderwerpen, bijvoorbeeld de geschiedenis van zijn woonplaats, zijn levensloop of eventuele hobby's. Na afloop van het interview neemt de SAND-veldwerker het gesprek over en stelt naar aanleiding van de gegeven antwoorden eventueel nog aanvullende vragen in (noodzakelijkerwijs) het standaardnederlands.

In België wordt voor een andere aanpak gekozen. Op basis van eerdere veldwerkervaringen (met voornamelijk lexicografisch onderzoek) laten de onderzoekers dialectsprekende taalkundigen alle mondelinge interviews afnemen: één taalkundige neemt circa tien verschillende meetpunten c.q. plaatsen voor zijn rekening met twee informanten per interview.

### 3.6 Werving hulpinterviewers en informanten

In eerste instantie worden bestaande informantenbestanden gescreend op potentiële kandidaten die in aanmerking komen als hulpinterviewer. Het informantenbestand van het Meertens Instituut omvat circa 850 personen die vragenlijsten op het terrein van de volkskunde en dialectologie ingevuld hebben, waaronder de informanten van het Goeman-Taeldemanproject (zie §4.1). Ook in België bestaan er informantenbestanden waaruit geput kan worden, namelijk die van de FAND (zie Taeldeman & Verleyen 1999) en bestanden ten behoeve van dialectwoordenboeken.

Informanten die in een plaats wonen die als meetpunt geselecteerd is en in aanmerking lijken te komen voor de functie van hulpinterviewer, worden telefonisch getest op het profiel zoals geschetst in §3.4 en op geschiktheid om te fungeren als hulpinterviewer. Iemand die erg normatief of streng is ten opzichte van dialectsprekers, bijvoorbeeld door meningen als ‘men spreekt hier geen goed dialect’, is niet geschikt voor deze taak.

Als binnen de bestaande informantenbestanden geen geschikte hulpinterviewers gevonden kunnen worden, wordt contact opgenomen met plaatselijke dialectverenigingen en/of heemkunde-kringen. Van leden van deze organisaties mag een algemene, actieve interesse in de streek/regio verwacht worden en eventuele bereidheid om mee te werken aan het onderzoek.

## 4 De elicitatietechnieken

In dit hoofdstuk bespreken we eerst de schriftelijke elicitatietechniek (zie Oxford 1982) die we in het SAND-project toepassen, namelijk de vragenlijst. Merk nogmaals op dat de data die met deze technieken verkregen worden, gerapporteerd en geen feitelijk gedrag representeren (zie § 2). Grote verschillen tussen feitelijke en gerapporteerde of geobserveerde data kunnen drie verschillende oorzaken hebben. In de eerste plaats vormen schriftelijke vragenlijsten een evident nadeel. Dialecten zijn natuurlijk spreektaal die niet of nauwelijks geschreven worden. Dit betekent dat de gevraagde variant altijd in de spelling van de standaardtaal aangeboden zal worden. Er wordt van de informant dus heel wat verwacht. Bij vertaalvragen waarin vertaling naar het dialect moet plaatsvinden, leest hij een variant in de standaardspelling, moet deze variant omzetten voor zichzelf in zijn dialect en vervolgens schrijven in een zelf uit te kiezen dialectspelling. Bij vragen van het type ‘komt constructie x voor in uw dialect?’, moet de informant ook nog weer terug vertalen naar de standaardvariant (Van Bree 1981).

Ten tweede hebben informanten bij schriftelijke vragenlijsten de neiging zich op het lexicale en op het uitspraakniveau te richten. Het grote gevaar is dus dat de informant de zin afkeurt om de inhoud vanwege lexicale, stilistische redenen maar niet om syntactische redenen, terwijl dit wel is wat de onderzoeker wil weten (Greenbaum 1973:210). In Figuur 1 is een voorbeeld van een deel van een schriftelijke vragenlijst te zien die in 1995 aan dialectsprekers in Rijnland in Duitsland is rondgestuurd.

Deze vragenlijst beoogt te achterhalen of er al dan niet reflexieve onpersoonlijke en adjunct mediumconstructies in het Rijnlandse dialect voorkomen (zie Cornips 1997,

2000). Bij onderdeel 3 wordt naar de Rijnlandse vertaling gevraagd.<sup>6</sup> De informant heeft bij (a) tot en met (d) overal ‘ja’ doorgestreept en aldus kenbaar gemaakt dat geen van de varianten in het Rijnlands dialect –in dit geval Aken – voorkomt. Bij de vertaaltvraag heeft deze informant ingevuld dat *herrlich* als dialectwoord totaal onbekend is in het Akense dialect (zie Figuur 1). De informant heeft daarom alle varianten – dus uit lexicale motieven – ‘syntactisch afgekeurd’ terwijl variant (c) in het Akense/Rijnlandse dialect volledig grammaticaal is.

Figuur 1: Voorbeeld uit een schriftelijke syntactische vragenlijst

		1		2
		kommt vor/		am
		ist manchmal zu hören		gebräuchlichsten
a	Dieser Stuhl sitzt herrlich	<del>ja</del>	nein	a
b	Dieser Stuhl sitzt sich herrlich	<del>ja</del>	nein	b
c	Es sitzt sich herrlich auf diesem Stuhl	<del>ja</del>	nein	c
d	Es sitzt herrlich auf diesem Stuhl	<del>ja</del>	nein	d
3	Übersetzung .....: “Herrlich” = total ungebräuchlich bezw. Nur aus der Bibel bekannt (Übersetzung nach Hermanns = brijlant.)			

Ten derde speelt in schriftelijke vragenlijsten het repetitie-effect een rol. Aangezien informanten in hun vertaling zich vooral op inhoud, lexicon en uitspraak in samenhang met spelling concentreren, zijn zij geneigd om de aangeboden zinnen in syntactisch opzicht slaafs te vertalen. Ten slotte is het volgorde-effect van belang. Informanten vinden constructies aanvaardbaarder naarmate zij meer soortgelijke constructies tijdens het invullen van een vragenlijst onder ogen krijgen (zie van Bree 1981, Gerritsen 1991:15, Cornips 1991 & 1994:46). In de SAND-vragenlijst is met al deze factoren die verschillen tussen het feitelijk en gerapporteerd gedrag kunnen veroorzaken zoveel mogelijk rekening gehouden.

In een pilot-onderzoek is de vragenlijst aan 31 taalkundigen, verspreid over Nederland en België, toegestuurd en vervolgens samen met één of twee SAND-uitvoerder(s) geëvalueerd. Uit de evaluatie blijkt dat alle bovengenoemde test-effecten in de vragenlijst een rol spelen. Ook het commentaar van de taalkundigen is in de uiteindelijke vragenlijst zo goed mogelijk verwerkt.

<sup>6</sup> De instructie luidde: “Auf der gestrichelten Linie unter III übersetzen Sie bitte denjenigen Satz, der Ihnen am gebräuchlichsten erscheint (umkringelt in Spalte II), in Ihren Dialekt.”



#### 4.1 De SAND-vragenlijst

Na de verwerking van de resultaten verkregen uit het pilot-onderzoek is de definitieve SAND-vragenlijst aan de informanten van het Meertens Instituut voorgelegd. De informanten van het Meertens Instituut (nu circa 850 personen) zijn 'geoeffende' informanten in de zin dat zij elk jaar vragenlijsten op het terrein van dialectologie en volkskunde invullen. Het voordeel van dit respondentenbestand is dat het onmiddellijk 'bruikbaar' is. Echter, aangezien de sociale gegevens zoals sekse, beroep/opleiding en leeftijd van de informanten onvolledig zijn, is niet altijd te achterhalen of de verschillen tussen de antwoorden van de informanten veroorzaakt worden door sociale variabelen en/of door de geografische spreiding van de informanten. Het informantenbestand telde eind 1996 505 mannen en 220 vrouwen; er zijn naar verhouding dus zeer weinig vrouwen en dit geldt met name voor België. De spreiding naar leeftijd is zeer onevenwichtig: de meeste informanten zijn tussen de 60 en 100 jaar oud. Bovendien is de geografische spreiding onevenwichtig; gaten in het netwerk zijn te vinden in België, in Zuid-Limburg, westelijk Drenthe en in Friesland.<sup>7</sup>

Op basis van deze vragenlijst zijn we in staat het design van de mondelinge interviews en de (in)directe elicitatietesten zo efficiënt mogelijk te maken. De gegevens van de vragenlijst dienen dus als input voor de latere mondelinge veldwerkkonden (zie §5). Alleen de mondeling verkregen gegevens zullen in de atlas verwerkt worden.

Met de schriftelijke vragenlijst krijgen we beter inzicht (i) in de geografische spreiding van de syntactische variabelen, (ii) of de methodiek van de vragenlijst in methodologisch en linguïstisch opzicht geschikt is voor de latere mondelinge elicitatietesten en (iii) waar in het Nederlandstalig gebied zich binnen de vier inhoudelijke thema's (zie §1) interessante syntactische verschijnselen voordoen. Tenslotte levert de schriftelijke vragenlijst wellicht potentiële kandidaten op voor de functie van hulpinterviewer (zie §3.5).

Het design van de schriftelijke vragenlijst houdt zoveel mogelijk rekening met de eerder genoemde testeffecten. Er vindt geen rechtstreekse bevraging plaats, met andere woorden er wordt niet gevraagd of de informanten bepaalde constructies in hun dialect (on)grammaticaal of (on)aanvaardbaar en/of fout/goed vinden. In Figuur 2 is te zien dat aan de informanten wel gevraagd is of bepaalde constructies (i) al dan niet in hun dialect voorkomen en (ii) hoe (on)gebruikelijk de constructies in hun dialect zijn.

Bij deze laatste vraag wordt de informanten gevraagd op een vijfpuntsschaal aan te geven hoe ongebruikelijk/gebruikelijk de zin in hun dialect is. Deze opzet is beter dan een keuzemogelijkheid uit drie opties bijvoorbeeld *komt voor – komt niet voor – weet niet* omdat op deze manier een gedifferentieerder beeld ontstaat van de gemarkeerdheid van de constructie. Bovendien hebben informanten de neiging om de extreme waarden van een schaal niet te gebruiken. In het voorbeeld worden de diverse mogelijke volgordes in een

<sup>7</sup> Tenslotte is het onduidelijk wie in het bestand als dialectspreker en wie als zegspersoon en informant optreedt. Een zegspersoon is iemand die geen vragenlijsten op zijn adres krijgt maar geraadpleegd wordt door een informant om antwoorden op de vragen te geven. Een informant is een persoon die op zijn adres vragenlijsten krijgt toegesonden maar niet noodzakelijk zelf de antwoorden produceert. Hij kan een zegspersoon raadplegen. Bovendien kunnen informanten/ zegspersonen/ dialectsprekers al lang niet meer in het geboortedorp of stad wonen omdat zij inmiddels al (jarenlang) verhuisd zijn maar zij vullen wel nog de vragenlijst voor het dialect van die plek in.

*Figuur 2: Voorbeeld van indirecte bevraging in de SAND-vragenlijst*

(i) Hieronder worden voor één zin verschillende woordvolgordes gegeven. Zou u willen aangeven welke van de onderstaande volgordes van *moet – kunnen – werken* in uw dialect voorkomen? Wilt u dat aangeven door *ja* of *nee* te omcirkelen?

(ii) Als u 'ja' hebt omcirkeld, wilt u dan aangeven hoe gebruikelijk of ongebruikelijk die volgorde is door een cijfer te omcirkelen op de schaal van 1 tot en met 5? Het cijfer 1 betekent dat de volgorde ongebruikelijk is, het cijfer 5 betekent dat de volgorde zeer gebruikelijk is in uw dialect.

	<u>Komt voor</u>	<u>ongebruikelijk &lt;=&gt; gebruikelijk</u>
a. Ik weet dat Jan hard moet kunnen werken.	ja / nee	1 – 2 – 3 – 4 – 5
b. Ik weet dat Jan hard moet werken kunnen.	ja / nee	1 – 2 – 3 – 4 – 5
c. Ik weet dat Jan hard kunnen moet werken.	ja / nee	1 – 2 – 3 – 4 – 5
d. Ik weet dat Jan hard kunnen werken moet.	ja / nee	1 – 2 – 3 – 4 – 5
e. Ik weet dat Jan hard werken kunnen moet.	ja / nee	1 – 2 – 3 – 4 – 5
f. Ik weet dat Jan hard werken moet kunnen.	ja / nee	1 – 2 – 3 – 4 – 5

(iii) Indien geen enkele van de bovenstaande volgordes in uw dialect voorkomt, wilt u dan opschrijven hoe u deze zin in uw dialect zou zeggen? Als er meer manieren zijn om een zin te vertalen wilt u dan alle vertalingen geven?

Vertaling:.....

driedelig werkwoordscluster afgevraagd. Met deze indirecte bevraging voorkomen we dat de informant teveel vanuit een normatief taalbesef antwoorden geeft (zie §2.1). In een dialect-standaardsetting waarin het dialect duidelijk structureel van de standaardtaal verschilt, uit het begrip normativiteit zich trouwens anders dan in een (sub)standaardsetting zoals in het westen van Nederland (zie §2) waar sprake kan zijn van sociaal wenselijke antwoorden of het negeren van gestigmatiseerde varianten. Daar waar het dialect structureel van de standaardtaal verschilt, bestaat de mogelijkheid dat de informanten de onderzoeker laten weten hoe het dialect in 'vroeger tijden' gesproken werd (zie Hinskens 1993:13).

