

GENERATIEVE GRAMMATICA EN PSYCHOLINGUISTIEK I  
INLEIDING IN DE GENERATIEVE GRAMMATICA

DOOR

DR. W. J. M. LEVELT

De term „psycholinguïstiek” wordt pas sinds rond 1950 regelmatig gebruikt om een bepaald wetenschappelijk probleemgebied aan te duiden. De eerste poging om onder dit hoofd tot een expliciete omschrijving te komen van de problemen die psychologen en linguïsten gemeen hebben is vervat in een overzicht door OSGOOD en SEBEOK, dat in 1954 verscheen. Sindsdien is er een omvangrijke literatuur ontstaan, waarin het woord „psycholinguïstiek” min of meer ter aanduiding van een programma wordt gebezigd. Een verhelderend overzicht over dit alles is samengesteld door DIEBOLD (1965).

Niet lang na haar geboorte echter, werd de psycholinguïstiek opgeschrikt door een stormachtige ontwikkeling in een van de ouders, de linguïstiek. In 1957 publiceerde NOAM CHOMSKY zijn *Syntactic Structures*, dat het ontstaan markeert van wat sindsdien de *generatieve of transformationele grammatica* heet. Chomsky's theorie groeide uit tot een van de belangrijkste bewegingen in de Amerikaanse taalwetenschap, en liet niet na een grote invloed uit te oefenen op het psycholinguïstische bedrijf. Dit laatste is vooral aan George Miller te danken die als eerste de psychologische betekenis van deze nieuwe stroming in de linguïstiek inzag. (zie MILLER & CHOMSKY, 1963).

Dit tweetal artikelen beoogt een overzicht te geven over de transformationele grammatica en de psychologische research die onder invloed ervan tot stand is gekomen. Dit eerste artikel is een *summiere* uiteenzetting van de beginselen van de generatieve linguïstiek. Vertrouwdheid met het begrippenapparaat van deze wetenschap is namelijk vereist om de psycholinguïstische research, zoals in het tweede artikel behandeld, op haar waarde te schatten. Alternatieve inleidingen kan men vinden in de oorspronkelijke literatuur (bijv. CHOMSKY, 1965; BACH, 1964; LEES,

1960; KATZ, 1966), of ook in het Nederlands in het proefschrift van KRAAK (1966) dat na de voltooiing van dit artikel verscheen.

## 1. INLEIDING

Deze introductie in de theorie van de generatieve grammatica is aan enkele essentiële beperkingen onderhevig. In de eerste plaats kan niet worden ingegaan op de relatie met andere linguïstische theorieën. Met name wordt voorbijgegaan aan de vele reacties die deze theorie binnen de algemene taalwetenschap heeft opgeroepen (voor reacties van Nederlanders zie bijv. REICHLING (1961) en UHLENBECK (1963); zie ook CHOMSKY's antwoord hier op (ter perse)). Dit overzicht zal beperkt blijven tot een summiere, doch zo objectief mogelijke beschrijving van de hoofdzaken van de generatieve grammatica. Een tweede beperking ligt in het feit dat de literatuur op dit gebied niet alleen reeds omvangrijk is, maar ook uiterst technisch van aard. De kracht van de generatieve grammatica is juist, dat elke grammaticale relatie door een verzameling van geformaliseerde regels volledig wordt beschreven. Elke „losse” argumentatie faalt dientengevolge in dit opzicht. Om dit te voorkomen zullen de essentialia van deze leer worden beschreven zonder technisch te worden verantwoord; hoogstens zal hier en daar een verduidelijkend voorbeeld worden gegeven. Ook in dit opzicht zal de literatuurlijst uitkomst moeten bieden voor geïnteresseerden (zie bijv. CHOMSKY, 1963).

## 2. TAALBEHEERSING EN TAALGEDRAG

De theorie van een natuurlijke taal dient een beschrijving te geven van de *competentie* van een vloeiende spreker. Met competentie is bedoeld datgene dat de spreker in staat stelt met andere sprekers van dezelfde taal te communiceren. Laten we deze competentie met het Nederlandse woord *taalbeheersing* omschrijven (zoals een doofstom kind toch beheersing van de Nederlandse taal kan hebben). Een spreker heeft over het algemeen weinig inzicht in deze taalbeheersing. De preciese aard ervan kan dan ook slechts worden afgeleid uit zijn *taalgedrag*. Hoewel de taalbeheersing een noodzakelijke voorwaarde is voor adequaat taalgedrag, kunnen deze twee niet worden gelijkgesteld; en dit niet alleen vanwege pathologische vormen als doofstomheid. Bijvoorbeeld, voor het verstaan van gesproken taal is niet alleen beheersing van de betreffende taal nodig, doch ook een bepaalde geheugencapaciteit om de binnenkomende informatie zolang in op te slaan gedurende de verwerking.

Ook zijn bepaalde verwerkingsstrategieën nodig die de zinnen in kleinere eenheden opsplitsen, enz. De taalbeheersing is dus slechts een van de voorwaarden voor aangepast taalgedrag.

De meest karakteristieke eigenschap van de taalbeheersing is dat het de spreker/luisteraar in staat stelt onbepaald veel zinnen te produceren en te verstaan. Het kan worden aangetoond dat vrijwel elke gesproken zin nieuw is, dwz. als combinatie van woorden nooit eerder bestaan heeft. Een grammatica-theorie dient dit creatieve aspect van de taalbeheersing adequaat te beschrijven (CHOMSKY, 1964).

### 3. DE GENERATIEVE REGELS

Meer specifiek dient de grammatica-theorie aan te geven op welke wijze de taalbeheersing een onbeperkte hoeveelheid zinnen toestaat. Een grammatica van een taal geeft dan in de eerste plaats de regels, nodig om alle welgevormde zinnen van die taal te produceren. Bovendien dienen alle niet-welgevormde zinnen door dezelfde regels te worden uitgesloten. De verzameling van welgevormde zinnen van een taal is theoretisch oneindig: voor elke gegeven welgevormde zin  $Z$ , kan men een langere zin  $Z'$  produceren die eveneens welgevormd is.

De grammatica kan nu twee vormen aannemen: zij kan eveneens *oneindig* zijn; dat is bijvoorbeeld het geval wanneer de grammatica een opsomming is van alle zinnen die welgevormd zijn. Dit is om voor de hand liggende redenen een wetenschappelijk onwenselijke situatie. Blijft over dat zij een *eindige* verzameling regels is.

De enige manier om met een eindige verzameling regels een oneindig aantal zinnen te genereren, is door althans een gedeelte van deze regels *recursief* te doen zijn. Dat wil zeggen, een zelfde groep regels is willekeurig vaak toepasbaar in het proces van genereren van een zin. Er zijn nu allerlei systemen van recursieve regels denkbaar die mogelijk in staat zouden zijn de zinnen van een taal voort te brengen.

