

Morfologische regels: ik weet, ik weet wat jij ook weet

Jan Don

Wie een spelletje ‘Scrabble’ speelt, gebruikt een woordenboek als scheidsrechter. Het antwoord op de vraag ‘bestaat dat woord eigenlijk wel’ wordt met een blik in het woordenboek snel beslist. Dit stukje is een pleidooi om die huiselijke praktijk te veranderen; iedere moedertaalspreker beschikt over een veel betere scheidsrechter dan welk woordenboek ook. Die scheidsrechter zit in ons hoofd en beschikt over een verzameling regels terwijl het woordenboek niets anders is dan een (wat lange) lijst.

Wat zijn precies de woorden van het Nederlands? We kunnen de woorden in het woordenboek opvatten als een lijst. We zouden elk woord een getal kunnen geven, tellend vanaf 0 en beginnend bij het woord ‘a’. Tegen de tijd dat we in Van Dale bij het woord ‘zymotisch’ zijn aangekomen hebben we zo’n 300.000 woorden geteld. Een aanzienlijke verzameling, maar wel een eindige verzameling. Bovendien zijn ons tijdens het tellen een aantal dingen opgevallen: in de eerste plaats staan er een heleboel woorden niet in het woordenboek. Het woord *tilbaar* staat wel in Van Dale, maar we zoeken tevergeefs naar *tilbaarheid*. De argeloze ‘Scrabble’-speler zou nu zo maar kunnen beslissen dat dit kennelijk geen woord van het Nederlands is. Maar waarom eigenlijk niet? Iedere spreker van het Nederlands weet wat dat woord betekent (‘de mogelijkheid om getild te kunnen worden’), hoe het wordt uitgesproken, dat de klemtoon op de eerste lettergreep ligt, dat het om een zelfstandig naamwoord gaat en dat het woord gecombineerd wordt met het lidwoord ‘de’ (en niet met ‘het’). Als sprekers van het Nederlands dergelijke specifieke en gedetailleerde kennis hebben over een woord (ook al zal niet iedereen die kennis op die manier expliciet kunnen maken), wordt het wel moeilijk om nog vol te houden dat *tilbaarheid* géén woord is van het Nederlands. Maakt Van Dale een vergissing? Is *tilbaarheid* per abuis weggelaten? Van Dale vermeldt, om nog eens een voorbeeld te noemen, wel het woord *scheepsarcheologie*, maar *scheepsarcheoloog* ontbreekt. Is *scheepsarcheoloog* geen woord? De vraag stellen is hem beantwoorden: kennelijk bevat Van Dale niet alle woorden van het Nederlands.

Maar al had *scheepsarcheoloog* er wel in gestaan, hoe waarschijnlijk was het geweest dat de redactie had besloten om ook *scheepsarcheoloogje*, *scheepsarcheologiemuseum*, *scheepsarcheologenopleiding*, etc. op te nemen? En als de redactie al tot zo’n besluit was gekomen, had die redactie dan niet ook moeten besluiten om het volgende rijtje op te nemen?

- (i) a. scheepsarcheologiemuseumdirecteur
- b. scheepsarcheologenopleidingsinstituut
- c. scheepsarcheologenopleidingsinstituutmedewerkers

En zelfs dan... u voelt het wel: er zit geen eind aan de lijst woorden die we op goede gronden moeten beschouwen als ‘woorden van het Nederlands’. Geen woordenboek is groot genoeg om deze woorden allemaal te herbergen. Altijd kunnen we wel weer een woord bedenken dat niet in de lijst staat: de verzameling woorden van het Nederlands is oneindig.

Geen lijst maar regels

De verzameling woorden van het Nederlands kan dan ook niet met een lijst worden weergegeven; daarvoor is de verzameling te groot. Wat we nodig hebben zijn regels. Als we de verzameling natuurlijke getallen willen opschrijven, schiet een lijst op dezelfde manier tekort: er is altijd wel weer een groter getal te vinden (door bijvoorbeeld 1 op te tellen bij het grootste getal in de lijst) dat niet in de lijst staat. Maar we kunnen wel een procedure of regel bedenken die gegarandeerd alle natuurlijke getallen oplevert (en geen andere!). De procedure in (ii) bijvoorbeeld is hiervoor heel geschikt:

- (ii) Als x een getal is behorend tot N , dan geldt dat $x + 1$ eveneens behoort tot N .