Andere bewerkingstesten in de SAND-vragenlijst zijn invuloefeningen zoals figuur 3 illustreert. In deze opdracht wordt *te*-plaatsing en de vorm van het voegwoord in een infinitiefzin/voorwaardelijke bijzin bevraged. Indien de bevragede zin onacceptabel is, hoeft de informant niets in te vullen:

*Figuur 3: Voorbeeld van een invuloefening in de SAND-vragenlijst*

Komen er in uw dialect één of meer woorden voor op de open plekken in de zin? Zo ja, wilt u dat woord of die woorden dan op de betreffende plek invullen? Als er meer mogelijkheden zijn, wilt u dan de andere mogelijkheden onder iedere zin in uw dialect opschrijven? Soms kan het zo zijn dat de zin goed is in uw dialect zonder iets in te vullen op één of meerdere open plekken. In dat geval laat u de open plek gewoon leeg.

Voorbeeld:

Testzin: Jan heeft de gewoonte \_\_\_\_\_ vroeg \_\_\_\_\_ bellen.

Ingevuld: Jan heeft de gewoonte om \_\_\_\_\_ vroeg te \_\_\_\_\_ bellen.

Heb je genoeg mensen \_\_\_\_\_ de maïs van het land \_\_\_\_\_ halen?

Andere mogelijkheid in uw dialect:

1) .....

2) .....

De meeste vragen betreffen echter reguliere vertaalvragen, zoals figuur 4 laat zien (vertalingen van reflexieven):

*Figuur 4: De vertaalvragen in de SAND vragenlijst*

Wilt u de volgende zinnen vertalen in uw eigen dialect? Als er meer manieren zijn om een zin te vertalen wilt u dan alle vertalingen geven? Denk er bij uw vertaling aan dat uw spelling uw uitspraak zo precies mogelijk moet weergeven.

a. Jan herinnert zich dat verhaal wel.

Vertaling:

b. Marie en Piet zien elkaar voor de kerk.

Vertaling:

c. Toon wast zich.

Vertaling:

d. De timmerman heeft geen spijkers bij zich.

Vertaling:

Sommige vertaalvragen benaderen in de standaardspelling de dialectvorm zoals de possessiefvormen in Figuur 5. De informanten dienen deze zinnen om te zetten in het dialect. In de instructie wordt al expliciet aangegeven dat de input de dialectvariant benadert omdat anders de structurele afstand tussen het dialect en de standaardtaal het omzetten of het vertalen erg lastig maakt:

*Figuur 5: De vertaalvragen waarin de input de dialectvariant benadert*

Wilt u onderstaande zinnen vertalen? Denk eraan dat de manier waarop u de zinnen noteert zo precies mogelijk de uitspraak moet weergeven.

a. Marie haar auto is kapot.

Vertaling:

b. Gisterens krant ligt onder de TV.

Vertaling:

c. Jan is Karolien en Kristien se/hun broertje.

Vertaling:

Opmerkingen:.....

In het extreme geval kunnen de vertaalvragen niet alleen een repetitie-effect oproepen maar kunnen zij ook interferentie of het gebruik van hyperdialectismen bevorderen (zie Hinskens 1993). We houden daar later bij de interpretatie van de antwoorden rekening mee doordat aan exacte 'syntactische' vertalingen zeer geringe evidentie wordt toegekend in verband met het repetitie-effect, maar sterk afwijkende vertalingen leveren daarentegen grote evidentie voor de validiteit van het antwoord (Carden 1976:101).

De vragenlijst is lang geworden en is zeker veel langer dan de vragenlijsten die de Meerstens-informanten gewoonlijk invullen. Het beantwoorden van de vragenlijst kost de taalkundigen in de pilot al twee uur tijd. De vragenlijst is daarom in drie delen opgesplitst en aan de informant is gevraagd om de drie delen op aparte tijdstippen in te vullen. De antwoorden op de vragen zijn gecodeerd en in een databank verwerkt. Om te voorkomen dat we alleen op het eerste deel van de vragenlijst antwoord krijgen, hebben we de drie delen in verschillende volgordes aangeboden. Van de 850 hebben 52 informanten te kennen gegeven dat zij wegens ziekte, ouderdom etc. geen schriftelijke vragenlijst kunnen terugsturen. 368 informanten hebben de vragenlijst (geheel of gedeeltelijk) ingevuld en teruggestuurd:

<i>Figuur 6: de non-respons op de schriftelijke vragenlijst</i>	
totaal aantal schriftelijke vragenlijsten uitgestuurd	850
minus non-respons (te oud, ziek, overlijden, etc)	798 (100%)
teruggekomen vragenlijsten	368 (47%)

De SAND-vragenlijst betreft een compromis tussen de theoretische doelstelling van het project en de eisen die aan de validiteit van het verzamelen van dialectdata gesteld moeten worden. Het is van belang om op te merken dat de vragenlijst inhoudelijk zeer vernieuwend is. Het is de eerste keer dat er in het Nederlandstalig gebied een syntactische vragenlijst gebaseerd op theoretische generatieve inzichten uitgestuurd is.

#### 4.2 De mondelinge (in)directe elicitatietesten in het SAND-project

De data afkomstig uit de mondelinge interviews en elicitatietesten vormen de directe input voor de syntactische atlas. Het grote voordeel van mondelinge versus schriftelijke bevraging is vooral dat het dialect gesproken kan worden, waardoor de dialectsyntax 'getriggerd' wordt. Bovendien is een grote controle op de kwaliteit van de verzamelde gegevens mogelijk omdat (i) het taalgebruik en de reacties van de informant ter plekke geobserveerd kan worden; (ii) bij twijfel aan de antwoorden van de informant doorgenvraagd kan worden en (iii) de interviewer kan controleren of de informant de vragen goed begrijpt en het 'juiste' antwoord geeft.

Indirecte elicitatietesten wil zeggen dat de testen geen natuurlijk taalgedrag van de informant oproepen en dat het uitvoeren van de testen niet zozeer een syntactisch maar een algemeen taalvermogen van de informanten vraagt. Een belangrijke eis bij de keuze van de testen moet zijn dat alle informanten en hulpinterviewers met eenvoudige instructies in staat moeten zijn om de testen goed uit te voeren. In Cornips (1994) is uitgebreid geëxperimenteerd met mondelinge indirecte elicitatietesten.

In de mondelinge veldwerkronden zullen we voornamelijk 'komt-voor'- en 'plaatjes'-vragen aan de informanten aanbieden (Greenbaum 1973, Van Bree 1981, Gerritsen 1991, Cornips 1994).

## ● 5 Veldwerkronden

In deze paragraaf wordt een voorlopig overzicht en fasering gepresenteerd van de verschillende schriftelijke en mondelinge veldwerkronden.

### 1 schriftelijke vragenlijst

Zomer 2000 is de schriftelijke (moeder)vragenlijst naar zo'n 850 Meertens-informanten opgestuurd met het verzoek om de lijst vóór 1 oktober 2000 te retourneren aan het Meertens Instituut. Ingevulde vragenlijsten zijn verwerkt in een elektronische databank. Op

basis van deze databank is een vragenlijst voor de mondelinge interviews gemaakt die (i) geografisch getrap is, zodat bijvoorbeeld reflexieve middel-constructies alleen in de provincies Limburg afgevraagd worden en daar waar zij ongrammaticaal zijn niet én die (2) implicationeel is in de zin dat als een bepaalde constructie niet grammaticaal is in een bepaald dialect, ook andere gerelateerde constructies niet bevroegd zullen worden.

### 2 *1e mondelinge interview*

Mei 2001 wordt er begonnen met de eerste mondelinge interviews. In de eerste ronde krijgen alle 250 hulpinterviewers in een vraaggesprek de geografisch getrapte en implicatieve vragenlijst voorgelegd. Zij spreken de constructies in het dialect in de DAT/minidisc recorder in. In de tweede ronde krijgen de 250 informanten deze opgenomen constructies te horen en worden zij verder ‘ondervraagd’ door de hulpinterviewer.

### 3 *schriftelijke of telefonische follow-up-studie*

Tijdens de analyses van het mondeling verzamelde materiaal kan blijken dat een gerichte schriftelijke of telefonische follow-up-studie wenselijk is. Het is in individuele gevallen een gemakkelijke manier om uitsluitsel te krijgen over een bepaalde kwestie. Het zal lang niet voor alle 250 meetpunten nodig zijn om aanvullende gegevens op te vragen.

### 4 *Eventueel 2e mondelinge interview*

Afhankelijk van de individuele wetenschappelijke interesses van de SAND-promovendi zullen misschien nog enkele informanten benaderd worden in een zeer gespecialiseerde schriftelijke of mondelinge ronde van dataverzameling. Deze data zullen echter niet gebruikt worden voor de atlas, maar dienen uitsluitend ter ondersteuning van de proefschriften.

## 6 Ten besluit

Uniek aan dit project is de samenwerking tussen dialectologen, sociolinguïsten en theoretisch syntactici. Zoals in de inleiding reeds benadrukt werd, zal het eindproduct van dit project een meerwaarde hebben ten opzichte van de meer traditioneel dialectologische atlassen omdat de Syntactische Atlas van de Nederlandse Dialecten mede gebaseerd is op theoretische generatieve inzichten. In de atlas wordt op systematische wijze variatie onderzocht aan de hand van de volgende thema's: woordvolgordevariatie in de rechterperiferie van de zin, woordvolgordevariatie in de linkerperiferie van de zin, variatie in pronominale verwijzing en variatie in negatie en kwantificatie.

Het SAND-project is bovendien vernieuwend omdat voor het eerst een classificatie gemaakt wordt van de Nederlandse dialecten op basis van syntactische in plaats van fonologische variatie. Lange tijd is syntactische variatie in de Nederlandstalige dialecten om uiteenlopende redenen onderbelicht gebleven, maar het SAND-project zal de syntactische variatie in het Nederlandstalige gebied letterlijk en figuurlijk op de kaart zetten.

## 7 Bibliografie

- Barbiers, S. (2000).** ms. *Variation in the Right-Periphery: Questions for the Syntactic Atlas of the Netherlands Dialects*. Meertens Instituut.
- Bennis, H. & J. Van der Auwera (1999).** *Een Syntactische Atlas van de Nederlandse Dialecten (SAND)* VNC-Aanvraag nr 205-41-105.
- Boves T. & M. Gerritsen (1995).** *Inleiding in de sociolinguïstiek*. Utrecht: Het Spectrum.
- Bree, C. van (1981).** *Hebben-constructies en datiefconstructies binnen het Nederlandse taalgebied: een taalgeografisch onderzoek*. Leiden: Dissertatie.
- Brouwer, D. (1989).** *Gender variation in Dutch*. Dordrecht: Foris Publications.
- Carden, G. (1976).** Syntactic and semantic data: replication results. *Language in Society* 5: 99-104.
- Cornips, L. (1991).** Syntactische variatie in Heerlen: de methodologie. In: Hout, R. van & E. Huls (eds.), *Artikelen van de Eerste Sociolinguïstische Conferentie*. Delft: Eburon, 89-103.
- Cornips, L. (1994).** *De syntactische variatie in het Algemeen Nederlands van Heerlen*. Amsterdam: Dissertatie, IFOTT 6.
- Cornips, L. (1997).** De betrouwbaarheid van de schriftelijke enquête Willems (1885): de adjunct middel in de Limburgse dialecten. *Gramma/TTT* 5.2, 61-76.
- Cornips, L. (2000).** Variatie in reflexieve middelen in het Heerlense dialect en in de Rijnlandse dialecten. In: Besten, H. den, E. Elffers & J. Luif (eds.), *Syntaxis van het Nederlands. Feestbundel voor Wim Klooster*.
- Cornips, L. & W. Jongenburger (1999).** Het TCULT-project. In: Huls, E. & B. Weltens (eds.), *Artikelen van de Derde Sociolinguïstische Conferentie*. Delft: Eburon, 99-110.
- Daan, J. (1999).** Dialects. In: Preston, E.R. (ed.), *Handbook of perceptual dialectology*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 9-30.
- Daan, J. & D. Blok (1977).** Van Randstad tot landrand. Toelichting bij de kaart: Dialecten en Naamkunde. Bijdragen en Mededelingen der Dialectencommissie XXXVII. Amsterdam.
- Gerritsen, M. (ed.) (1991).** *Atlas van de Nederlandse dialectsyntaxis (AND)*. Amsterdam: P.J. Meertens-Instituut voor Dialectologie, Volkskunde en Naamkunde.
- Giles, H. & P. Powesland (1975).** *Speech Style and Social Evaluation*. Cambridge: Cambridge University Press.
- Goeman, A. & J. Taeldeman (1996).** Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval* 48-1, 38-59.
- Goosens, J. (1977).** *Inleiding tot de Nederlandse Dialectologie*. Wolters-Noordhoff, Groningen.
- Greenbaum, S. (1973).** Informant elicitation of data on syntactic variation. *Lingua* 31, 201-212.
- Hinskens, F. (1993).** *Dialect levelling in Limburg. Structural and sociolinguistic aspects*. Dissertatie Nijmegen.
- Hout, R. van (1989).** *De structuur van taalvariatie*. Dordrecht: Foris.