#### 3.1. *Het finite state model*

Een voor psychologen welbekend systeem, dat ook in de linguïstiek een zekere opgang heeft gemaakt is het zgn. *finite state model*. De zin wordt met dit model beschreven als het resultaat van een Markov-proces. Voor elk woord (of andere eenheid, dat doet hier niet ter zake) van een taal is door de regels gespecificeerd door welke woorden dit woord kan worden gevolgd (eventueel met een specificatie van de

waarschijnlijkheden van deze overgangen). Wanneer eenzelfde woord in dit proces weer kan terugkeren is het proces recursief en kan oneindig veel verschillende zinnen opleveren. Op deze basis is bijvoorbeeld HOCKETT's linguïstische theorie geschoeid (1955). Psychologen kennen dit model over het algemeen in de vorm van „zinnen”, die een zoveelste orde benadering van Nederlands zijn (voor voorbeelden, zie VAN DE GEER (1957, p. 310 e.v.). CHOMSKY (1956) kon bewijzen dat dit model principieel niet in staat is een natuurlijke taal als Engels te beschrijven.

### 3.2. *Het phrase structure model*

Een ander mogelijk model is het *phrase structure model*; we zullen het hier *constituenten grammatica* noemen. Dit model is veel algemener in het voortbrengen van zinnen dan het Markov-proces. Het kan namelijk onder andere alle sequenties genereren, die het Markov-proces voortbrengt, maar bovendien nog meer. In zijn eenvoudige vorm (de contextvrije, hierover straks meer) is dit model echter evenmin voldoende om alle zinnen van een natuurlijke taal voort te brengen. In een iets gecompliceerdere vorm (de contextgevoelige vorm) echter, vormt het een belangrijk onderdeel van de transformationele grammatica. Dit rechtvaardigt een nadere beschrijving van het type regels en een voorbeeld.

De regels van de constituenten-grammatica zijn zgn. *herschrijfregels*. Zij zijn van de vorm:

$$(1) \quad XAY \rightarrow XBY.$$

Wanneer de regels syntactische regels zijn kan A een grammaticale categorie aanduiden als „naamwoordsgroep” (NG), „werkwoordsgroep” (WG), „voorzetselgroep” (VZG), „naamwoordelijke bepaling” (NB), „naamwoord” (N), „werkwoord” (W), „voorzetsel” (VZ), „lidwoord” (L), enz. B, X en Y zijn rijen symbolen die zowel van het type *grammaticale categorie* als van een ander type kunnen zijn, dat *formatief* heet. Het verschil tussen beide is dat grammaticale categorieën zowel links als rechts van de pijl kunnen voorkomen (op de A-, resp. de B-plaats) en formatieven alleen rechts (op de B-plaats). De regel (1) zegt nu dat de grammaticale categorie A kan worden herschreven tot de symbolenrij B wanneer A zich tussen X en Y bevindt. De regel wordt ook wel weergegeven als:

$$(1a) \quad A \rightarrow B/X - Y.$$

X en Y vormen samen de context waarbinnen A kan worden herschreven tot B. X en Y mogen ook ontbreken. Doen ze dat beiden, dan heet de regel *contextvrij*. De regel luidt dan  $A \rightarrow B$ , en betekent dat A in *elke* context als B herschreven kan worden. In alle andere gevallen heet de regel *contextgevoelig*. In de syntactische regels mag B niet gelijk zijn aan het nul-symbool, dwz. er mogen geen constituenten spoorloos verdwijnen (in de fonologische regels zijn hierop echter uitzonderingen). Als voorbeeld zullen we nu met contextvrije regels de volgende zinnen genereren:

(2) *de psycholoog ontmoet de linguist*

(3) *oefening baart kunst*

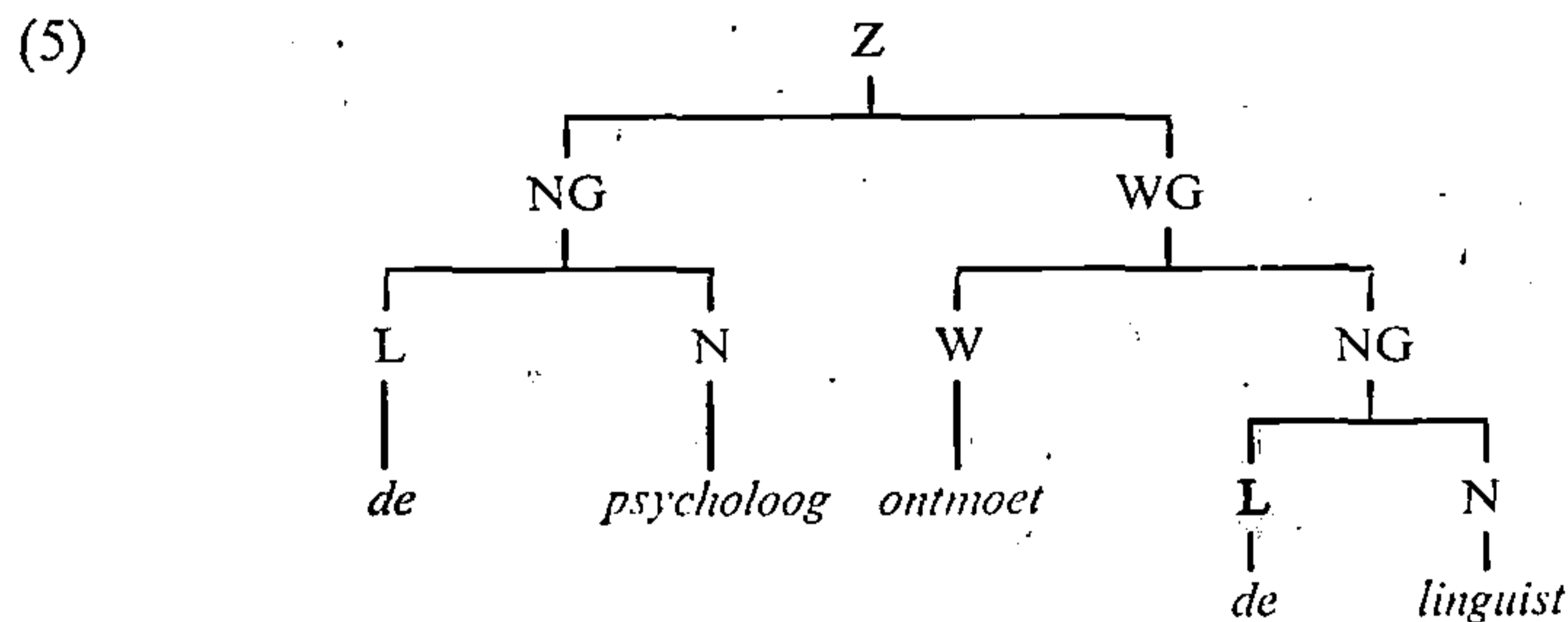
De grammatica hiervoor is de volgende:

