

De graad van produktiviteit van het suffix *-ing*

Harald Baayen*

1. Inleiding: een kwantitatieve benadering van produktiviteit

In de klassieke benadering van het verschijnsel produktiviteit, te vinden in onder meer het werk van Uhlenbeck en Schultink, worden produktieve woordvormings-procédés onder meer gekarakteriseerd door het feit dat de aantallen door deze procédés voortgebrachte verschillende formaties, de typen, in principe niet telbaar zijn. Produktieve morfologische categorieën kenmerken zich door hun principiële uitbreidbaarheid. Terwijl een inventarisatie van de aantallen typen bij improduktieve categorieën mogelijk is in het licht van de beperkte omvang van dergelijke categorieën, heeft de principiële uitbreidbaarheid van de produktieve morfologische categorie tot gevolg dat een telling van het aantal typen geen uitsluitsel kan bieden ten aanzien van de produktiviteitsvraag. Zoals blijkt uit de tellingen in tabel 1 van de aantallen typen zoals die gevonden worden in het corpus Uit den Boogaart (1975), is het onmogelijk om op grond van het aantal typen, voortaan aan te duiden met *V*, te bepalen of een procédé produktief dan wel improduktief is.

produktief	<i>V</i>	improduktief	<i>V</i>
		-nis	28
		-te	39
-sel	44		
-schap	64		
		nomina actionis met vocalische alternantie	100
-heid	466		
-ing	943		
-je	1031		
		ongelede nomina	1495

Tabel 1: Produktiviteit en *V*

Zo verschijnt het suffix *-sel* (*zaagsel*) met nagenoeg hetzelfde aantal typen als het improduktieve suffix *-te* (*zwakte*). Voor het produktieve suffix *-schap* (*ballingschap*) tellen we minder typen dan voor de improduktieve categorie van nomina actionis met vocalische alternantie (*ingrijpen* — *ingreep* — **ingrijping*). En terwijl de suffixen *-heid* (*goedheid*), *-ing* (*telling*) en *-je* (*woordje*) produktief zijn, tellen we hier toch minder typen dan in het geval van de verzameling van ongelede nomina, een 'categorie' die niet met behulp van morfologische regels uitbreidbaar is en dus

per definitie als improductief moet worden aangemerkt. Het zal duidelijk zijn dat het aantal getelde typen niet voldoet als een kwantitatieve maat voor produktiviteit. De tellingen van de aantallen typen gaan voorbij aan het uitbreidbare karakter van de produktieve categorieën, met als gevolg dat de aantallen mogelijke, maar in het corpus niet gerealiseerde typen ten onrechte niet verdisconteerd worden. In het geval van de improductieve categorie van nomina actionis met vocalische alternantie is het gemiste aantal typen gering, in het geval van *-sel* en *-schap* liggen deze aantallen beduidend hoger (9 versus 81 en oneindig, afgaande op de schattingen in Baayen 1989). Omdat we met een telling van het aantal typen in lexicon of corpus geen greep krijgen op het aantal mogelijke woorden, ligt de conclusie voor de hand dat het verschijnsel produktiviteit zich wezenlijk aan kwantificatie onttrekt (vergelijk Schultink 1961). Anderzijds moeten we vaststellen dat met het hanteren van een begrip als 'graad van produktiviteit', een begrip dat in de theorie van Schultink (1961) het mogelijk maakt het ene produktieve affix als produktiever dan een ander in te schalen, het belang van een objectieve produktiviteitsmaat impliciet gegeven is. Zo doen bijvoorbeeld de getelde aantallen typen vermoeden dat *-ing* (943) produktiever is dan *-sel* (44). Een maat voor produktiviteit, met name in de nauwere interpretatie van produktiviteitsgraad, kan hier zinvol zijn, mits voor deze maat niet het aantal typen V genomen wordt.

In Baayen (1989) heb ik laten zien dat het mogelijk is op principiële wijze de produktiviteit van morfologische categorieën te kwantificeren, mits we voor elk type beschikken over de gebruiks- of tokenfrequentie van dat type. De tokenfrequenties van de typen kunnen we aan de hand van een corpus vaststellen, bijvoorbeeld het corpus Uit den Boogaart (1975), dat aan de in dit artikel genoemde tellingen ten grondslag ligt. Zij nu n_1 het aantal typen dat precies éénmaal in het corpus wordt geteld, en zij N de som van alle tokenfrequenties van alle typen, dan is

$$P = \frac{n_1}{N},$$

een goede benadering van de (mathematische) kans dat bij vergroting van het corpus nieuwe typen gevonden worden. In tegenstelling tot het aantal typen V , is de maat P wel degelijk in staat om produktieve en improductieve categorieën van elkaar te scheiden, zoals blijkt uit tabel 2.

Zowel voor de nomina actionis met vocalische alternantie als voor de ongelede nomina is de kans op nieuwe typen minimaal. De improductieve suffixen *-nis* en *-te* sluiten zich bij de groep van ongelede nomina aan — zij zijn statistisch niet van de ongelede nomina te onderscheiden (zie Baayen 1989). De produktieve suffixen kenmerken zich daarentegen door P -waarden die significant hoger liggen dan die van de ongelede nomina: het betreft hier dan ook principiële uitbreidbare categorieën.

