

W. J. M. Levelt

Het semantisch weefgetouw

G. Kempen *Memory for word and sentence meanings*, Nijmegen 1970

Zowel vanuit linguïstische als psychologische hoek is er in Nederland een sterk toenemende activiteit op taalpsychologisch gebied. De toon van publicaties verraadt meestal onmiddellijk uit welk van deze tradities het onderzoek is voortgekomen. Zo is het, ook zonder het schutblad te lezen, direct duidelijk dat Kempen's dissertatie een psychologische is. Maar van een aantal recente psychologische bijdragen aan de psycholinguïstiek is die van Kempen toch het meest direct door linguïstische theorie geïnspireerd, en men kan daaraan toevoegen: het minst door psychologische.

De dissertatie, die het semantisch geheugen voor woorden en zinnen tot object heeft, bestaat uit vijf hoofdstukken. Het eerste geeft een overzicht over psychologische betekenis-theorieën. Het is een korte samenvatting van de neo-behavioristische en de door Katz geïnspireerde componentiële opvattingen over betekenis. Het tweede hoofdstuk geeft het geheugenmodel, dat 'set-feature' model heet. Het derde hoofdstuk verdedigt de linguïstische zinnigheid van het model. Het vierde relateert het aan andere (psychologische) geheugenmodellen en het vijfde beoogt een aantal psychologische implicaties van het model experimenteel te toetsen. We zullen ons hier hoofdzakelijk beperken tot een bespreking van de hoofdstukken twee, drie en vijf, d.w.z. het model en zijn linguïstische, resp. psychologische plausibiliteit. Over het vierde hoofdstuk volstaan we met de opmerking dat vooral hierop onze bovenstaande opmerking van toepassing is, dat het onderzoek niet sterk vanuit psychologische theorieën geïnspireerd is. De auteur gaat vrijwel geheel voorbij aan de zeer omvangrijke literatuur over de computer-simulatie van het semantisch geheugen, en aan de mathematische theorieën over het woord-geheugen.

Het model

Een model van het semantisch geheugen heeft drie essentiële aspecten: een informatie-verwerkend gedeelte dat aangeboden informatie vertaalt in

een code die acceptabel is voor het tweede gedeelte, het eigenlijke betekenisgeheugen, waarin de informatie wordt opgeslagen volgens bepaalde organisatie-principes, en tenslotte een 'output'-mechanisme dat informatie kan terugzoeken en in gedecodeerde vorm naar buiten afgeven.

Kempen's set-feature model is in essentie de beschrijving van de interne organisatie van het tweede gedeelte, het eigenlijke geheugen. Uit deze beschrijving volgt vanzelf de code waarin de informatie wordt gepresenteerd. Kempen (p. 35) stelt dat deze wijze van representeren linguïstisch acceptabel moet zijn.

Hoe stelt Kempen zich semantische representaties voor? Ik meen de formele eigenschappen van het model geen geweld aan te doen met het volgende beeld: het geheugen is een soort weefgetouw. Elk woord in de taal is gekoppeld aan één draad van de schering en aan één draad van de inslag. Wanneer we een gegeven inslagdraad volgen, passeren we achtereenvolgens alle schering-draden. Telkens wanneer de schering-draad iets zegt over de betekenis van de inslagdraad, worden schering en inslag op die plaats met een knoopje vastgebonden. Zo'n knoopje heet een 'memory location' (ML). De rij van knoopjes die langs een scheringdraad te vinden zijn bepaalt de betekenis van het inslagwoord; die betekenis is nl. de verzameling van de scheringwoorden die met het inslagwoord verknoopt zijn. Deze verzameling van woorden wordt opgevat als de karakteristieke verzameling van betekenis-kenmerken voor het betreffende woord: 'features' zijn woorden, en woorden kunnen features zijn. Op dit punt distancieert Kempen zich uitdrukkelijk van Katz.