- (4) (i)  $Z \rightarrow NG + WG$  (v)  $L \rightarrow de$   
 (ii)  $NG \rightarrow L + N$  (vi)  $N \rightarrow linguist, psycholoog,$   
 (iii)  $NG \rightarrow N$  *oefening, kunst*  
 (iv)  $WG \rightarrow W + NG$  (vii)  $W \rightarrow ontmoet, baart$

De generatie van de eerste zin (2) kan als volgt verlopen:

$$Z \rightarrow NG + WG \rightarrow L + N + WG \rightarrow L + N + W + NG \rightarrow L + N + W + L + N,$$

waarna de regels (v), (vi) en (vii) (de *lexicale regels*) nu het gewenste resultaat kunnen geven: *de psycholoog ontmoet de linguist*. Dit proces kan ook in diagramvorm worden weergegeven:



Een boomdiagram als dit heet een *Phrase-marker*, ook wel kortweg *P-marker* (in het Nederlands zou men van een *zinsdeler* kunnen spreken).

De P-marker kan in plaats van als boom ook door middel van benoemde haakjes worden voorgesteld:

$$(((de)_L(psycholoog)_N)_{NG}((ontmoet)_W((de)_L(linguist)_N)_{NG})_{WG})_Z.$$

Deze representaties zijn volkomen isomorf.

Zin (3) kan met dezelfde grammatica (4) als volgt worden gegenereerd:

$$Z \rightarrow NG + WG \rightarrow NG + W + NG \rightarrow N + W + NG \rightarrow N + W + N,$$

hetgeen na toepassing van de lexicale regels het gewenste

$$(((oefening)_N)_{NG}((baart)_W((kunst)_N)_{NG})_{WG})_Z$$

kan opleveren.

De generatieve capaciteit van de grammatica (4) blijkt nu uit het feit dat niet alleen de zinnen (2) en (3) door deze grammatica worden opgeleverd, maar ook welgevormde zinnen als *de linguist baart de psycholoog* en nog vele anderen. Toch is het aantal zinnen dat door deze grammatica kan worden voortgebracht eindig, nl. 144, omdat er nog geen recursieve regels in zijn opgenomen. Verderop zal blijken dat een belangrijk soort recursieve phrase structure-regel in de transformationele grammatica is opgenomen. Deze recursieve regels zijn gekenmerkt door het voorkomen van het symbool Z („zin”) *rechts* van de pijl. Dat betekent dat in de loop van een afleiding het aanvangssymbool Z opnieuw kan verschijnen. Deze Z wordt dan weer herschreven door de cyclus van regels opnieuw te doorlopen, waarbij Z weer opnieuw kan opduiken, enz. Zo'n recursief proces ligt ten grondslag aan zinnen als:

(6) *de man die de ratten ving ging naar Hameln,*

waarin de zin *de man ving de ratten* op bepaalde wijze is ingebed in de zin *de man ging naar Hameln*. Ook meervoudige inbedding komt veel voor. De volgende zin is een voorbeeld van viervoudige inbedding, dat de schrijver ervan (MORSBACH, 1896, p. 49) spontaan uit de pen schijnt te zijn gevloeid <sup>1)</sup>:

(7) *Später jedoch, als man einen glatteren versbau erzielte und die fehlende senkung am verschluss als zu hart und das metrum zerstörend vermieden wurde, gab man . . . die so bequemen reimwörter, unter denen die mit schweren ableitungssilben gebildeten wegen ihres häufigen vorkommens in der sprache die grösste rolle spielten keineswegs preis.*

<sup>1)</sup> Dit voorbeeld dank ik aan Prof. M. Halle.

Een constituentengrammatica alleen bleek onvoldoende om bekende natuurlijke talen adekwaat te beschrijven (CHOMSKY, 1961; POSTAL, 1964). In dit verband is het van belang de betekenis van het begrip *structuurindex* uiteen te zetten. Een eerste doel van een grammatica is, zoals opgemerkt, het volledig voortbrengen van alle zinnen van de taal. Er is echter een tweede doelstelling. De grammatica moet ook het uitgangspunt kunnen worden voor allerlei linguïstische generalisaties. Met name de generatie-geschiedenis van een zin, welke zijn structuurindex wordt genoemd, moet aanknopingspunten bieden voor de juiste interpretatie van de zin. Bijvoorbeeld, het is het natuurlijke gevoel met betrekking tot een zin als

(8) *de man was oud, vermoeid, lang, . . . , maar vriendelijk*

dat de adjectieven *oud, vermoeid, lang, enz.* niet gegroepeerd zijn in subgroepen. Elk tweetal adjectieven kan worden verwisseld zonder de betekenis van de zin merkbaar te beïnvloeden. Deze realiteit dient in de structuurindex op een of andere wijze terug te vinden te zijn. Om nu met een constituenten-grammatica een willekeurige rij adjectieven als in (8) te genereren moet een recursieve herschrijffregel worden geïntroduceerd, bijvoorbeeld van de vorm

(9) 
$$\text{Adj} \rightarrow \text{Adj} + \text{Adj}.$$

Door deze herhaald toe te passen kan nu weliswaar de zin (8) worden gegenereerd, doch de structuurindex faalt nu met betrekking tot de realiteit van ongegroepeerdheid van deze adjectieven. Regel (9) resulteert nl. essentieel in groepering, zoals bijvoorbeeld de volgende afleiding laat zien:

$$\text{Adj} \rightarrow (\text{Adj} + \text{Adj}) \rightarrow ((\text{Adj} + \text{Adj}) + \text{Adj}) \rightarrow (((\text{Adj} + \text{Adj}) + \text{Adj}) + \text{Adj}) \rightarrow (((\text{Adj} + \text{Adj}) + \text{Adj}) + (\text{Adj} + \text{Adj}))$$

of soortgelijk. De adjectieven zijn hier hiërarchisch gegroepeerd in deelgroepen die geen enkele linguïstische realiteit hebben. De constituenten-grammatica levert hier teveel structuur, kan men zeggen. In dit geval kan een constituenten-grammatica een oplossing zonder zo'n zinloze interne structuur slechts leveren met een oneindige verzameling regels:

$$\text{Adj} \rightarrow \text{Adj} + \text{Adj}$$

$$\text{Adj} \rightarrow \text{Adj} + \text{Adj} + \text{Adj}$$

$$\text{Adj} \rightarrow \text{Adj} + \text{Adj} + \text{Adj} + \text{Adj} \text{ enz.}$$