We zijn nu in staat om de begrippen produktiviteit en produktiviteitsgraad van een exacte, kwantitatieve interpretatie te voorzien. De produktiviteitsgraad van een morfologisch procédé meten we met behulp van P . Daarmee interpreteren we de produktiviteitsgraad als een kans, in strikte zin de kans op nieuwe typen bij

produktief	<i>P</i>	improduktief	<i>P</i>
		nomina actionis met vocalische alternantie	0.004
		ongelede nomina	0.008
		<i>-nis</i>	0.013
		<i>-te</i>	0.013
<i>-ing</i>	0.038		
<i>-sel</i>	0.080		
<i>-schap</i>	0.109		
<i>-heid</i>	0.114		
<i>-je</i>	0.253		

Tabel 2: Produktiviteit en *P*

uitbreiding van het gebruikte corpus, in wat bredere zin de toepassingswaarschijnlijkheid van de betreffende morfologische regel. De globale of algehele produktiviteit van een morfologisch procédé karakteriseren we door naast *P* ook *V*, het aantal getelde typen, in de analyse te betrekken, waarbij we *V* de specifieke interpretatie van pragmatische bruikbaarheid geven. Verschillen in *V*, zoals bijvoorbeeld bij nomina op *-sel* (44) en nomina op *-ing* (943), weerspiegelen althans ten dele de mate van inzetbaarheid in het taalgebruik. (Voor nadere motivatie van deze interpretatie van *V* zie Baayen 1989, 1990.) De globale produktiviteit van een morfologische categorie bepalen we dus in termen van het aantal typen *V* in het corpus aangevuld met de kans om buiten het corpus nog nieuwe typen aan te treffen.

Bij uitstek produktief zijn die processen waarvoor zowel *P* als *V* groot zijn. Het diminutiefsuffix (*P* = 0.25, *V* = 1031), en ook de categorie van N+N samenvoegingen (*P* = 0.23, *V* = 4277) zijn voorbeelden van bijzonder produktieve woordvormingsprocédés. Produktief is ook een suffix als *-sel*, dat weliswaar met vrij weinig typen (44) verschijnt, maar toch een hoge *P*-waarde heeft. De globale produktiviteit van *-sel* is dus geringer dan die van het diminutiefsuffix, maar op grond van de gemeten graad van produktiviteit kunnen we vaststellen dat het niet improduktief is.

Bij uitstek improduktief zijn die processen waarvoor zowel *P* als *V* zeer lage waarden aannemen. Meestal gaat een lage waarde van *P* gepaard met een gering aantal typen *V*, zoals in het geval van *-nis* en *-te*, maar het aantal typen kan relatief hoog uitvallen, vergelijk bijvoorbeeld de nomina actionis met vocalische alternantie in tabel 1.

In deze bijdrage ga ik nader in op de produktiviteitsgraad van het suffix *-ing*, zoals we die meten aan de hand van het corpus Uit den Boogaart (1975). We tellen weliswaar 943 typen voor *-ing*, maar het verschijnt met een opmerkelijk lage graad van produktiviteit, namelijk 0.038, de laagste waarde van *P* in tabel 2. Met deze waarde kenmerkt *-ing* zich nog wel als produktief, 0.038 verschilt significant van 0.008, de *P*-waarde van ongelede nomina, maar het contrast tussen

-ing enerzijds (*V* hoog en *P* laag) en *-heid* en *-je* anderzijds (*V* en *P* hoog) vraagt om nadere explicatie. Nu heeft Van Haeringen (1971) erop gewezen dat er bij *-ing* geen sprake is van onbeperkte produktiviteit. Hij acht de mogelijkheden van *-ing* weliswaar numeriek onbegrensd, maar laat vervolgens zien dat de mogelijkheden voor suffigering van *-ing* geenszins onbegrensd zijn. Hij vraagt met name aandacht voor het de produktiviteit remmende effect van blokkering door nominalisaties met vocalische alternantie, en op de beschikbaarheid van alternatieve woordvormingsstrategieën, zoals bijvoorbeeld nominale infinitieven van het type het komen. De produktiviteitsgraad die we voor *-ing* vinden is echter zodanig laag dat het onwaarschijnlijk mag worden geacht dat de door Van Haeringen besproken factoren alléén hiervoor verantwoordelijk kunnen worden gesteld. De aanvullende verklaring die ik in Baayen (1989) heb voorgesteld berust op de observatie dat in het Nederlands een relatief gering aantal grondwoorden beschikbaar is voor suffigering met *-ing*. In deze bijdrage bespreek ik eerst enkele gegevens die deze verklaring ondersteunen. Vervolgens behandel ik een tweede factor die mijns inziens mede verantwoordelijk is voor de lage produktiviteitsgraad van *-ing*, namelijk het sterk 'verbale' karakter van dit nominaliserende suffix.

2. De toevoer van grondwoorden

De verklaring die ik in Baayen (1989) ontwikkel voor deze lage graad van produktiviteit van *-ing* is gebaseerd op de idee dat er niet voldoende verba in het Nederlands beschikbaar zijn voor suffixatie met *-ing* om een hoge produktiviteitsgraad te garanderen bij een groot corpus. Deze verklaring berust op de volgende overwegingen.

Zoals boven is uiteengezet, is het, althans in theorie, karakteristiek voor een productief morfologisch procédé dat het een in principe niet telbaar aantal nieuwe woorden kan voortbrengen. Voorwaarde is dan wel dat de woordvormingsregel ofwel zelf-recursief is (zoals het prefix *her-*, vergelijk *herherverdelen*), of dat de woordvormingsregel beschikt over een oneindige verzameling grondwoorden. Daar *-ing* niet zelf-recursief is, is het voor het genereren van een niet telbaar aantal nieuwvormingen afhankelijk van de grootte van de verzameling grondwoorden. Juist hier doet zich het probleem voor dat in het Nederlands het aantal mogelijke verba sterk begrensd is, en niet adequaat als 'oneindig' gekarakteriseerd kan worden. Voor *-ing* heeft dit tot gevolg dat de aanvoer van voldoende nieuwe typen zodanig stagneert dat de produktiviteitsgraad er merkbaar door beïnvloed wordt, een mogelijkheid die door Uhlenbeck (1953) in het algemene geval reeds is onderkend.