Waarbij N staat voor de verzameling (niet negatieve) gehele getallen. Op dezelfde manier kunnen we voor het Nederlands zeggen:

- (iii) Als x een woord is behorend tot NL, en y is een woord behorend tot NL, dan geldt dat $x+y$ eveneens een woord is behorend tot NL.

Waarbij NL staat voor het Nederlands. Als *museum* een woord is en *directeur* is een woord dan is *museum+directeur* ook een woord. Als moedertaalsprekers van het Nederlands hebben we een regel als (iii) in ons hoofd. Ze behoort tot de kennis die we van onze moedertaal hebben en waar we als taalgebruikers - of 'Scrabble'-spelers - dus een beroep op kunnen doen.

Een woordenboek is een lijst axioma's

Nu bestaat zo'n procedure of regel als (iii) bij de gratie van een startpunt, een axioma. Als we willen vaststellen of *scheepsarcheologiemuseumdirecteur* een woord is, moeten we eerst kijken of *scheepsarcheologiemuseum* een woord is en of *directeur* een woord is. Dat eerste kunnen we bepalen door eerst vast te stellen of *scheepsarcheologie* een woord is en of *museum* een woord is. En om dat vast te kunnen stellen, moeten we weten of *scheep* een woord is en of *archeologie* een woord is. We zien dus dat de procedure *recursief* is: de procedure herhaalt zichzelf voor steeds kleinere delen van het oorspronkelijke woord. Totdat we eindigen met een lijst woorden die niet verder opgedeeld kunnen worden. De woorden *museum*, *directeur*, *archeologie* etc. zijn allemaal niet verder op te splitsen in kleinere delen en om vast te stellen of dergelijke woorden werkelijk woorden van het Nederlands zijn, hebben we dus wel degelijk een lijst nodig. Het woordenboek vormt een lijst axioma's van de taal. Dat *museum* een woord is van het Nederlands met de betekenis 'museum' volgt nergens uit; het is een feit dat we als moedertaalsprekers van het Nederlands ooit hebben moeten leren en onthouden.

Zonder dergelijke axioma's is het lastig praten. Hieronder staat het eerste couplet van een Nederlandse vertaling van Lewis' Carroll *Jabberwocky* (*De Krakelwok* van Ab Westervaarder & René Kurpershoek):

't Was bradig en de slijp'le torfs
Driltolden op de wijde weep:
Misbrozig stonden borogorfs,
't Verdoolde grasvark schreep.

We herkennen aardig wat Nederlands in zo'n couplet: *'t, was, op, de* etc. allemaal bekende Nederlandse woorden, maar geen woorden die je snel geneigd bent in het woordenboek op te zoeken. Dit soort elementen vormen slechts het cement tussen de bakstenen. De bakstenen zelf daarentegen, de inhoudswoorden, zijn allemaal niet terugvindbaar in een Nederlands woordenboek. Geen idee wat *bradig* is of *weep*. *Misbrozig* lijkt erg op *mistroostig*, maar het is het niet. En *grasvark* zal wel een dier zijn, en zou heel goed gevormd kunnen zijn met regel (iii), maar bij het vaststellen of *vark* een woord is, geeft het Nederlandse woordenboek niet thuis. Kortom, deze 'Scrabble-speler' overspeelt zijn hand en wordt terecht teruggefloten.

Recursie

De regel maakt het woordenboek niet overbodig. Dat strookt precies met ons gevoel dat het woord *museumdirecteur* voorspelbaar is, als we de woorden *museum* en *directeur* kennen, terwijl het bestaan van de woorden *museum* en *directeur* zelf niet ergens uit afgeleid kan worden. De scheidsrechter in ons hoofd heeft een eindige lijst van verder niet op te splitsen woorden (*museum*, *scheep*, *directeur*) en een procedure als (iii) om te bepalen of iets een woord is van het Nederlands of niet. Sommige ongelede woorden hebben zelfs twee vormvarianten: *schip* en *scheep* bijvoorbeeld. We moeten aannemen dat beide in de lijst zijn opgenomen. De Van Dale is ten opzichte van zo'n scheidsrechter altijd in het nadeel; immers, de procedure werkt voor alle woorden in (i) en ook voor verdere uitbreidingen van die woorden omdat de procedure recursief is en in principe een oneindige verzameling woorden kan karakteriseren. En dat is precies wat we nodig hebben.