- Jongenburger, W. & J. Aarssen (te verschijnen).** Linguistic and cultural diversity in a multi-ethnic neighbourhood. A survey study. *Journal of Multilingual and Multicultural Development*.
- Jongenburger, W. & R. Vousten (2000).** Dialect use in the Netherlands: certified millennium proof. Lezing gehouden op Sociolinguistics Symposium 2000, Bristol.
- Kessler, B. (1995).** Computational Dialectology in Irish Gaelic. Proceedings of the European Association for Computational Linguistics, EACL, Dublin, 60-67.
- Labov, W. (1972).** *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Nerbonne, J. & W. Heeringa (1998).** Computatieve vergelijking en classificatie van dialecten. *Taal en Tongval* 50, 2, 164-193.
- Oxford, R.L. (1982).** Technical issues in designing and conducting research on language skill attrition. In: Lambert, R. & B.F. Freed (eds.), *The loss of language skills*. Massachusetts: Rowley, 119-137.
- Schatz, H. (1986).** Plat Amsterdams in its social context: A sociolinguistic study of the dialect of Amsterdam. Amsterdam: Publicaties van het P.J. Meertens Instituut.
- Taaldeman, J. & G. Verleyen (1999).** De FAND: een kind van zijn tijd. *Taal en Tongval* 51-2, 217-240.
- Velde, H. Van de (1996).** *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. Dissertatie, KUN.
- Weijnen, A.A. (1966).** *Nederlandse dialectkunde*. Assen.
- Wijngaerd, G. Vanden (2000)** ms. *Vragen aan de Linkerperiferie*, K.U.Brussel, Mei 2000.



# Forum

## Taalkunde en Taalbeheersing: What's in a name?

BREGJE HOLLEMAN  
JOOST SCHILPEROORD

In de Forumrubriek van dit tijdschrift (nr. 6.2) pleit Henk Verkuyl (onder de titel 'Taalkunde en taalbeheersing: hoe verder?') voor een herbezinning op de relatie tussen taalkunde en taalbeheersing. In zijn visie zijn de twee disciplines ooit gescheiden om historische redenen, maar zijn ze (nog steeds) onderdeel van 'de' taalwetenschap. De theoretische taalkunde neemt onder druk van de taalpsychologie inmiddels de relatie tussen taal en andere cognitieve vermogens serieuzer, terwijl de taalbeheersing allang niet meer gaat over 'het beheersen van taal'. De taalbeheersing bestudeert de relatie tussen taal en cognitie met een focus op teksten, terwijl de taalkunde dat onder andere doet voor kleinere teksteenheden: de syntactici met een nadruk op de zin als eenheid en de semantici met het woord als uitgangspunt. Dit zo zijnde, stelt Verkuyl een disciplinaire fusie voor, waarbij de werkverdeling zo is dat de semantici zich bezighouden met woorden, de syntactici met zinnen en de taalbeheersers met teksten.

Behalve deze inhoudelijke redenen, heeft Verkuyl ook een meer pragmatisch argument voor een herverkaveling: de term 'taalbeheersing' doet niet alleen geen recht meer aan wat taalbeheersers feitelijk doen, maar taalbeheersers maken zich op internationale congressen ook nog eens belachelijk door te moeten zeggen dat ze zich bezighouden met 'language mastering'. Waarom noemen zij zich dan toch geen taalkundige, of linguïst? Kripke, aldus Verkuyl: taalbeheersers willen liever niet met taalkunde 'in engere zin' geassocieerd worden.

Vooropgesteld: een integratie van beide discipli-

nes is wat ons betreft een nastrevenswaardige zaak. Sterker nog, we hebben de indruk dat de werkverdeling zoals Verkuyl die voorstelt, thematisch beschouwd al(lang) realiteit is. Om deze reden ook is taalbeheersingsonderzoek, althans in Leiden en Utrecht, ondergebracht in een onderzoeksinstituut voor taalwetenschap. De dwarsverbanden in de organisatie van het onderzoek, zijn echter nog niet overtuigend zichtbaar in feitelijke samenwerking, noch in een integratie van *onderwijsprogramma's*. Hoe komt dat?

Verkuyl houdt het erop dat emotionele en historische oorzaken de belemmering vormen voor een verdere (onderwijs)integratie van taalkunde en taalbeheersing. Wij zullen in dit stukje echter uiteenzetten dat die belemmeringen naar ons oordeel elders gezocht moeten worden. Het is niet zozeer de thematische verkaveling die op (emotionele of historisch bepaalde) problemen zal stuiten, maar een nogal fundamenteel verschil in benadering en methodologie binnen de disciplines dat synthese van afdelingen in de weg staat. Wij willen dat demonstreren door het hypothetische onderwijsprogramma wat concreter te maken aan de hand van een aantal voorbeelden. Die voorbeelden dienen vooral om zichtbaar te maken hoe taalbeheersers het onderdeel 'tekst' daarin ingevuld zouden willen zien, en welke discrepanties dat oplevert in de benadering van 'teksten' enerzijds, en die van 'zinnen' en 'woorden' anderzijds.

Alvorens dat te doen, willen we kort ingaan op de terminologische kwestie die Verkuyl opwerpt: het gebruik van de term 'taalbeheersing', of

‘communicatiekunde’ kan namelijk wel degelijk inhoudelijk worden gemotiveerd, en heeft meer dan (alleen) emotionele redenen.

*‘Language experts’ versus ‘language masters’*

Zelfs wanneer taalbeheersers hun vak in het buitenland als ‘language mastering’ aan zouden duiden, lijkt dat ons nog geen grond het vak maar met ‘de’ taalkunde samen te laten gaan. Maar taalbeheersers noemen hun vak niet zo. Zij noemen zich ‘text linguist’, of ‘cognitive linguist’. Zij houden zich bezig met ‘speech communication’, met ‘text interpretation’, met ‘communication studies’ of met ‘document design’. Kortom, in de (internationale) taalwetenschap allemaal gangbare aanduidingen die zicht geven op de taalwetenschappelijke en dikwijls multidisciplinaire terreinen die de taalbeheerser tot de zijne rekent. In het Nederlands gaat de voorkeur overigens vaak uit naar de term ‘communicatiekunde’ om zo de preoccupatie te benadrukken die taalbeheersers hebben met taalgebruik, met taal als communicatiemiddel, en dus met taal, tekstelementen en teksten in natuurlijke gebruikcontexten. Maar goed, laten we van dit ‘word play’ geen ‘sword play’ maken, en ons verder richten op de inhoudelijke kant van Verkuyls voorstel. Het probleem daarvan is dat de voorgestelde werkverdeling niet het juiste vertrekpunt is om samenwerking in onderwijs te realiseren. Dit zullen we laten zien door een concreet beeld te geven van een mogelijk gezamenlijk onderwijsprogramma.

***Woorden, zinnen, teksten: compositionele betekenisrelaties?***

De kern van zo’n gezamenlijke programma Nederlandse Taalkunde zou volgens Verkuyl gevormd worden door ‘woorden, zinnen, teksten’. Vermoedelijk zou dat er inhoudelijk ongeveer als volgt uitzien. De student die het programma gaat volgen zou bij ‘woorden’ leren dat die definieerbare betekenissen hebben. Die betekenissen kunnen worden geformaliseerd volgens een of andere symbolische notatie waarmee een relatie gelegd wordt met het cognitief systeem

van de moedertaalspreker. Als de formalisering steeds goed toepasbaar blijkt op soortgelijke woorden, of juist met systematische aanpassingen op andersoortige woorden, dan, zo is het idee, zou daarmee een onderdeel van dat cognitieve systeem weergegeven of blootgelegd kunnen worden.

Over zinnen leert de student dat er verbanden tussen woorden of woordgroepen zijn die voortvloeien uit een universele mentale grammatica. Woorden blijken niet alleen te kunnen worden beschreven in termen van betekenissen, maar ook in termen van de syntactische posities of eigenschappen die ze kunnen in- of aannemen. De betekenis van een zin is dus niet alleen een kwestie van de betekenissen van de woorden die er deel van uitmaken, maar is ook compositioneel bepaald. Het thema ‘woorden’ kan dus na het thema ‘zinnen’ worden uitgebreid: semantiek is niet alleen een beschrijving van woordbetekenis, maar bevat ook per woord een beschrijving of formalisering van de syntactische eigenschappen. Ook de beschrijving van zinnen gaat door middel van een formele notatie, waarbij de aanname weer is dat indien deze goed toepasbaar blijkt op een gedefinieerde set van zinnen, daarmee dan een stuk cognitie is beschreven of blootgelegd.

Over teksten, ten slotte, leert de student dat er ook weer relaties tussen zinnen kunnen bestaan, en dat de betekenis (of ‘functie’) van een zin voor een deel bepaald wordt door de tekstuele context ervan: de discoursstructuur en de opgebouwde mentale representaties van de tekst zoals die tot dusver verwerkt is. Ook deze tekstuele context kan weer worden geformaliseerd in termen van geïntroduceerde referenten en relaties daartussen.

Het bovenstaande lijkt ons een plausibele uitwerking van hetgeen Verkuyl zich voorstelt bij een gezamenlijk onderwijsprogramma van taalkunde en taalbeheersing. Het zou, gegeven de gangbare praktijk van onderzoek en onderwijs in de taalkunde dan ook een mooie cursusreeks ‘Nederlandse taalkunde’ vormen. Bezien we het pro-

gramma echter vanuit de gangbare praktijk van onderzoek en onderwijs binnen de taalbeheersing, dan is het hier geschetste onderwijsprogramma niet erg adequaat. Taalbeheersers zien *feitelijk taalgebruik* als het centrale object van hun vakgebied. De aandacht gaat dus niet uit naar wat er allemaal mogelijk is in taal, maar naar wat mensen feitelijk *doën* met taal – omdat dat inzicht biedt in wat het betekent om een ‘talig wezen’ te zijn. Juist dat taalgebruiksaspect speelt in bovenstaande taakverdeling geen enkele rol. De concentratie op taalgebruik heeft verstrekken de implicaties voor de vraagstellingen en de methodologie binnen de taalbeheersing, en daardoor is, zo vrezen we, het verband tussen woorden, zinnen en teksten veel minder ‘clear cut’ als hierboven geschetst. We lichten dit toe met behulp van een aantal voorbeelden.

*Woorden, zinnen, teksten en hun context*

Een centrale aanname in de taalbeheersing is dat de context een cruciale rol speelt in de interpretatie van reëel taalgebruik. Waarom, bijvoorbeeld, is het zo dat zoveel, ook hoog-opgeleide taalgebruikers, moeite hebben met het correct invullen van een formulier? De woorden zijn duidelijk, de zinnen zijn duidelijk, en toch functioneert de tekst niet goed. Dit type vraag wordt gesteld door taalbeheersers die gericht zijn op de *optimalisering* van teksten. Zo’n vraag kan niet worden beantwoord vanuit de semantiek of de syntaxis, maar alleen op basis van combinaties van theorieën die iets zeggen over de vraag hoe het functioneren van een tekst begrepen kan worden vanuit kenmerken van taalgedrag enerzijds, en vanuit de context en het doel van de tekst anderzijds. Ander voorbeeld. Hoe komt het dat een reclame wordt begrepen als een aansporing om iets te kopen, zonder dat dit expliciet in de tekst staat? In heel veel alledaags taalgebruik tellen betekenissen van woorden en zinnen blijkbaar niet compositioneel en lineair op tot de betekenis van een tekst.

Deze voorbeelden maken duidelijk dat noties als ‘sociale context’ en ‘genrekennis’ nodig zijn

om te beschrijven en verklaren wat taalgebruikers ‘doen’ met een tekst. Een bijzondere vorm van invloed vanuit de ‘context’ op het verwerken en begrijpen van tekst, bestaat hem daarin dat die context in sommige gevallen ook (kennis van) de taal zelf kan zijn. We illustreren dat aan de hand van de volgende voorbeeldzin, afkomstig uit het elektronische Volkskrantcorpus, ons door Arie Verhagen onder de aandacht gebracht (waarvoor dank!).

Het gedicht dat je greep en dat gelezen werd en herlezen, en na verloop van tijd van buiten gekend, het baande zich *een klankpad* in je hoofd [...]

Veel taalkundigen gaan uit van de premisse dat de informatie die wij uit een zin halen, ook allemaal in die zin besloten ligt. Als we een zin dus correct ‘begrijpen’, dan hebben we er de semantische informatie ‘uit’ gedistilleerd die er ook ‘in’ zit. Daarmee is het verwerken en begrijpen van een zin dus een context-onafhankelijk proces, en een formele theorie die ons begrip ervan verantwoordt doet dan ook geen beroep op contextuele kennisbronnen (behalve in geval van pronomina, maar daar komen we straks nog over te spreken).

Maar nu de voorbeeldzin. We stellen allereerst vast dat het voor ons relevante deel van deze zin is gebouwd op het ‘zich een weg banen’-stramien. Die constructie komt in het Nederlands veel voor: ‘Jan baande zich een weg door de menigte’ begrijpen we moeiteloos. Toch is er met die zin iets raars aan de hand. Allereerst wijst corpusonderzoek uit, dat het werkwoord ‘banen’ eigenlijk alleen maar voorkomt in combinatie met ‘weg’, of een prototypische variant ervan (‘pad’, bijvoorbeeld). Maar wat betekent ‘banen’ dan? Eigenlijk kun je dat, behoudens (zeldzame) gevallen met ‘letterlijke’ betekenis, zoals in ‘Hij baande (‘vereffenen’) een weg door het koren’, niet goed zeggen. Wat de constructie als geheel aan betekenis uitdrukt, is volgens Constructiegrammatica een bepaald *traject* dat het subject aflegt langs een *pad* dat wordt uitgedrukt door de

voorzetselconstituent. Dat we zinnen die op dit stramien gebouwd zijn zo begrijpen, kun je niet zo goed verantwoorden op basis van de lexicale elementen 'weg' dan wel 'banen'. Met andere woorden, het betekenisaspect is geassocieerd met de constructie *als geheel*; de zin drukt een 'event' van (het veroorzaken van) een beweging uit. Hoewel het werkwoord 'banen' in geval van conventioneel gebruik van de constructie daar nog wel een indicatie voor geeft (Hij baande (zich) een weg door het koren), is dat betekenisaspect ook aanwezig wanneer het karakter van 'banen' veel minder duidelijk is (Hij baande zich een weg naar de top van het management). In veel gevallen heeft overigens ook 'weg' niet de gecodeerde betekenis [entiteit WEG]. Sterker nog, er blijkt zelfs in veel zinnen die op het stramien gebouwd zijn geen relevante argumentrelatie (meer) te bestaan tussen werkwoord en direct object; je kunt er wel een boom boven tekenen, maar die vertelt je weinig over hoe je die zin begrijpt.