In wat volgt ga ik eerst in op de redenen om aan te nemen dat het aantal mogelijke verba in het Nederlands begrensd is. Vervolgens bespreek ik de relatie tussen de grootte van het corpus en de graad van produktiviteit, om tenslotte de in Baayen (1989) ontwikkelde analyse met additionele observaties te ondersteunen.

Het Nederlands kent een vijftal manieren om werkwoorden te vormen: door suffixatie, door toevoeging van een partikel, door samenstelling, door impliciete transpositie en door prefixatie. Willen deze verschillende procédés een oneindig

aantal verba kunnen genereren, dan moeten zij produktief zijn, terwijl er tevens voldoende recursiviteit in dit morfologisch subsysteem moet zijn ingebouwd. De produktiviteitsgraad van een morfologische categorie wordt namelijk met name gedragen door die formaties die gevormd zijn op basis van meervoudig gelede woorden, vergelijk het paar *onovergankelijk* — *onovergankelijkheid*. Onder deze formaties vinden we de meeste typen met tokenfrequentie 1, waardoor $P = n_1/N$ hoge waarden kan aannemen. (Voor de relatie tussen deze observatie en de wet van Zipf zie Baayen 1989). Is het aantal beschikbare meervoudig gelede grondwoorden voor suffigering met *-ing* gering, dan moet dit de produktiviteitsgraad van *-ing* negatief beïnvloeden. Als we nu de verschillende werkwoordvormende procédés in het Nederlands langslopen, dan vinden we dat de mogelijkheden voor vorming van meervoudig gelede verba inderdaad gering zijn, terwijl ook de produktiviteitsgraad van de betreffende procédés geregeld relatief laag blijkt uit te vallen.

Ten aanzien van de inheemse woordvormingsprocédé's kunnen we vaststellen dat het Nederlands geen gelede verba vormt met behulp van suffixen. In het licht van de voorkeur van talen in het algemeen voor suffixatie — de lexicale toegang tot geprefigeerde woorden verloopt moeizamer dan bij gesuffigeerde woorden, zie Cutler e.a. (1985) — is de afwezigheid van suffixatie in de verbale morfologie van het Nederlands op zich reeds tekenend. Anderzijds is in het niet-inheemse deel van het lexicon het suffix *-eer* wel produktief (vergelijk Baayen 1990), hetgeen de verhoudingsgewijs hoge productiviteitsgraad van nominalisaties van het type democratisering ($P = 0.103$) mogelijk maakt (zie Baayen 1989).

Combinaties van werkwoord en partikel, het type *opbellen*, zijn produktief, maar lenen zich nauwelijks voor verdere prefigering. We kunnen nog wel *heropbellen* vormen, maar een finiete verbale vorm als **herbelt* op is onwelgevoemd. Voorts hechten deze partikels zich bij voorkeur aan monosyllabische grondwoorden. De partikels *door*, *weg* en *aan* verschijnen in het corpus Uit den Boogaart (1975) slechts in drie gevallen in combinatie met een geleed grondwoord op een totaal van 145 typen. We moeten constateren dat dit procédé een marginale bijdrage levert aan de vorming van meervoudig gelede verba.

Door samenstelling kunnen eveneens verba worden gevormd. Naast A+V samenstellingen van het type *goedkeuren* zijn alleen de scheidbare N+V samenstellingen, het type *ademhalen*, produktief (de Vries 1975, Baayen 1990). In vergelijking met nominale samenstellingen is deze produktiviteit echter marginaal. Afgezien van de produktiviteitsvraag, zijn ook hier de mogelijkheden voor verdere verbale derivatie (en samenstelling) beperkt. Composita als *dichtplakken*, *blootstaan* en *pianospelen* lenen zich doorgaans niet als input voor een werkwoordvormende morfologische regel. Ook blijkt dat het met name ongelede verba zijn die als rechter lid van de samenstelling worden gebruikt. Zo is in alle in het corpus Uit den Boogaart (1975) voorkomende A+V en N+V samenstellingen het werkwoord ongeleed.

Werkwoorden komen ook tot stand via impliciete transpositie, een produktief proces. Toch levert het geen substantieel aantal grondwoorden aan voor affigering met *-ing*. Impliciete transpositie is niet zelf-recursief (de Vries 1975), en werkt voornamelijk op ongelede nomina. Afgeleide nomina lenen zich niet voor

impliciete transpositie, al doet een enkele samenstelling, *voetbal* bijvoorbeeld, dit wel. Dit procédé levert dus geen bijdrage aan de vorming van meervoudig gelede grondwoorden als basis voor suffigering met *-ing*. Bovendien kunnen we opmerken dat, aangezien *-ing* zich met name hecht aan gelede grondwoorden (Van Haeringen 1971), het effect van conversie op de produktiviteit van *-ing* minimaal is.

Tenslotte is er de mogelijkheid van prefigering. Prefixen als *be-* (*beschrijven*), *ont-* (*ontploffen*), *her-* (*herzien*) en *ver-* (*verplaatsen*) leveren veel grondwoorden voor *-ing*-suffigering aan. Maar ook hier vinden we dat de vorming van grondwoorden van drie of meer morfemen sterk beperkt is. Verba van het type (*herhervredelen*, *beangstigen*, *vervijvoudigen*) zijn schaars, met als gevolg dat formaties met de complexiteit van bijvoorbeeld *beangstiging* vrij zelden worden aangetroffen.

Uit dit alles blijkt dat de mogelijkheid van vorming van veelvuldig gelede werkwoorden in het Nederlands weliswaar niet ontbreekt, maar dan toch zeer beperkt is: affixstapelning komt nauwelijks voor. Ook de produktiviteit van werkwoordvormende morfologische procédés is veel beperkter van aard dan de produktiviteit van naamwoordvormende procédés, een kwestie waar we in de volgende paragraaf nog op terug komen. Is het theoretisch wellicht mogelijk met de prefixen *her-*, *be-* en *ver-* een oneindig aantal complexe verba te genereren, in de praktijk wordt het gebruik door factoren als conceptualiseerbaarheid, psychologische verwerkbaarheid en pragmatische bruikbaarheid tot een fractie van het theoretisch mogelijke ingeperkt.