Hoe ziet zo'n knoopje eruit? Het ene woord 'zegt iets' over het andere. Kempen noemt dit een 'minimal proposition' (MP). Als het inslagwoord 'paard' is en het schering-woord 'dier', dan specificceert het knoopje dat 'paard' een element is van de verzameling 'dier', en niet andersom. Het knoopje geeft zonnodig de richting aan van de inclusierelatie, stelt Kempen. Deze dominantierichting is het enige dat over de aard van de relatie tussen de twee woorden wordt vastgelegd. Tot zover de representatie van 'minimal propositions' zoals 'het paard is een dier', 'Pascal is een uitvinder', enz. Om complexe zinnen voor te stellen is natuurlijk meer nodig. Bijvoorbeeld voor 'Pascal is de uitvinder van de rekenmachine': Kempen doet dit door een soort tweede-orde geheugen-plaatsen te creëren. In dit geval is het een ML die in verbinding staat met enerzijds het knoopje dat 'Pascal' en 'uitvinder' verbindt en anderzijds het knoopje dat 'rekenmachine' en 'uitvinding' verbindt.

Elke ML kan op deze wijze 'naar boven' worden verbonden met één en slechts één andere ML. De tweede-orde ML kent niet de dominantie-specificatie van de elementaire ML. Het ligt voor de hand nu ook hogere-orde ML's toe te laten, die op hun beurt elk in verbinding staan met twee (en slechts twee) lagere-orde ML's. Zo bestaan er hiërarchieën van geheugenplaatsen. Met betrekking tot het model zelf heb ik vooral moeite met de volgende twee punten.

1. De formalisering. Wil een beschrijving van complexe informatienetwerken niet uit de hand lopen, dan moet men een geschikt systeem van beschrijving kiezen. Zulke systemen bestaan. Men kan ze ontleen aan de 'graph-theory', of men kiest de vorm van een computerprogramma. Vooral de laatste mogelijkheid is aantrekkelijk. De verzameling van kenmerken voor een woord wordt dan een lijst van andere woorden, die op hun beurt weer de adressen vormen voor andere lijsten. Dit is equivalent met het weefgetouw, maar het is op deze manier veel gemakkelijker, om opberg- en terugvind-operaties te definiëren en ze te toetsen op hun 'real-time'-kenmerken. Eén voorbeeld ter adstructie. Kempen stelt dat een ML de mogelijkheid heeft om naar slechts één andere ML te verwijzen. Stel dat de hoorder de volgende informatie heeft opgeborgen: 'Pascal is de uitvinder van de rekenmachine'. Hem wordt nu tevens meegedeeld dat Pascal de uitvinder is van de luchtpomp. Men zou zeggen dat dankzij het voorhanden zijn van de eerste informatie, de tweede gemakkelijk opneembaar moet zijn. Dat is niet het geval. Het knoepje waarin de relatie Pascal/uitvinder ligt vastgelegd is reeds verbonden met een tweede-orde geheugenplaats en kan niet meer worden verbonden met een andere tweede-orde ML. Toch is dat nodig om nu de verbinding Pascal/uitvinder—uitvinding/luchtpomp tot stand te brengen. De dissertatie is niet expliciet op dit punt. Het lijkt me in de geest van het set-feature model te zijn om nu aan te nemen dat er niet één eerste-orde geheugenplaats is voor Pascal/uitvinder, maar een *verzameling* van geheugenplaatsen. Ik leid dit af uit het feit dat Kempen de geheugenlocatie voor 'Pascal is de uitvinder van de rekenmachine' voorstelt als de intersectie van twee *verzamelingen* (p. 32); de ene verzameling komt voort uit de intersectie van 'Pascal' en 'uitvinder', de andere verzameling uit die van 'uitvinding' en 'rekenmachine'. Als er voor Pascal/uitvinder niet slechts één, maar een hele verzameling van ML's bestaat, is er geen reden waarom dat niet ook voor alle andere eerste-orde ML's het geval zou zijn. Er is dus niet één eerste-orde weefsel, maar een hele stapel van isomorf geknoopte weefsels.