### 3.3. *Het transformationele model*

De vraag naar een adekwate structuurindex, alsmede andere overwegingen (zie CHOMSKY, 1957) leidden tot de introductie van een type regels dat wezenlijk van de herschrijfregel (1) verschilt. Deze nieuwe operaties worden *grammaticale transformaties* genoemd. De grammaticale transformatie geeft de mogelijkheid om in een serie symbolen termen te doen vervallen, toevoegen of verwisselen. Evenals de herschrijfregel (1) in de constituenten-grammatica kan de grammaticale transformatie in zijn algemene gedaante formeel worden weergegeven. Exacte weergave ervan vereist echter behandeling van allerlei technische details. Om lengte en verstaanbaarheid van dit overzicht niet teveel geweld aan te doen, volstaan we met een verwijzing naar CHOMSKY & MILLER (1963, pp. 301–302) en CHOMSKY (1961, pp. 131–132) voor een exacte formulering en beperken we ons hier tot het geven van enkele voorbeelden om de transformatieoperatie te verduidelijken. Beschouw de zin:

(10) *Jan schopt de bal*

Er is een aantal zinnen dat hieraan nauw verwant is, zoals: *schopt Jan de bal?*, *de bal wordt door Jan geschopt*, *wie schopt de bal?*, *Jan schopt de bal niet*, *wordt de bal door Jan geschopt?*, enz. Zoals gezegd is het aantrekkelijk, wanneer de structuurindices van deze verschillende zinnen, dat is hun generatiegeschiedenis, deze kennelijke verwantschap doet uitkomen. Dit is moeilijk binnen het kader van de constituenten-grammatica, doch kan eenvoudig worden verwezenlijkt met transformationele operaties. Bijvoorbeeld, generatie van *Jan schopt de bal* en van *de bal wordt door Jan geschopt* zijn vrijwel identiek tot op een punt, waar een passieve transformatie  $T_{pas}$  verschijnt in de generatie van de tweede zin. De transformatie verwisselt symbolen en voegt nieuwe in. Op essentiële punten zijn de structuurindices gelijk voor de twee zinnen. Soortgelijk voor *schopt Jan de bal?* Hier heeft een vraagtransformatie  $T_{vr}$  plaatsgevonden. Deze transformaties eenmaal gegeven, kan nu ook de zin *wordt de bal door Jan geschopt?* worden gegenereerd, door zowel  $T_{pas}$  als  $T_{vr}$  te laten werken.

Verwisselingen, toevoegingen en weglatingen van symbolen is aan strikte beperkingen onderhevig. Met name gebeurt dat nooit onafhankelijk van de P-marker zoals die op het moment van de transformatie gegeven is. In feite veranderen transformaties P-markers in nieuwe P-markers, waarbij „takken” verwisseld, toegevoegd of weggelaten



worden. Verderop zal de preciese plaats van de transformatie in de grammatica nader worden omschreven.

#### 4. DE DRIE COMPONENTEN VAN DE GRAMMATICA

Phrase structure regels en transformaties vormen de operaties binnen een transformationele grammatica. Zij zijn werkzaam in drie wel onderscheidbare domeinen: de *syntactische*, de *fonologische* en de *semantische component* van de grammatica. Deze zullen nu achtereenvolgens kort worden beschreven.

##### 4.1. *De syntactische component*

Dit is de generatieve kern van de grammatica. Alle bovenstaande regels behoorden tot de syntactische component. Zo genereerden de regels (4) de zin *de psycholoog ontmoet de linguïst*. De beschrijving daarvan ging stilzwijgend voorbij aan het feit dat de termen rechts van de pijl in de lexicale regels (v), (vi) en (vii), ofwel de onderste regel in de P-marker (5), nog niet de fonetische vorm van de zin aangeven. Deze termen kunnen daarom het best *formatieven* worden genoemd. De syntactische component genereert oneindig veel van zulke rijen formatieven, terwijl elke rij formatieven eenduidig een zin bepaalt. Het leidt over het algemeen niet tot verwarring om zo'n rij formatieven een *zin* te noemen. In deze betekenis levert de syntactische component oneindig veel zinnen, plus voor elke zin een structuurindex, een „verslag” van de generatiegeschiedenis. Uit de structuurindex blijkt bijvoorbeeld hoe de zin hiërarchisch is onderverdeeld in kleinere eenheden, de constituenten.

Zinnen en structuurindices behoren aldus tot de output van de syntactische component. Deze output vormt de input voor zowel de fonologische als de semantische component van de grammatica, maar op verschillende wijze zoals verderop zal blijken.

##### 4.2. *De fonologische component*

Deze bepaalt de zgn. fonetische interpretatie van de zin, dwz. de fonologische regels geven aan aan welke beperkingen de uiteindelijke rij van *fonen* die deze zin representeert onderhevig is. Men moet zich dit concreter als volgt voorstellen: De *lexicale formatieven* (er zijn ook zgn. *grammaticale*, zoals woordgrenzen enz.) van een zin, bijv. *de, linguïst* enz. in (2) bestaan uit segmenten: /d/, /e/ etc. Elk segment kan worden beschreven als een bundel systematische kenmerken, de zgn.

*distinctive features*. Zo kan de /d/ tot aan irrelevante vrije variatie geheel worden gekarakteriseerd door de verzameling distinctive features, zoals weergegeven in (11).

(11)	— vocalic + consonantal — grave + diffuse — strident — nasal — continuant + voiced	(12)	— vocalic + consonantal — grave + diffuse — strident — nasal — continuant — voiced	(13)	— vocalic + consonantal — grave + diffuse — strident + nasal — continuant + voiced
------	---	------	---	------	---

Hierin staan + en — respectievelijk voor de al of niet aanwezigheid van de betreffende feature. Soms is een feature ongespecificeerd, dat wil zeggen vrij. Er is hier geen ruimte om in te gaan op de preciese betekenis van de features in (11). Een enkel voorbeeld ter verduidelijking moet volstaan:

Over het algemeen is een tiental features voldoende om de verschillende segmenten in een taal van elkaar te onderscheiden. Zo is de /t/ van de /d/ onderscheiden op slechts één feature: [voiced]. De feature-kolom voor /t/ is gegeven in (12). De specificatie voor /n/ is gegeven in (13). De /n/ verschilt van /d/ in de feature [nasal]. /n/ en /t/ verschillen aldus in twee features: [nasal] en [voiced].

De formatieven die uit de syntactische component komen bestaan zo onder andere uit segmenten als (11), (12) en (13) en kunnen worden opgevat als feature matrices, waarin de kolommen segmenten zijn en de rijen features. De cellen bevatten +, — of niets. Deze feature matrices, output van de syntactische component worden *classificatorische matrices* genoemd.