Op welke wijze beïnvloedt het eindige karakter van de verzameling verba in het Nederlands de produktiviteitsgraad P van *-ing*? Om deze vraag te kunnen beantwoorden moeten we ons eerst realiseren dat P wiskundig gezien een functie van N , de som van de tokenfrekwenties van alle typen, is. Naarmate N (en dus ook V) groter wordt, daalt P . Deze daling, die op het eerste gezicht wat vreemd aandoet, is bij nader inzien niet meer dan logisch. Beschouw N als de omvang van een steekproef van woorden van een gegeven morfologische categorie. Bij een kleine steekproef (N klein) nemen we slechts een gedeelte van de in de populatie aanwezige typen waar. Bij vergroting van de steekproef tellen we een aantal nieuwe typen, met als gevolg dat het aantal nog niet waargenomen typen in de populatie kleiner is geworden. De kans om dan vervolgens weer nieuwe typen te vinden is daarmee gedaald. Het feit dat de typen verschillende, en deels zeer lage tokenfrekwenties hebben, heeft tot gevolg dat P zelfs bij oneindige populaties wel moet dalen. Wel is het zo dat bij oneindige populaties, karakteristiek voor (zeer) produktieve woordvormingsregels, de waarde van P maar langzaam daalt. Is de onderliggende verzameling typen echter eindig en bovendien klein, dan zal de waarde van P bij vergroting van de steekproef snel lagere waarden aannemen. Improduktieve processen, die gekenmerkt worden door populaties grondwoorden die klein zijn, verschijnen dan ook met minimale waarden van P .

We zijn nu in staat om de vraag naar het effect van de begrensdheid van het aantal mogelijke werkwoorden op de produktiviteitsgraad van *-ing* nader aan te scherpen. Voor het suffix *-ing* tellen we 1025 typen, bij een steekproef van 8326 to kens. Daarbij hebben we nominalisaties afgeleid van verba op *-eer* meegerekend.

Hierdoor komt de waarde van P op 0.042 te liggen. Van de in Baayen (1989) geanalyseerde derivatieve categorieën is dit de grootste steekproefgrootte. We weten dat P daalt bij groeiende steekproefgrootte. Als het aantal beschikbare grondwoorden beperkt is, heeft dit tot gevolg dat de daling van P zich versneld zal doorzetten. De vraag die we ons dus moeten stellen is of de lage gemeten waarde van P voor *-ing* voortvloeit uit het feit dat er 'te weinig' grondwoorden zijn, of dat deze waarde reeds verklaard kan worden op grond van het dalende karakter van P , en dus ook bereikt zou zijn als er wel oneindig veel grondwoorden beschikbaar zouden zijn. In Baayen (1989) neem ik aan dat het relatief beperkte aantal grondwoorden de daling van P versnelt, zonder daarvoor nadere evidentie aan te voeren.

We kunnen deze verklaring van de lage P -waarde van *-ing* echter toetsen door na te gaan hoe de graad van produktiviteit van *-ing* zich verhoudt tot die van bijvoorbeeld *-er* in een kleinere steekproef. Het deverbale *-er* (*gever*) verschijnt in het corpus Uit den Boogaart (1975) met 1755 tokens, 321 typen en een produktiviteitsgraad van 0.076, waarbij we de allomorfen op *-aar* (*weigeraar*) hebben meegeteld. De P -waarde van *-er* is dus bijna tweemaal zo hoog als die van *-ing* (0.042), maar *-er* verschijnt dan ook met veel minder typen en tokens. Als het inderdaad zo is dat het eindige karakter van de verzameling Nederlandse werkwoorden, in combinatie met de grootte van de steekproef, mede verantwoordelijk is voor de lage produktiviteitsgraad van *-ing* in vergelijking met *-er*, dan mogen we verwachten dat dit verschil in de graad van produktiviteit minder wordt bij een kleinere steekproef. De idee is dat er bij een kleinere steekproef van 'uitputting' van de verzameling verbale grondwoorden nog geen sprake is. Het suffix *-er* heeft vanuit deze optiek in het corpus Uit den Boogaart met slechts 1755 tokens nog geen of nauwelijks last van deze uitputting, hetgeen een relatief hoge graad van produktiviteit mogelijk maakt. Het suffix *-ing*, daarentegen, ondervindt bij 8326 tokens wel hinder van de uitputtingsfactor. Als we nu het totale aantal tokens op *-ing* beperken tot de orde van grootte van die op *-er*, dan mag de uitputtingsfactor geen bepalende rol meer spelen. Met andere woorden, als we de aantallen woordvoorkomens (N), de aantallen typen (V) en de produktiviteitsgraden (P) bepalen voor *-ing* en *-er* in een subcorpus van het corpus Uit den Boogaart (1975), bijvoorbeeld het deelcorpus Cdbl, dat materiaal uit de dagbladers bevat, dan voorspellen we dat *-ing* verschijnt met een graad van produktiviteit die dichter komt te liggen bij die van *-er*.

Tabel 3 bevestigt ons vermoeden: in het volledige corpus Uit den Boogaart (1975) is de P -waarde van *-ing* iets meer dan de helft van die van *-er* (0.552), in het deelcorpus Cdbl is P_{ing} $4/5$ (0.804) van P_{er} .