Als we—niet zonder reden—aannemen dat dit wordt bedoeld, dan staan we voor een interessant dilemma. We kunnen enerzijds veronderstellen dat nieuwe schering- en inslagdraden pas worden gespannen als dat nodig is, dus bijvoorbeeld in het onderhavige geval waar de knoop Pascal/uitvinder reeds 'bezet' is. Dat zou betekenen dat het opnemen van nieuwe informatie in zo'n geval extra tijdrovend moet zijn. 'Pascal is uitvinder van de luchtpomp' is moeilijk te leren als we al weten dat Pascal uitvinder is van de rekenmachine. Dit kan worden getoetst, maar het lijkt me een onaannemelijke predictie.¹ We kunnen anderzijds veronderstellen dat eerste-orde ML's direct als verzamelingen ontstaan; dat we meteen op een hele batterij weefgetouwen werken, zogezegd. Er is dan altijd wel een onbezette knoop te vinden die met een tweede-orde ML verbonden kan worden. De realisering hiervan kost ons echter een waterhoofd. Want wat we eisen voor eerste-orde ML's (nl. dat ze verzamelingen zijn), moeten we om analoge reden ook voor 2e-orde en hogere orde ML's eisen. Mét de orde neemt het aantal noodzakelijke verzamelingen exponentieel toe. Zonder in gedetailleerde berekeningen te vervallen is het duidelijk dat er dan een geheugen van onvoorstelbare omvang nodig is.

Wanneer het model even was uitgeprobeerd op de computer zou zo'n probleem direct zijn opgevallen. En opgelost: door multipele adressering naar tweede-orde lijsten toe te staan.

Het gebrek aan formalisering bemoeilijkt over het algemeen de evaluatie van linguïstische en psychologische implicaties zoals we nog een paar maal zullen zien.

2. Typen van relaties. Kempen onderstelt voor de eerste-orde ML's een inclusie-relatie (met richting) en voor hogere orde ML's een verder ongespecificeerde relatie. Semantische beschrijving vergt echter zonder twijfel een grotere variëteit van relaties. Hoe wil men het begrip 'watersnood' alleen met inclusie-relaties en zonder causale relaties beschrijven, het begrip 'geweer' zonder finale, het begrip 'doos' zonder spatiële, het begrip 'slecht' zonder oppositionele, het begrip 'emeritus' zonder temporele, etc.?

Er moet, kortom, meer variëteit in de knopen van het weefsel komen. Een uitbreiding van het model in die richting lijkt alleszins mogelijk.

Linguïstische implicaties

In hoofdstuk III stelt Kempen dat het model houdbaar moet zijn uit linguïstisch oogpunt. Eerlijk gezegd betwijfel ik of deze eis gesteld moet worden, zolang linguïsten niet duidelijker zijn over wat ze met 'competence'

bedoelen. Wanneer een grammatica slechts van marginaal belang zou blijken te zijn voor de beschrijving van actueel taalgedrag, dan hoeft men zich niet te veel om linguïstische plausibiliteit te bekommeren. Zonder hierop verder in te gaan,² neem ik Kempen's doelstelling echter even over om te zien wat ervan terecht komt. Het dient wél opgemerkt dat de recensent hiervan als psycholoog vermoedelijk geen evenwichtig beeld geeft—met excuses aan schrijver en lezer. Hoofdstuk III geeft set-feature beschrijvingen van een aantal linguïstische verschijnselen. Eerst het soortelijke versus concrete gebruik van naamwoorden ('een paard is een dier' vs. 'ik zag een paard'). Kempen postuleert twee verzamelingen voor zo'n naamwoord, één voor het soortelijke en één voor het concrete. Waar schering- en inslagdraad elkaar kruisen legt hij een knoop die de inclusie aangeeft van de concrete vorm in de soortelijke. Als ik linguïst was zou ik een syntactische oplossing boven een semantische prefereren. Het probleem is nl. identiek voor alle naamwoorden van dit type—het lijkt me erg oneconomisch onder die omstandigheden de lezing voor elk zelfstandig naamwoord te verdubbelen.

