



Regionale variatie in de postverbale distributie van presentatief *er*

Stefan Grondelaers*

Dirk Spielman*

An Carbonez**

*Departement Linguïstiek K.U.Leuven

**Universitair Centrum voor Statistiek K.U.Leuven

PUBLICATIEDATUM 14 JUNI 2001

ARTIKELNUMMER 01.04

SAMENVATTING

Met behulp van een uitgebreid corpus formeel en informeel Belgisch en Nederlands Nederlands behandelt deze studie de complexe distributie van niet-anaforisch *er* in bepalingssinitieële presentatieve zinnen zoals *Op de hoek van de straat is (er) een winkel* (verder “bepalingssinnen”). Volgens de ANS zijn voor deze distributie “geen strikte regels te geven. Het kan facultatief zijn, er kan semantisch of stilistisch verschil in het spel zijn, én er is vooral veel individuele, soms ook regionale variatie in het gebruik (1997: 473).” Om deze bewering te toetsen confronteerden we twee taalstructurele factoren – *bepalingstype* en *verbale specificiteit* – en twee contextuele factoren – *regio* en *register* – in een regressie-analyse van *er*’s distributie in de 1905 bepalingssinnen uit het corpus. Die statistische analyse toont dat de ANS inaccuraat en veel te pessimistisch is i.v.m. *er*’s gedrag. Dat de taalstructurele en de contextuele factoren door elkaar genoemd worden in de geciteerde ANS-passage, verhult om te beginnen de veel grotere impact van de taalstructurele factoren op *er*’s distributie. Daarnaast conflicteert het predictieve succes van het *er*-model dat we uit de data distilleren (de Gamma-index is 84.8 %) opvallend met het “geen strikte regels”-pessimisme van de ANS. De belangrijkste conclusie van deze studie is echter dat *er*’s distributie in het Belgische en Nederlandse materiaal door proportioneel en structureel verschillende modellen verklaard wordt. De belangrijkste praktische consequentie van deze bevinding is dat de ANS aparte lemma’s zou moeten wijden aan *er*’s distributie in het Belgische en het Nederlandse Nederlands. Op methodologisch vlak, tenslotte, is deze studie een pleidooi voor het gebruik van niet-geëliciteerde, spontane taaldata bij de studie van syntactische variatie.

SUMMARY

Building on an extensive written corpus of formal and informal Belgian and Netherlandic Dutch, this study tackles the complex distribution of non-anaphoric er “there” in adjunct-initial presentative sentences such as Op de hoek van de straat is (er) een winkel At the corner of the street (there) is a shop. The Dutch standard grammar ANS maintains that for this distribution “no strict rules can be given. It can be optional, there may be semantic or stylistic differences, and there is a lot of individual, sometimes also regional variation (1997: 473).” In order to test this view, we confronted two language-structural factors – adjunct type and verbal specificity – and two contextual factors - region and register – in a regression analysis of the use of er in the 1905 adjunct-initial presentative sentences in the corpus. This statistical analysis demonstrates that the ANS is inaccurate and far too pessimistic as far as presentative er’s postverbal distribution is concerned. The fact that language-structural and contextual factors are put on a par in the quoted passage inadequately reflects the far greater impact of the structural factors on er’s distribution. In addition, the predictive success of the global er-model distilled from the data (the Gamma-index is 84.8 %) is strikingly at odds with the “no strict rules”-pessimism of the ANS. The most important discovery in this paper, however, is that er’s distribution in the Belgian and Netherlandic materials is accounted for by proportionally and structurally different models. The practical consequence of this finding is that the ANS should devote separate entries to er’s distribution in Belgian and Netherlandic Dutch. On a methodological level, this paper argues strongly in favour of spontaneous, non-elicited language data as the empirical basis for the study of syntactic variation.

1. INLEIDING

Hoewel *er* qua omvang één van de meest bescheiden woorden van de Nederlandse taal is, blijkt het aantal linguïstische beschrijvingen van het woordje en de moeizame vooruitgang die in die beschrijvingen geboekt wordt omgekeerd evenredig met die bescheiden gestalte: het aantal vragen i.v.m. *er* is nog altijd veel groter dan de bevredigende antwoorden. De *er*-variant die in weerwil van massieve theoretische aandacht nog altijd het minst begrepen lijkt, is het presentatieve *er* in zinnen zoals (1)-(3)¹:

1. Er waren geen slaappillen in huis. (DC81)²
2. Er gleed nog een laatste scherf uit de lijst. (DA129)
3. Er zijn natuurlijk ook andere middelen. (DC310)

Dat presentatieve *er* kan ook zinsintern verschijnen, zoals in (4)-(6). In dat laatste geval bevindt het zich vlak na de persoonsvorm:

4. a Een paar weken geleden was er ook geen maan. (DA50)
b *Een paar weken geleden was ook geen maan.
5. a In Nederland zijn er meer symfonie-orkesten. (NRC
16/05/1992)
b In Nederland zijn meer symfonie-orkesten.
6. a ?Op de plaats van de Orstkommandatur stond er nu een
nieuw bankgebouw.
b Op de plaats van de Orstkommandatur stond nu een nieuw
bankgebouw. (DA220)

Presentatieve zinnen met de bepaling voorop (verder “bepalingszinnen”) genieten nauwelijks enige theoretische status als onafhankelijke presentatieve constructie; het is dan ook het zogenaamde “optionele” karakter van het postverbale *er* dat taalkundigen in verband met (4)-(6) intrigeert. Nu is de distributie van het presentatieve *er* buiten de eerste zinsplaats inderdaad bijzonder lastig beregelbaar, want de voorkeur voor *er* is maar zelden absoluut. Alleen van (4) kan gesteld worden dat *er* absoluut noodzakelijk is voor alle sprekers van het Nederlands; in de andere zinnen is de noodzakelijkheid van *er* veeleer een kwestie van meer

1 De Algemene Nederlandse Spraakkunst (1997: 464 e.v.) onderscheidt het niet-verwijzende presentatieve *er* van drie verwijzende gebruiken: het locatieve *er* (Woont hij in Den Haag? Hij woont *er* al jaren), het prepositionele *er* (Dat ongeluk is al zo lang geleden gebeurd, ik denk *er* nooit meer aan), en het kwantitatieve *er* (Heeft u ook rozen? Ik heb *er* nog tien).

of minder. Daarnaast lijkt de voorkeur voor *er* ook nog eens regionaal bepaald: in het Belgische Nederlands, zo werd reeds vaker vastgesteld, bestaat een grotere voorkeur voor handhaving van *er*.

Volgens de eerste en de tweede editie van de Nederlandse standaardgrammatica ANS zijn voor de aan- of afwezigheid van het presentatieve *er* buiten de eerste zinsplaats dan ook "geen strikte regels te geven. Het kan facultatief zijn, er kan semantisch of stilistisch verschil in het spel zijn, én er is vooral veel individuele, soms ook regionale variatie in het gebruik" (1997: 473, maar zie ook 1984: 820). Op p. 477 van de tweede editie, specifiek in verband met de weglaatbaarheid van postverbaal *er* in bepalingzinnen, luidt het: "In de standaardtaal wordt *er* na een plaatsbepaling (...) gemakkelijker weggelaten dan in andere gevallen. Overigens verschilt de voorkeur voor het al dan niet gebruiken van *er* buiten de eerste zinsplaats per geval; duidelijke regels zijn hier vooralsnog niet te geven. Er zijn ook nogal grote individuele verschillen in het *er*-gebruik. Wel is er een duidelijke geografisch bepaalde tendentie: *er* staat (al dan niet na een plaatsbepaling) sterker in België (behalve Oost-Limburg) en eveneens, zij het in mindere mate, in het zuidelijk deel van Nederland (vooral Zeeland) dan in de rest van het taalgebied".

Alleen De Rooij (1991) gaat dieper in op mogelijke factoren die het postverbale gedrag van *er* bepalen. In het derde artikel uit een reeks waarin hij regionale variatie in het gebruik van *er* bestudeert, rapporteert De Rooij over een enquête waarmee hij bij Amsterdamse en Leuvense informanten naar de gewenstheid van *er* in zinnen met een plaatsbepaling peilt. Voorbeelden (7)-(13) zijn enkele testzinnen van De Rooij (1991: 119). Het cijfer tussen haakjes kwantificeert de voorkeur voor de aan- of afwezigheid van *er* (+*er* resp. -*er*) bij 26 Amsterdamse (A) en 28 Leuvense informanten (L); het theoretisch maximum is + *er* 4 en -*er* 4.

7. In Italië zijn (er) bergen die vuur spuwen. (A -er 0.69/L +er 2.03)
8. Ook in Suriname waren (er) autoriteiten die dat ingezien hebben. (A +er 0.73/L +er 0.82)
9. Ook in Indonesië is (er) sprake van een "generation gap". (A +er 0.69/L +er 0.64)

2 De afgekorte corpusnamen bij de voorbeelden (1)-(6) verwijzen respectievelijk naar *De Coltmoorden* (Jef Geeraerts 1980), *De Aanslag* (Harry Mulisch 1982) en het NRC Handelsblad.

10. Hier in Athene is (er) een gisting in de atmosfeer. (A +er 0.27/L +er 1.04)
11. Vanmiddag komt (er) iemand een pakje brengen. (A +er 0.15/L +er 0.65)
12. Nee, met kegels wordt (er) niet meer gespeeld. (A -er 0.42/L +er 0.32)
13. Op de hoek van onze straat is (er) een bloemenwinkel. (A -er 2.84/L +er 1.40)

De enquête bevestigt niet alleen dat er regionale verschillen zijn (merk in dat verband de grote discrepantie tussen A en L op in zinnen 7 en 13), ze laat tevens zien dat (i) de verschillen qua *er*-voorkeur zelfs tussen de verwante zinnen (7)-(10) - die alle een eigenaam in de plaatsbepaling hebben - vrij groot zijn en (ii) dat zinnen met een niet-plaatsbepaling voorop niet noodzakelijk meer *er* nodig hebben dan zinnen met een plaatsbepaling (want De Rooijs zinnen 11 en 12 komen in de hiërarchie van “*er*-noodzakelijkheid” maar op plaats 12 en 13 van de 18). Een en ander bewijst volgens De Rooij duidelijk dat het plaatsaanduidende karakter van de bepaling niet de enige factor kan zijn die invloed uitoefent op de aan- of afwezigheid van *er*. Een mogelijke andere factor naast de aard van de bepaling is volgens De Rooij de aard van het werkwoord: zo blijken zinnen met het verbum finitum *zijn* over het algemeen moeilijker zonder *er* te kunnen (idem: 124). Ondanks dit alles besluit De Rooij dat “nog steeds geldt wat in de ANS over repletief *er* buiten de eerste zinsplaats in het algemeen wordt opgemerkt: “Voor de aan- of afwezigheid van *er* zijn geen strikte regels te geven” (ANS 1984:820). Misschien kunnen de hier vermelde enquêteresultaten en de mogelijke aanzetten tot verklaring daarvan als uitgangspunten en werkhypothesen dienst doen voor een werkelijk grondig onderzoek van dit probleemcomplex” (1991:127).

In dit artikel trachten we tegemoet te komen aan deze verwachting door de drie parameters die *er*'s postverbale distributie blijktens De Rooijs enquêtegegevens beïnvloeden – *regio* (*er* staat sterker in België dan in Nederland), *bepalingstype* (*er* komt vaker voor na vooropgeplaatste tijdsbepalingen dan na plaatsbepalingen) en *werkwoordelijke specificiteit* (*er* komt vaker voor na het hoofdwerkwoord *zijn* dan na specifiekere werkwoorden) – als uitgangspunt te nemen voor een grondig kwantitatief onderzoek op basis van, zoals De Rooij zelf voorstelt (o.c.: 116-117), “een (omvangrijk) corpus waaruit *alle* zinnen met een [plaats]bepaling op de eerste zinsplaats geselecteerd zijn, met en zonder *er*”.

Dergelijk empirisch onderzoek – zo zal nog blijken – levert evenmin de strikte regels of glasheldere onderscheidingen op die de ANS graag aan haar lezers aanbiedt; aan het einde van dit stuk houden we een op tendensen gebaseerd model over dat *er*'s gedrag echter veel preciezer kan voorspellen dan het op introspectie en enquêtes gebaseerde ANS-lemma. Dit artikel heeft dan ook de bijkomende ambitie aan te tonen dat corpusonderzoek een interessante aanvulling biedt op introspectief verworven gegevens. Op het vlak van de syntactische variatie is introspectie namelijk een riskante onderneming, en dat geldt ook voor de collectieve introspectie die een enquête is: onze intuïties zijn maar zelden zo ijl en instabiel als bij de beslissing of een bepalingszin nu wel of niet *er* moet bevatten. Het is in dit opzicht interessant dat Labov (1972: 103) fonologen die een beroep doen op intuïties van native speakers bij het beoordelen van minimale klankenparen waarschuwt voor twee types afwijkingen die die gegevens kunnen corrumpen: “(...) native speakers can make distinctions in minimal pairs that they do not make in actual speech: normative responses often preserve fanciful, archaic or mythical distinctions. Omgekeerd, “(...) native speakers can fail to recognize or register distinctions which they regularly make in natural speech (Labov 1970): their self-reports often reflect the patterns of younger speakers rather than their own, or blur regular phonetic distinctions that are too subtle to withstand the glare of conscious examination”. Om dezelfde reden hanteren Geeraerts, Grondelaers & Bakema (1994: 17 e.v.) bij hun onderzoek naar lexicale variatie corpusdata, want “what people think they do with words is not necessarily the same as what they actually do, in the sense that our conscious awareness of the flexibility with which we use the resources of the language may well be rather restricted”.