Als we *museum+directeur* op de aangegeven manier aaneenschakelen weten we als taalgebruikers heel wat meer dan alleen maar dat het een woord is. Bekijk daartoe eens de volgende 'omgewisselde' paren:

(iv)	soepvlees	vleessoep
	computerspel	spelcomputer
	ochtendgebed	gebedsochtend

Soepvlees is een soort vlees en *vleessoep* is een soort soep. Een *computerspel* is een soort spel, terwijl *spelcomputer* een soort computer is. Hetzelfde geldt voor de andere voorbeelden: steeds bepaalt het rechterdeel wat de betekenis van het samengestelde woord is. Bovendien kunnen we het geslacht van het samengestelde woord op dezelfde manier voorspellen: het komt altijd overeen met het geslacht van het rechterdeel van het samengestelde woord. Dus is het *het soepvlees* (want het is *het vlees*), maar *de vleessoep* (want *de soep*). We zeggen dat het rechterdeel het hoofd van de samenstelling is. Regel (iii) moet dan dus eigenlijk als volgt luiden (waarbij we voor het gemak de verwijzing naar het Nederlands maar even hebben weggelaten):

(iii)' Als x een woord is en y is een woord, dan is $x+y$ een woord waarvan y het hoofd is.

Dat geeft ons een antwoord op de vraag waarom we van woorden die we nooit eerder hebben gehoord meteen weten of het een *de*-woord of een *het*-woord is. Als je de woorden in (i) nog eens bekijkt, zie je dat (i-a) een *de*-woord, (i-b) een *het*-woord en (i-c) weer een *de*-woord is. En ook weet je ogenblikkelijk dat (i-a) verwijst naar een soort directeur, (i-b) naar een soort instituut en (i-c) naar een soort medewerkers. Die kennis heb je omdat regel (iii)' tot je kennis behoort.

Achterborsten en tilbaarheid

Jonge kinderen geven er blijk van al op vroege leeftijd regel (iii)' te kennen. Op de website 'Ouders Online' (<http://www.ouders.nl/>) is een verzameling woorden van kinderen bijeengebracht die ouders op de een of andere manier opvallend vonden. Op die lijst vind je woorden als *achterborsten* (voor schouderbladen), *appelbotten* (voor klokhuis), *beensnor* (haar op de benen) etc. Geen woorden die de kinderen van hun ouders of anderen hebben gehoord, maar zelf gebouwde uitbreidingen van de woordenschat volgens regel (iii)'.

Keren we terug naar *tilbaarheid*. Dat woord staat niet in Van Dale, maar is voor iedere moedertaalspreker een woord waarvan de betekenis, het geslacht (het is een 'de'-woord), en de uitspraak meteen duidelijk zijn. Maar *tilbaarheid* is niet afleidbaar uit regel (iii)'. Immers, *heid* noch *baar* zijn woorden van het Nederlands. Kennelijk hebben we behalve (iii) nog een andere procedure in ons hoofd om woorden van het Nederlands te maken. Het verschil tussen *tilbaarheid* en de woorden die volgens (iii) gevormd kunnen worden is dat *tilbaarheid* bestaat uit een stam met twee achtervoegsels, terwijl de woorden die met (iii)' gemaakt kunnen worden geheel uit woorden bestaan. Behalve dat we in het Nederlands woorden kunnen maken door ze achter elkaar te schakelen, samenstellingen genoemd, beschikken we daarnaast over achtervoegsels als *-baar* en *-heid* waarmee we nieuwe woorden kunnen maken.

Om te onderzoeken hoe die woordvorming in elkaar steekt, bekijken we eerst het volgende rijtje:

(iv)	zuinigheid	kortademigheid
	vrolijkheid	vrouwvriendelijkheid
	breekbaarheid	onberekenbaarheid

Dit zijn allemaal zelfstandig naamwoorden en als we *-heid* in gedachten weglaten, houden we allemaal bijvoeglijke naamwoorden (*zuinig*, *vrolijk*, etc.) over. We kunnen van een bijvoeglijk naamwoord een zelfstandig naamwoord maken door er *-heid* achter te plakken. Bovendien blijkt dat alle zelfstandige naamwoorden hetzelfde geslacht hebben – ze krijgen allemaal het lidwoord *de* (de *zuinigheid*, de *kortademigheid* etc.). Dat kan geen toeval zijn, maar is een gevolg van de woordvormingsregel. We noemen *het*-woorden voor het gemak neutraal en *de*-woorden niet-neutraal. We kunnen dan voor de woorden met *-heid* de volgende regel opstellen:

(v) Als x een bijvoeglijk naamwoord is, dan is $x+heid$ een niet-neutraal zelfstandig

naamwoord.