De behandeling van werkwoorden in termen van minimale proposities gaat—kortweg—als volgt: Intransitieve werkwoorden vragen één MP: 'John laughed → MP: John - laughing' (p. 37), transitieve drie: 'John kills Bill: MP₁: John - killing, MP₂: Bill - killed, MP₃: MP₁-MP₂' (p. 37). Net als zelfstandige naamwoorden hebben transitieve werkwoorden blijkbaar eveneens een dubbele semantische representatie nodig—ook hier zou ik de gebruikelijke syntactische oplossing hebben geprefereerd. Complexere zinnen resulteren in omvangrijke MP-hiërarchieën. Men zou bij het zien van al die voorbeelden graag weten of die representaties equivalent zijn met P-markers. De schrijver merkt op dat er geen links-rechts ordening is—dat is echter voor phrase-structure representatie evenmin een noodzakelijke eis. Ook komt het me voor dat de representaties niet steeds tussen subject en object kunnen onderscheiden. Voor 'The naughty children were punished' is een representatie MP₁: naughty-children, MP₂: MP₁-punished. Hoe wordt dan "The naughty children punished" gerepresenteerd, of om een wat gelukkiger voorbeeld te kiezen: hoe verschilt de representatie voor 'de deugnieten stelen' van die voor 'de deugnieten worden gestolen'?

De linguïstische paragrafen uit dit boek kunnen ieder op zichzelf worden bekeken. Wat mijn voornaamste bezwaar is tegen dit hoofdstuk, is dat de auteur geen *generatieve*, dat is expliciete beschrijving geeft van de MP-structuren. Er is geen expliciete test of een MP of MP-hiërarchie welgevormd is of

niet. Daarvoor zou een stelsel regels nodig zijn, welk geheel ontbreekt. Slechts op basis van zo'n expliciet stelsel van organisatie-principes zou vergelijking met bestaande grammatica's goed mogelijk zijn.

Psychologische implicaties

Het set-feature model is een performance-model ter beschrijving van taalgedrag, waarin het semantisch geheugen een rol speelt. Hoofdstuk v geeft een aantal experimenten die hypothesen toetsen welke uit het model worden afgeleid. Worden ze uit het model afgeleid? Eerder merkten we op dat een volledig model een input-, een geheugen-, en een output-component moet hebben. Kempen beperkt zich—en dat is zijn goed recht—tot de beschrijving van het geheugen-gedeelte. Bij experimentele toetsing zal men echter noodzakelijk met input (codering) en output (decodering) van informatie te maken krijgen en is het vaak niet eenvoudig te beslissen welke effecten aan welke componenten toegeschreven moeten worden. Kempen's hypothesen zijn soms geen eenduidige afleidingen uit het geheugenmodel, zoals we zullen zien.

Daar staat tegenover dat Kempen's hypothesen allerm minst triviaal zijn. Het zijn telkens bijzonder sterke voorspellingen, en Kempen toont zich steeds een zeer inventief experimentator, wanneer hij zijn hypothesen operationaliseert. Het eerste experiment toont aan dat een associatiewoord relatief snel wordt gegeven wanneer het associatiewoord via een eerste-orde MP met het stimulus-woord is verbonden. Tragere reacties treden op wanneer een hiërarchie van MP's tussen de woorden medieert. Door een interessante procedure weet Kempen aan te tonen dat dit ook geldt wanneer de woordparen gelijke associatieve frekwentie hebben. De mediërende MP-structuur wordt op intuïtieve³ wijze afgeleid uit zinnnetjes die de proefpersonen na hun associatietaak maakten voor elk woordpaar (bijv.: stimulus-woord: 'lopen', reactie proefpersoon: 'reizen', zinnnetje proefpersoon: 'lopen is een vorm van reizen'). Ofschoon dit resultaat met het model overeenkomt, zou het vermoedelijk ook door Quillian's model⁵ worden voorspeld. Men kan zelfs helemaal zonder informatie-hiërarchieën: Wanneer men gewoon het aantal woorden telt dat stimulus- en reactiewoord van elkaar scheiden in de zin (i.e. 4 voor *lopen-reizen*), dan blijkt de reactietijd korter te zijn naarmate de woorden dichter bijeen liggen in de zin. Hetzelfde geldt trouwens als zinlengte als parameter wordt genomen.