In de beide geciteerde gevallen betreft de bestudeerde variatie de niet-graduele, binaire keuze tussen twee identificeerbare en onderscheidbare (fonetische of lexicale) varianten. Bij onze *er*-variatie daarentegen wordt van de informanten een oordeel gevergd over de flexibiliteit waarmee ze een grammaticaal morfeem van de allervluchtigste substantie hanteren, vaak in zinnen waarvan de grammaticaliteit noch de betekenis ingrijpend verandert met de toevoeging of verwijdering van *er*. alleen in zinnen zoals (4) en (6) is de aanwezigheid van *er* grammaticaal vereist of overbodig; de grote meerderheid van bepalingszinnen vormt een schemerzone tussen die comfortabele extremen waarin *er*'s aanwezigheid een kwestie van meer of minder is. En over exact hoeveel meer of minder kunnen taalgebruikers nauwelijks bewust en accuraat rapporteren. De Rooijs belangrijkste

conclusie is wellicht dat naast de grote verschillen in beoordeling tussen de zinnen “ook de verschillen per informant groot zijn” (o.c.:128). Ook de enquête waarover in Grondelaers (2000: 256) gerapporteerd wordt, verdringt massieve inter-participantvariantie ondanks het rigide design alle andere conclusies.

In Grondelaers & Brysbaert (1996) en Grondelaers (2000: 197-258) werd aangetoond dat de methodologische gevaren van introspectie in een zorgvuldig geconstrueerd psycho-experimenteel design nagenoeg volledig onder controle gehouden kunnen worden. In dit artikel willen we zoals gezegd demonstreren dat de observatie van niet-geëliciteerd taal materiaal in de vorm van een corpusanalyse een interessante aanvulling op introspectieve gegevens biedt. Dat laatste betekent geenszins dat corpusgebaseerd onderzoek geheel gevaarloos is. Bij de beschrijving van de gebruikte materiaalverzameling in paragraaf 2 wijzen we op een aantal “valkuilen” waarmee we bij de samenstelling van het corpus geconfronteerd werden, en bespreken we de veiligheidsmaatregelen die we in acht genomen hebben om die valkuilen te vermijden.

We organiseren het materiaal in dit artikel als volgt. In de volgende sectie lichten we de structuur van het corpus toe waarop het onderzoek gebaseerd is, met bijzondere aandacht voor de Internettaal in de materiaalverzameling. In de derde sectie bekijken we de invloed van de factoren *regio*, *bepalingstype* en *werkwoordelijke specificiteit* afzonderlijk op basis van absolute frequenties en proporties, en identificeren we een vierde relevante factor, i.e. *register*. In sectie 4 introduceren we vervolgens de logistieke-regressieanalyse, waarmee niet alleen de invloed van afzonderlijke factoren gemeten en vergeleken kan worden, maar ook eventuele interacties tussen de factoren. De regressiegegevens beantwoorden in sectie 5 de belangrijkste vragen die in verband met *er*'s postverbale distributie gesteld kunnen worden. Om na te gaan of het pessimisme van de ANS en van De Rooij omtrent de beregelbaarheid van die distributie gerechtvaardigd is, bekijken we om te beginnen het *er*-voorspellende succes van een globaal verklaringsmodel met afzonderlijke factoren en interacties tussen factoren. Belangrijker nog is de vraag naar eventuele regionale variatie: op basis van de regressiegegevens kan nagegaan worden of de *er*-variatie in het Belgische tekstmateriaal verklaard kan worden met hetzelfde model als de *er*-variatie in het Nederlandse materiaal; als dat niet het geval is, dan moeten voor de Belgische zowel als voor de Nederlandse *er*-distributie aparte regelsystemen ontworpen worden. In sectie 6

vatten we samen, en formuleren we de taalnormatieve consequenties van onze bevindingen in termen van een ANS-regelsysteem.

2. DE MATERIAALVERZAMELING

Voordat we de structuur en de samenstelling van het corpus toelichten dat voor dit onderzoek gebruikt werd, moeten we ingaan op de vraag of moeizaam verworven corpusgegevens zoveel waardevoller zijn dan de introspectieve oordelen van de getrainde linguïst. Weegt de wetenschappelijke winst die met het gebruik van corpusgegevens geboekt wordt op tegen het bijzonder arbeidsintensieve karakter van het verzamelen ervan? Corpus-sceptici zoals Verkuyl (1998:63 e.v.) vinden alvast van niet, vooral niet wanneer de verzamelde gegevens kwantitatief benut worden voor theoretische doeleinden, bijvoorbeeld in frequentietellingen ter ondersteuning van introspectief veronderstelde tendensen. De onderzoeker heeft volgens Verkuyl namelijk geen enkele controle over de aard en de kwaliteit van de teksten in het corpus³; bovendien kunnen de fenomenen die hij op corpusbasis tracht te bestuderen best “idiosyncrasieën” zijn van de taalredacteurs- en correctoren van de kranten waaruit het corpus in hoofdzaak bestaat.

Dat we onze corpusgegevens in dit artikel toch statistisch zullen aanwenden betekent geenszins dat we Verkuyls waarschuwingen in de wind slaan. De structurele risico's die hij noemt zijn echter beide vermijdbaar. Zo doen we om te beginnen geen beroep op reeds beschikbare materiaalverzamelingen zoals de INL-corpora, maar maken we gebruik van het CONDIV-corporus, een omvangrijke tekstenbank die de eerste auteur (mede) samenstelde ten behoeve van het VNC-project “Convergentie en divergentie in de Nederlandse woordenschat”, waarvan hij de Leuvense uitvoerder is. Tabel 1 bevat een overzicht van de structuur en de omvang van het CONDIV-corporus:⁴

3 Verkuyl verwijst daarbij specifiek naar het INL-corporus.

4 De eerste auteur dankt zijn mede-uitvoerders Katrien Deygers, Hilde Van Aken en Vicky Van den Heede en hun projectleiders voor hun toestemming om delen van het corpus te mogen gebruiken voor eigen onderzoek. Meer informatie over het CONDIV-corporus is te vinden in Grondelaers, Deygers, Van Aken, Van Den Heede & Speelman (2000), en in Grondelaers, Van Aken, Speelman & Geeraerts (ter perse).

		zeer informeel (geen redactionele controle)		zeer formeel (veel redactionele controle)		
		Internet		Kranten		
		IRC	Usenet	Massakranten		Kwaliteitskranten
				Regionaal	Nationaal	
N	(6.965.291)	(7.748.436)	<i>De Limburger</i> (1.680.636)	<i>De Telegraaf</i> (1.590.581)	<i>NRC Handelsblad</i> (1.520.064)	
B	(8.207.007)	(4.980.780)	<i>Het Belang van Limburg</i> (3.012.330)	<i>Het Laatste Nieuws</i> (3.486.911)	<i>De Standaard</i> (3.228.910)	
			<i>De Gazet van Antwerpen</i> (3.068.405)			

Tabel 1
Overzicht van de geëxcerpeerde bronnen en hun omvang

Het CONDIV-corpus bestaat uit twee soorten taalmateriaal. In de sectie krantentaal hebben we een onderscheid gemaakt tussen *kwaliteitskranten* zoals *De Standaard* en het *NRC Handelsblad* – die doorgaans voor de maatschappelijke en culturele bovenlaag van het krantenlezende publiek bedoeld zijn –, en *massakranten* zoals *De Telegraaf* en *Het Laatste Nieuws* – die voor een breder en minder veeleisend lezerspubliek bestemd zijn. Er zijn *nationale* massakranten – zoals *Het Laatste Nieuws* en *De Telegraaf* –, maar ook *regionale* massakranten⁵ zoals *Het Belang van Limburg*, *De Gazet van Antwerpen* of *De Limburger*, die zich thematisch vooral tot het nieuws in één regio beperken.

Naast krantenmateriaal bevat het CONDIV-corpus ook Internettaal. Uit Geeraerts, Grondelaers & Spielman (1999) weten we dat het incorrect is het Nederlands als een monostrataal gegeven te beschouwen. Tussen standaardtaal en dialect bevindt zich namelijk tenminste één tussenniveau waarop een hogere graad van informaliteit samengaat met een toenemende mate van geografische specialisatie: naarmate taalgebruikers in minder formele communicatiesituaties terechtkomen drukken ze zich in een regionaler klinkend register uit dat nochtans niet helemaal dialectisch is. In recente publicaties wordt dat register onder

5 In verband met de Belgische en Nederlandse regionale massakranten moet opgemerkt worden dat de eerste wel maar de tweede (meestal) niet nationaal verspreid worden.

meer “tussentaal” (Taeldeman 1992: 33-52), “verkavelings-Vlaams” (Van Istendael 1993: 116), of “soap-Vlaams” genoemd (Geeraerts 1999: 232)⁶.

Omdat het onmogelijk is a priori vast te stellen hoeveel tussenstrata precies onderscheiden moeten worden, gaan we niet in de eerste plaats op zoek naar taalgebruik dat mogelijk representatief is voor een bepaald stratum, maar vertrekken we veeleer van verschillende types *taalsituaties* als onafhankelijke variabele, en beschouwen we het daarin gebruikte *taalgebruik* als afhankelijke variabele. Stratificationeel-stilistische variatie accommoderen we dan door taal materiaal te vergelijken uit vijf stilistisch verschillende taalsituaties, die we bovenaan in tabel 1 kunnen uitzetten op een as van “zeer informeel” tot “zeer formeel”. De drie types krantenmateriaal – regionale massakranten, nationale massakranten en kwaliteitskranten – nemen op die informaliteitsas de hoogste plaatsen in. Recht evenredig met de formaliteit van een taalsituatie is overigens de mate van redactionele controle op de taalproductie in die situatie. Die is in de kranten uiteraard het grootst, omdat een krantenartikel door eindredacteuren of taalcorrectoren nagelezen wordt die in sommige gevallen ingrijpen in spelling en formulering. De nadelige gevolgen van die ingrepen voor de corpusrepresentativiteit beperken we in de eerste plaats door aan de materiaalverzameling twee types spontaan tot stand gekomen, niet-gecontroleerde of -gereviseerde Internettaal toe te voegen, die tevens als vertegenwoordigers van het informele Nederlands fungeren.

IRC – Internet Relay Chat – is een module waarmee on-line en synchroon gecommuniceerd wordt in babbelkanalen zoals #Vlaanderen of #Leuven.⁷ Vermits IRC-gebruikers hun schriftelijke communicatie zoveel mogelijk trachten aan te passen aan de wetten van gesproken conversatie, genereert IRC een “geschreven gesproken Nederlands” dat bukt van dialectismen en spreektaalfenomenen. Nog afgezien van het feit dat IRC een zeer informele taalvariant genereert, is het voor taalkundigen ook interessant omdat het homogeen Nederlandse en Belgische corpora oplevert; een typisch kenmerk van IRC is namelijk regionaal antagonisme tussen Nederlandse en Belgische gebruikers, die aparte

6 Voor zover we kunnen nagaan bestaan er geen specifieke benamingen voor het Noord-Nederlandse tussenregister: de term “Poldernederlands” (Stroop 1998) is niet helemaal equivalent met de opgegeven termen omdat hij voornamelijk naar een substandaard uitspraakvariant van het Nederlandse Nederlands verwijst.

7 IRC is on-line en *synchroon* omdat de gebruiker onmiddellijk kan repliceren op een boodschap die hij ziet verschijnen in het tekstvenster van het babbelkanaal waarop hij ingelogd is.

kanalen hebben waaruit ze vreemde gebruikers gewoonlijk weren (voor een overzicht van de sociologische en linguïstische kenmerken van Internet Relay Chat, zie onder meer Ko (1996), Bays (1998), Hentschel (1998), Paolillo (1999) en vooral Herring (1996)).