Inspecteer je nu Nederlandse of Engelse corpora, dan blijkt dat er op dat 'banen' duchtig gevarieerd kan worden, ook met werkwoorden die conventioneel niet het idee van 'beweging' oproepen. Je vindt voorbeelden als 'zich een weg naar de top van het bedrijf *slijm*', of 'She *knitted* her way across the Atlantic'. Zulke extensies (want dat zijn het) van 'banen' concretiseren de manier waarop het causale proces voltrokken wordt, daar waar 'banen' dat in het vage laat. Maar we begrijpen die concretisering alleen maar onder cruciale referentie aan de gekende 'weg-banen'-constructie (in de Constructiegrammatica heet het dat de extensies de cruciale eigenschappen van de basisconstructie *beërven*), en er is dus meer aan de orde dan alleen een aardig opgeschreven stilistische variant.

De voorbeeldzin is extra bijzonder omdat de variatie niet het werkwoord 'banen' betreft, maar juist het constructie-onderdeel 'weg', waarmee er dan ineens 'klankpad' opduikt. Maar wat is een 'klankpad' dan? Buiten de context van deze specifieke zin niets, volgens ons; een nonsensicaal woord. Echter, *binnen* deze zin kunnen we er wel

degelijk chocola van maken, en wel op grond van een subtiel samenspel van informatie binnen en buiten de zin. 'Buiten', omdat we weten dat de zin een extensie is van de 'weg-banen'-constructie die we al kennen en wel zo dat hier het element 'weg' geconcretiseerd wordt: een klankpad 'is' een soort weg zou je kunnen zeggen. 'Binnen' omdat 'klankpad' in het kader van het gedicht waarover de zin ook rept, wel begrijpelijk is. De betekenis zou, lijkt ons, veel lastiger te achterhalen zijn geweest als er in plaats van 'klankpad' iets als 'klankspel' had gestaan, of wanneer we het woord 'klankpad' in een andere context zouden aantreffen.

### *Hoe formaliseren; en wat?*

De voorbeelden maken onder andere duidelijk dat een nevenschikking van vraagstellingen vereist is: het is niet nodig of wenselijk eerst alles van het woord of de zin te weten voordat de stap naar grotere teksteenheden of naar taal-in-gebruik gemaakt mag worden. Daarnaast is cruciaal voor het welslagen van Verkuyls herverkavelingsvoorstel dat er een afstemming tussen de onderzoeksmethodologieën van de taalbeheersing en de taalkunde plaatsvindt: de verschillen in opvattingen, benaderingen en definities van centrale concepten moeten worden geslecht. Van groot belang zal daarbij vooral de vraag zijn of we kunnen komen tot een principiële nevenschikking van hetgeen binnen beide disciplines onder 'formaliseren' verstaan wordt. De besproken voorbeelden maken duidelijk dat een strakke, statische formalisering veel van wat er in en met taal gebeurt niet aan het licht kan brengen; niet alle talige fenomenen blijken met een zelfde wijze van formaliseren te kunnen worden beschreven. De studie van taalgebruik impliceert dat aan de hand van analyses van tekstcorpora (zie het 'weg'-voorbeeld), of aan de hand van concrete taalgebruikssituaties (zie het formuleren-voorbeeld) wordt beschreven 'wat mensen kennelijk doen met taal'. Omdat de resulterende generalisaties en theorieën niet (altijd) worden weergegeven met behulp van formalismen die ontleend

zijn aan de logica, zijn taalkundigen, zo is onze ervaring, nogal eens geneigd te denken dat taalbeheersers een broertje dood hebben aan formaliseren. Dat is onjuist. Eén van de centrale methoden van onderzoek in de taalbeheersing is dat hypothesen over de (talige) werkelijkheid volgens de experimentele methodologie worden getoetst aan de hand van reële taalgebruik. Zo'n statistische toetsing is alleen mogelijk op basis van een formalisering van de onderzoeksvraag, en dus van (een deel van) de theorie. Een statistische regressievergelijking – om maar eens zo'n formalisering te noemen – is weliswaar een ándere manier van formaliseren dan die gangbaar zijn in de taalkunde, maar eentje die minstens zo sterk is geworteld in wetenschappelijke onderzoekstradities als formalismen die ontleend worden aan de logica. We besluiten dit betoog met het bespreken van een voorbeeld dat dit kan verhelderen.

Bij het woord 'discourse' of 'tekst' denken taalkundigen, zo is onze ervaring, al snel aan de Discourse Representatie Theorie (Kamp, 1984). Zo ook Verkuyl. DRT is een model-semantische theorie van referentialiteit en coreferentialiteit. Hoe worden in teksten referenten geïntroduceerd?; hoe wordt daar naar terugverwezen?; en hoe wordt erover geprediceerd? De theorie veronderstelt een 'stack' waarin alle predicaties over referenten worden weggeschreven. Zoals gezegd: taalbeheersers zijn zeer geïnteresseerd in teksten, en daarmee ook in de wijze waarop taalgebruikers een tekst verwerken. Toch denken zij bij het woord 'discourse' of 'tekst' niet meteen aan DRT. Daar is een reden voor. De vraag is namelijk of DRT voor het onderzoeken van tekstverwerkingsprocessen een adequaat instrumentarium biedt. Om vraagstellingen rondom tekstverwerking te beantwoorden, ontcom je niet aan een bezoek aan de meer psychologisch georiënteerde afdelingen van de bibliotheek. Daar wordt duidelijk dat over processen van tekstverwerking en interpretatie reeds het nodige bekend is. Een van de merkwaardigste zaken is dat het leesproces, het verwerken dus van een tekst, in hoge mate selectief is. Dat betreft zowel het lezen zelf, als datgene wat lezers uiteindelijk van de tekst ont-

houden. Vraag je mensen bijvoorbeeld 'Met hoeveel dieren van elke soort ging Mozes de Ark in?', dan antwoorden zij in de regel 'twee', zonder op te merken dat de vraag zo gesteld niet beantwoordbaar is. Zo'n fenomeen kan verklaard worden vanuit de selectieve aandacht tijdens het lezen. De hypothese luidt dan: lezers besteden (relatief) meer aandacht aan de 'nieuwe informatie' ('met hoeveel dieren van elke soort'), en niet aan de 'gegeven' informatie ('ging Mozes de Ark in'). (Voor de minder bijbelvasten onder ons: het was Noach, en niet Mozes, die de ark inging.) Of zo'n hypothese stand houdt kan alleen volgens de experimentele methode worden onderzocht, en de vraag zelf kan dus alleen volgens deze wijze worden geformaliseerd.

Iets dergelijks geldt voor het leesproduct. Selectiviteit van leesprocessen speelt een rol bij wat mensen van teksten *onthouden*. Lezers 'kopiëren' de informatie in een tekst niet in hun geheugen, zoals een scanner dat doet, maar ze bouwen ervan een interne, mentale representatie op, waar, om te beginnen lang niet alles in 'zit' wat in de tekst staat, maar juist ook veel meer dan wat in de tekst staat. Selectie en inferentie zijn dus procesbepalende factoren van het lezen. Hoe komt dat? Daar zijn nogal wat redenen voor te bedenken (interesse en voorkennis van de lezer bijvoorbeeld), maar een belangrijk aspect is de beperkte verwerkings- en opslagcapaciteit van het menselijk werkgeheugen. Het is juist deze factor die binnen een DRT-kader niet de plaats kan krijgen die het toekomt. Kintsch' *construction-integration*-model (Kintsch, 1998) is een formalisering van de processen die bepalen wat we nu precies wel en niet uit teksten onthouden. Het doet toetsbare, en inmiddels getoetste en houdbaar bevonden uitspraken.

Kortom, de keuze voor methoden en formalisering wordt bepaald door wat je te weten wilt komen van teksten en taalgebruik. B geen van de methoden kan op voorhand superieur aan de andere worden genoemd. Met andere woorden, het is niet het geval dat een DRT-geïnspireerde onderzoeker een 'slechte' vraagstelling heeft, en de Kintsch-aanhanger een 'goede', of vice versa.

Maar omdat de taalbeheerser uitgaat van de beperkingen die het menselijke werkgeheugen oplegt aan het proces van verwerking, en dus vragen stelt die beter bij het Kintsch-model passen, rijst het probleem dat de uitkomsten van hun onderzoek minder makkelijk te koppelen zijn aan die van DRT-geïnspireerden, laat staan aan de analyses op woord- en zinsniveau van een taalkundige – om de eenvoudige reden dat een DRT-model geheel geen werkgeheugen veronderstelt.

**Taalkunde en taalbeheersing: zo (steeds) verder**

Terug naar onze student. Die krijgt enerzijds te maken met de logische formalisering die met name toegepast worden bij de beschrijving van woorden en zinnen, en anderzijds met statistische of analytische formalisering die worden toegepast bij de beschrijving van taalgedrag. Beiden claimen weliswaar in het algemeen een onderdeel van ‘de’ cognitieve realiteit te weer spiegelen, maar de invulling van het begrip ‘cognitief’ is verschillend – alleen al omdat de ene benadering de nadruk legt op structuren (het potentiële), en de andere op processen (het reële). Met enige goede wil zal voor ervaren onderzoekers het ‘diepere’ verband tussen beide benaderingen nog wel te zien zijn, maar of de student nog enige samenhang in dit alles ontwaart, is zeer de vraag. Tussen de benaderingen, de methodologieën en de opvattingen van wat ‘kennis’ is in de taalkunde en de taalbeheersing bestaan (nog steeds) zulke grote fricties, dat onze student uiteindelijk tamelijk schizofreen zijn of haar bul in ontvangst zou komen nemen.

Kortom, het al dan niet slagen van Verkuyls voorstel hangt niet op de naamgeving van de disciplines. Taalbeheersers voelen zich inhoudelijk op hun plaats in een onderzoeksinstituut voor taalkunde. Maar voordat een organisatorische samenvoeging plaats kan vinden in het onderwijs, en de term ‘taalbeheersing’ afgeschaft kan worden, is meer nodig dan de verkaveling van

thema's die Verkuyl voorstelt. Er dient een principiële nevenschikking plaats te vinden van onderzoeksvragen, benaderingen, methoden en formaliseringswijzen zoals die in beide disciplines gehanteerd worden. Concreet betekent dit dat vragen over teksten en tekstverwerking niet pas gesteld mogen worden als die over woorden en zinnen zijn beantwoord, dat vragen over de psychologische processen bij tekstverwerking en – productie niet pas onderzocht kunnen gaan worden als de DRT-ers hun theorie afhebben; dat een statistische formalisering of een formalisering ‘in woorden’ evenwaardig is aan een logisch-formele, en dat een statisch-compartmentele en een dynamische-integratieve benadering van het menselijke cognitieve systeem even noodzakelijk zijn om meer te weten te komen over dat systeem *en de werking ervan*.

Laat dus de beide disciplines toch maar (naast elkaar) bestaan – maar laat de beroepsmatige beoefenaars van één van die disciplines met elkaar in gesprek raken en blijven om de verbanden te zoeken en te verhelderen. En of een taalkundige zich dan op een buitenlands congres als ‘language expert’ wenst te introduceren, is aan haar of hem.

**Literatuur**

- Kamp, H. (1984)**, A theory of truth and semantic representation. In: J.A.G. Groenendijk, T.M.V. Janssen & M.J.B. Stokhof (eds.), *Truth, interpretation and information*. Dordrecht: Foris.
- Kintsch, W. (1998)**, *Comprehension; A paradigm for cognition*. Cambridge MA: Cambridge University Press

*Bregje Holleman, UU, Taalbeheersing,  
Trans 10, 3512 JK Utrecht  
Joost Schilperoord, KUB, Tekstwetenschap,  
Postbus 90153, 5000 LE Tilburg*

# Boekbeoordelingen

**Henk Verkuyl, *Semantiek. Het verband tussen taal en werkelijkheid*.** Amsterdam: Amsterdam University Press, 2000. ISBN 90 5356 4152 fl 19,90.

Op Henk Verkuyl is de keuze gevallen om, in het kader van de serie *Weten*, het algemene publiek vertrouwd te maken met de stand van zaken binnen de semantiek. Verwonderlijk is deze keuze geenszins. Verkuyl is mede-auteur van twee semantische publicaties: het ambitieuze tweedelige *Gamut* (1982) en het bescheiden *De Jong, Oversteegen en Verkuyl* (1990). En hij gaat ook schuil achter *Verschuyf* (1988), een boek dat verschillende semantische onderwerpen ter sprake brengt. Verkuyls belangstelling voor semantiek is ook waarneembaar in zijn wetenschappelijk werk. Boeken, artikelen, boekbesprekingen en proefschrift getuigen van zijn semantische belangstelling en kunde. Hij geldt terecht als een van de vooraanstaande semantici met een solide internationale reputatie waar Nederland zo rijk aan is.

Verkuyl concentreert zich in dit boek op enkele semantische verschijnselen die zich op woordniveau afspelen. De bespreking van deze verschijnselen beslaat vier hoofdstukken. Deze worden gevolgd door een nawoord met literatuurverwijzingen en dankzeggingen. De uitspraken over de toekomst van het vak die de redactie belooft als een vast onderdeel van de boeken in de serie, ontbreken jammer genoeg.