Op dit punt moeten we ons wel realiseren dat de wijze waarop we de reductie in

	N	V	P
<i>-ing</i>	1994	493	0.115
<i>-er</i>	551	155	0.143

Tabel 3: N , V en P voor *-ing* en *-er* in het subcorpus Cdbl

de steekproefgrootte tot stand hebben gebracht, namelijk door van het volledige corpus Uit den Boogaart (1975) (geschreven taal) over te gaan op een deelcorpus, het neveneffect met zich mee brengt dat het aantal tokens in *-ing* sterker is gereduceerd dan het aantal tokens in *-er* ($1994/8326=0.24$ in het geval van *-ing*, $551/1755=0.31$ in het geval van *-er*). We moeten daarom nog nagaan of bij een reductie van *-er* tot ongeveer een kwart van het aantal tokens het nog steeds zo is dat de produktiviteitsgraden van *-ing* en *-er* dichter bij elkaar liggen dan in het volledige corpus. Voor $0.24 \times 1755 = 420$ tokens op *-er* in het Cdbl blijkt het aantal typen 135 te zijn en de produktiviteitsgraad 0.171. Ook onder deze heranalyse is de produktiviteitsgraad van *-ing* nog altijd $2/3$ (0.672) van die van *-er*. We kunnen dus concluderen dat het eindige karakter van de verzameling grondwoorden naar alle waarschijnlijkheid dankzij het grote aantal typen de produktiviteit van *-ing* merkbaar negatief beïnvloedt.

3. Het sterk verbale karakter van *-ing*

Opvallend is dat *-er* in tabel 3 nog steeds een hogere produktiviteitsgraad kent dan *-ing*. Dit verschil nodigt uit tot nader onderzoek: al zijn *-ing* en *-er* geen rivaliserende affixen, hetgeen het vergelijken van hun produktiviteitsgraden bemoeilijkt, toch loont het de moeite om te trachten te achterhalen welke factor voor dit verschil verantwoordelijk is. Vermoedelijk moeten we deze factor in verband brengen met het sterk verbale karakter van *-ing*. Eerst laat ik zien dat nomina en verba zich kwantitatief verschillend gedragen, en dat het frekwentiegedrag van *-ing* overeenkomsten vertoont met dat van verba. Vervolgens ontwikkel ik de idee dat nominalisaties op *-ing* meer verbale karakteristieken hebben dan nominalisaties op *-er*, en dat het aan dit sterk verbale karakter van *-ing* te wijten is dat dit suffix een lagere graad van produktiviteit heeft dan *-er*.

Een klassiek onderscheid tussen nomina en verba is dat nomina typisch verwijzen naar (fysische) entiteiten, en dat verba prediceren, dat wel zeggen entiteiten met elkaar in verband brengen. Afgezien van het feit dat het wellicht mogelijk is om nomina en verba in semantische termen consequent van elkaar te onderscheiden, zoals betoogd door Langacker (1987), kunnen we vaststellen dat werkwoorden, in tegenstelling tot nomina van het type *stoel* en *huis*, een argumentstructuur hebben. In dit opzicht zijn verba dus wezenlijk complexer dan nomina. Onderzoek naar de interconceptuele organisatie van nomina en verba in het mentale lexicon bevestigt dit. Graesser e.a. (1987) laten zien dat nomina zich relatief goed lenen voor een hiërarchische, taxonomische organisatie. Zo laten nomina als *dier*, *zoogdier*, *walrus*, *insekt*, *koel*, *hond* en *vlieg* zich goed ordenen met behulp van een boomstructuur. Bij sommige deelverzamelingen verba, bijvoorbeeld bij werkwoorden van beweging (*lopen*, *kruipen*, *schuifelen*, *vliegen*, *hinkelen*) is een taxonomische ordening eveneens mogelijk. Doorgaans echter leidt een matrixorganisatie, waarbij extensieve kruisklassifikatie toegepast wordt, bij verba tot betere ordeningen. Deze verschillen tussen nomina en verba hangen samen met het referentiële karakter van nomina, dat nomina doorgaans tot 'stand alone' concepten maakt, en met het predicerende karakter van verba, dat met zich

meebrengt dat verba in een groot aantal semantische velden participeren en voor hun interpretatie voor een belangrijk deel van hun argumenten afhankelijk zijn (vergelijk *enter a city* en *enter into a computer file*).

Nomina en verba onderscheiden zich eveneens door hun frekwentiegedrag. Dit blijkt onder meer als we ongelede verba en ongelede nomina met elkaar vergelijken.

	<i>N</i>	<i>V</i>	<i>N/V</i>	<i>P</i>
Nederlands (600.000)				
nomina	37836	1495	25.31	0.008
verba	169602	1757	96.53	0.002
Engels (18.000.000)				
nomina	2781258	6582	422.56	0.0001
verba	3660693	2581	1418.32	0.0000

Tabel 4: Frekwentiegegevens voor ongelede nomina en verba in het Nederlands en het Engels.
Bronnen: Baayen (1989, 1990), Baayen & Lieber (1990).

Tabel 4 illustreert dit aan de hand van het Nederlands (corpus Uit den Boogaart (1975), 600.000 tokens) en het Engels (Cobuild corpus (Renouf 1987) zoals verwerkt op CELEX, 18.000.000 tokens). Het gemiddeld aantal tokens per type, *N/V*, ligt voor ongelede verba veel hoger dan bij ongelede nomina. De *P* waarden laten hier zien dat de kans om bij uitbreiding van het Uit den Boogaart corpus nog nieuwe ongelede woorden tegen te komen voor nomina substantieel groter is dan voor verba. In het geval van het veel grotere Cobuild corpus is deze laatste kans zelfs verwaarloosbaar (0.000007). Met andere woorden, bij de ongelede verba vallen we eerder en vaker 'in herhaling' dan bij ongelede nomina.