In de internetmodule *Usenet* wordt in verschillende “newsgroups” offline⁸ en asynchroon gedebatteerd door middel van e-mails die de gebruiker aan een bestaande “thread” van e-mailboodschappen toevoegt. Omdat e-mail niet on-line is, waardoor debatteerders hun bijdrage kunnen nalezen alvorens ze te versturen, is het talige register dat op Usenet gebruikt wordt veel minder informeel dan op IRC.⁹ Voorbeeld (14) bevat een uittreksel uit het Belgische discussieforum “be.politics”:

```
14. {CD:er}
{CD:br}
{CD:bc}Newsgroups: be.politics,nl.politiek,soc.culture.belgium
{CD:bc}Subject: Re: klacht tegen Wim Elbers
{CD:bc}From: paul@indigoproductions.be (Paul Goris)
{CD:bc}Date:veThis_>364F0C5A.7CCD@l.vb.net
<3653337a.106881606@news.online.be>
<72p186$c3m$1@news3.Belgium.EU.net>
<365067b4.24874836@news.online.be>
<72rgs5$jkb$6@mach.vub.ac.be>
In article <72rgs5$jkb$6@mach.vub.ac.be>, apardon@rc4.vub.ac.be
(Antoon Pardon) wrote:
```

```
[.....]
{CD:bc}>
{CD:bc}>Je kan niet alles via de wet oplossen. Als je alles
{CD:bc}>waarvan je denkt dat het verboden zou moeten zijn
{CD:bc}>inderdaad gaat verbieden dan heb je ofwel ook een hele
{CD:bc}>hoop dingen verboden waarvan mensen gaan denken hoe
{CD:bc}>men het in hemelsnaam in zijn hoofd kon halen om zoiets
{CD:bc}>te verbieden ofwel heb je zo'n gedetailleerde lijst van wat
{CD:bc}>mag en wat niet mag dat niemand nog gerust kan zijn dat
{CD:bc}>wat hij doet legaal is want er zou maar eens ergens een
{CD:bc}>element in de lijst zijn waar men net niet aan denkt.
```

8 Usenet is *offline* en *asynchroon* omdat er “propagation delay” (Paolillo 1999) is tussen het opstellen, nalezen, verzenden, ontvangen, lezen en beantwoorden van een e-mailboodschap.

9 Over de linguïstische aspecten van Usenet en andere discussieforums, zie onder meer Harrison (1998).

In het algemeen is *dit* waar. Volgens de logica volgt de onoirbaarheid van racisme uit andere wetten. Door de veelheid van inbreuken of een speciale sociale relevantie op een gegeven tijdstip kan het invoeren/handhaven van een antiracisme wet beargumenteerd worden. Voor mij is het niet essentieel.

Het onderzoek naar *er*'s postverbale distributie is niet op het gehele CONDIV-corpus gebaseerd. Om te beginnen werd het IRC-materiaal uit de analyse geweerd. De typische interactionele kenmerken van IRC - "temporality and immediacy" (Bays 1998) – vereisen namelijk specifieke formuleringstechnieken van de IRC-ers "to augment the speed and the capacity of information transfer" (idem). De belangrijkste strategieën in dat opzicht zijn "abbreviation, elipsis and a telegraphic style, which reduce the quantity of words that need to be typed, sent and read" (idem, maar zie ook Hentschel 1998). Het is niet onwaarschijnlijk dat veel *er*'s sneuvelen in dat streven naar een gecompriëerde, telegrafische stijl (in het bijzonder daar waar ze niet strikt nodig zijn voor grammaticaliteit),¹⁰ en daarom beperken we de analyse tot bronnen waarin factoren zoals productiesnelheid geen rol spelen.

Om het gevaar op "taalcorrectoridiosyncrasieën" waarvan Verkuyl (1998) gewag maakt verder te beperken, maken we in deze analyse bovendien alleen gebruik van taal materiaal uit kranten die ons desgevraagd uitdrukkelijk verzekerden geen specifiek taalbeleid t.a.v. het gebruik van *er* te voeren. De Belgische kranten – behalve *De Gazet van Antwerpen* (die geen antwoord gaf op onze vraag) – bleken geen van alle *er*-vijandig. Bij de Nederlandse kranten kregen we alleen van *De Telegraaf* en *NRC* een antwoord; de respectieve taalredacteurs verzekerden ons dat in hun bladen geen speciale aandacht aan postverbaal *er* geschonken wordt.

Tenslotte moet nog vermeld worden dat we het Belgische en het Nederlandse Usenet-corpus terwille van de onderlinge kwantitatieve vergelijkbaarheid en de vergelijkbaarheid met het krantenmateriaal beperken tot forums over cultuur, politiek, wetenschap, sport en televisie. Het Nederlandse Usenet-corpus bevat dan 2.287.648 woorden, het Vlaamse 2.449.193.

10 Grondelaers (2000:193-196) bevat evidentie waaruit blijkt dat het *er*-gebruik in babbelkanalen wel degelijk de sporen draagt van de telegrammatische en elliptische stijl van babbelkanalen.

Uit dit gereduceerde corpus extraheerden we met behulp van Dirk Speelmans *Abundantia Verborum* (1997)¹¹ de targetconstructie; omdat alle bepalingsszinnen met een prepositionele bijwoordelijke bepaling zoals *in de gracht* of *op het dak* of met een afzonderlijk bijwoord zoals *overal* of *morgen* beginnen – en preposities zowel als bijwoorden een niet al te omvangrijke gesloten klasse van “onveranderlijke” woorden (ANS 1997: 451) vormen –, en omdat die bijwoordelijke bepalingen en bijwoorden in bepalingsszinnen nagenoeg altijd door een leesteken of een nevenschikkend voegwoord voorafgegaan worden (eveneens beperkte categorieën van onveranderlijke tekens), konden we een digitale query formuleren waarmee uit het corpus 1905¹² bepalingsszinnen geëxtraheerd werden. De distributie van de geëxtraheerde observaties over de verschillende brontypes is weergegeven in tabel 2.

	Usenet	Massakranten		Kwaliteitskrante n
nl	n = 192	<i>Telegraaf</i> n = 227		<i>NRC</i> n = 263
be	n = 225	<i>Het Belang van Limburg</i> n = 397	<i>Het Laatste Nieuws</i> n = 198	<i>De Standaard</i> n = 403

Tabel 2
Distributie van de observaties over de verschillende brontypes

Merken we, alvorens we in de volgende paragrafen de analyse van de afzonderlijke factoren aanvatten, nog op dat de afhankelijke variabele tenzij anders aangegeven de postverbale aanwezigheid van *er* is, gekwantificeerd als de ratio tussen de absolute frequentie van de bepalingsszinnen in een bron of bronnengroep die met postverbaal *er*

11 *Abundantia Verborum* bevat een module om digitale bestanden tot bruikbare linguïstische corpora te transformeren, een zoekmachine die de targetdata uit die corpora extraheert, een module om de resulterende gegevens te labelen, classificeren en manipuleren, en een component waarin statistische analyses op de gegevens uitgevoerd kunnen worden. Meer informatie over *Abundantia Verborum* is te vinden op <http://www.ling.arts.kuleuven.ac.be/genling/abundant>.

12 We beperkten ons bij de selectie van bepalingsszinnen tot actieve hoofdzinnen die – als ze *er* bevatten – de ondubbelzinnig presentatieve variant bevatten. De rechtvaardiging van deze beperking is te vinden in Grondelaers (2000: 81-85). De efficiëntie van de gebruikte query leiden we af uit het feit dat uit een digitale versie van Harry Mulisch' *De Aanslag* 114 van de 118 bepalingsszinnen die bij vroeger onderzoek (cf. Grondelaers & Brysbaert 1996) manueel geëxcerpeerd werden, correct geëxtraheerd worden door de query. De vier niet-geëxtraheerde zinnen werden voorafgegaan door kwantificerende bijwoorden (*iets* en *wat*) of door het focuspartikel *ook*.

geattesteerd werden, en de totale frequentie van de bepalingssinnen in die bron of bronnengroep. De relatieve frequentie van *er* die we aldus berekenen biedt een indirecte statistische maat voor de flexibiliteit waarmee *er* in een bepaalde regio, bron, bronnengroep of stijlregister gebruikt wordt.

Daarnaast moet nog gewezen worden op een fundamenteel onderscheid tussen de “taalstructurele” factoren *bepalingstype* en *werkwoordelijke specificiteit*, en de “contextuele” factor *regio*; de respectievelijke gevolgen van beide types factoren voor de analyse verschillen namelijk aanzienlijk. Als bijvoorbeeld blijkt dat de factor *werkwoordelijke specificiteit* invloed uitoefent op de voorkeur voor *er*, dan zijn we één stap dichterbij ons einddoel, i.e. een regelsysteem voor *er*'s distributie. Indien daarentegen blijkt dat *er*'s distributie contextueel bepaald is, dan impliceert dat gewoon dat ons regelsysteem slechts voor een beperkt taalgebied of stijlregister geldt, en dat compliceert de situatie aanzienlijk. Nemen we bijvoorbeeld de factor *regio*: als in de volgende paragrafen aan het licht zou komen dat *er* in Nederland niet enkel minder – zoals algemeen aangenomen wordt – maar ook anders gebruikt wordt dan in België, dan impliceert dat in feite dat we om alle variatie te beregelen een Nederlands naast een Belgisch model nodig hebben. Dat betekent op zijn beurt dat de postverbale distributie van *er* in België en Nederland in afzonderlijke ANS-lemma's beschreven moet worden, zoals momenteel al het geval is bij de beschrijving van de aard van de hoeveelheidsaanduiding bij kwantitatief *er* (ANS 1997: 480-485).

3. DE DRIE FACTOREN AFZONDERLIJK BESCHOUWD

3.1. Regio

	Nederland		België	
	abs	rel	abs	rel
-er	383	56,16	603	49,3
+er	299	43,84	620	50,7
totaal	682		1223	

Tabel 3

De distributie van bepalingszinnen met en zonder *er* in het Nederlandse en het Belgische materiaal

Als aanvulling op de geëliciteerde gegevens die suggereren dat postverbaal *er* sterker staat in het Belgische dan in het Nederlandse Nederlands, bevat tabel 3 de absolute en relatieve frequentie van *er* in het Nederlandse en het Belgische corpus, dat zoals gezegd uit niet-geëliciteerd, spontaan taal materiaal bestaat. Op de verticale as contrasteren we *er* met zijn afwezigheid, op de horizontale as plaatsen we de Nederlandse en de Belgische bronnen tegenover elkaar.

In ons materiaal blijkt de frequentie van *er* in de Belgische bronnen wel degelijk hoger dan in de Nederlandse bronnen (50.7 % > 43.84 %); het verschil is in een chi-kwadraattest significant op het .01-niveau. Voorlopig bevestigen onze gegevens dus de traditionele taalkundige opvatting over de prominentere status van *er* in België. Die gegevens moeten echter op twee belangrijke manieren genuanceerd worden. Om te beginnen lijkt het reële verschil tussen Nederland en België qua *er*-gebruik in ons materiaal – 6.86 % – veel minder uitgesproken dan de regionale verschillen in De Rooijs (1991: 123) enquêtegegevens. We gaan hier niet verder in op de kwantitatieve maat waarin De Rooij zijn gegevens vertaald heeft, maar het valt wel op dat de Leuvense voorkeur voor *er* in tenminste 70 % van zijn materiaal twee keer zo groot is als de Amsterdamse. Onze gegevens manifesteren niet die uitgesproken verschillen. De methodologische gevaren van enquête-elicitering in acht genomen lijkt het dus enigszins voorbarig om op basis van De Rooijs gegevens in de tweede

editie van de ANS (1997: 477) gewag te maken van “een duidelijke geografisch bepaalde tendentie: *er* staat (al dan niet na een plaatsbepaling) sterker in België (...).”¹³

Daarnaast is het *er*-aandeel ook niet constant over de bronnen of bronnengroepen die het Belgische en het Nederlandse materiaal constitueren. De tabellen 4 en 5 bevatten respectievelijk de absolute en relatieve frequentie van *er* in de Nederlandse en de Belgische bronnen¹⁴:

	totaal		usenet		massa		kwal	
	abs	rel	abs	rel	abs	rel	abs	rel
-er	383	56,2	99	51,6	121	53,3	163	62,0
+er	299	43,8	93	48,4	106	46,7	100	38,0
totaal	682		192		227		263	

Tabel 4

Absolute en relatieve frequenties van *er* in het totale Nederlandse materiaal, en per aparte Nederlandse bron

	totaal		usenet		massa		kwal	
	abs	rel	abs	rel	abs	rel	abs	rel
-er	603	49,3	76	33,8	278	46,7	249	61,8
+er	620	50,7	149	66,2	317	53,3	154	38,2
totaal	1223		225		595		403	

Tabel 5

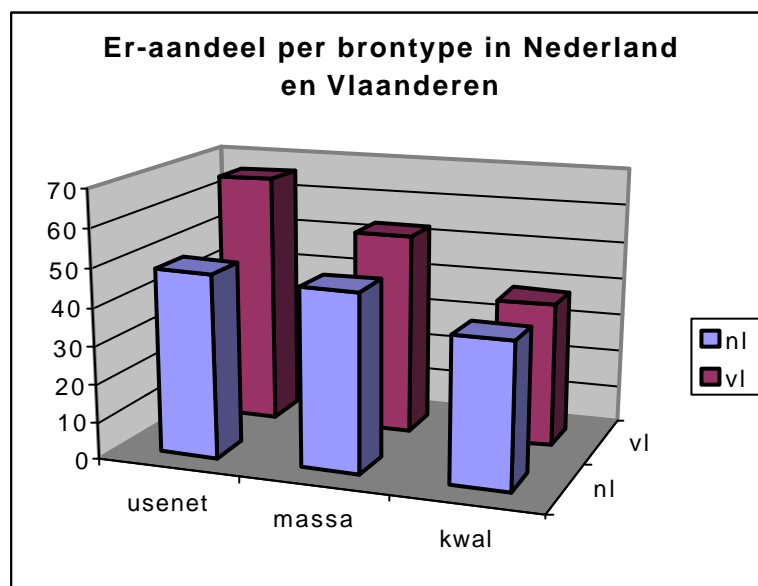
Absolute en relatieve frequenties van *er* in het totale Belgische materiaal, en per aparte Belgische bron (*Het Belang van Limburg* en *Het Laatste Nieuws* werden samengevoegd)

In figuur 1 wordt de verhouding tussen de Belgische en Nederlandse *er*-aandelen in de drie brontypes aanschouwelijk weergegeven. Twee tendensen vallen meteen op. Van links naar rechts in figuur 1 is er voor beide regio's een uitgesproken dalende *er*-tendens.