De kern van dit boek wordt gevormd door de eerste twee hoofdstukken. Daarin ontwikkelt de auteur een semantisch model voor kennisrepresentatie. Verkuyl maakt van dit model gebruik om een uiteenzetting te geven van de semantische strekking en de verschillen tussen

de termen *stereotype* en *prototype*. De bladzijden die hij aan dit onderwerp wijdt, zijn een welkom geschenk. Zelfs een doorgewinterde schrijver laat zich wel eens verleiden tot het misleidend identificeren van deze noties (vergelijk Lyons 1995:96). Dit deel van het boek bevat ook een bespreking van metaforen. Verkuyl gaat uit van enkelvoudige zinnen die de semantische structuur  $f(A, B)$  vertonen. De zin *een larius danst* bijvoorbeeld, correspondeert met de semantische structuur EEN(LARIKS, DANSEN). De eerste term, de restrictie, verwijst naar de verzameling lariksen in de werkelijkheid. De tweede term, het bereik, verwijst naar de verzameling objecten die dansen. De doorsnede van deze verzamelingen is leeg. Maar, stelt Verkuyl vast, in feite treedt hier een andere interpretatieprocedure in werking. Om deze zin als een zinvolle mededeling te kunnen opvatten, is het voldoende dat we de verwijzing van het bereik aanpassen aan de verwijzing van de restrictie. Op deze manier ontstaat een afwijkende lezing: het bewegen van de takken van de boom wordt opgevat als dansen. Deze voorbeeldzin vraagt om een aanpassing van de referentie van het bereik. In andere gevallen zal het nodig zijn de restrictie aan het bereik aan te passen.

Andere onderwerpen die hier op heldere wijze besproken worden, betreffen de berekenbaarheid van woordbetekenis, de grenzen van compositionaliteit en de relatieve transparantie van collocaties, alsmede de noties homonymie, synonymie, antonymie en hyponymie

Verkuyls model bestaat uit vier subsystemen: twee individugebonden en twee niet-individueel gebonden systemen. De eerste twee zijn het conceptuele systeem, *C*, en het lexicaal systeem, *L*. Met de concepten die *C* bevat, correspondeert

ren lexemen uit L. Deze subsystemen worden gekarakteriseerd als de individuele componenten van het model. Tegenover deze individugebonden systemen stelt Verkuyl twee systemen die een niet-individueel karakter hebben: een systeem dat informatie over taalvormen bevat, *W*, en een systeem dat encyclopedische kennis over de werkelijkheid bevat, *E*. De informatie die deze systemen bevatten, is voor iedereen toegankelijk.

De wisselwerking tussen deze vier subcomponenten, C, L, E, W, geeft een dynamisch beeld van semantische verschijnselen. Er loopt bijvoorbeeld een verbindingslijn tussen de individuele componenten L en C via de sociale componenten W en E. De weg van L naar W kan gezien worden als een "variant van het raadplegen van een woordenboek" die "ons direct verbindt met informatie in E die het mogelijk maakt het woord toe te passen op de werkelijkheid". De gewone sociale informatie wordt dan opgeslagen in C. Maar iemand kan ook starten met C en via E en W in de positie komen waar zij een begrip kan verwoorden, dat is in L belanden. Deze route beschrijft derhalve een lexicaliseringsproces.

Verkuyls gedachtegang over de werking van zijn model is gestructureerd tegen de achtergrond van een drietal opposities. Woordenboeken versus encyclopedieën, kennis van taal (*W* + *L*) versus kennis van de niet-talige werkelijkheid (*C* + *E*) en sociale kennis (*I*) versus individuele kennis (*K*). Het onderscheid tussen individuele en sociale kennis is wellicht het belangrijkste. Reeds in Putnam (1985) werd rekening gehouden met de sociale aspecten van betekenis. In Putnams visie toetsen taalgebruikers hun betekenisinzichten aan die van de experts, de dragers van semantische kennis. Maar Verkuyl volgt Putnam alleen in schijn. In de eerste plaats is bij Verkuyl de kennisdrager een abstracte entiteit en niet een groep van uitverkooren taalgebruikers. In de tweede plaats streeft Verkuyl naar een dynamische representatie van het proces van vergaren van semantische infor-

matie. En dit aspect krijgt bij hem de overhand. Niet zozeer het sociale aspect maar het dynamische typeert Verkuyls model van kennisrepresentatie.

Centraal in dit model zijn twee relaties die kennis structureren: de inclusierelatie die het mentale lexicon mede structureert, en de meronymische relatie die hetzelfde doet met het conceptuele apparaat. Twee interessante kenmerken van deze twee relaties zijn de volgende. In de eerste plaats laat Verkuyl zien dat het gebruikelijk is dat karakterisering van entiteiten een beroep doen op beide relaties. Karakteriseren we bijvoorbeeld tijgers als vierpotig, dan voegen we het meronymische paar *poot-tijger* en het inclusiepaar *tijger-vierpotig* aan ons kennisbestand toe. Een driepotige tijger tast de inclusierelatie aan, maar niet de meronymische relatie: we laten deze gelden zelfs bij ontstentenis van de andere. Met andere woorden, een meronymisch paar kan verbonden worden met verschillende inclusierelaties. Dit inzicht is enigszins verwant aan Putnams stelling dat een stereotype verbonden kan worden met termen die verschillende extensies hebben.

In de tweede plaats maakt Verkuyls model een synthese mogelijk tussen twee betekenis-theorieën: een afgezwakte versie van de componentieanalyse en de prototypische theorie. Hij bereikt deze synthese door een fijnere analyse van het C-deel van zijn model te presenteren. Een woord uit het mentale lexicon *hengst* correspondeert met het begrip in HENGST. Dit begrip is verbonden met een verzameling van begrippen die noodzakelijke voorwaarden belichamen. Deze zijn uiteraard begrippen die HENGST omvatten. En ieder begrip is verbonden met zintuiglijke informatie: de prototypische informatie. De noodzakelijke voorwaarden worden dan als het ware aangevuld met de prototypen, opgevat als organisatorische principes van onze zintuiglijke ervaringen.

Verkuyls beschrijving van zijn model wordt ontsierd door een filosofische faux pas.



“Betrachte z. B. einmal die Vorgänge, die wir Spiele nennen. Was ist allen diesen gemeinsam?”, vroeg de gekwelde Wittgenstein en hij voegde aan zijn vraag een mismoedige verwachting toe: “... schau, ob ihnen allen etwas gemeinsam ist. – Denn wenn du sie anschaust, wirst du zwar nicht etwas sehen, was allen gemeinsam wäre. . .” Deze passage heeft Verkuyl in gedachten wanneer hij Wittgensteins “aanval op inclusie als een regulerend principe” meent te kunnen pareren. Hij vindt Wittgensteins verwachting onterecht en zijn probleem een schijnprobleem. Verkuyls verweer is echter verdacht eenvoudig. Er is volgens hem een “zeer abstracte categorie waarin alle spelen thuishoren: bezigheden, de genusaanduiding. Het kan dus wel”. Maar dit kan niet correct zijn. Wittgenstein zit niet om genusaanduidingen verlegen. Spelen zijn immers *Vorgänge*. Deze onkarakteristieke vergissing van Verkuyl brengt een tweede fout met zich mee. Verkuyl beschouwt het probleem van afwijkende entiteiten eveneens als een schijnprobleem. Zoals hij de zaken stelt, levert een driepotige tijger geen probleem op voor de definitie van tijger als vierpotig. Een driepotige tijger is inderdaad een tijger. Zo gesteld is het probleem inderdaad een schijnprobleem. Maar de originele vraagstelling is een andere, namelijk of aan een entiteit die alle eigenschappen van een tijger bezit, behalve de cruciale eigenschap P, de naam van *tijger* toekomt. Verkuyl ontwijkt deze vraag omdat hij van afwijkende tijgers uitgaat. De filosofen die zich over deze zaken buigen, gaan eerder uit van een denkbeeldige wereld waarin de extensie van tijger niet een collectie van dieren is, maar een collectie van robots, of waarin de extensie van water niet H<sub>2</sub>O is maar een andere substantie. En het probleem is dan niet langer een schijnprobleem, maar een conceptueel probleem. Deze filosofische misstappen ontseries het boek, maar de schade is beperkt. Het blijft een helder geschreven uiteenzetting van enkele semantische onderwerpen die voor buitenstaanders toegankelijk is.

Dit boek kan echter geen dienst doen als een verslag omtrent de semantische state of the art. Was dit de bedoeling geweest, dan hadden veel meer onderwerpen aan de orde moeten komen, dan hadden concurrerende semantische theorieën een uiteenzetting verdiend. De ingelichte lezer zoekt hier tevergeefs informatie over de verschillende semantische behandelingen van tempus en aspect, over de juiste analyse van kwantificerende uitdrukkingen, over de semantische waarde van analyses over thematische relaties. Verkuyl gaf kennelijk eerder de voorkeur de lezer een proeve van semantische analyse te geven dan hem een overzicht van de verworvenheden en de problemen van de semantiek aan te bieden. We kunnen Verkuyl derhalve terecht verwijten dat hij een van de doelstellingen van de serie *Weten* niet gehaald heeft. Het resultaat is echter een boeiend verhaal over een zich steeds ontwikkelende discipline.

*Victor Sánchez-Valencia*

## Bibliografie

- De Jong, F.M.G., E. Oversteegen en H.J. Verkuyl (1990).** *Betekenis en Taalstructuur: inleiding in de formele semantiek*. Dordrecht: Foris.
- Gamut, L.T.F. (1982).** *Logica, taal en betekenis. Deel I: inleiding in de logica. Deel II: Intensionele logica en logische grammatica*. Utrecht/Antwerpen: Het Spectrum.
- Lyons, J. (1995).** *Linguistic Semantics*. Cambridge: Cambridge University Press.
- Putnam, H. (1975).** The meaning of “meaning”. In: H. Putnam, *Mind, Language and Reality*, Ch. 12.
- Verschuyll, H.J. (1988).** *Cryptogrammatica: het cryptogram als taalspel*. Utrecht: Kosmos.

**Nicoline van der Sijs en Jaap Engelsman, *Nota Bene. De invloed van het Latijn en Grieks op het Nederlands*.** Utrecht/Amsterdam: Sdu Uitgevers, 2000. 257 blz. ISBN 90 5797 067 8. fl 25,00.

In *Nota Bene* beschrijven Nicoline van der Sijs en Jaap Engelsman welke invloed het Latijn en het Grieks op het Nederlands hebben gehad. Het eerste en meest omvangrijke deel van dit populair-wetenschappelijke boek is geschreven door Nicoline van der Sijs en handelt over de aan de klassieke talen ontleende woorden. In het tweede gedeelte gaat Jaap Engelsman in op nog altijd gangbare citaten en uitdrukkingen die teruggaan op het Grieks of Latijn.

Nicoline van der Sijs is geen onbekende op het gebied van de leenwoorden. Eerder schreef zij *Leenwoordenboek. De invloed van andere talen op het Nederlands* (1996) en *Geleend en uitgeleend. Nederlandse woorden in andere talen & andersom* (1998). Het deel dat Van der Sijs in *Nota Bene* voor haar rekening neemt, is gebaseerd op het genoemde *Leenwoordenboek*. Zij heeft de passages die hierin voorkomen over Griekse en Latijnse leenwoorden voor *Nota Bene* sterk uitgebreid met cultuur-historische informatie.

*Nota Bene* is geen droge opsomming van klassieke woorden en uitdrukkingen in het Nederlands. Om bij het eerste deel te beginnen: hierin komen niet alleen Nederlandse woorden aan de orde die rechtstreeks teruggaan op het Latijn en Grieks, maar ook de omstandigheden die tot overname hebben geleid. Op deze wijze krijgt de lezer niet alleen zicht op de Latijnse en Griekse leenwoorden, maar ook op de grote cultuur-historische uitstraling van de klassieke wereld. Van een uitputtende opsomming kan overigens geen sprake zijn, daarvoor is het aantal Griekse en Latijnse leenwoorden te groot.

Van der Sijs behandelt de ontleningen in chronologische volgorde, te beginnen bij de klassieke wereld zoals onze Germaanse voorouders die leerden kennen. Dat levert naast een grondige historische schets ook leuke weetjes

op. Dat de Grieken en Romeinen de Germanen als barbaren beschouwden, is algemeen bekend. Maar waren de Germanen dan zo onbeschaafd? Van der Sijs meldt ons dat het Griekse *barbaros* werd gebruikt voor iemand die geen Grieks sprak, maar een onverstaanbaar taaltje brabbelde dat de Grieken als *bar-bar* in de oren klonk (p. 22).

In alle hoofdstukken zijn de leenwoorden naar thema ingedeeld. Aan elk thema is vervolgens een afzonderlijke paragraaf gewijd. Dat komt niet alleen de overzichtelijkheid ten goede, maar ook de toegankelijkheid van het boek. De paragrafen zijn bovendien zo opgezet dat men ze ook afzonderlijk kan lezen.

Na het hoofdstuk over de Romeinse tijd volgen de hoofdstukken 'Middeleeuws Latijn', 'Kerklatin' en 'De Renaissance'. Voor de lezer die na de inleiding een chronologische ordening per periode verwacht, is deze indeling niet direct doorzichtig. In de inleiding van het hoofdstuk over het middeleeuws Latijn weet Van der Sijs het onderscheid met het Kerklatin weliswaar helder uiteen te zetten, maar een opmerking in de algemene inleiding over de aparte positie van het Kerklatin (en daarmee van het gelijkname hoofdstuk) was op zijn plaats geweest. Nu wordt de chronologie stilzwijgend doorbroken en bevindt de lezer zich iets over de helft van het eerste deel plots weer bij de Romeinen. Dat neemt overigens niet weg dat het een goede keuze is geweest om de opkomst van het christendom en Latijn van de kerk in een apart hoofdstuk te behandelen. De lezer krijgt zo een duidelijk beeld van de taalkundige 'gevolgen' van de kerstening en van de wisselwerking tussen het Latijn en de volkstaal in de vorm van betekenisontleningen en leenvertalingen.