Het is op deze classificatorische matrices dat de regels van de fonologische component werken. De regels zijn van de vorm (1) en worden meestal in de (1a) notatie weergegeven. Zij veranderen de classificatorische matrices in zgn. *fonetische matrices*, die de eigenlijke articulatorisch-akoestische beschrijving van de zin zijn. Ook de fonetische matrix heeft distinctive features als rijen, alleen zijn de waarden ervan nu niet meer noodzakelijk binair (+ of —), er is ruimte voor een graduele specificatie. Deze specificatie wordt door de fonologische regels gegeven. Tevens vullen deze de lege cellen in en kunnen features van teken of waarde doen veranderen al naar gelang de omgeving waarin het be-

treffende segment zich bevindt. Ook kan een segment verdwijnen. De B in regel (1) kan in de fonologische regels het nul-symbool zijn.

Nu een enkel voorbeeld voor het effect van de context in een fonologische regel van het Nederlands. In de Nederlandse dialecten die stemhebbende continuënten hebben, zoals /z/, /v/ (en in sommige gevallen ook /ʒ/, de stemhebbende g) worden deze meestal stemloos, dwz. /s/, /f/ en /x/ als zij worden voorafgegaan door /p/, /t/, /k/, /s/, /f/ of /č/. Zo bijvoorbeeld de /z/ in *zand* in contexten als 't *zand*, de /v/ van *vee* in *koop vee*, de /ʒ/ van *geld* in *zakgeld*. De distinctive features vormen de mogelijkheid om al deze verschijnselen in één abstracte regel samen te vatten. Dit pleit voor hun linguïstische realiteit (zie hierover vooral HALLE, 1962, 1964). De stemhebbende continuënten zijn namelijk gekenmerkt door de features [+consonantal], [+continuent], [+voiced], terwijl de hele groep /p/, /t/, /k/, /s/, /f/, /č/ van de andere medeklinkers verschilt op slechts twee features, zij zijn namelijk [-nasal] en [-voiced]. Alle bovenstaande verschijnselen kunnen nu in de volgende algemene regel van de vorm (1a) worden weergegeven:

$$\left[ \begin{array}{l} + \text{ consonantal} \\ + \text{ continuent} \\ + \text{ voiced} \end{array} \right] \rightarrow \left[ \begin{array}{l} + \text{ consonantal} \\ + \text{ continuent} \\ - \text{ voiced} \end{array} \right] / \left[ \begin{array}{l} - \text{ nasal} \\ - \text{ voiced} \end{array} \right] -$$

Zonder informatieverlies wordt dit ook wel eenvoudiger geschreven als:

$$\left[ \begin{array}{l} + \text{ consonantal} \\ + \text{ continuent} \end{array} \right] \rightarrow [- \text{ voiced}] / \left[ \begin{array}{l} - \text{ nasal} \\ - \text{ voiced} \end{array} \right] -$$

Voor de fonologische regels speelt niet alleen de onmiddellijke omgeving van de segmenten een rol, maar soms ook de structuurindex van de zin. Bijvoorbeeld de toekenning van klemtoon is hiervan gedeeltelijk afhankelijk. Hierover is overigens nog weinig bekend.

#### 4.3. De semantische component

Zoals de fonologische component een fonologische representatie van de zin produceert, geeft de semantische component een semantische interpretatie van de zin. De semantische regels, de *projectieregels*, beschrijven hoe de betekenis van de zin bepaald is door de betekenis van de afzonderlijke formatieven en de structuurindex. Er worden op het ogenblik pogingen gedaan om de betekenis van zo'n formatief te beschrijven als een bundel *semantic features* (KATZ & FODOR, 1964),

bijvoorbeeld *eerlijk* = [+ human, + moral, enz.], *man* = [+ human, + masculine, + adult, enz.]. Dat een voldoende lange rij van zulke features de betekenis van een formatief zo ver kan inperken dat het geïdentificeerd kan worden, blijkt uit radioprogramma's als „Het hangt aan de muur en het tikt”. Er is niettemin serieuze twijfel over deze procedure als mogelijke basis voor een semantische theorie (zie bijv. BOLINGER, 1965). Hoe het ook zij, de projectieregels dienen te beschrijven hoe zulke betekenissen kunnen worden gekoppeld tot samengestelde betekenissen (bijv. *eerlijke man*). De structuurindex van de zin speelt hierbij een essentiële rol. Dit niet alleen met het oog op wat aan wat wordt gekoppeld, maar ook met betrekking tot hoe de koppeling dient te geschieden. Dit laatste hangt namelijk af van de grammaticale relaties die er tussen de betreffende constituenten bestaan. Een zelfde rij formatieven kan op die manier een verschillende betekenis hebben, al naar gelang de structuurindex, bijv.

(14) *het roepen van kinderen vulde de lucht*

Deze zin is dubbelzinnig. De eerste betekenis komt voor in: *het roepen van kinderen door hun moeders vulde de lucht*. De tweede zin is vervat in *het geroep van kinderen vulde de lucht*. Voor de semantische interpretatie is het daarom van groot belang of de structuur van de zin zo is dat de kinderen onderwerp van het roepen zijn dan wel de moeders. Dit moet zichtbaar zijn in de structuurindex van de zin. Dat dit hier inderdaad het geval is zal verderop worden aangetoond: de structuurindex is verschillend voor de twee betekenissen van de zin. Al naar gelang de betekenis van de zin zijn de woorden op verschillende wijze tot samengestelde eenheden (constituenten) gegroepeerd. Bovendien bepaalt de hiërarchische ordening die constituenten plegen te hebben (zie bijv. de P-marker (5)) de orde waarin de projectieregels worden toegepast. De projectieregels beginnen „onder in de boom”, dwz. bij de kleinste samengestelde constituenten (*de psycholoog*), (*de linguïst*). Als deze zijn geïnterpreteerd, herhaalt de cyclus zich op het volgende niveau, voor (5) leidt dat tot de interpretatie van de constituent (*ontmoet (de linguïst)*), en tenslotte wordt de grootste constituent, nl. de zin (*(de psycholoog) (ontmoet (de linguïst))*) geïnterpreteerd. De structuurindex bepaalt op deze wijze de opvolgende eenheden voor de projectieregels.