13 In de eerste editie van de ANS wordt slechts gewag gemaakt van een regionale voorkeur voor handhaving van *er* (1984: 822). De veel concretere substantiëring van die tendens in de tweede editie van de ANS is – voor zover we dat kunnen nagaan – volledig gebaseerd op de bevindingen in De Rooij (1991).

14 In tabel 5 zijn de Belgische massakranten *Het Belang van Limburg* en *Het Laatste Nieuws* voor het gemak van de regionale vergelijking van brontypes samengevoegd.

Daarnaast wordt de afstand tussen de Nederlandse en de Belgische *er*-voorkeur van links naar rechts gaandeweg kleiner: bedraagt het verschil tussen het *er*-aandeel in het Belgische en het Nederlandse Usenet-materiaal nog 17.8 % (significant op het .001-niveau in een chi-kwadraattest), tussen de Belgische en Nederlandse massakranten zit er slechts 6.6 % verschil (significant op het .01-niveau). In de Belgische en Nederlandse kwaliteitskranten is het *er*-aandeel even groot. Door de aanvankelijk grotere afstand tussen de *er*-aandelen in Nederland en België, is de dalende *er*-tendens aan Belgische kant veel opvallender.



Figuur 1
Verhouding tussen de Nederlandse en Belgische *er*-aandelen in de drie brontypes

Het ziet er dus naar uit dat de voorkeur voor *er* in de Nederlandse maar vooral in de Belgische bronnen deels *stilistisch* bepaald is: de voorkeur voor *er* daalt naarmate de formaliteit van een bron of bronnengroep toeneemt. Hoewel we de taalnormatieve gevolgen van de kwantitatieve gegevens pas in sectie 5 grondig onder de loep zullen nemen, bespreken we in dit verband al een mogelijke aanvulling op de laatste paragraaf van het lemma dat de ANS (1997: 477) aan de postverbale aanwezigheid van *er* in zinnen met een bepaling op de eerste zinsplaats wijdt; in zijn huidige vorm stelt die paragraaf dat er een “duidelijke geografisch bepaalde tendentie [is]: *er* staat (al dan niet na een plaatsbepaling) sterker in België (behalve Oost-Limburg) en eveneens, zij het in mindere mate, in het

zuidelijk deel van Nederland (vooral Zeeland) dan in de rest van het taalgebied.” Op basis van de voorlopige bevindingen dient die paragraaf als volgt herschreven te worden:

“Wel is er een geografische en een stilistische tendentie: *er* staat niet alleen sterker in België dan in Nederland, maar komt ook vaker in informeel dan in formeel register voor. Globaal gezien is het verschil in *er*-voorkeur tussen Nederland en België het sterkst op substandaardtaalniveau; in de standaardtaal is het gebruik van *er* in België en Nederland nagenoeg vergelijkbaar.”

3.2. Bepalingstype

De enige factor die door nagenoeg alle beschrijvers als een determineerder van *er*'s postverbale distributie beschouwd wordt is het type vooropgeplaatste bepaling. De ANS merkt in dit verband op dat “als de bepaling op de eerste zinsplaats een plaatsbepaling is, *er* in de standaardtaal bij voorkeur weggelaten [wordt]” (ANS 1984: 822). In de tweede editie (1997: 477) is die opinie afgezwakt tot: “in de standaardtaal wordt *er* na een plaatsbepaling (...) gemakkelijker weggelaten dan in andere gevallen.”¹⁵

Om de hypothese te testen dat *er* makkelijker achterwege blijft in zinnen met een locatieve bepaling voorop dan in zinnen met een andere bepaling, worden de geattesteerde zinnen in plaatsbepalingszinnen en temporele bepalingzinnen geclassificeerd¹⁶. De tabellen 6 en 7 bevatten de absolute en relatieve frequenties van *er* per parameterwaarde in resp. het Nederlandse en het Belgische materiaal.

15 Voor vergelijkbare visies op het *er*-inhiberende karakter van locatieve bepalingen, zie Bech (1952: 18), Paardekoper (1971: 60), Van Es & Van Caspel (1971: 79 e.v.), Jordens (1974: 168 en noot 8), De Schutter (1974:347), De Schutter & Van Hauwermeiren (1983: 84), Zwart (1990: 484 noot 14), De Rooij (1991: 114) en Barbier (1993: 10).

16 Uit geattesteerde zinnen zoals *Bij die ramp vielen 34 doden (hbvl/5400)* blijkt dat het onderscheid tussen locatief en temporeel niet binair is. *Bij die ramp* kan niet zonder meer als locatief of temporeel gecategoriseerd worden, want de prepositie roept een locatieve interpretatie ("op de plaats van de ramp") zowel als een temporele interpretatie ("tijdens de ramp") op. Omdat het weinig frequent voorkomt – en dus oninteressant is als individuele parameterwaarde – verwijderen wij dit intermediaire bepalingstype uit de analyse. Zinnen met niet-situerende bepalingstypes werden a priori uit de materiaalverzameling verwijderd.

	totaal			usenet			massa			kwal		
	-er	+er	% er	-er	+er	% er	-er	+er	% er	-er	+er	% er
temp	218	84,2	3	59	95,2	19	82	81,2	19	77	80,2	
loc	342	81	19,1	96	34	26,2	102	24	19,0	144	23	13,8
totaal	383	299	43,8	99	93	48,4	121	106	46,7	163	100	38,0

Tabel 6

Absolute en relatieve frequenties van *er* per bepalingstype in het gehele Nederlandse materiaal en per aparte Nederlandse bron

	totaal			usenet			massa			kwal		
	-er	+er	% er	-er	+er	% er	-er	+er	% er	-er	+er	% er
temp	76	357	82,4	9	72	88,9	36	192	84,2	31	93	75,0
loc	527	263	33,3	67	77	53,5	242	125	34,1	218	61	21,9
totaal	603	620	50,7	76	149	66,2	278	317	53,3	249	154	38,2

Tabel 7

Absolute en relatieve frequenties van *er* per bepalingstype in het gehele Belgische materiaal en per aparte Belgische bron

De globale gegevens in de *totaal*-kolommen van tabellen 6 en 7 bevestigen zelfs pre-statistisch het uitgesproken effect van locatieve bepalingen op de distributie van *er* in bepalingstypen. Temporele bepalingstypen bevatten in maar liefst 84.2 en 82.4 % van de geattesteerde Nederlandse en Belgische bepalingstypen een postverbaal *er*, terwijl locatieve bepalingstypen in slechts 19.1 % en 33.3 % *er* bevatten.

Blijkens de gegevens in deze tabellen beperkt de regionale variatie die in sectie 3.1 geïdentificeerd werd zich vooral tot locatieve bepalingstypen: het grootste deel van de *er*-variatie aan Nederlandse kant wordt verklaard door de kwaliteit van de zinsinitiale bepaling, en dat – zo suggereert de *totaal*-kolom van tabel 6 – op opvallend binaire en symmetrische wijze: locatieve zinnen bevatten in slechts 19.1 % van alle gevallen *er*, en een vergelijkbare portie temporele zinnen (15.8 %) bevat geen *er*. Aan Belgische kant daarentegen wordt de *er*-variatie blijkens tabel 7 veel minder goed verklaard door de bepalingstypenfactor, en de symmetrie waarvan sprake in het Nederlandse materiaal is opvallend afwezig: net zoals Nederlandse temporele bepalingstypen bevatten Belgische zinnen met een temporele

bepaling in ongeveer 16 % van de geattesteerde gevallen geen *er*, maar Belgische plaatsbepalingszinnen worden in maar liefst 33.3 % van de geattesteerde gevallen wél met *er* geconstrueerd. Belgische bepalingzinnen hebben m.a.w. niet genoeg aan een locatieve bepaling om zonder *er* te kunnen.

De regionale variatie beperkt zich dus niet enkel tot het informele stijlregister, maar ook tot het locatieve zinstype; het beeld van de Noord/Zuid-variantie op het gebied van de voorkeur voor *er* wordt dus progressief complexer met de toevoeging van nieuwe factoren.

3.3. Werkwoordelijke specificiteit

Naast het effect van *regio* en *bepalingstype* op *er*'s frequentie wordt in de *er*-literatuur ook gewag gemaakt van de mogelijke *er*-inhiberende invloed van werkwoorden die het proces of de relatie die ze coderen met enige graad van precisie beschrijven. De ANS (1997: 472) merkt in dit verband op dat "in het algemeen geldt dat een zin zonder presentatief *er* meer voorkeur heeft naarmate het gezegde een sterker handelingskarakter vertoont en/of de referent van het onderwerp een meer actieve rol speelt in het gebeuren." Volgens Van Es & Van Caspel (1971: 82 e.v.) valt *er* bij voorkeur weg als de bepalingzin een ander hoofdwerkwoord dan *zijn* bevat, en De Rooij (1991: 124) biedt empirische evidentie voor die stelling: zijn stimuli A-K – die een vorm van het zelfstandig werkwoord *zijn* bevatten – staan helemaal bovenaan in de hiërarchie van *er*-noodzakelijkheid en hebben een positieve *er*-coëfficiënt (die aangeeft dat +*er* de voorkeur geniet bij de geënquêteerden), terwijl de zinnen met een ander werkwoord dan *zijn* onderaan in de hiërarchie staan en – op één na – een negatieve *er*-coëfficiënt hebben, waaruit blijkt dat *er* volgens de informanten overbodig is.

In tegenstelling tot *regio* en *bepalingstype* laat de factor *werkwoordelijke specificiteit* zich niet makkelijk in een empirisch implementeerbare parameter vertalen. Bepalingzinnen laten slechts een beperkt aantal verbale klassen toe – in Levins (1993: 92 e.v.) terminologie VERBS OF EXISTENCE, VERBS OF APPEARANCE en DISAPPEARANCE, en een aantal MOTION VERBS –, maar de specificiteit van die werkwoorden kan betrekking hebben op elk van de drie conceptuele ingrediënten die de verbale semantiek van deze werkwoorden veronderstelt: nagenoeg alle werkwoorden in bepalingzinnen coderen een temporeel geïntantieerde *relatie* tussen het *subject* en de locatieve of temporele *setting* waarnaar de zinsinitiële bepaling verwijst. In het licht van de conceptuele onscheidbaarheid van het verbale proces van zijn subject en setting

kan men zich afvragen of een globale taxonomische classificatie van de werkwoorden in bepalingszinnen mogelijk en zinvol is. Liever dan op één lineaire classificatiedimensie af te beelden wat in de linguïstische realiteit drie afzonderlijke, moeilijk onderscheidbare variabelen blijken, grondvesten wij onze taxonomie van de verbale specificiteit op de enige (enigszins) beheersbare parameter van die specificiteit (die niet al in andere analyses aan bod komt), i.e. de omvang van de mogelijke klasse van subjecten die een werkwoord in een presentatieve zin subcategoriseert.

De werkwoorden in onze bepalingszinnen kunnen op basis van dat criterium in tenminste drie specificiteitsniveaus onderverdeeld worden. Op het laagste niveau bevindt zich het werkwoord *zijn*, dat in termen van subjectbeperking het minst restrictief is, omdat het de aanwezigheid van een entiteit in een bepaalde locatie *maximaal schematisch* karakteriseert, en dus geen enkele beperking oplegt aan subject, setting, en de interactie tussen beiden. Het is in dit opzicht dan ook niet verwonderlijk dat nagenoeg alle werkwoorden die we in bepalingszinnen aantreffen hyponiemen van *zijn* vertegenwoordigen.