In het hoofdstuk 'De Renaissance' is uitgebreid aandacht voor het Humanistenlatijn, en komt ook het Grieks om de hoek zetten. Tot nu toe is daar weinig aandacht aan besteed, omdat het Grieks tot aan de Renaissance slechts invloed heeft gehad via het Latijn. Dat de

Griekse invloed niet onderschat moet worden, blijkt uit het register waarin is aangegeven wanneer een Latijns woord op het Grieks teruggaat. Van de in het boek genoemde Latijnse leenwoorden blijkt maar liefst een derde Griekse wortels te hebben. Directe ontleningen uit het Grieks betreffen vooral literaire en taalkundige termen, zoals *acrostichon*, *litotes*, *epenthesis* en *prothesis* (p. 132).

Taalkundige uiteenzettingen worden in *Nota Bene*, ondanks het populair-wetenschappelijke karakter van het boek, niet geschuwd, zoals blijkt uit het hoofdstuk 'Moderne nieuwvormingen op basis van het Latijn en Grieks', waarin uitvoerig wordt uitgelegd op welke manieren nieuwvormingen worden gemaakt. In 'Aanpassing van Latijnse en Griekse leenwoorden' gaat Van der Sijs nog een stapje verder door aan de hand van klankveranderingen in het Latijn en Nederlands uit te leggen hoe de ouderdom van Latijnse leenwoorden kan worden bepaald. Hierdoor is het boek niet alleen interessant voor de leek die meer wil weten van de invloed van de klassieke talen op het Nederlands, maar ook voor de meer taalkundig onderlegde lezer.

'Gevleugelde woorden van Griekse en Latijnse oorsprong' is de titel van het tweede deel van *Nota Bene*. Jaap Engelsman behandelt hierin een selectie van nog steeds gangbare citaten uit de klassieke Oudheid. Wie zegt te leven *tussen hoop en vrees* of iets betitelt als *een storm in een glas water* citeert respectievelijk Vergilius en Cicero. Citaten die uit hun verband worden gehaald, gaan een eigen leven leiden en veranderen in de loop van de tijd van vorm en betekenis. De nu nog gangbare klassieke citaten gaan dan ook slechts ruwweg terug op hun bron. Engelsman beschrijft telkens waar een uitdrukking vandaan komt, hoe deze zich heeft ontwikkeld en geeft de nodige achtergrondinformatie. Ook hij heeft gekozen voor een chronologische opzet en neemt ons van de Grieken mee naar de Romeinen, en van de Middeleeuwen naar de Renaissance en later. De hoofd-

stukken over de Grieken en de Romeinen vormen het zwaartepunt. De middeleeuwers grepen niet terug op de klassieken (het hoofdstuk over de Middeleeuwen telt dan ook slechts twee bladzijden), en in de Renaissance maakte het Grieks, dat in de Middeleeuwen weinig bekend was, geen echte come-back en werd het Latijn steeds verder teruggedrongen tot het domein van de motto's, deviezen en lijfspreuken.

Niet elke uitdrukking die Jaap Engelsman behandelt, is even bekend. *Je kunt niet tweemaal in dezelfde rivier stappen* van de Griekse filosoof Heraclitus is daarvan een voorbeeld (p. 191). Een groot bezwaar is dat niet, omdat Engelsman kort en bondig op de uitdrukking ingaat en er bovendien genoeg overbekende uitdrukkingen behandeld worden.

Achter in het boek vindt de lezer een register met de besproken Griekse en Latijnse leenwoorden. Dit register is tevens een handig hulpmiddel voor wie wil weten op welk Grieks of Latijns woord een ontlening teruggaat. Achter elk leenwoord is namelijk de oorspronkelijke vorm gegeven. Gaat het om een Latijnse vorm die op zijn beurt weer teruggaat op een Grieks woord, dan is dat aangegeven met het teken •. Hadden de auteurs in die gevallen ook nog de Griekse vorm vermeld, dan was het overzicht compleet geweest. Het had het register echter ook flink doen uitdijen, omdat er nogal wat Griekse woorden zijn die wij via het Latijn hebben leren kennen. Wellicht is daarom de keuze gemaakt om het bij een markering te houden.

Wie *Nota Bene* gelezen heeft, kan niet anders dan onder de indruk zijn van de kennis van zaken die Van der Sijs en Engelsman bezitten en van de heldere wijze waarop zij de invloed van het Latijn en Grieks op het Nederlands beschrijven. Die invloed blijkt *nota bene* zo groot te zijn, dat we ons kunnen aansluiten bij de conclusie van de auteurs "dat de hedendaagse Nederlandse maatschappij, wetenschap en cultuur er zonder de Romeinse en Griekse erfenis onherken-

baar anders uit zouden zien, en dat als we alle Latijnse en Griekse woorden uit onze woordenschat zouden schrappen, communicatie onmogelijk zou zijn" (p. 12).

*Corrie de Haan*

### **Bibliografie**

- Van der Sijs, N. (1996).** *Leenwoordenboek. De invloed van andere talen op het Nederlands.* Den Haag/Antwerpen: Sdu Uitgevers/Standaard Uitgeverij.
- Van der Sijs, N. (1998).** *Geleend en uitgeleend. Nederlandse woorden in andere talen en andersom.* Amsterdam/Antwerpen: Contact.

# Signalementen

## *Zestiende-eeuws taalgidsje hertaald*

De productiviteit van Nicoline van der Sijs is al vaak geprezen. Zij blijkt bovendien bij elk boek dat verschijnt, nog veelzijdiger dan we bij eerdere gelegenheden al dachten. Haar recentste boek, *Wie komt daar aan op die olifant?*, is weliswaar uitgegeven bij Het Taalfonds van uitgeverij L.J. Veen, maar lijkt toch vooral interessant om buitentaalkundige – met name historische – redenen.

Het omslag vermeldt dat dit de 151e editie is van de taalgids die Noël van Berlaimont in 1527 in Antwerpen uitgaf in twee talen. Het boekje bevatte een beknopte woordenlijst en enkele aanwijzingen over uitspraak en grammatica, maar de kern werd gevormd door een zevental dialogen en enkele modelbrieven. De methode van Berlaimont werd in heel Europa zo populair dat er nu dus meer dan 150 edities van verschenen zijn. Het boek werd bovendien in twaalf talen vertaald. Op dezelfde leest werden bovendien tal van andere werken geschoeid. Volgens Van der Sijs is bijvoorbeeld de beroemde reeks *Wat en Hoe?* geïnspireerd door Van Berlaimont; en aan het eind van de zestiende eeuw breidde de koopman Frederik de Houtman het model van Van Berlaimont uit met dialogen die speciaal waren toegesneden op de koopman die naar Oost-Indië wilde en daar in het Maleis wilde handelen, de weg vragen, peper wegen en met een sultan spreken.

In *Wie komt daar aan op die olifant?* worden de taalgidsjes van Van Berlaimont en De Houtman, hertaald in modern Nederlands, opnieuw uitgegeven. Ook het verslag dat De Houtman voor prins Maurits schreef over zijn spectaculaire

eerste reis naar Atjeh, waar hij twee jaar gevangen gehouden werd en met de dood werd bedreigd, is hier hertaald opgenomen. Van der Sijs geeft bovendien in een uitgebreide inleiding biografische schetsen van de twee auteurs en een gedegen verantwoording van haar bronnen en de keuzen die ze maakte tijdens het hertalen.

Alles bij elkaar geeft dit boek daarmee een interessant doorkijkje in het alledaagse leven van de Nederlander in de zeventiende eeuw, zowel thuis als in de Oost. De historische sensatie, het gevoel even heel dicht bij de medemens van bijna vier eeuwen geleden te staan, is daarmee op bijna elke bladzijde voelbaar.

Wel rijst een beetje de vraag wat een dergelijk boekje doet in een serie die Het Taalfonds heet. Er wordt nauwelijks over taal gesproken in de dialogen, en doordat de dialogen zelf heel adequaat hertaald zijn valt er ook nauwelijks iets uit op te maken over het taalgebruik van de zeventiende-eeuwer. Van der Sijs heeft het tot haar principe gemaakt om oude boeken die in Het Taalfonds opnieuw worden uitgegeven, in modern Nederlands om te zetten. In het algemeen is dat misschien te verdedigen: op die manier worden de werken toegankelijk gemaakt voor een groot publiek. Maar juist in dit geval had het me aardig geleken als af en toe de oorspronkelijke tekst van de dialogen had kunnen worden afgedrukt naast de vertaling, precies zoals in de oorspronkelijke boeken van Van Berlaimont en De Houtman het Frans of het Maleis naast het Nederlands stond. De dialoogvorm leent zich daar uitstekend voor; in dat geval had dit boek de eenentwintigste-eeuwse geïnteresseerde lezer niet alleen inzicht gege-

ven in de zeventiende-eeuwse maatschappij, maar ook in de taal zoals deze enkele honderden jaren geleden ongeveer geklonken moet hebben.

**Bibliografische gegevens:** *Wie komt daar aan op die olifant? Een zestiende-eeuws taalgidsje voor Nederland en Indië, inclusief het verhaal van de avontuurlijke gevangenschap van Frederik de Houtman in Indië.* Hertaald en toegelicht door Nicoline van der Sijs. Amsterdam/Antwerpen: Veen, 2000. 220 blz. ISBN 90 204 5890 6 fl 34,90.

Marc van Oostendorp

### *Geschiedenis van de dialectgeografie in het Nederlandse taalgebied*

Dit boek is ondanks de ruime titel vooral een geschiedenis van het 'Dialectenbureau', dat wil zeggen de voormalige afdeling Dialectologie, en zijdelings ook de andere afdelingen die het KNAW-instituut vormden dat nu het Meertens Instituut is. In een kort hoofdstuk aan het begin leest men wat er vóór het Dialectenbureau is gebeurd, en het laatste hoofdstuk (voor het Slotwoord) wat er naast het Dialectenbureau is gedaan. Dat hoofdstuk heet 'De universiteiten', en blijft beperkt tot de Nederlandse universiteiten. Het voorbijgaan aan alles wat er aan Vlaamse universiteiten gedaan is op het gebied van de dialectgeografie in het Nederlandse taalgebied, maakt de eerste titel tot een vlag die de lading niet dekt.

De tweede titel – rondom Kloeke en het Dialectenbureau – is meer op zijn plaats. De oprichting van het Dialectenbureau blijkt gepaard te gaan met een professorenstrijd, die ertoe leidt dat niet Kloeke, de meest vernieuwende dialectoloog van dat moment, de leiding gaat nemen, maar de nauwelijks in dialecten geïnteresseerde Meertens (die, dat voeg ik er te

zijner verdediging aan toe, toen al wel een aardige studie over het Utrechts op zijn naam had staan). Het heeft tot gevolg dat het Dialectenbureau van begin af aan niet de leidende positie in de dialectgeografie heeft ingenomen die van een taalgebied-brede dialectcentrale verwacht had mogen worden. De aan anderen dienstbare, wat achter de ontwikkelingen aanlopende houding is het Dialectenbureau altijd wat blijven aankleven, zo is de indruk die men uit Daans relaas van de geschiedenis krijgt.

Het geschiedkundige overzicht eindigt in de jaren zestig ('Na de oorlog', p. 68), maar dat wil allerminst zeggen dat er over de jaren daarna geheel gezwegen wordt. De volgende hoofdstukken ('Publicaties en activiteiten', 'Het Dialectenbureau en de taalkundige theorieën' en het al genoemde 'De universiteiten') hebben echter niet het chronologische karakter van de eerste negen hoofdstukken.

Het boek wordt afgesloten met een 'Register van persoonsnamen', dat niet alleen aangeeft in welk hoofdstuk iemand ter sprake komt, maar ook het geboortejaar, eventueel het jaar van overlijden en de functie noemt. De vermelding 'arts' achter Winkler is weliswaar niet onjuist, maar geeft ook weer niet direct zijn grootste betekenis weer. De medische achtergrond van de auteur van het *Algemeen Nederduitsch en Friesch Dialecticon* is meer een aardig weetje. Wie van dit soort weetjes houdt, komt in Daans boek niets tekort. Ze vormen, tezamen met de vele historische foto's, de charme van dit werk.

**Bibliografische gegevens:** Jo Daan, *Geschiedenis van de dialectgeografie in het Nederlandse taalgebied. Rondom Kloeke en het Dialectenbureau.* Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen, 2000. 139 blz. ISBN 90-6984-289-0 NLG 25,00. Te bestellen bij de uitgever: 020-5510780.

Harrie Scholtmeijer

### *Woordenboek van de Kamper Taal*

Op 19 juni 2000 werd in de stampvolle raadszaal van het Kamper stadhuis het woordenboek van de Kamper Taal gepresenteerd. Dit woordenboek is samengesteld door een werkgroep afkomstig uit Kampen, Kampereiland en Kamperveen, op basis van de woordenlijst in Gunninks dissertatie *Het dialect van Kampen en omgeving* (1908). Opvallend is dat Gunnink als uitgangspunt van zijn beschrijving niet het stad-Kampers nam, maar het Kampereilands, omdat dat naar zijn mening het zuiverst was. In het huidige woordenboek komen alledrie de genoemde dialecten aan bod. Als een woord of betekenis tot één of twee plaatsen beperkt is, wordt dat met een vermelding van die plaatsen aangegeven.