Het is tenslotte duidelijk dat voor de semantische interpretatie van de kleinste constituenten een soort woordenboek nodig is, een *dictionnaire*, waarin de betekenissen der formatieven zijn weergegeven. Hier wordt

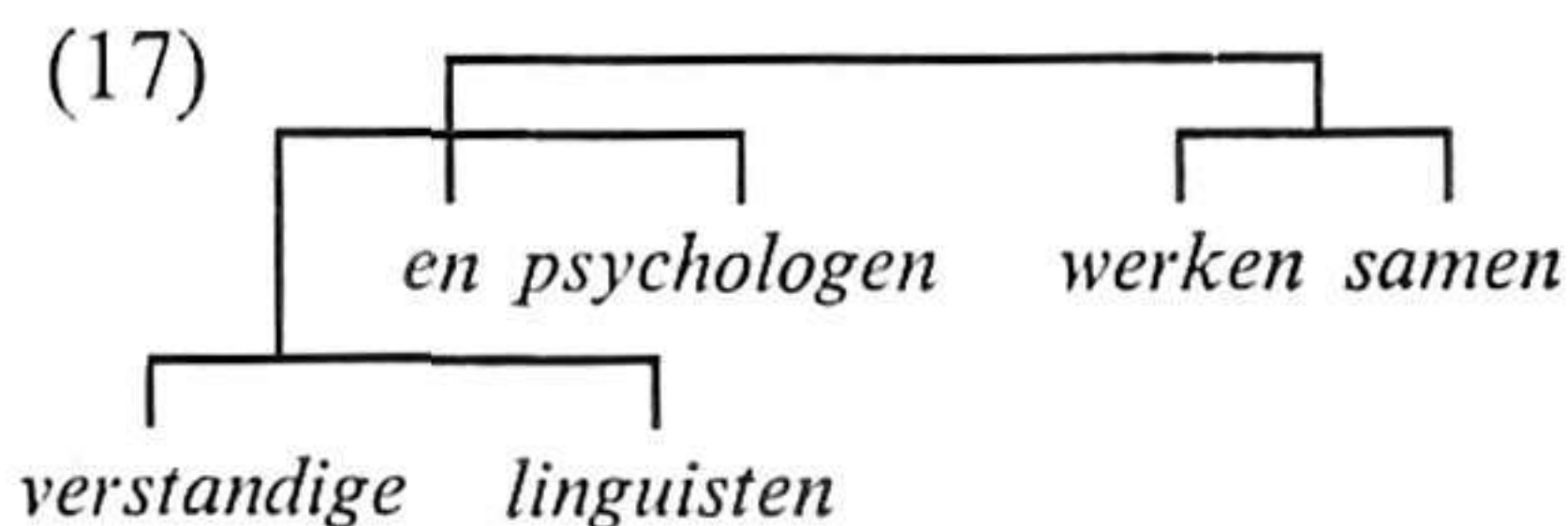
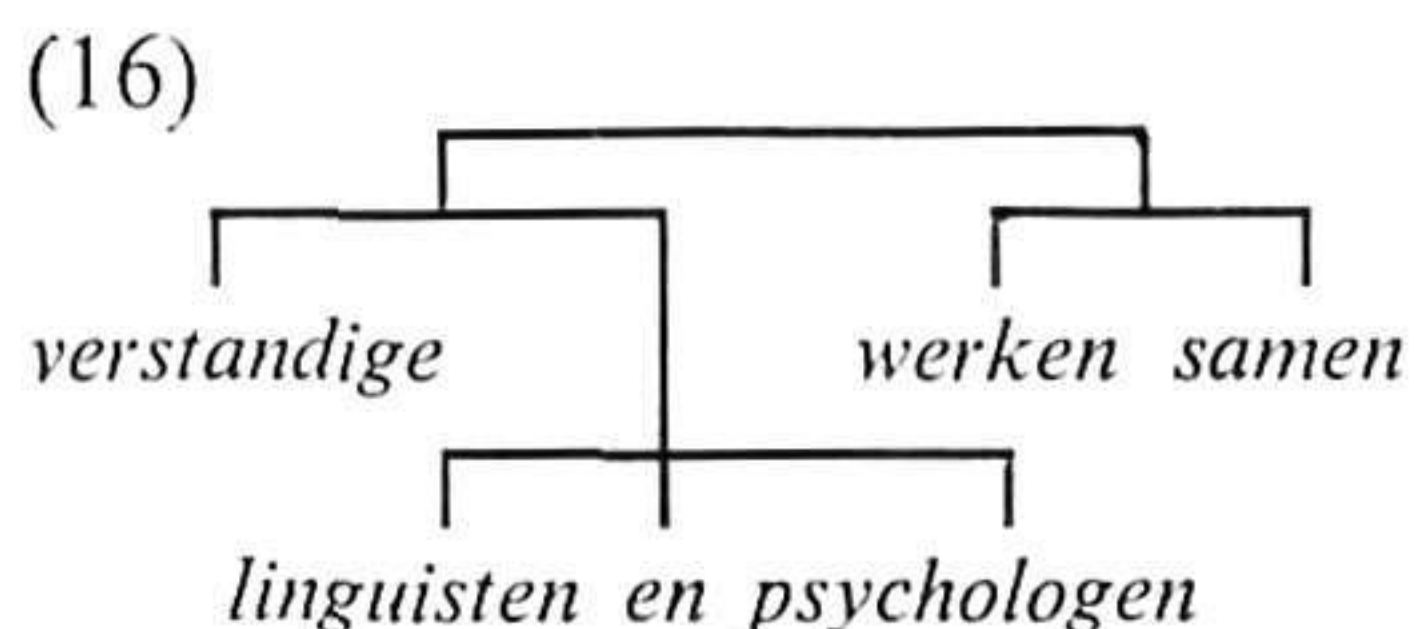
verder niet ingegaan op de eigenschappen van zo'n dictionaire, die met de projectieregels een integraal onderdeel vormt van de semantische component (zie KATZ, 1966).

5. OPPERVLAKTE- EN DIEPTESTRUCTUUR

De onderste regel in de P-marker (5) geeft de formatieven van de zin *de psycholoog ontmoet de linguist* in de goede vorm en in de goede volgorde. Afgezien van de fonetische interpretatie is dit de zin. Wanneer dit het geval is noemen we zo'n P-marker de *oppervlaktestructuur* van de zin. De oppervlaktestructuur is de input voor de fonologische component, dwz. zij is voldoende om de fonetische representatie van de zin nader te laten bepalen door de fonologische regels. Soms laat de oppervlaktestructuur ook zien hoe de zin semantisch geïnterpreteerd moet worden. Zo kan de dubbelzinnigheid van een zin wel eens blijken uit de oppervlaktestructuur. Dit is bijvoorbeeld het geval voor de volgende ambigue zin.