Op het hoogste niveau van de specificiteitshiërarchie treffen we werkwoorden aan die slechts met een beperkte klasse van onderwerpen geattesteerd worden. Het gaat daarbij in de eerste plaats om verba uit de subgroepen van Levins (1993: 250) VERBS OF EXISTENCE die “typical of certain entities” zijn. We hadden het in dat verband reeds over VERBS OF ENTITY-SPECIFIC MODES OF BEING zoals *vloeien*; andere voorbeelden zijn *branden* en *smeulen* (van vuur), en *waaien* en *gieren* (van wind). Even subjectbeperkend zijn de VERBS OF MODES OF BEING INVOLVING MOTION (o.c.: 251) zoals *wapperen* (van vlaggen) en de MEANDER VERBS. VERBS OF SOUND EXISTENCE (o.c.: 252) zoals *klinken* en *echoën* beperken hun subject tot geluidsproducerende entiteiten, terwijl VERBS OF GROUP EXISTENCE (o.c.: 253) zoals *zwemmen*, *dansen* en *krioelen* in hun typische betekenis op het bestaan van resp. vissen, bijen en mieren wijzen. Op dat hoogste niveau bevindt zich nog een andere subgroep van de VERBS OF EXISTENCE, i.e. Levins (1993: 255) VERBS OF SPATIAL CONFIGURATION, die als zodanig niet typisch voor bepaalde entiteiten zijn (zoals de net behandelde werkwoordsklassen), maar die omdat ze “the spatial configuration of an entity with respect to some location” beschrijven hoge eisen aan de uitgedrukte relatie en locatie stellen, en op die manier dus ook de mogelijke onderwerpen beperken. Typische voorbeelden zijn *zitten*, *staan*, *liggen* en *hangen*. Hetzelfde geldt voor de tweede grote klasse van werkwoorden die we op het hoogste niveau aantreffen: de VERBS OF APPEARANCE, DISAPPEARANCE AND OCCURRENCE (o.c.: 258-261)

verwijzen resp. naar “the appearance of an entity on the scene” (cf. *verschijnen, landen, opkomen*, etc.) “the disappearance or going out of existence of some entity” (cf. *vergaan, omkomen* etc.)¹⁷, en “the occurrence of an event” (*plaatsvinden, aan de gang zijn*, etc.). Twee subcategorieën van VERBS OF MANNER OF MOTION tenslotte – die Levin (1993: 264-267) ROLL- en RUN-VERBS noemt omdat ze resp. de niet door een protagonist gecontroleerde bewegingen van inanimate entiteiten en de manieren waarop animate entiteiten bewegen beschrijven – komen ook in bepalingszinnen voor, maar vooral bij de RUN-verbs kan men zich volgens Levin (o.c.: 267) de vraag stellen of ze in presentatieve zinnen niet gewoon als VERBS OF EXISTENCE beschouwd moeten worden. *Drijven, glijden* en *rollen* zijn typische ROLL-verbs, terwijl *rennen* en *springen* frequent voorkomende voorbeelden van RUN-verbs zijn.

Anders dan Van Es & Van Caspel (1971) en De Rooij (1991) beperken we ons bij het classificeren van de werkwoordelijke specificiteit niet tot een tweedeling tussen *zijn* enerzijds en specifiekere werkwoorden anderzijds. Wij voegen nog een tussenniveau toe met daarop een beperkte groep van (over het algemeen frequent geattesteerde) werkwoorden die aan één van hun conceptuele ingrediënten een minimale beperking opleggen, en in onze taxonomie dus één niveau specifiek zijn dan het schematische *zijn*. Het werkwoord *bestaan* is net iets specifiek dan *zijn* omdat het zijn onderwerp binnen de metafysische grenzen van onze wereld situeert. *Ontstaan* voegt aan de betekenis van *zijn* een inchoatief aspect toe, *blijven* een imperfectief aspect, en *voorbij gaan* en *eindigen* een perfectief aspect. Het vaak geattesteerde werkwoord *heersen* verleent aan zijn onderwerpen (ook de niet-animate) een iets grotere agentiviteit. Wat al deze werkwoorden naast hun weinig specifieke betekenis gemeen hebben, is het feit dat ze in tegenstelling tot *zijn* minimale beperkingen opleggen aan mogelijke subjecten.

Alvorens we de resultaten bespreken moet nog gewezen worden op de transitieve werkwoorden in het materiaal, die niet ondergebracht kunnen worden in de net voorgestelde hiërarchie – omdat ze niet één maar twee participanten veronderstellen. Mede omwille van hun geringe frequentie (n=78) worden ze in de tabellen 8 en 9 buiten beschouwing gelaten. In de linkerkolom verwijst “zijn” (uiteraard) naar vormen van het hoofdwerkwoord *zijn*.

17 Op het vlak van de DISAPPEARANCE-werkwoorden verschillen het Engels en het Nederlands, omdat het Engels in principe geen presentatieve zinnen met die werkwoorden toelaat (cf. Levin 1993: 260), terwijl dat in het Nederlands wel kan. Vgl. *Op de Noordzee verging gisteravond een Maltese tanker*, of *Er verging gisteravond een Maltese tanker op de Noordzee*.

Werkwoorden zoals *bestaan*, *ontstaan* of *blijven* zijn intermediair (“int”) qua specificiteit; de meest specifieke verba die we in bepalingzinnen aantreffen worden in wat volgt EXISTENCE & APPEARANCE-WERKWOORDEN (“e.&a.”) genoemd.

	totaal			usenet			massa			kwal		
	-er	+er	% er	-er	+er	% er	-er	+er	% er	-er	+er	% er
<i>zijn</i>	30	163	84,5	13	61	82,4	7	44	86,3	10	58	85,3
int.	195	111	36,3	50	26	34,2	59	50	45,9	86	35	28,9
e.&a.	158	25	13,7	36	6	14,3	55	12	17,9	67	7	9,5
totaal	383	299	44,8	99	93	43,6	121	106	50,0	163	100	41,2

Tabel 8

Absolute en relatieve frequenties van *er* per verbale specificiteitsrang in het totale Nederlandse materiaal en per aparte Nederlandse bron

	totaal			usenet			massa			kwal		
	-er	+er	% er	-er	+er	% er	-er	+er	% er	-er	+er	% er
<i>zijn</i>	20	363	94,8	4	88	95,7	10	195	95,1	6	80	93,0
int.	260	181	41,0	30	36	54,5	111	89	44,5	119	56	32,0
e.&a.	323	76	19,0	42	25	37,3	157	33	17,4	124	18	12,7
totaal	603	620	51,6	76	149	62,5	278	317	52,3	249	154	45,9

Tabel 9

Absolute en relatieve frequenties van *er* per verbale specificiteitsrang in het gehele Belgische materiaal en per aparte Belgische bron

De progressief afnemende *er*-aandelen in de *totaal*-kolommen van het Nederlandse en Belgische materiaal suggereren dat de semantische specificiteit van het hoofdwerkwoord een uitgesproken invloed op de voorkeur voor *er* uitoefent. In tegenstelling tot *bepalingstype* lijkt de factor *werkwoordelijke specificiteit* echter nauwelijks geografisch bepaald: er zijn geen systematische verschillen tussen de kolommen van de tabellen 8 en 9.

In de volgende paragrafen trachten we met behulp van een aangepaste statistische techniek klaarheid te scheppen in het complexe beeld dat de vier besproken variabelen en hun interacties van *er*'s distributie schetsen.

4. LOGISTIEKE REGRESSIE

Tabel 10 toont de simultane invloed op *er*'s distributie van de contextuele factoren *regio* – “Nederland” vs. “België” op de horizontale as – en *register* – “usenet” vs. “massakranten” vs. “kwaliteitskranten” op de horizontale as –, en de taalstructurele factoren *bepalingstype* – “temporeel” vs. “locatief” op de verticale as – en *werkwoordelijke specificiteit* – ZIJN vs. “intermediaire werkwoorden” vs. “werkwoorden van EXISTENCE en APPEARANCE” op de verticale as:

			nl						vl						avg. + <i>er</i>	
			use		mas		kwa		use		mas		kwa			
			abs	rel	abs	rel	abs	rel	abs	rel	abs	rel	abs	rel		
em	zijn	-er	0	0,0	0	0,0	1	2,3	0	0,0	0	0,0	0	0,0	99,6	
		+er	35	100,0	32	100,0	43	97,7	40	100,0	111	100,0	48	100,0		
	int.	-er	3	13,6	11	22,4	14	33,3	1	5,0	16	21,9	18	35,29		78,1
		+er	19	86,4	38	77,6	28	66,7	19	95,0	57	78,1	33	64,7		
	e&a	-er	0	0,0	8	40,0	4	40,0	8	38,1	20	45,5	13	52,0		64,1
		+er	5	100,0	12	60,0	6	60,0	13	61,9	24	54,5	12	48,0		
oc	zijn	-er	13	33,3	7	36,8	9	37,5	4	7,7	10	10,6	6	15,8	76,4	
		+er	26	66,7	12	63,2	15	62,5	48	92,3	84	89,4	32	84,2		
	int.	-er	47	87,0	48	80,0	72	91,1	29	63,0	95	74,8	101	81,5	20,4	
		+er	7	13,0	12	20,0	7	8,9	17	37,0	32	25,2	23	18,5		
	e&a	-er	36	97,3	47	100,0	63	98,4	34	73,9	137	93,8	111	94,9	6,9	
		+er	1	2,7	0	0,0	1	1,6	12	26,1	9	6,2	6	5,1		
avg + er			61,4		53,5		49,6		68,7		58,9		53,4			

Tabel 10
Absolute en relatieve frequenties van *er* als functie van *regio*, *register*, *bepalingstype* en *werkwoordelijke specificiteit*

Tot hiertoe hebben we de impact van de factoren *regio*, *register*, *bepalingstype* en *werkwoordelijke specificiteit* op de zinsinterne distributie van *er* louter op individuele basis bestudeerd. Om het collectieve effect van die factoren te meten, maken we gebruik van een complexere statistische techniek, die tevens een antwoord biedt op drie aanvullende vragen. Zo willen we om te beginnen graag een statistische maat waarmee de respectievelijke invloed van de afzonderlijke factoren *hiërarchisch* kan uitgedrukt worden: welke factor heeft

de grootste impact op het gebruik van *er*? Voor een gefundeerde evaluatie van de traditionele benaderingen van de postverbale distributie van *er* is het daarnaast essentieel dat we een statistische maat vinden voor het verklarende succes van het model dat de vier afzonderlijke factoren combineert: volstaat een basismodel met onze vier factoren om de geobserveerde variatie in het postverbale gedrag van *er* te voorspellen, of moeten we andere factoren toevoegen? In het licht van de vaststelling en neutralisering van contextuele variatie in de postverbale distributie van *er* is het tenslotte onontbeerlijk te achterhalen of de *er*-variatie in het Belgische materiaal door hetzelfde model verklaard wordt als de *er*-variatie in Nederlandse teksten.

Voor een statistisch onderbouwd antwoord op de nieuwe vragen voeren we een *logistische-regressieanalyse* op het materiaal uit. Logistische regressie is een speciale vorm van regressieanalyse. Een toelichting ervan vertrekt best van de eenvoudigste variant, *enkelvoudige lineaire regressie*. Het vertrekpunt van een enkelvoudige regressieanalyse is een correlatie tussen twee reeksen van veranderlijke gegevens. Bijvoorbeeld, als werknemers een periodieke salarisverhoging krijgen, dan is er een relatie tussen hun salaris en hun leeftijd (specifieker: hun salarisanciënniteit). Als de werknemers in kwestie bijvoorbeeld om de twee jaar een salarisverhoging van 2 % krijgen, dan correleert de stijging van het loon op een systematische manier met de stijging van de leeftijd, en die correlatie kan in een formule uitgedrukt worden. We moeten er natuurlijk rekening mee houden dat de relatie tussen salarisanciënniteit en salaris niet van nul begint, maar in principe zijn aanknopingspunt neemt bij het aanvangssalaris van de werknemers. In het algemeen bestaat hun salaris op een bepaald moment dan uit het aanvangssalaris, plus de periodieke verhogingen die op grond van een groeiende salarisanciënniteit zijn toegekend.

In de formule $y = a + bx$ is y de afhankelijke variabele (of responsvariabele) “salaris” en x de onafhankelijke variabele “salarisanciënniteit”; de constante a is het aanvangssalaris, en b de mate waarin de salarisanciënniteit bijdraagt tot het salaris op een gegeven moment. De “+” in de formule toont dat de salarisanciënniteit van het voorbeeld een positieve invloed heeft op het salaris. Het is in theorie echter ook mogelijk dat oudere werknemers vanaf een bepaalde leeftijd jaarlijks 1 % salaris moeten inleveren; in dat geval herschrijven we de formule uiteraard als $y = a - bx$.