Ook bij gelede nomina en verba vinden we dit verschil in frekwentiegedrag terug. Tabel 5 somt de frekwentiegegevens van een representatief aantal nominale en verbale categorieën in het Nederlands op (zie verder Baayen 1989, 1990). Twee dingen vallen op. In de eerste plaats missen we bij de verba derivatieve processen die grote aantallen typen opleveren, zoals *-tje* (1031), *-heid* (466) en *-ing* (943). In de tweede plaats valt op dat een prefix als *ver-*, dat een redelijk aantal typen oplevert, gekenmerkt wordt door een vrij lage graad van produktiviteit (0.035). Omgekeerd vinden we een hoge graad van produktiviteit alleen bij *door-*, maar nu is het aantal typen bijzonder gering (30). De algehele produktiviteit van de verbale morfologie, zo moeten we vaststellen, is veel minder uitbundig dan die van de nominale morfologie, waar *P* en *V* gezamenlijk grote waarden kunnen aannemen.

We hebben gezien dat verba cognitief complexer zijn dan nomina, en dat zij in vergelijking met nomina veel hogere gebruiksfrekwenties kennen. Deze twee feiten

	<i>N</i>	<i>V</i>	<i>P</i>
<i>-tje</i>	2580	1031	0.253
<i>-heid</i>	2251	466	0.114
<i>-ing</i>	8049	943	0.038
<i>-sel</i>	261	45	0.080
<i>-schap</i>	265	64	0.109
<i>ver-V</i>	953	118	0.035
<i>-eer</i>	358	79	0.073
<i>be-N</i>	725	64	0.028
<i>door-</i>	109	30	0.119
<i>ont-N</i>	124	24	0.097

Tabel 5: Nominale en verbale categorieën in het Nederlands (corpus Uit den Boogaart 1975).

staan vermoedelijk niet los van elkaar: hoge gebruiksfrekwenties bevorderen het leren en onthouden van woorden, en dragen er toe bij dat de conceptueel moeilijker woorden, de verba, geleerd en onthouden kunnen worden. Vermoedelijk is deze conceptuele complexiteit er mede verantwoordelijk voor dat de algehele produktiviteit van verbale categorieën gedempt is ten opzichte van nominale categorieën.

We keren terug tot de produktiviteit van de nominaliserende suffixen *-ing* en *-er*. Dit soort nominalisaties staan bekend als typische tegenvoorbeelden voor semantische definities van de woordsoorten *N* en *V*. Immers, deze nominalisaties erven aspecten van de argumentstructuur van hun grondwoorden. Zo vinden we in het Cdbl de NP

- (1) haar afkeuring van Ben Goerion's pro-Israëliische propagandistische opzet van het proces

die gerelateerd kan worden met de zin

- (2) Zij keurt Ben Goerion's pro-Israëliische propagandistische opzet van het proces af.

Nominalisaties op *-ing* delen klaarblijkelijk in de conceptuele complexiteit van hun grondwoord. Als de boven geformuleerde hypothese dat conceptuele complexiteit en intensief gebruik hand in hand gaan juist is, dan moet het verbale karakter van nominalisaties op *-ing* mede verantwoordelijk gesteld worden voor de relatief lage graad van produktiviteit van *-ing*.

Ook nominalisaties op *-er* vertonen verbale trekken. De NP

- (3) de stichter van de Katendrechtse speeltuinvereniging en het Katendrechtse volkshuis

kunnen we beschouwen als corresponderend met

- (4) Hij heeft de Katendrechtse speeltuinvereniging en het Katendrechtse volkshuis gesticht.

Er is echter een cruciaal verschil in vergelijking met het boven gegeven voorbeeld met *afkeuring*. In het geval van *stichter* is het externe argument van het verbale grondwoord gebonden en dus niet beschikbaar voor nadere specificatie in de NP. Bij *afkeuring* is het externe argument vrij, het kan apart in de NP verschijnen. De argumentstructuur van *-er* is beknot ten opzichte van die van *-ing*. Met andere woorden, *-er* is nominaler van karakter dan *-ing*. Dit doet vermoeden dat het feit dat *-ing* met een lagere graad van produktiviteit verschijnt dan *-er* verband houdt met een verschil ten aanzien van het verbale karakter tussen deze suffixen.

Interessant is in dit verband dat intransitieve verba voor atmosferische verschijnselen (*hagelen, donderen, onweren*) zich, zoals Van Haeringen (1971) opmerkt, niet lenen voor nominalisatie met *-ing*. Hij verklaart dit door een beroep te doen op het blokkerende effect van de nomina *hagel, donder* en *onweer*. Zijn analyse maakt echter niet duidelijk waarom nominale infinitieven als *het onweren* en *het donderen* wel mogelijk zijn en niet geblokkeerd worden. In het licht van het voorafgaande kunnen we deze analyse gemakkelijk aanscherpen. Bij deze 'meteorologische' verba is, in de terminologie van Talmy (1985), sprake van lexicalisatie van de 'figure' in het ongelede werkwoord. Anders gezegd, het externe argument is hier afwezig en in het werkwoord gelexicaliseerd, hetgeen het gebruik van onpersoonlijke constructies afdwingt. Dit heeft tot gevolg dat een nominalisatie met *-ing* geen argumentstructuur meer kan erven. Zo'n formatie zou derhalve volledig nominaal zijn, met als gevolg dat de nomina *hagel, donder* en *onweer* een bijzonder sterk blokkerend effect kunnen uitoefenen. Dat nominale infinitieven niet geblokkeerd worden wordt inzichtelijk als we met Van Haften e.a. (1985) aannemen dat nominale infinitieven verbaler van karakter (in kenmerken: [+V]) zijn dan nominalisaties op *-ing* ([-V]). Daar (*het*) *hagelen*, (*het*) *bliksemen* en (*het*) *donderen* het kenmerk [+V] dragen, terwijl *hagel, bliksem* en *donder* gespecificeerd zijn als [-V], kan blokkering zich niet voordoen.