Aanvankelijk schrapten de samenstellers de thans verouderde woorden uit de lijst-Gunnink. In de definitieve versie van het Woordenboek keerden ze toch weer terug, voorzien van een asterisk. Lastig is dat dit asteriskteken ook (volkomen overbodig) gebruikt wordt voor vormvarianten (bijv. *\*elkstiek* = z. *elestiek*). Het =-teken in dit voorbeeld is in het Woordenboek overigens achter elke ingang geplaatst: overbodig en typografisch lelijk.

De totstandkoming, het doel en de inrichting van het Woordenboek komen ter sprake in de Inleiding, van de hand van Philomène Bloemhoff-de Bruijn. Uitvoerige paragrafen over de spelling (gebruikt is de zogeheten IJsselacademiespelling) en de uitspraak sluiten deze Inleiding af. Daarna volgt de eigenlijke woordenlijst. Opvallend is hierin het ontbreken van de *h* (op één woord na *hubberig* = z. *'ubberig* [huiuverig]; de vorm met *h* is Kampereilands). In de Inleiding wordt het weglaten van de *h* verklaard als een overblijfsel van de gewoonte onder de Kamper elite in de Franse tijd om in hun taal het Frans (dat immers ook geen *h* kent) te imiteren. Als gezonken cultuurgood leeft dit taalkenmerk, net als de uvulaire R, in de huidige volkstaal van Kampen en omgeving voort.

Het *Woordenboek van de Kamper Taal* is mooi uitgegeven (hard kaft, glanzend papier). Veel foto's geven het verleden weer. In boeken over dialect is dat geen uitzondering, maar het verdraagt zich toch niet helemaal goed met de eerste doelstelling zoals die in de Inleiding (p. 10) genoemd wordt: 'de woordenschat van het Kampers *zoals dat nu nog bestaat*, voor het nageslacht vast te leggen' (cursivering van mij). Het historische karakter geldt ook voor de kinderrijmpjes, recepten e.d. die door het boek zijn gestrooid. Als van een spotliedje wordt gezegd: 'werd begin deze eeuw gezongen op de bewaarschool van juffrouw Klaucke' (p.79), mogen we aannemen dat bedoeld is: begin vorige eeuw.

**Bibliografische gegevens:** A Fien (+), J.B. Keuter, A. van der Linde-van der Weerd, P. van Mierlo, K. van 't Oever, *Woordenboek van de Kamper Taal*. Aangevuld, bewerkt en ingeleid door Ph. Bloemhoff-de Bruijn. Kampen: IJsselacademie,, 2000. 191 blz. Publicaties van de IJsselacademie nr. 121 ISBN 90-6697-109-6 NLG 30,00, voor donateurs van de IJsselacademie NLG 25,50. Te bestellen bij de uitgever: 038-3315235.

*Harrie Scholtmeijer*

### *Taal in de klas*

In 1997 promoveerde Dora Dolle-Willemsen in Tilburg op het proefschrift *Gezien onderscheid naar sekse in het basisonderwijs. Interactie als invalshoek*. Het onderzoek toonde aan dat leerkrachten jongens en meisjes in hun klas verschillend benaderen. Recentelijk verscheen het boekje *Taal in de klas*, een populair wetenschappelijke versie van dit proefschrift.

Veel leerkrachten ontkennen dat zij onderscheid maken tussen jongens en meisjes. Een speciaal observatiesysteem (Feedbackresearch Observation Guide, FROG) werd gebruikt om

de leerkrachten te confronteren met de werkelijkheid als het gaat om sociale interactie tussen henzelf en de leerlingen. De volgende vragen konden bijvoorbeeld met behulp van dit observatiesysteem worden beantwoord. Hoeveel tijd is de docent aan het woord? Hoeveel vragen stelt de leerkracht aan jongens en hoeveel aan meisjes? Wat zijn dit voor vragen: inzichtvragen of kennisvragen? Hoe lang wordt aan een leerling de tijd gegund om na te denken over een antwoord?

De uitkomsten van de studie blijken, met name voor de leerkrachten zelf, onthutsend. Jongens krijgen meer aandacht in de klas. Ze krijgen meer beurten en andere vragen dan de meisjes. Inzichtvragen worden vaker aan jongens gesteld, kennisvragen meestal aan meisjes.

De leerkracht heeft de mogelijkheid om verschillende vragen te stellen: rechtstreeks aan een leerling of een Wie-weet-het-vraag (vraag aan de hele klas). Jongens zijn eerder geneigd Wie-weet-het-vragen te beantwoorden. Aan jongens wordt ook vaker gevraagd een correctie te geven op vragen die door een medeklasgenoot niet goed zijn beantwoord (correctie-beurten).

Het onderzoek laat bovendien zien dat het tempo in de klas erg hoog ligt. Een leerling heeft na een vraag nauwelijks de tijd om na te denken over het antwoord.

Het onderzoek had het doel om vast te stellen hoe het taalgedrag van leerkrachten ten opzichte van hun leerlingen is. Anders dan in soortgelijke studies in het buitenland is echter in deze studie ook de coaching van leerkrachten na de observatie opgenomen. De docenten werden door middel van video-opnames geconfronteerd met hun eigen taalgedrag en zochten met een begeleider naar mogelijkheden tot verbetering. Wat betreft het tempo bijvoorbeeld zijn direct bij de tweede video-opname al positieve veranderingen te zien: de leerlingen krijgen meer tijd om een antwoord te formuleren en vooral de meisjes doen hier hun voordeel mee. Het geven van meer correctiebeurten aan

meisjes is een andere manier om hen evenwichtiger bij de interactie in de klas te betrekken. Dit soort beurten zijn volgens Dolle-Willemsen erg belangrijk voor de ontplooiing van kinderen. Als de leerkracht vertrouwen in jou heeft, geeft dat zelfvertrouwen. De onderzoeker trekt dit hogere zelfvertrouwen verder door: meisjes schatten zichzelf uiteindelijk hoger in en zouden het zelfs beter aandurven een als moeilijk bekend staande exacte studie te kiezen.

We hebben hier te maken met een interessant onderzoek, dat zeer leesbaar is beschreven in het boekje *Taal in de klas*. Een aanrader voor een ieder die al dan niet beroepsmatig te maken heeft met kinderen.

#### **Bibliografische gegevens:**

Dora Dolle-Willemsen en Agnes Verbiest: *Taal in de klas*. Amsterdam/Antwerpen, Contact, 1999. 60 blz. ISBN 90-254-9565-6 fl. 16,90.

*Henny van der Neut*

#### ***Nieuwe woorden en uitdrukkingen***

In het boekje *De taal van het jaar* behandelt taalhistoricus en journalist Ewoud Sanders bijna vijfhonderd nieuwe woorden en uitdrukkingen van het jaar 1999. Het aardige van dit boekje is dat de neologismen in kortere of langere verhaaltjes worden gepresenteerd. De verhaaltjes vermelden de definitie en vindplaats van het nieuwe woord of de uitdrukking zoals in de meeste neologismenoverzichten gebruikelijk is. Daarnaast zijn in *De taal van het jaar* ook de maatschappelijke omstandigheden die tot de geboorte van een woord hebben geleid, opgenomen. Veelal zijn kranten- of tijdschriftfragmenten opgenomen om details of achtergrondinformatie toe te voegen.

Sanders verantwoordt in zijn uitgebreide voorwoord de keuze voor opname van nieuwe



woorden of uitdrukkingen in het Nederlands taalgebied. Eenmalige neologismen vielen af. Woorden uit 1999 die geregeld in krantenbestanden voorkwamen, staan er wel in. Maar ook woorden die slechts een enkele keer voorkwamen, maar die veel zeggen over de tijd waarin we leven. Ook zijn woorden of uitdrukkingen opgenomen die weliswaar eerder het daglicht zagen, maar die in 1999 algemeen bekend werden. Voorbeelden hiervan zijn 'onder de pet houden' en 'eclipsbril'. 'Onder de pet houden' werd in februari 1999 door de Bijlmerenquête plotseling bekend (430 vindplaatsen), maar dateert al uit de jaren zeventig. De 'eclipsbril' werd in 1996 een keer aantreffen, maar werd in 1999 200 maal in kranten en tijdschriften geteld. Sanders maakte onder meer gebruik van 21 digitaal beschikbare krantenbestanden, van nieuwsgroepen en grote databestanden die via Internet bereikbaar zijn.

De meeste neologismen werden geleverd door bedrijfsleven, economie, handel en financiën (25) bijvoorbeeld 'bedrijfsmasseur', 'cur-susmoe', 'midlife-miljonair', 'retourpinnen'. De nieuwe media leverden twintig nieuwe woorden. Sanders gaat in zijn inleiding in op het effect van de televisie op de taal. Wat betreft 1999 had het televisieprogramma Big Brother een grote invloed met woorden als 'BB' en 'dagboekkamer', terwijl 'effe lekker knuffelen' en 'laat je niet gekmaken' gevleugelde uitdrukkingen werden; oorlog en politiek (bijvoorbeeld 'betonbom', 'dioxinekip' en 'Blair-baby') leverden achttien nieuwe woorden en de medische wetenschap zestien, zoals 'bijlmersyndroom', 'televisiefiets' of 'wegkwijn-ziekte'; sport ('sporthoofddoekje' en 'Tae bo') ten slotte zorgde net als verkeer en vervoer ('grachtengordelbus', 'calamiteitenvaartuig', 'herriobarometer') voor elf neologismen. Nogal wat woorden, dertien procent, komen uit het Engelse taalgebied. Dit is iets meer dan het percentage Engelse woorden in de Grote Van Dale. Sanders verklaart dit vanwege het feit dat de taalgemeenschap nog nauwelijks de tijd heeft

gehad om Nederlandse tegenhangers te verzinnen. Er is volgens hem dan ook geen sprake van een grotere invloed van het Engels. Voorbeelden van neologismen uit het Engels zijn 'barebacking', 'bullshit bingo', 'domaingrabbing', 'hacktivist', 'page jacking', 'palletscan' en 'shockvertising'.

*De taal van het jaar* geeft een overzicht van markante gebeurtenissen in het millenniumjaar en is vooral door de gekozen opzet een aardig boekje geworden.

### **Bibliografische gegevens:**

Ewoud Sanders: *De taal van het jaar. Nieuwe woorden en uitdrukkingen editie 2000*. Amsterdam/Antwerpen, Contact, 2000. 176 blz. ISBN 90-254-9707-1 fl. 24,90.

*Henny van der Neut*

# Uit de tijdschriften

De rubriek *Uit de tijdschriften* geeft kort weer wat er in andere tijdschriften op het gebied van de taalkunde is verschenen. Momenteel worden in deze rubriek de volgende tijdschriften behandeld: *Anéla, Driemaandelijke Bladen, Leuvense Bijdragen, Naamkunde, Nederlands van Nu, Neerlandica Extra Muros, Ons Erfdeel, Onze Taal, Spiegel, Southern African Linguistics and Applied Language Studies, Taal en Tongval, Taalkundig Bulletin, Tijdschrift voor Nederlandse Taal- en Letterkunde, VDW-berichten, Verslagen en Mededelingen van de KANTL, De Woordenaar*. Uitgevers van niet vermelde taalkundige periodieken die hun tijdschrift besproken willen zien, wordt verzocht contact op te nemen met de redacteur van deze rubriek:

Dr. J. Nijen Twilhaar, Oerdijk 35, 7433 AG Schalkhaar, tel.: 0570-608080  
e-mail: jnt@xs4all.nl

## Nederlands van nu

49 (2001), nr. 2

In het tweede nummer van deze jaargang vinden we onder meer een artikel van Katrien Deygers over de eigenaardigheden van chatten via het internet, dat verloopt in de zogenaamde IRC (internet relay chat)-taal: *De chatter spreekt een eigen taal*. In *Schrijft Knack Belgisch? (1)* neemt Siegfried Theissen het taalgebruik van het tijdschrift *Knack* onder de loep, waarin telkens terugkomende lexicale afwijkingen van het Standaardnederlands voorkomen. Marc De Coster bespreekt enkele nieuwvormingen in *Kraak maar raak!* Frans Debrabedere bekijkt de etymologie van *Tantemanekes, Tantewannekes en Tantefeer*.

Verder zijn er de kleinere bijdragen en de vaste rubrieken.

## Neerlandica Extra Muros

XXXIX (2001), 2

Het tweede nummer van deze jaargang wordt geopend door Emmeken van der Heijden met een artikel over *De ANS in het buitenland: over anderstalige inbreng in de ANS*. Robert S. Kirsner gaat in op het gebruik van de finale partikels *hoor* en *hè* en hun interactie met de gebiedende wijs en de infinitivus pro imperativo: *De deur wel dichtdoen, hè? en andere mysterieuze zaken*. De volgende bijdrage is van J.A. van Leuvensteijn: *Abraham Blankaart aan het woord. Positiebepaling door middel van aanspreekvormen*.

## Uit de tijdschriften

### Ons Erfdeel

44 (2001), nr. 2

De tweede aflevering van deze jaargang bevat geen taalkundige bijdragen.

nr. 3

In het derde nummer staat een kort artikel van de hand van A.M. Hagen waarin hij een positieve beschrijving geeft van *Het Meertens-Instituut anno 2001*, dat een mooi tegenwicht vormt tegen het beeld dat van dit instituut wordt opgeroepen in de romans van Voskuil. In zijn bijdrage *Een zondagspak?* bespreekt Dirk Geeraerts het Nederlands in Vlaanderen: gedrag, beleid, attitudes. Verder geeft Dirk van Assche in de rubriek *Taal- en cultuurpolitiek Een staal-kaart van de neerlandistiek in de Franstalige wereld*.