Een enkelvoudige regressieanalyse levert twee soorten informatie op. Op basis van de verzamelde gegevens van een aantal werknemers berekent de computer om te beginnen

niet alleen een *schatter* voor b , maar ook voor a . In de meeste gevallen ligt de optimale waarde voor a namelijk niet a priori vast: zelfs in ons eenvoudige salarisvoorbeeld kan a alleen als aanvangssalaris geïdentificeerd worden bij werknemers met een stijgende salarisanciënniteit; bij werknemers die periodiek moeten inleveren wordt a beter gelijkgesteld met het hoogste salaris dat de betreffende werknemers ooit mochten ontvangen. En dus is het veiliger de computer een schatting voor a te laten maken, die we a posteriori zinvol trachten te interpreteren.

Daarnaast bepaalt de computer ook de *significantie* van a en b . De statistische maat hiervoor is de p-waarde (of “overschrijdingswaarde”). De afspraak is dat we een p-waarde van minder dan 0.05 (het significantieniveau dat in de linguïstiek gebruikelijk is¹⁸) als een voldoende indicatie mogen beschouwen dat de relatie die we aantreffen niet op toeval berust. In dat geval kunnen we de gegevens van de steekproef (de onderzochte groep werknemers) veilig extrapoleren naar de volledige populatie (alle werknemers in dezelfde situatie).

Als we de regressietechniek toepassen op een situatie met meerdere onafhankelijke variabelen spreken we van een *meervoudige lineaire regressie*. Betrekken we naast de periodieke salarisverhogingen bijvoorbeeld ook de prijscompensaties in ons salarismodel, dan neemt dat model de vorm aan van de formule $y = a + bx_1 + cx_2$, waarbij x_2 uiteraard ingevuld wordt door de prijscompensatievariabele. Dit model¹⁹ is het eenvoudigste meervoudige regressiemodel met twee afzonderlijke factoren, waarvoor de computer net zoals bij de enkelvoudige regressie schatters berekent (a , b en c) die op significantie getest worden. Veel complexer worden de modellen als men ook de mogelijkheid van *interacties* tussen factoren aftast; het is namelijk niet ondenkbaar dat het gecombineerde effect van twee factoren op y afwijkt van het verwachte cumulatieve effect van de individuele factoren. In ons voorbeeld zullen zowel salarisanciënniteit als prijscompensaties voor een salarisstijging zorgen, maar als prijscompensaties voornamelijk toegekend worden aan werknemers die een periodieke salarisverhoging van meer dan 5 % genieten, dan moet naast de afzonderlijke factoren “salarisanciënniteit” en “prijscompensatie” ook de interactie tussen beide factoren in het model opgenomen worden. Dat laatste heeft dan de vorm $y = a$

18 Met een significantieniveau van 0.05 zijn we 95 % zeker dat de geobserveerde relatie niet op toeval berust.

+ $bx_1 + cx_2 + dx_1x_2$, waarbij d de schatter is die de computer voor de invloed van de interactie tussen x_1 en x_2 berekent. Naarmate het aantal onafhankelijke variabelen toeneemt wordt het aantal mogelijke interacties progressief groter: twee onafhankelijke variabelen laten slechts één interactie toe, drie onafhankelijke variabelen leveren vier mogelijke interacties op, terwijl er bij vier variabelen al meteen zes interacties mogelijk zijn.

Hoewel omvattende modellen waarin alle mogelijke interacties zijn opgenomen in principe een grotere verklarende kracht hebben, bestaat bij die modellen vaak het probleem dat mathematisch mogelijke interacties linguïstisch soms oninterpreteerbaar zijn. Bij meervoudige regressie is het dan ook noodzakelijk een model te vinden met een zo groot mogelijke verklarende kracht én interpreteerbare interacties.

Klassieke regressietechnieken zijn ontwikkeld voor interval- en ratiovariabelen²⁰, maar er zijn ook regressietechnieken die bruikbaar zijn voor nominale afhankelijke variabelen met waarden 0 en 1, zoals $-er$ en $+er$. Om technische redenen is het in dit geval niet mogelijk om zoals bij het salarisvoorbeeld een lineair verband te zoeken tussen y (de kans op er) en allerlei onafhankelijke variabelen x_1 en x_2 , want de voorspelde y -waarden hebben de onintuïtieve eigenschap dat ze buiten het bereik $[0...1]$ kunnen vallen. Statistici vangen deze technische ongemakken op door een model te gebruiken dat een niet-lineair verband zoekt in de vorm van een *logistische curve*, vandaar ook de naam *logistische regressie* voor deze techniek²¹. De responsvariabele y wordt daarin gemodelleerd als $P(Y=1)$, i.e. “de kans dat y een bepaalde waarde heeft”, in ons geval “ $+er$ ”. Omwille van technische details (die in deze toelichting niet relevant zijn), gebruiken we bij logistische regressie een mathematische afgeleide van $P(Y=+er)$, nl. de logit van er .

19 Onder “model” verstaan we een verzameling verklarende factoren.

20 Statistici onderscheiden doorgaans vier *meetniveaus* voor variabelen. *Nominale* variabelen zoals *regio* – met waarden “Nederland” en “België” – hebben geen enkele inherente ordening; *ordinale* variabelen zoals zevenpunts-grammaticaliteitsschalen met als polen “perfect grammaticaal” en “volstrekt onaanvaardbaar” hebben een inherente ordening die zinvol vertaald kan worden in een klassieke getalorderrelatie (1, 2, ... tot 7), maar de afstand tussen 1 en 2 is niet noodzakelijk dezelfde als de afstand tussen 2 en 3. *Interval*variabelen zoals de Celsius temperatuurschaal zijn eveneens inherent geordend, en daar is het verschil tussen 20 en 40 (graden Celsius) wél hetzelfde als het verschil tussen 40 en 60 (graden Celsius). Het verschil tussen *interval*- en *ratio*variabelen, tenslotte, ligt in het feit dat de eerste een arbitrair nulpunt hebben – vandaar ook het feit dat op de Celsiusschaal negatieve temperaturen mogelijk zijn – terwijl *ratio*variabelen (zoals de leestijden in Grondelaers & Brysbaert (1996)) een inherent nulpunt hebben (want een leestijd kan 0 zijn).

21 De precieze technische details zijn hier niet relevant; voor een uitgebreidere toelichting, zie onder meer Rietveld & Van Hout (1993: 330 e.v.).

Het niet-lineaire verband tussen afhankelijke en onafhankelijke variabelen maakt dat logistische regressie veel moeilijker te interpreteren is dan lineaire regressie; dat geldt niet alleen voor de individuele schatters voor de onafhankelijke variabelen, maar ook voor de samenvattende grootheden waarin het globale succes van een model wordt uitgedrukt.

De statistische gegevens worden als volgt gepresenteerd. De modelformule

$$\text{logit ER} = a + xb + yc + zb*c\dots$$

bevat naast de constante a de namen van de significante variabelen (b , c , etc.) en/of interacties ($b*c$), en de *schatters* (x , y , etc.) waarin het statistiekpakket SAS het belang van die variabelen uitdrukt. Merk op dat die grootheden niet meteen interpreteerbaar zijn als een indicatie van de respectievelijke invloed van die variabele. Daarvoor dient de Odds Ratio: variabelen met een positieve schatter hebben een Odds Ratio boven 1 die aangeeft met welke factor het gebruik van *er* t.o.v. het niet-gebruik van *er* in het besproken model stijgt onder invloed van de betreffende variabele; variabelen met een negatieve schatter hebben een Odds Ratio onder 1 die aangeeft met welke factor het gebruik van *er* t.o.v. het niet-gebruik van *er* onder invloed van de onafhankelijke variabele in kwestie daalt. Een Odds Ratio van 6 betekent dus dat we op basis van de data kunnen verwachten dat het gebruik van *er* t.o.v. het niet-gebruik van *er* zes maal zal toenemen; een Odds Ratio van 0.2 geeft aan dat we verwachten dat het gebruik van *er* t.o.v. het niet-gebruik van *er* vijf maal afneemt. Odds Ratios zijn overigens uitsluitend bruikbaar om de hiërarchische verhouding tussen de onafhankelijke variabelen binnen hetzelfde model te bepalen; voor de vergelijking van het effect van dezelfde variabele in verschillende modellen gebruiken we het *betrouwbaarheidsinterval* (confidence interval) van de Odds Ratio, die het domein afbakent waarin de Odds Ratio zich, gegeven de variantie in de data en gegeven een bepaald significantieniveau (in ons geval zoals gezegd 0.05), bevindt.

De SAS-output bevat ook grootheden waarmee de globale kwaliteit van een model in zijn geheel geëvalueerd kan worden. Een eerste manier om het succes van een model te kwantificeren is nagaan *hoeveel* van de totale variatie in de data verklaard wordt door het model. Het *Akaike Information Criterion* (AIC) geeft twee grootheden op. In de eerste wordt m.b.v. een log likelihood ratio (de mathematische achtergrond is hier opnieuw niet relevant) de totale hoeveelheid variatie uitgedrukt in het *intercept only-model*, het “nulhypothese”-

model zonder de onafhankelijke variabelen die *er*'s distributie verklaren. De tweede grootheid drukt de hoeveelheid variatie uit die we niet hebben kunnen verklaren met het *intercept and covariates-model*, het model waarin de *er*-verklarende factoren wel zijn opgenomen. Hoe kleiner de tweede t.o.v. de eerste grootheid, hoe beter het model.

Daarnaast levert SAS ook grootheden op die de voorspellende kracht van een model reflecteren. Voor het bepalen van die grootheden gaat SAS uit van alle mathematisch mogelijke koppels van observaties in de materiaalverzameling waarbij de eerste observatie *er* bevat, en de tweede observatie geen *er*, i.e. koppels van de structuur $\{+er, -er\}$. *Concordant* (C) zijn alle koppels waarbij het bestudeerde model voor de *+er*-observatie een grotere kans op *er* voorspelt dan voor de *-er*-observatie. *Discordant* (D) zijn alle koppels waarbij het bestudeerde model voor de *+er*-observatie een kleinere kans op *er* voorspelt dan voor de *-er*-observatie. De *gamma index* – de genormaliseerde maat voor de verhouding tussen concordant en discordant die we in dit artikel gebruiken – berekenen we op basis van de formule $(C-D)/(C+D)$. Het spreekt vanzelf dat modellen met een hogere gamma-index een grotere predictieve kracht hebben.

5. DE GLOBALE INVLOED VAN DE DRIE VARIABELEN

Beginnen we de logistische-regressieanalyse met de *er*-variatie in de globale database. We vertrekken daarbij onmiddellijk van een model dat niet alleen de afzonderlijke variabelen bevat, maar ook de mogelijke twee-aan-twee interacties (i.e. de interacties tussen twee factoren). De SAS-output voor het krachtigste model met afzonderlijke variabelen en interacties is de volgende:

$$15. \text{ logit ER} = -3.8992 + 3.6941 \text{ bepalingstype} + 4.2317 \text{ werkwoordelijke} \\ \text{specificiteit 1} + 1.1831 \text{ werkwoordelijke specificiteit 2} + 1.2121 \text{ regio} + \\ 0.9873 \text{ register 1} + 0.4407 \text{ register 2} - 1.2791 \text{ bepalingstype*regio}$$

In deze notatie worden voor de variabelen met drie waarden – *werkwoordelijke specificiteit* en *register* – telkens twee schatters opgegeven, een eerste schatter voor de invloed op *er*'s distributie van het verschil tussen de waarden “1” en “3” (tussen, in het geval van *werkwoordelijke specificiteit*, hoofdwerkwoord *zijn* en een hoofdwerkwoord van het EXISTENCE & APPEARANCE-type), en een tweede schatter voor de invloed op *er*'s distributie van

het verschil tussen de waarden “2” en “3” (het verschil tussen intermediaire werkwoorden en EXISTENCE & APPEARANCE-werkwoorden).

Merken we om te beginnen op dat elk van de variabelen en interacties sterk significant is ($p < 0.0001$, behalve voor *register 2* $p = 0.0051$). Voor *bepalingstype* en *werkwoordelijke specificiteit* is die uitstekende significantiescore geen verrassing, want die leverden bij de individuele analyse al opvallende proportieverschillen op. Ook de beide *register*-variabelen zijn sterk significant in het globale materiaal, al bleek hun effect in 3.1. in hoofdzaak beperkt tot het Belgische staal. De variabele *regio* – die in sectie 3.1. slechts een verschil van 6.86 % tussen het Belgische en het Nederlandse *er*-gebruik aan het licht bracht – blijkt in het dit model even significant als de andere variabelen, en dat maakt voorzichtigheid bij de interpretatie van de regioneutrale gegevens voortaan noodzakelijk, in het bijzonder omdat *regio* in dit model ook nog eens in een significante interactie met *bepalingstype* opduikt (een interactie die we al met het blote oog geïdentificeerd hadden in 3.1.).