(15) *Verstandige linguisten en psychologen werken samen*

Er zijn nu twee oppervlakte- P-markers die deze zin kunnen domineren, nl. de P-marker (16) en de P-marker (17) (de constituent-aanduidingen zijn hier weggelaten):



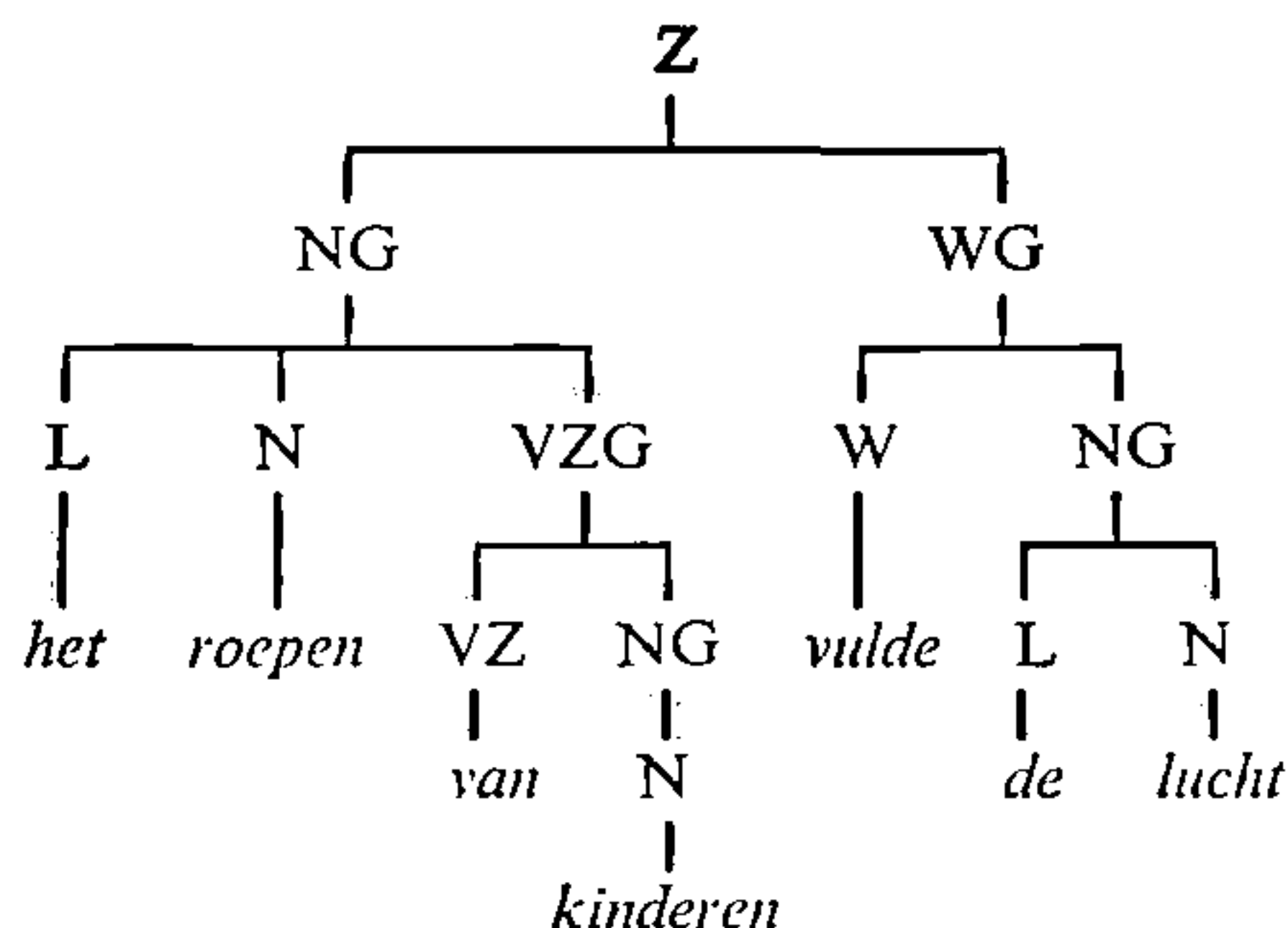
In het geval (16) betekent de zin dat verstandige linguisten met verstandige psychologen samenwerken, in het geval (17) dat verstandige linguisten met psychologen samenwerken. (16) en (17) zijn oppervlaktestructuren omdat zij de formatieven van de zin in de uiteindelijke volgorde en vorm weergeven die zij in de zin hebben. Alleen de fonetische representatie moet nog volgen. De dubbele semantische interpretatiemogelijkheid is hier in de oppervlaktestructuur zichtbaar.

De oppervlaktestructuur is echter meestal onvoldoende voor een juiste semantische interpretatie. De reden hiervoor is dat de oppervlaktestructuur slechts zelden het resultaat is van herschrijfregels alleen. Zoals uiteengezet zijn herschrijfregels vaak onvoldoende om zinnen van een

natuurlijke taal te genereren. De oppervlakte-P-markers zijn meestal het resultaat niet alleen van herschrijfregels, maar ook van transformatieoperaties. Dit wil eenvoudig zeggen dat een oorspronkelijke, door herschrijfregels gegenereerde P-marker door transformaties (d.i. weglaten, toevoegen of verwisselen van takken) in de P-marker van de oppervlaktestructuur is omgezet. Daarom heet de oppervlakte P-marker *afgeleide P-marker*. De oorspronkelijke P-marker stelt de *dieptestructuur* van de zin voor. De oppervlaktestructuur resulteert uit de dieptestructuur door transformaties. De dieptestructuur wordt voortgebracht door herschrijfregels.

Dit onderscheid tussen oppervlakte- en dieptestructuur is onder andere ingegeven door het feit dat een zin met slechts één mogelijke oppervlakte-P-marker niettemin syntactisch dubbelzinnig kan zijn. (Dat is dus anders dan in zin (15), waar twee verschillende P-markers dezelfde rij formatieven konden beschrijven.) Het is nu mogelijk om zo'n geval als volgt te beschrijven: Er zijn twee verschillende dieptestructuur-P-markers die echter door onderling verschillende transformaties in dezelfde afgeleide P-marker resulteren. Zonder de transformatiegeschiedenis is het dan niet meer na te gaan hoe de oorspronkelijke dieptestructuur was. Een voorbeeld van deze situatie is zin (14), *het roepen van kinderen vulde de lucht*. De oppervlaktestructuur is voor beide betekenissen van de zin iets als (18).