We keren terug tot het kwantitatief vergelijken van *-er* en *-ing*. Is er, naast het gemeten verschil in produktiviteit (0.042 versus 0.076 in het hele corpus Uit den Boogaart, 0.115 versus 0.143 in het deelcorpus Cdbl) nog meer kwantitatieve evidentie aan te voeren dat deze analyse ondersteunt? Met andere woorden, zijn er complementaire aanwijzingen dat *-ing* verbaler is dan *-er*? Laten we om deze vraag te beantwoorden nagaan hoe vaak *-ing* en *-er* zich daadwerkelijk 'verbaal' gedragen in het deelcorpus Cdbl. Daarbij maken we onderscheid tussen twee typen constructies: constructies met incorporatie, bijvoorbeeld *luchtverontreiniging* en *werkgever*, en constructies waarbij de argumenten syntactisch gerealiseerd worden als prenominale dan wel postnominale genitivische NP's (zie de voorbeelden (1) en (3) boven.) In het vervolg verwijs ik naar deze constructies met de afkortingen INC en SYN.

Tabel 6 vat de frekwentiegegevens van *-ing* en *-er*, uitgesplitst naar INC en SYN, samen. Er is uitgesplitst naar constructietype — 'NP's V-ing van NP' versus 'NP-V-ing' — waarbij geabstraheerd is van de lexicale inhoud van de NP's.

Bij het tellen van de SYN constructies doet zich nog het probleem voor, dat

	<i>-ing</i>		<i>-er</i>	
	SYN	INC	SYN	INC
<i>N</i>	438	176	99	197
<i>V</i>	222	74	39	45
<i>n_i</i>	139	42	26	25
<i>P</i>	0.32	0.24	0.26	0.13

Tabel 6: Frekwentiegegevens van *-ing* en *-er* in het deelcorpus Cdbl, uitgesplitst naar het gebruik van syntactische (SYN) en incorporerende (INC) constructies.

nominalisaties als *woning* en *bekleding* hun argumentstructuur hebben verloren. We moeten dus rekening houden met de mogelijkheid dat *van* in een NP met de opbouw 'V-ing van NP' een algemene genetivische relatie tot uitdrukking brengt, zoals bijvoorbeeld in *de koers van het schip*. Of er al dan niet sprake is van een echt argument is niet altijd even duidelijk. Bij mijn tellingen heb ik de soms complementaire criteria gehanteerd dat (1) de NP een natuurlijk syntactisch analogon moet hebben wil er van een argument sprake zijn, en (2) dat het bestaan van een vergelijkbare constructie met incorporatie eveneens voor de aanwezigheid van een argument pleit. In *deze speciale aflevering van de gids* is *de gids* geen argument van *aflevering*, maar in *de aflevering van het schip* is *het schip* dat wel. Lastiger is de keuze bij *de voorzitter van de raad*, waar analoog aan *de kapitein van het schip* sprake kan zijn van een algemene genetivische relatie. Op grond van de overweging dat van *de voorzitter van de raad* geldt dat hij de raad voorziet, en op grond van het feit dat het Nederlands vergelijkbare gevallen met incorporatie kent, zoals bijvoorbeeld *fractievoorzitter*, heb ik *de raad* als argument van *voorzitter* gehonoreerd.

Gegeven dat de tellingen op de juiste manier zijn uitgevoerd, kunnen we uit tabel 6 aflezen dat *-ing* produktiever is dan *-er* in beide typen constructies. Zowel wat de aantallen typen als wat de graad van produktiviteit betreft overtroeft het *-er*. Nu kunnen we de vergelijking aanscherpen door alleen die gevallen mee te nemen waar *-ing* een constructie aangaat met het interne argument. Immers, het suffix *-er* bindt het externe argument, hetgeen de SYN constructie bij de

	<i>-ing</i>		<i>-er</i>	
	SYN	INC	SYN	INC
<i>N</i>	316	147	99	197
<i>V</i>	177	60	39	45
<i>n_i</i>	114	34	26	25
<i>P</i>	0.36	0.23	0.26	0.13

Tabel 7: Frekwentiegegevens van *-ing* en *-er* in het deelcorpus Cdbl, uitgesplitst naar het gebruik van syntactische (SYN) en incorporerende (INC) constructies. Alleen constructies met een intern argument zijn meegerekend.

intransitieve grondwoorden van *-er* onmogelijk maakt. Als *-er* intrinsiek nominaler is dan *-ing*, dan zal de globale produktiviteit van *-er* wat de INC en SYN constructies betreft ook geringer moeten uitvallen dan die van *-ing* als we *-ing* dezelfde beperking opleggen die bij *-er* met het taalsysteem gegeven is. Beperken we dus de vergelijking tot formaties afgeleid van transitieve verba, dan blijkt dat *-er* inderdaad minder produktief blijft dan *-ing*.

Tabel 7 vertoont hetzelfde beeld als tabel 6: *-ing* blijft dominant, zowel wat betreft het aantal typen als wat betreft de graad van produktiviteit, voor beide constructies. Ons vermoeden dat *-ing* intrinsiek verbaal is dan *-er* is hiermee bevestigd.

Dit resultaat hebben we verkregen door te abstraheren van de vorm van de NP's. Bij de syntactische constructies ligt dit voor de hand. Als we daar differentiëren naar de vorm en het lexicale materiaal in de NP's, blijkt bijna elk token van een nominalisatie een type op te leveren. Bij *-ing* zijn er slechts 13 gevallen waar precies dezelfde NP meer dan eens als argument van dezelfde nominalisatie verschijnt (bijvoorbeeld *wijziging der statuten*). Bij *-er* gaat het om één enkel geval (*de voorzitter van de raad*). De SYN constructie krijgt bij deze wijze van tellen een produktiviteitsgraad van 0.941 in het geval van *-ing* ($N = 438$, $V = 425$, $n_1 = 412$), en een produktiviteitsgraad van 0.980 in het geval van *-er* ($N = 99$, $V = 98$, $n_1 = 97$). Onder deze telling vallen verschillen in produktiviteit weg, hetgeen te verwachten is: de ongekende mogelijkheden van syntactische variatie binnen de NP voeren de maat voor de graad van produktiviteit op tot een waarde nagenoeg gelijk aan 1.0.