Ondanks hun hoge significantie oefenen niet alle bestudeerde variabelen dezelfde invloed uit op de *er*-variatie. De Odds Ratio 40.2 voor *bepalingstype* suggereert dat het gebruik van *er* t.o.v. het niet-gebruik van *er* zowat 40 keer hoger ligt in zinnen met een temporele bepaling dan in zinnen met een locatieve bepaling. De Odds Ratio 68.8 voor *werkwoordelijke specificiteit 1* duidt aan dat het gebruik van *er* t.o.v. het niet-gebruik van *er* zelfs meer dan 68 keer toeneemt wanneer een EXISTENCE & APPEARANCE-werkwoord vervangen wordt door een vorm van *zijn*, terwijl de Odds Ratio 3.27 voor *werkwoordelijke specificiteit 2* aangeeft dat het gebruik van *er* t.o.v. het niet-gebruik van *er* slechts zo'n 3 maal toeneemt wanneer een EXISTENCE & APPEARANCE-werkwoord een intermediair werkwoord wordt. De Odds Ratios voor *werkwoordelijke specificiteit* bewijzen m.a.w. dat het hoofdwerkwoord *zijn* de belangrijkste *er*-trigger is in een bepalingszin: de kans op *er* wordt het efficiëntst beperkt door een specifiek hoofdwerkwoord dan *zijn* te gebruiken; welk specifiek werkwoord dat is (intermediair of EXISTENCE & APPEARANCE) heeft blijkens deze gegevens minder belang.

De Odds Ratio van 3.36 voor *regio* reflecteert dat het gebruik van *er* t.o.v. het niet-gebruik van *er* meer dan drie maal hoger ligt in België dan in Nederland. De impact van de *register*-variabelen is het beperktst: de Odds Ratio 2.684 voor *register 1* duidt aan dat het gebruik van *er* t.o.v. het niet-gebruik van *er* zo'n 2,5 keer stijgt tussen het kwaliteitskrantencorpus en

het Usenet-corpus; blijkens de Odds Ratio 1.554 voor *register 2* stijgt het gebruik van *er* t.o.v. het niet-gebruik van *er* 1,5 keer tussen kwaliteits- en massakranten.

De Odds Ratio van de interactie *regio*bepalingstype* is moeilijker interpreteerbaar, omdat de stijging in het gebruik van *er* t.o.v. het niet-gebruik van *er* o.i.v. beide variabelen op twee manieren teruggekoppeld kan worden naar het materiaal. Om te beginnen kunnen we deze interactie interpreteren zoals in 3.1.: de regionale variatie in het gebruik van *er* is grotendeels tot de locatieve zinnen beperkt. De gegevens in tabel 10 laten echter ook toe de interactie als volgt te interpreteren: het effect van de variabele *bepalingstype* is het grootst in het Nederlandse materiaal.

Bekijken we vervolgens de grootheden waarmee we de waarde van het model als geheel kunnen evalueren. De grote voorspellende kracht van dit model – de gamma-index van 84.8 % kan zeer hoog genoemd worden – staat in schril contrast met het pessimisme – “duidelijke regels zijn hier vooralsnog niet te geven” – dat in de ANS (1984: 820; 1997: 477) en De Rooij (1991: 127) m.b.t. de beregelbaarheid van *er* heerst. In de overgrote meerderheid van alle presentatieve bepalingsszinnen is de distributie van *er* met een eenvoudig algoritme voorspelbaar: het gebruik van *er* in een temporele bepalingsszin met hoofdwerkwoord *zijn* levert, net zoals de weglating van *er* in een locatieve bepalingsszin met een specifiekere werkwoord dan *zijn*, nooit opvallend ongrammaticale bouwsels op; dat geldt – zoals we in de vorige paragrafen al zagen – in hogere mate voor het Nederlandse dan voor het Belgische materiaal.

Rest ons nog de vraag of de *er*-variatie in het Nederlandse en het Belgische materiaal door verschillende modellen verklaard wordt, en dus apart voor Nederland en België beschreven moet worden in de ANS. Het feit dat de invloed van de contextuele variabele *regio* en van de interactie tussen *regio* en *bepalingstype* significant was in het globale model is een krachtige indicatie dat de Nederlandse en Belgische situatie verschillen. Dat leiden we eveneens af uit de vergelijking in tabel 11 van het Belgische en het Nederlandse model met afzonderlijke variabelen; de kolommen 1 en 4 daarin bevatten resp. de Nederlandse en Belgische Odds Ratios voor de schuingedrukte variabelen (*beptype* = bepalingstype en *wwspec* = werkwoordelijke specificiteit; voor de variabelen *werkwoordelijke specificiteit* en *register* – die drie variabele waarden hebben – zijn er telkens twee Odds Ratios, de eerste voor de vergelijking van de 1ste en de 3de waarde, en de tweede voor de vergelijking van de 2de en de 3de waarde):

	Nederland			België		
	O.R.	Betrouwbaarheids-intervallen		O.R.	Betrouwbaarheids-intervallen	
<i>beptype</i>	31,667	18,703	53,615	11,871	8,26 7	17,045
<i>wwspec 1 vs 3</i>	47,177	22,41	99,318	91,937	52,3 4	161,493
<i>wwspec 2 vs 3</i>	3,204	1,718	5,976	3,299	2,27 3	4,788
<i>register 1 vs 3</i>	1,869	1,033	3,382	3,418	2,10 7	5,545
<i>register 2 vs 3</i>	1,65	0,945	2,881	1,515	1,04	2,206

Tabel 11

Odds Ratios en betrouwbaarheidsintervallen in het Nederlandse en Belgische basismodel ter verklaring van *er*-variatie

Aangezien in de modellen in tabel 11 geen interacties opgenomen zijn, bevatten ze precies dezelfde ingrediënten – i.e. de afzonderlijke factoren – en zijn ze variabele per variabele vergelijkbaar, waardoor proportionele verschillen onmiddellijk aan het licht komen. Hoewel ze ter wille van de vergelijkbaarheid bewust eenvoudig (i.e. interactieloos) gehouden werden, vertegenwoordigen beide basisconfiguraties in tabel 11 modellen die nauwelijks minder krachtig zijn dan de complexere varianten met interacties die verderop besproken worden (de respectieve gamma-indexen voor het Nederlandse en het Belgische basismodel in tabel 11 zijn 86.5 % en 85.1 %).

In verband met het onderscheid tussen het Nederlandse en het Belgische basismodel in tabel 11 moet om te beginnen opgemerkt worden dat de tweede register-variabele in het Belgische materiaal wel, maar in het Nederlandse materiaal niet significant is ($p = 0.0780$). De register-gegevens bevestigen dus de geringere impact van bronvariatie in het Nederlandse materiaal. Tabel 11 confirmeert tevens dat *bepalingstype* en *werkwoordelijke specificiteit 1* (net zoals in het globale materiaal) de belangrijkste variabelen zijn in de Belgische en Nederlandse teksten, al is de verhouding tussen deze factoren asymmetrisch in de beide materiaalverzamelingen: in het Belgische materiaal hebben *bepalingstype* en *werkwoordelijke specificiteit 1* blijkens de Odds Ratios 31.667 en 47.177 een vergelijkbare

invloed op *er*'s distributie; in het Nederlandse materiaal daarentegen lijkt de invloed van bepalingstype beperkter dan die van werkwoordelijke specificiteit (11.871 vs. 91.937).

De verschillende rol die *bepalingstype* in het Belgische en Nederlandse materiaal speelt blijkt het best uit een vergelijking van de betrouwbaarheidsintervallen van de respectievelijke Odds Ratios voor *bepalingstype*: wanneer de Nederlandse en Belgische betrouwbaarheidsintervallen voor *bepalingstype* niet overlappen, is er sprake van een proportioneel verschil tussen beide modellen. De gegevens in tabel 11 tonen aan dat dat het geval is: de betrouwbaarheidsinterval van de Odds Ratio voor *bepalingstype* in het Nederlandse materiaal (18.703 - 53.615) overlapt niet met de betrouwbaarheidsinterval van de Odds Ratio voor *bepalingstype* in het Belgische materiaal (8.267 – 17.045). We kunnen dus met enige mate van confidentie besluiten dat er tenminste een proportioneel verschil is tussen het Nederlandse en het Belgische *er*-model.

Dat verschil wordt nog duidelijker als we samen met SAS op zoek gaan naar het sterkste model voor de afzonderlijke Belgische en Nederlandse situatie. Als input voor die analyse geven we niet alleen de aparte variabelen *bepalingstype*, *werkwoordelijke specificiteit 1 & 2*, en *register 1 & 2* op, maar ook de mogelijke twee-aan-twee interacties tussen beide. Dat levert de modellen in (16) (voor Nederland) en (17) (voor België) op:

$$\begin{aligned} 16. \text{ logit ER} = & -4.2904 + 4.5131 \text{ bepalingstype} + 4.8934 \text{ werkwoordelijke} \\ & \text{specificiteit 1} + 2.4305 \text{ werkwoordelijke specificiteit 2} - 0.8474 \\ & \text{bepalingstype*werkwoordelijke specificiteit 1} - 1.9544 \\ & \text{bepalingstype*werkwoordelijke specificiteit 2} + 1.5235 \\ & \text{bepalingstype*register 1} + 0.438 \text{ bepalingstype*register 2} \end{aligned}$$

$$\begin{aligned} 17. \text{ logit ER} = & -2.7857 + 2.4741 \text{ bepalingstype} + 4.5211 \text{ werkwoordelijke} \\ & \text{specificiteit 1} + 1.1935 \text{ werkwoordelijke specificiteit 2} + 1.2291 \text{ register 1} + \\ & 0.4154 \text{ register 2} \end{aligned}$$

Wat bij de vergelijking van de modellen in (16) en (17) met de modellen in tabel 11 meteen opvalt, is dat het Belgische basismodel zonder interacties in tabel 11 ook het krachtigste model is dat SAS voor de Belgische distributie van *er* vindt (met gamma-index 85.1 %). Voor de Nederlandse situatie liggen de zaken iets ingewikkelder: door toevoeging van de interacties *bepalingstype*werkwoordelijke specificiteit 1 & 2* en *bepalingstype*register 1 & 2* wordt blijkens het AIC meer variatie verklaard dan in het basismodel (*Intercept + covariates* 472.839 < 477.64), en blijkens de gamma-index ook meer variatie voorspeld dan

in het basismodel (89.5 % > 85.1 %), maar de toegevoegde interacties zijn niet allemaal significant. Weglating van de niet-significante interacties *bepalingstype*wwspec 1* ($p = .5177$) en *bepalingstype*register 2* ($p=.2685$) resulteert echter in een statistisch onaanvaardbaar model.

In verband met de vergelijking tussen het Nederlandse model in (16) en het Belgische model in (17) moet verder opgemerkt worden dat enkel het Belgische model de *register*-variabelen als hoofdeffect handhaaft. In het Nederlandse model wordt het effect van *register* geconditioneerd door de factor *bepalingstype*.

Uit de vorige paragrafen is inmiddels afdoende duidelijk geworden dat de beste Belgische en Nederlandse modellen voor *er*'s distributie uit verschillende ingrediënten bestaan, zodat er niet alleen een proportioneel verschil is tussen de Belgische en Nederlandse *er*-voorkeur, maar ook een structureel verschil. *Er* wordt m.a.w. niet enkel minder gebruikt in Nederland, maar ook anders. De taalnormatieve consequenties van dat verschil bekijken we in de slotparagraaf.

6. CONCLUSIES EN TAALNORMATIEVE CONSEQUENTIES

We vatten samen. In deze studie hebben we met behulp van een omvangrijk, stilistisch en geografisch gecontroleerd corpus van geattesteerd hedendaags Nederlands een afzonderlijke en globale analyse uitgevoerd van de factoren waarvan op voorhand bekend was dat ze de postverbale distributie van niet-anaforisch *er* in presentatieve zinnen met een vooropgeplaatste bepaling (in dit artikel “bepalingszinnen” genoemd) kunnen beïnvloeden. De statistisch geconfirmeerde resultaten van die analyse hebben consequenties voor zowel de taalwetenschappelijke als de normatieve beschrijving van presentatief *er*.

Op het taaldescriptieve niveau hebben we met behulp van een logistische-regressieanalyse vastgesteld dat de taalstructurele factoren *bepalingstype* (de kans op *er* wordt kleiner na een plaatsbepaling) en *werkwoordelijke specificiteit* (de kans op *er* wordt groter na het hoofdwerkwoord *zijn*), net zoals de contextuele factoren *regio* (*er* staat sterker in België dan in Nederland) en *register* (*er* staat sterker naarmate de taalsituatie informeler wordt – een factor waarvan het belang tot hertoe onderschat werd) een significante invloed op *er*'s distributie uitoefenen. Daarnaast blijkt er een significante interactie tussen de factoren *bepaling* en *regio* te bestaan: het verschil tussen het Nederlandse en het Belgische *er*-

gebruik is nagenoeg tot locatieve bepalingssinnen beperkt. De invloed op *er*'s distributie is echter niet bij alle factoren even uitgesproken: terwijl de taalstructurele factoren een massieve impact hebben op de voorkeur voor *er*, blijkt de invloed van de contextuele factoren wat minder uitgesproken.