### Onze Taal

70 (2001), nr. 4

Het vierde nummer van deze jaargang begint met een bijdrage van Peter Burger over *De huisstijl van de kwaliteitskranten*. Dick Smakman gaat in op de waardering van moderne uitspraakverschijnselen in *Poldernederlands of Gooise R?* Leendert Plug verdiepte zich in de geluidwegval in het Nederlands: *Mag et wa minde weze?* De bijdrage *Een taal is een taal als haar sprekers dat willen* van Marc van Oostendorp gaat over de erkenning van het Zeeuws als streektaal. Riemer Reinsma leert ons in zijn artikel *'Álitalia bruid van klm'* iets over de huwelijksmetaforen waarvan het bedrijfsleven gebruik maakt bij fusies.

Verder zijn er de kleinere bijdragen en de vaste rubrieken.

nr. 5

Deze aflevering begint met een bijdrage van Renée van Bezooijen, getiteld *Régionale omroepen, ádequate voorzieningen*, waarin de auteur

zich afvraagt of de klemtoon echt steeds vaker naar voren verschuift. Jan Erik Grezel heeft een interview met de taalkundige Jan W. de Vries, die zich bijna dertig jaar met het Nederlands in het buitenland bezighield en die onlangs afscheid nam als hoogleraar aan de Leidse universiteit. Jan Don gaat in op het (on)voorspelbare geslacht van Nederlandse woorden: *Dé afwas gaat in hét sop*.

Verder zijn er de kleinere bijdragen en de vaste rubrieken.

nr. 6

Dit nummer begint met een bijdrage van Ewoud Sanders over de heel eigen taal van het populaire communicatiemiddel sms: *Hoest? Veel suc6!* Jan Erik Grezel heeft een interview met Siemon Reker, bijzonder hoogleraar Gronings. In de nieuwe maandelijkse rubriek *Etymologica* gaat Noline van der Sijs in op *Het ontstaan van de etymologie*. De bijdrage van Frank Jansen gaat over communiceren via de kleding: *Sprekende kleding*. In *Niet bang voor meertaligheid* voert Marc van Oostendorp een vraaggesprek over taalpolitiek met parlementariër Judith Belinfante

Verder zijn er de kleinere bijdragen en de vaste rubrieken.

### Southern African Linguistics and Applied Language Studies

18 (2000), nr. 1

Dit is het eerste nummer dat verschijnt onder de bovengenoemde titel. Dit tijdschrift is een combinatie van het tijdschrift *South African Journal of Linguistics*, dat deel uitmaakte van deze rubriek, en de periodiek *Southern African Journal of Applied Language Studies*. Deze eerste aflevering wordt geopend door Jacobs A. Naude met een artikel over *Diachronic syntax and language change: The case of Qumran Hebrew*. Nerina Bosman gaat in op *Die motiveringsteorie en idioome*

met *hand*. De bijdrage van Elizabeth de Kadt gaat over de *English language proficiency in South Africa at the turn of the millennium*. Marie Spruyt geeft 'n *Kognitiewe benadering tot tydsdifferensiasie in diskoers*. Het artikel van Truida van der Merwe gaat over *Die funksies en toepassings van retoriese vrae*. Nkonko M. Kamwangamalu bespreekt *The state of codeswitching research at the dawn of the new millennium (2): Focus on Africa*. De bijdrage die daarop volgt is van Vivian de Klerk en Sally Hunt: *Discourse domination? The role of gender in seminar interaction*. De laatste bijdrage van dit nummer is van Gilles-Maurice de Schryver en D.J. Prinsloo en gaat over *The compilation of electronic corpora, with special reference to the African Languages*.

### Spiegel

17/18 (1999/2000), nr. 1-2

In dit dubbelnummer dat twee jaargangen bestrijkt, vinden we onder meer een artikel van Sjaak Kroon en Jan Sturm: *'De taal is gansch het volk': een zoektocht naar de oorsprong en het gebruik van een sententia*.

### Taal en Tongval

52 (2000), nr. 2

Het tweede nummer van deze jaargang begint met drie in memoriams, respectievelijk voor Valeer Frits Vanacker, A. Sassen en André de Groote. Het eerste artikel is van de hand van A. Weijnen: *Van een haander en een vrijs*. Sera de Vriendt behandelt *Diminutieven in het Brussels*. Katrien Deygers en Vicky Van Den Heede gaan in op *Belgisch Nederlandse 'klassiekers' als variabelen voor lexicaal variatie-onderzoek: een evaluatie*. *Waren de geuzen Goten?* is wat Michiel de Vaan zich te slotte afvraagt.

In de rubriek *Boekbesprekingen* komen de volgende werken aan de orde: J. Goossens, J. Taeldeman, G. Verleyen, *Fonologische Atlas van de*

*Nederlandse Dialecten*, deel I (door A. Weijnen), A.M. Duinhoven, *Middel nederlandse Syntaxis, synchroon en diachroon*. 2. *De werkwoordgroep* (door G. De Schutter), en Frans Debrabandere, *Kortrijks Woordenboek* (door Hugo Ryckeboer).

### Tabu

30 (2000), nr. 3-4

In deze dubbele aflevering staan vier artikelen. Het nummer wordt geopend door Helen de Hoop met een bijdrage over *Otje en scrambling in het Nederlands*. Jack Hoeksema brengt verslag uit van zijn onderzoek naar *Verplichte topicalisatie van kale enkelvoudens: de feit is dat-constructie*. Ron van Zonneveld en Roelien Bastiaanse verdiepten zich in *Frasale recursie: de syntaxis van de gelaagde VP*. In *Ikke en de default-nominatief: een reactie op Hoeksema* reageert Jan-Wouter Zwart op een eerder in deze jaargang verschenen publicatie van Jack Hoeksema.

### De Woordenaar

5 (2001), nr. 1

In deze aflevering vinden we onder meer een bespreking door Jan van Donselaar van C.A. Backers *Verklarend woordenboek van wetenschappelijke plantennamen*, een heruitgave van van dit werk uit 1936.

# Ontvangen boeken

**J. J. Spa**, *De dialecten van Groot-IJsselmuiden. Klank- en vormleer*. Kampen: IJsselacademie, 2000. 148 blz. ISBN 90 6697 115 0. f 29,95.

**Ineke van der Craats**, *Conservation in the Acquisition of Possessive Constructions*. Dissertatie Katholieke Universiteit Brabant 2000. X + 399 blz. ISBN 90 74029 14 0. f xx,xx.

**Karina van Dalen-Oskam en Marijke Mooijaart**, *Bijbels lexicon. Woorden en uitdrukkingen uit de bijbel in het Nederlands van nu*. Amsterdam: Prometheus, 2000. 416 blz. ISBN 90 5333 923 x. f 49,50.

**J. Goossens, J. Taeldeman en G. Verleyen**, *Fonologische Atlas van de Nederlandse Dialecten*. Deel II en deel III. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde, 2000. XXX + 330 blz. ISBN 90 72474 36 8. 123 EUR (4.962 BEF).

**J.B. Berns en J. van Marle (red.)**, *Overzees Nederlands*. Amsterdam: Meertens Instituut, 2000. 102 blz. ISBN 90 70389 63 0. f 29,50.

**Reinier Cozijn**, *Integration and inference in under-standing causal sentences*. Dissertatie Katholieke Universiteit Brabant 2000. II + 159 blz. ISBN 90 74029 11 6.

**Carina Gadourek**, *Albert Camus. Uitleg en aantekeningen bij zijn werken*. Assen: Van Gorcum, 2000. VI + 142 blz. ISBN 90 232 3602 5. f 29,50.

**Foolen, Ad, en Frederike van der Leek (red.)**, *Constructions in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins, 2000. VI + 338 blz. ISBN 90 272 3684 4. f 150,00.

**Rita Beyers (red.)**, *Van vader- naar moedertaal. Latijn, Frans en Nederlands in de dertiende-eeuwse Nederlanden*. Brussel: Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis, 2000. 196 blz.

**Reinhild Vandekerckhove**, *Structurele en sociale aspecten van dialectverandering. De dynamiek van het Deerlijkse dialect*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde, 2000. 352 blz. ISBN 90 72474 26 0. 21 EUR (847 BEF).

**Nienke Bakker**, *Guido Gezelle. Opbouw en analyse van zijn Bastaardwoordenboek*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde, 2000. 218 blz. ISBN 90 72474 35 X. 22 EUR (887 BEF).

**Jelle de Vries**, *Onze Nederlandse spreektaal*. Den Haag: Sdu Uitgevers, 2001. 564 blz. ISBN 90 12 09038 5. f 69,90.

**Manita Koop**, *Dat is goud. Tuinbouwtaal uit 't Westland*. De Lier: Adequaat, 2000. 51 blz. f 24,95.

**Jaap Hagedoorn**, *Overdracht van meer dan letters... Van Gorcum 1800-2000*. Assen: Van Gorcum, 2000. 191 blz.

**Ewoud Sanders**, *De taal van het jaar. Nieuwe woorden en uitdrukkingen. Editie 2001*. Amsterdam/Antwerpen, Veen, 2001. 143 blz. ISBN 90 204 5716 0. f 24,90.

**Willem Zonneveld**, *Van Afflighem en Chaucer: Het leven van Sinte Lutgart als jambisch gedicht*. Amsterdam: Stichting Neerlandistiek VU Amsterdam, 2000. 104 blz. ISBN 90 72365 63 1. f 30,00.

**Willy Vandeweghe e.a. (red.),** *Polyfonie. Opstellen voor Paul Van Hauwermeiren*. Gent: Mercator Hogeschool, 2001. 236 blz.

**Madeleine Hulsen,** *Language Loss and Language Processing. Three Generations of Dutch Migrants in New Zealand*. Dissertatie Katholieke Universiteit Nijmegen 2000. XII + 261 blz. ISBN 90 9013990 7. f 35,00.

**Grenzen. Verhalen en gedichten in de streektaal.** Kampen: IJsselacademie, 2001. 104 blz. ISBN 90 6697 126 6. f 49,50.

**Heleen Strating,** *Lexicale ontwikkeling in het Nederlands van autochtone en allochtone kleuters*. Dissertatie Katholieke Universiteit Brabant 2000. X + 173 blz. ISBN 90 74029 13 2. f 40,00.

**Susan C. Herring, Pieter van Reenen en Lene Schösler (eds.),** *Textual Parameters in Older Languages*. Amsterdam/Philadelphia: Benjamins, 2000. X + 448 blz. ISBN 90 272 3702 6. f 190,00.

**Hinke van Kampen en Ruud Stumpel,** *Dutch for self-study. Nederlands voor anderstaligen*. Met twee CD's. Utrecht: Het Spectrum, 2001. 272 blz. ISBN 90 274 7379 X.

**Eric Tiggeler,** *Vraagbaak Nederlands. Van spelling tot stijl: snel een helder antwoord op praktijkvragen over taal*. Den Haag: Sdu Uitgevers, 2001. 306 blz. ISBN 90 12 08996 4. f 52,50.

**Veronique De Tier en Ann Marynissen (red.),** *Van de streek. De weerspiegeling van dialecten in familienamen*. Groesbeek: Stichting Nederlandse Dialecten, 2001. 416 blz. ISBN 90 73869 06 4. f 25,00.

**A.A. Weijnen en A.P.G.M.A. Ficq-Weijnen,** *Nederlands*. Utrecht: Het Spectrum, 2001. 565 blz. ISBN 90 274 7204 1. f 16,97.

**J.A.H. van Gent,** *Duits-Nederlands*. Utrecht: Het Spectrum, 2001. 473 blz. ISBN 90 274 7196 6. f 16,97.

**G.A.M.M. van der Linden,** *Nederlands-Duits*. Utrecht: Het Spectrum, 2001. 510 blz. ISBN 90 274 7192 4. f 16,97.

**F.J.J. Baars, M.E. Pieterse-van Baars, en J.G.J.A. van der Schoot,** *Engels-Nederlands*. Utrecht: Het Spectrum, 2001. 558 blz. ISBN 90 274 7198 3. f 16,97.

**G.J. Visser, A.F.M. de Knegt, en C. de Knegt-Bos,** *Nederlands-Engels*. Utrecht: Het Spectrum, 2001. 543 blz. ISBN 90 274 7198 3. f 16,97.

**P.M. Maas en A.M. Maas,** *Frans-Nederlands*. Utrecht: Het Spectrum, 2001. 511 blz. ISBN 90 274 7201 7. f 16,97.

**H.W.J. Gudde,** *Nederlands-Frans*. Utrecht: Het Spectrum, 2001. 517 blz. ISBN 90 274 7202 5. f 16,97.

**L. Schram-Pighi en G. Visser-Boezaardt,** *Italiaans-Nederlands*. Utrecht: Het Spectrum, 2001. 409 blz. ISBN 90 274 7195 9. f 19,60.

**G. Visser-Boezaardt en L. Schram-Pighi,** *Nederlands-Italiaans*. Utrecht: Het Spectrum, 2001. 525 blz. ISBN 90 274 7197 5. f 19,60.

**S.A. Vosters,** *Spaans-Nederlands*. Utrecht: Het Spectrum, 2001. 526 blz. ISBN 90 274 7200 9. f 19,60.

**S.A. Vosters,** *Nederlands-Spaans*. Utrecht: Het Spectrum, 2001. 492 blz. ISBN 90 274 7199 1. f 19,60.

Omvang nu 78 pagina's,  
dus nog twee pagina's over