Anders ligt het als we de gevallen van incorporatie zo tellen dat *luchtvervuiling* en *watervervuiling* als twee typen in plaats van als twee tokens van één type geteld worden. Een dergelijke telling kan gemotiveerd worden op grond van de overweging dat *luchtvervuiling* en *watervervuiling* twee verschillende woorden zijn.

INC	<i>N</i>	<i>V</i>	n_1	<i>P</i>
<i>-er</i>	197	96	73	0.37
<i>-ing</i>	176	124	103	0.59

Tabel 8: Tellingen van nominalisaties op *-ing* en *-er* met incorporatie, met differentiatie naar het geïncorporeerde materiaal. (Cdbl)

Zoals tabel 8 laat zien, bewijst ook onder deze meer lexicale invalshoek het suffix *-ing* zich als produktiever dan het suffix *-er*: zowel het aantal typen als de graad van produktiviteit ligt wederom beduidend hoger.

4. Conclusies

Van Haeringen (1971) heeft erop gewezen dat de produktiviteit van *-ing* verre van onbegrensd is. Hij vestigde met name de aandacht op een aantal verschijnselen die de blokkering van nominalisaties op *-ing* van ongelede verba betreffen. Dat het

hier om een significante factor gaat blijkt uit het feit dat van de ongeveer 490 nominalisaties op *-ing* in het Cdbl er slechts 109 zijn afgeleid van een ongeleed werkwoord, terwijl er zo'n 157 formaties van het type *waarneming* voorkomen (afgeleid van 118 verschillende verba), waarvan het grondwoord in dit subcorpus niet met *-ing* gesuffigeerd gevonden wordt.

In deze bijdrage heb ik beargumenteerd dat de relatief lage produktiviteitsgraad van nominalisaties op *-ing* daarnaast terug te voeren valt op twee factoren die met name nominalisaties van gelede werkwoorden betreffen. In de eerste plaats wordt bij een voldoende grote steekproef merkbaar dat de voorraad verbale grondwoorden in het Nederlands beperkt is. In de tweede plaats blijkt het sterk verbale karakter van *-ing*, dat met name in vergelijking met het deverbale *-er* duidelijk wordt, remmend op de produktiviteit te werken. Dat de aard van het grondwoord een factor kan zijn die de morfologische produktiviteit mede bepaalt is reeds lang bekend. In deze bijdrage heb ik laten zien dat het effect van deze factor meetbaar is.

Ter afsluiting volgt nog een tweetal opmerkingen van meer methodologische aard. (1) Bij het inschatten van de produktiviteit van woordvormingsprocédés beperkt men zich doorgaans tot het vergelijken van rivaliserende affixen. De hier geboden analyse van *-er* en *-ing* laat zien dat een kwantitatieve analyse de mogelijkheden van het onderzoek naar factoren die de produktiviteit bepalen op dit punt verruimt. (2) De tabellen 6—8 laten zien dat zowel bij *-ing* als bij *-er* de SYN constructie produktiever is dan de INC constructie. Ook dit resultaat ligt in de lijn der verwachting: hoe morfologischer van aard de constructie, des te lager zal de graad van produktiviteit uitvallen.

NOOT

* Ik ben Geert Booij, Piet van Reenen en Ariane van Santen erkentelijk voor hun commentaar op eerdere versies van dit artikel.

BIBLIOGRAFIE

- Baayen, R.H. (1989) *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*. Diss. Vrije Universiteit, Amsterdam.
- Baayen, R.H. (1990) Corpusgebaseerd onderzoek naar morfologische produktiviteit, *Spektator* 19.3.
- Baayen, R.H. en Lieber, R. (1990) *Productivity and English Derivation*, ms. Vrije Universiteit.
- Cutler, A. en Hawkins, J.A. en Gilligan, G. (1985) The Suffixing Preference: a Processing Explanation, *Linguistics* 23, 723—758.
- Graesser, A.C. en Hopkinson, P.L. (1987) Differences in Interconcept Organization between Nouns and Verbs, *Journal of Memory and Cognition* 26, 242—253.
- Haaften, T. van, en Kerke, S. van der, en Middelkoop, M. en Muysken, P. (1985) Nominalisaties in het Nederlands, *Glott* 8, 67—104.
- Haeringen, C.B. van (1971) Het achtervoegsel *-ing*. Mogelijkheden en beperkingen, *De Nieuwe Taalgids* 64, 449—468.
- Langacker, R.W. (1987) Nouns and Verbs, *Language* 63, 53—94.
- Renouf, A. (1987) Corpus Development, in J.M.Sinclair (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing*, Collins.

- Schultink, H. (1961) Produktiviteit als morfologisch fenomeen, *Forum der Letteren* 2, 110—125.
- Talmy, L. (1985) Lexicalisation Patterns: Semantic Structure in Lexical Forms, in: T. Shopen (ed.), *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*, Cambridge, Cambridge University Press.
- Uhlenbeck, E.M. (1953) The Study of Word-Classes in Javanese, *Lingua* 3, 322—354.
- Uit den Boogaart, P.C. (ed.) (1975) *Woordfrequenties in gesproken en geschreven Nederlands*, Utrecht, Oosthoek, Scheltema & Holkema.
- Vries, J.W.de (1975) *Lexicale morfologie van het werkwoord in modern Nederlands*, Leiden, Universitaire Pers.