Het tweede experiment toont aan dat bij het reproduceren van een onthouden woordenlijst woorden die een korte afstand hebben in de MP-hiër-

archie de neiging hebben om in groepjes te worden gereproduceerd.⁴ Het derde experiment toont aan dat zinnen die hetzelfde object hebben ('Zij onderstrepen de twee', 'Zij omcirkelen de twee') in een soortgelijke geheugentaak vaker in directe opeenvolging worden gereproduceerd, dan zinnen die dezelfde adverbiale bepaling hebben ('Zij onderstrepen het tweemaal', 'Zij omcirkelen het tweemaal'). Ook hier weer vanwege het aantal MP-stappen tussen werkwoord en object resp. bepaling. De afwezigheid van een redelijk complete grammatica waaruit MP-hiërarchieën expliciet, niet intuïtief, kunnen worden afgeleid, maakt het onmogelijk om te stellen dat de voorgestelde hiërarchieën voor de experimentele zinnen uit het model volgen. Gegeven het model zijn ze intuïtief plausibel, meer niet.

Het vierde experiment gebruikt op interessante nieuwe wijze de procedure van Savin en Perchonok voor het meten van de geheugenbelasting van zinnen. De resultaten van dit experiment zijn helaas wat teleurstellend. Het enige wat blijkt is dat zinnen als 'De soldaat grijpt het geweer' minder geheugenruimte beslaan wanneer de pp. deze zin al eerder gezien heeft, dan wanneer hij de zin 'hij grijpt het' eerder gezien heeft (vertrouwdheid met individuele woorden van de zin werd constant gehouden). Dit resultaat zou door veel theorieën voorspeld worden—de relatie met het set-feature model is weinig specifiek.

In het vijfde en laatste experiment kreeg de pp. van zinnen die hij geleerd had steeds één woord aangeboden. Hij moest dan zo snel mogelijk het volgende woord zeggen. De reactietijden werden opgemeten. Twee typen zin werden in dit experiment gebruikt. Voorbeeld: 'Deze mensen stelen $\left\{ \begin{array}{l} \text{platen} \\ \text{altijd} \end{array} \right\}$ tijdens feesten'. Kempen voorspelde dat zij die de adverbiale zin (met 'altijd') hadden geleerd op het subject-naamwoord (mensen) sneller met het werkwoord (stelen) zouden reageren dan zij die de object-zin hadden geleerd (met 'platen'). Volgt deze—overigens niet vanzelfsprekende—voorspelling uit het model? Nee, Kempen maakt hiertoe een ad hoc hypothese over het output-mechanisme, nl. dat een woord pas kan worden afgegeven wanneer ook het andere woord waarmee het in een MP is verbonden is geactiveerd. 'Stelen' wordt moeilijk gereproduceerd omdat het op 'platen' moet wachten waarmee het in één MP is verknoopt. Dat geldt niet voor 'stelen' en 'altijd', volgens Kempen. Deze ad hoc hypothese wordt getest en sterk bevestigd.

Samenvattend geeft Kempen's dissertatie een aantal interessante psycho-

linguïstische experimenten, die een niet steeds overtuigende verbinding hebben met een model dat verdere formalisering en uitbouw behoeft. Het boek wekt goede verwachtingen met betrekking tot toekomstig experimenteel werk van deze auteur, dat naar wij hopen tot verdere verrijking van het model zal leiden.

1 Kempen zelf maakt in hoofdstuk v § 4 de omgekeerde predictie: Wanneer iemand 'The man killed him' heeft geleerd dan heeft hij minder moeite met het onthouden van 'The man killed the girl', dan wanneer hij 'He killed him' heeft geleerd, terwijl toch in het eerste geval de ML man/killed bezet is door een schakeling aan 'him', en in het tweede geval man/killed nog vrij is.

2 Zie over dit onderwerp Levelt en Schils: Psychologie en Linqüistiek in *Psychologie temidden van de wetenschappen*. Papers voor het congres ter opening van het Psychologisch Laboratorium, Nijmegen 1970.

3 Zie opmerking bovenaan pagina 65 van de dissertatie.

4 Ik heb wat moeite met de categorisering van *maan/landing* als een 'long distance'.

5 Quillian, M. R., Word concepts: A theory and simulation of some basic semantic capabilities: *Behavioral Science*, 1967, 12, 410-430.