De bespreking in de vorige paragraaf van het globale, regio-neutrale effect van de bestudeerde variabelen heeft enkel zin wanneer aan taalstructurele en contextuele factoren dezelfde variationele status wordt toegekend. In navolging van Geeraerts, Grondelaers & Bakema (1994), Geeraerts, Grondelaers & Speelman (1999) en Grondelaers & Geeraerts (ter perse) beschouwen wij contextuele variabelen echter als factoren die – extreem geformuleerd – het effect van variatieverklarende taalstructurele factoren tot één bepaalde context – d.w.z. tot één regio of één stijlregister - beperken. Het significante proportionele en structurele verschil tussen het Belgische en Nederlandse *er*-gebruik noopt ons er dan ook toe voor elk van beide regio's een apart model op te stellen voor de distributie van *er*. Uit de (interpreteerbare) gegevens waarover we momenteel beschikken blijken de bepalingfactor en de verbale factor in het Nederlandse model absoluut dominant: Nederlandse bepalingssinnen hebben doorgaans genoeg aan een locatieve bepaling om zonder *er* te kunnen; de verbale factor beperkt *er*'s frequentie in de weinige gevallen waar de locatieve bepaling dat niet al gedaan heeft. De *er*-beregellende efficiëntie van de bepalingfactor en de werkwoordelijke factor wordt in het Nederlandse materiaal ook nauwelijks beperkt door registervariatie: het *er*-verklarende succes van bepalingstype en verbale specificiteit is in nagenoeg elk brontype even groot. Het gevolg daarvan is dat *er* in Nederlandse locatieve bepalingssinnen met een EXISTENCE & APPEARANCE werkwoord slechts uitzonderlijk (n=2) voorkomt.

In het Belgische model daarentegen is het respectievelijke effect van *bepalingstype* en *verbale specificiteit* minder uitgesproken, en de factoren zijn ook nauwelijks aan elkaar gekoppeld. In combinatie met de registervariatie die in het Belgische materiaal wél een rol speelt, levert die beperktere bereikbaarheid in de minst formele bron – i.e. Usenet - een *er*-residu van maar liefst 26.1 % op in de context die *er*'s voorkomen het meest zou moeten beperken. Voor het Belgische *er*-gebruik is dan ook bijkomend onderzoek nodig.²²

22 In Grondelaers (2000: 134 e.v.) worden specifiek i.v.m. *er*'s distributie in Vlaamse locatieve bepalingssinnen drie bijkomende factoren voorgesteld – *bepalingstopicaliteit*, *spatiële verankering van*

Nochtans laten de Nederlandse en de Belgische *er*-distributie zich getuigen de regressiegegevens veel beter beregelen dan het pessimisme in de ANS en De Rooij (1991) de gebruiker doen geloven: dat duidelijke regels i.v.m. *er*'s distributie "vooralsnog niet te geven" zijn (ANS 1984: 820; 1997: 477) en De Rooij (1991: 127)) blijkt eigenlijk nauwelijks uit het hier gepresenteerde onderzoek. De taalnormatieve consequenties van onze bevindingen formuleren we dan ook in termen van mogelijke alternatieven voor de bestaande ANS-lemma's over *er*'s distributie. In de huidige versie van de ANS vindt de gebruiker op twee plaatsen informatie over *er*'s postverbale distributie. De tweede daarvan is specifiek gewijd aan postverbaal *er* in zinnen met een bepaling op de eerste zinsplaats (1997: 477):

"In de standaardtaal wordt er na een plaatsbepaling (...) gemakkelijker weggelaten dan in andere gevallen. Overigens verschilt de voorkeur voor het al dan niet gebruiken van er buiten de eerste zinsplaats per geval; duidelijke regels zijn hier vooralsnog niet te geven. Er zijn ook nogal grote individuele verschillen in het er-gebruik."

"Wel is er een duidelijke geografisch bepaalde tendentie: er staat (al dan niet na een plaatsbepaling) sterker in België (behalve Oost-Limburg) en eveneens, zij het in mindere mate, in het zuidelijk deel van Nederland (vooral Zeeland) dan in de rest van het taalgebied."

De meest dringende aanpassing die dit lemma behoeft is een splitsing in aparte versies voor de Nederlandse en de Belgische situatie, zoals momenteel al het geval is bij de beschrijving van de aard van de hoeveelheidsaanduiding bij kwantitatief *er* (o.c.: 480-485). Het Nederlandse lemma zou er als volgt kunnen uitzien:

"In Nederlandse zinnen met een temporele bepaling voorop staat bij voorkeur er, zeker na een vorm van zijn als hoofdwerkwoord. In Nederlandse zinnen met een plaatsbepaling op de eerste plaats kan er meestal weggelaten worden, zeker als de zin een ander hoofdwerkwoord dan zijn bevat."

In verband met de situatie in België kan in dit stadium van de analyse slechts met zekerheid vastgesteld worden dat (i) *er* vaker voorkomt na een temporele dan na een locatieve bepaling, dat (ii) *er* meer nodig is na *zijn* dan na specifiekere werkwoorden, en dat

de bepalingreferent, en positie van de verbale betekenis-kern –, waarmee het residu van "onvoorspelde" *er*'s in die zinnen naar gemiddeld 3.7 % wordt teruggebracht.

(i) *er* minder gebruikelijk is naarmate de gesprekssituatie formeler wordt. Het is vooralsnog niet opportuun deze set tendensen in termen van een ANS-lemma te formuleren. Wat inmiddels wél met enige confidentie geponeerd kan worden is dat verder onderzoek naar het complexe gedrag van presentatief *er* bij voorkeur aangevuld wordt met statistisch ondersteund on- of off-line onderzoek.

BIBLIOGRAFIE

- Barbier, I. (1993). "On the syntax of Dutch 'er'". In R.L. Lippi-Green & J.C. Salmons (eds.), *Germanic Linguistics*, 65-83. Amsterdam: Benjamins.
- Bays, H. (1998). "Framing and face in internet exchanges: a socio-cognitive approach". On-line publicatie op <http://viadrina.eu.v-frankfurt-o.de/~wjournal/bays.html>
- Bech, G. (1952). "Über das Niederländische Adverbialpronomen *er*". *Travaux du cercle linguistique de Copenhague* 8, 5-32.
- Butler, C. (1985). *Statistics in linguistics*. Oxford & New York: Blackwell.
- De Rooij, J. (1991). "Regionale variatie in het gebruik van *er* III". *Taal en tongval* 43, 113-136.
- De Schutter, G. (1974). *De Nederlandse zin. Poging tot beschrijving van zijn structuur*. Brugge: De Tempel.
- De Schutter, G. & P. van Hauwermeiren (1983). *De structuur van het Nederlands*. Malle: De Sikkel.
- Es, G.A. van & P.P.J. van Caspel (1971). *De patronen van de zinspotente groepen; grondtype A en zijn varianten II*. Publicaties van het archief voor de Nederlandse syntaxis. Groningen: Rijksuniversiteit.
- Geeraerts, D. (1999). "Noch standaard, noch dialect. 'Tussentaal' in Vlaanderen en Nederland". *Onze Taal* 68, 232-235.
- Geeraerts, D. & P. Bakema (1993). "Materiaalverzamelingsmethodes in lexicologie en lexicografie". In A. van der Veen (ed.), *Op je woorden passen. Voordrachten gehouden tijdens het symposium van het Instituut voor Nederlandse Lexicologie in Antwerpen op 18 januari 1993*, 10-22. Leiden: Instituut voor Nederlandse Lexicologie.
- 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 voetbalnamen*. Amsterdam: Meertens Instituut.

Geeraerts, J. (1980). *De Coltmoorden*. Antwerpen: Manteau.

Geerts, G., W. Haeseryn, J. de Rooij & M.C. van den Toorn (1984). *Algemene Nederlandse Spraakkunst*. Groningen: Wolters-Noordhoff.

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 K.U.Leuven.

Grondelaers, S. & M. Brysbaert (1996). "De distributie van het presentatieve er buiten de eerste zinsplaats". *Nederlandse Taalkunde* 1/4, 280-305.

Grondelaers, S. & D. Geeraerts (1998). "Vagueness as a euphemistic strategy", in A. Athanasiadou & E. Tabakowska (eds.), *Speaking of emotions: conceptualization and expression*, 357-374. Berlijn: Mouton de Gruyter.

Grondelaers, S., K. Deygers, H. Van Aken, V. Van den Heede & D. Speelman (2000). "Het CONDIV-corpus geschreven Nederlands". Verschijnt in *Nederlandse Taalkunde* 5/4.

Grondelaers, S. & D. Geeraerts (ter perse). "Towards a pragmatic model of cognitive onomasiology". Verschijnt in H. Cuyckens & Dominiek Sandra (eds.), *Cognitive perspectives on lexical semantics*. Berlijn: Mouton de Gruyter.

Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn (1997). *Algemene Nederlandse Spraakkunst*. Groningen en Deurne: Martinus Nijhoff – Wolters Plantyn.

Harrison, S. (1998). "E-mail discussions as conversation: moves and acts in a sample from a listserv discussion". On-line publicatie op <http://viadrina.euv-frankfurt-o.de/~wjournal/harrison.htm>

Hentschel, E. (1998). "Communication on IRC". On-line 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.
- Istendael, G. Van (1993). *Het Belgisch labyrint. Wakker worden in een ander land*. Amsterdam: De Arbeiderspers.
- Jordens, P. (1974). "Das Deutsche es und die Niederländischen Entsprechungen er/het. Eine kontrastive Analyse". In: H. Pörnbacher (ed.), *Festgabe des Deutschen Instituts der Universität Nijmegen Paul B. Wessels zum 65. Geburtstag*, 157-189. Nijmegen.
- Kirsner, R.S. (1979). *The problem of presentative sentences in Dutch*. Amsterdam: North Holland Publishing Company.
- Ko, Kwang-Kyu (1996). "Structural characteristics of computer-mediated language: a comparative analysis of Interchange discourse". *Electronic Journal of Communication* 6/3. On-line publicatie op <http://ublib.buffalo.edu/libraries/eresources/ejournals/records/ejc.html>
- Labov, W. (1970). "Proposal for continuation of research on sound changes in progress", submitted to the National Science Foundation (NSF-GS-3287).
- Labov, W. (1972). "Some principles of linguistic methodology". *Language in Society* 1, 97-120.
- Levin, B. (1993). *English verb classes and alternations. A preliminary investigation*. Chicago & London: The Chicago University Press.
- Moro, A. (1991). "The raising of predicates: copula, expletives and existence". *M.I.T. Working Papers in Linguistics* 15, 119-181.
- Mulisch, H. (1982). *De aanslag*. Amsterdam: De Bezige Bij.
- Nieuwborg, E. (1968). *De distributie van het onderwerp en het lijdend voorwerp in het huidige geschreven Nederlands in zijn A.B.-vorm*. Antwerpen: Plantijn.

- Paardekooper, P.C. (1971⁴). *Beknopte ABN-syntaksis*. Den Bosch: Malmberg.
- Paolillo, J. (1999). "The virtual speech community: social network and language variation on IRC". *Journal of Computer-Mediated Communication* 4/4. On-line publicatie op <http://jcmc.huji.ac.il/paolillo.htm>
- Rietveld, T. & R. van Hout (1993). *Statistical techniques for the study of language behaviour*. Berlijn: Mouton de Gruyter.
- Schermer-Vermeer, E.C. (1985). "De onthullende status van *er* in de generatieve grammatica". *Spektator* 15/2, 65-84.
- Schermer, I. (1986). "Er was eens...". *Onze Taal* 55, 48-49.
- Speelman, D. (1997). *Abundantia Verborum. A computer tool for carrying out corpus-based linguistic case studies*. Doctorale dissertatie K. U. Leuven.
- Stroop, J. (1998). *Poldernederlands. Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.
- Swiggers, P. & K. Van den Eynde (1985). "Distributie- en combinatiemogelijkheden van Nederlands *er*: een studie in syntactische classificatie". *Linguistics in Belgium* 7, 67-86.
- Swiggers, P. & K. Van den Eynde (1987). "Over *er*". *Forum der Letteren* 28 (Themanummer ANS), 129-132.
- Taeldeman, J. (1992). "Welk Nederlands voor de Vlamingen?". *Nederlands van Nu* 40, 33-52.
- Verkuyl, H.J. (1998). "O corpora, O mores". *Nederlandse Taalkunde* 3, 60-63.
- Zwart, C. J-W. (1992). "Dutch expletives and small clause predicate raising". *Proceedings N.E.L.S.* 22, 477-491.