Voor woordvorming met het achtervoegsel *-baar* kunnen we op dezelfde manier een regel ontdekken. De woorden in (vi) zijn allemaal met *-baar* gevormd:

(vi)	aantoonbaar	oplosbaar
	draagbaar	strafbaar
	huwbaar	toonbaar

We zien dat de woorden waar *-baar* aan hecht (stammen van) werkwoorden zijn, terwijl de woorden op *-baar* zelf bijvoeglijke naamwoorden zijn. We kunnen dus de volgende regel formuleren:

(vii) Als *x* een werkwoord is, dan is *x+baar* een bijvoeglijk naamwoord.

Omdat regel (vii) bijvoeglijke naamwoorden oplevert, terwijl regel (v) juist bijvoeglijke naamwoorden als startpunt neemt, verwachten we dat we eerst regel (vii) kunnen toepassen, gevolgd door regel (v). Die verwachting komt uit blijkens de mogelijkheid van woorden als *aantoonbaarheid*, *strafbaarheid*, *tilbaarheid*, etc.

Misbrozig

Nu wordt ook duidelijk dat we, hoewel we de woorden uit de vertaling van ‘Jabberwocky’ niet in het woordenboek kunnen opzoeken, toch een flink aantal dingen over die woorden weten. *Bradig* is een bijvoeglijk naamwoord, net als *harig*, *kalig*, *wollig* en *misbrozig*. Die kennis komt niet uit het woordenboek maar kunnen we als moedertaalsprekers uit iets anders afleiden. Er is parallel aan regel (vii) nog een andere regel voor de vorming van bijvoeglijke naamwoorden:

(vii) Als *x* een bijvoeglijk of zelfstandig naamwoord is, is *x-ig* een bijvoeglijk naamwoord.

En ook de uitkomst van regel (vii) kan weer het startpunt zijn voor de regel (v): *harigheid*, *kaligheid*, *wolligheid* en als we het woordenboek vergeten: *misbrozigheid*.

De regels die we tot dusverre hebben besproken zijn verre van compleet: het Nederlands heeft veel meer van dergelijke zogenoemde morfologische regels, maar het is ons hier niet om volledigheid te doen; het gaat ons er om een beeld te geven van het soort kennis dat we als moedertaalsprekers van het Nederlands over onze woordenschat bezitten: onze morfologische kennis. En dat is aanzienlijk meer dan in een woordenboek past. De regels zijn in nog een ander opzicht incompleet: de betekenis ontbreekt. We hebben die met opzet weggelaten omdat de betekenis van woorden nu eenmaal niet zo eenvoudig met woorden is weer te geven. Maar het is duidelijk dat als we weten wat *haar* betekent, we ook weten wat *harig* betekent; en als we weten wat het werkwoord *aantonen* betekent, dan weten we ook wat *aantoonbaar* betekent. Kortom, de betekenis van veel woorden die met behulp van dergelijke morfologische regels worden gevormd, is afleidbaar uit de regel.

Ieder jaar zo rond de jaarwisseling worden er lijstjes gemaakt van de ‘nieuwe woorden van het jaar X’. Dat is altijd even een moeilijk moment voor de morfoloog, want elke dag worden er door iedereen ‘nieuwe’ woorden gemaakt die daarvóór nog niet gemaakt waren. Alleen al op de voorpagina van *NRC Handelblad* van 17 februari 2006 prijken de volgende woorden die allemaal niet in Van Dale staan: *cellencomplex*, *zorgpremie*, *achtervolgingsploeg*, *middagtemperatuur*, *boycotmaatregelen*, *chipmachinefabrikant*, *vogelgriepvirus*, etc. Woorden die ons als krantenlezers niet eens opvallen als uitzonderlijk of nieuw. We kunnen ze net zo gemakkelijk begrijpen als andere woorden. Eendagsvliegen die niet worden opgemerkt. Elk van die woorden op zichzelf genomen is niet bijster interessant. Maar het vermogen dat schuilgaat achter de productie van die woorden is dat juist in hoge mate.