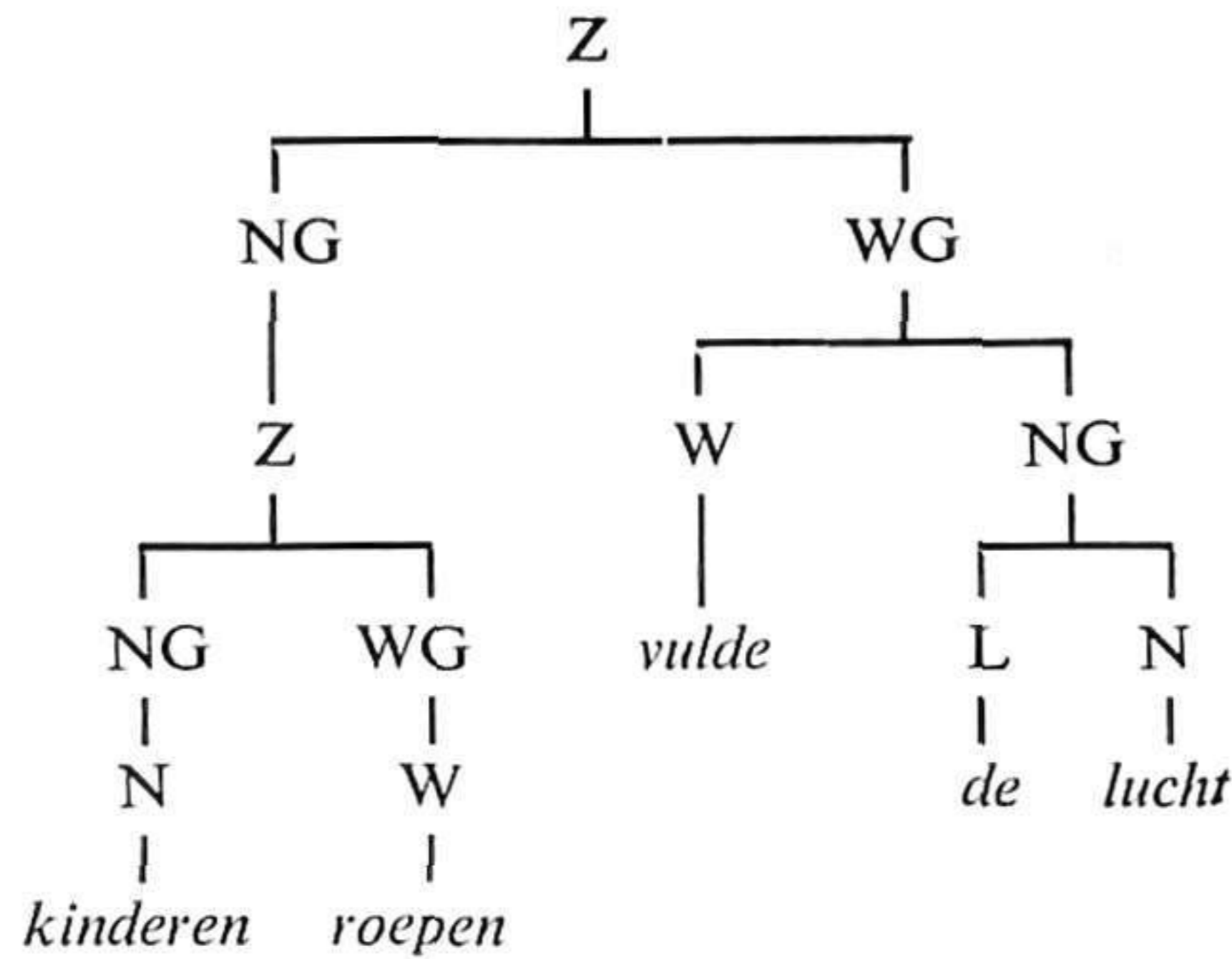
(18)



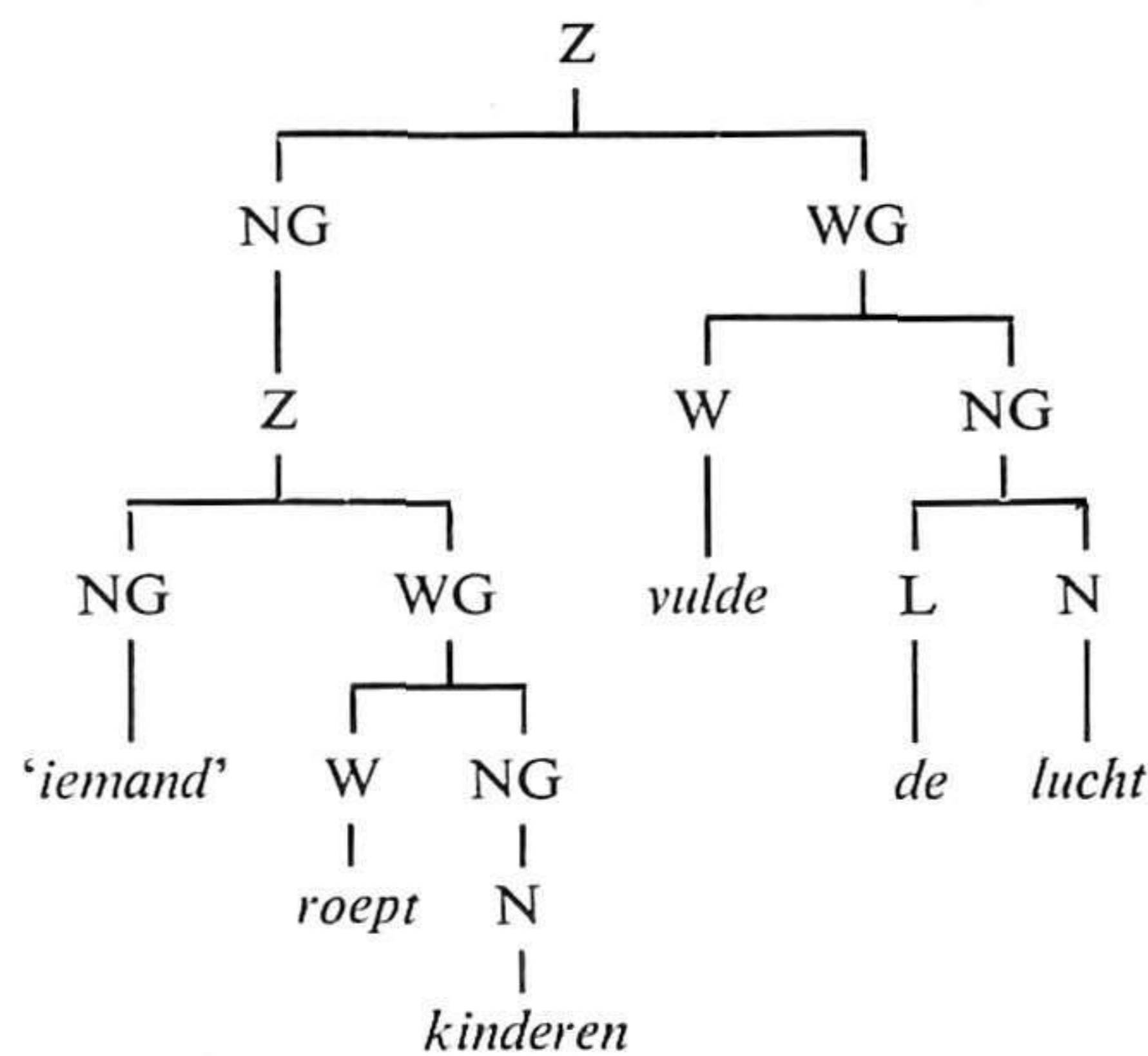
De formatieven staan hier in vorm en volgorde die voldoende is voor de fonologische representatie van de zin. Fonetisch is de zin dan ook gelijk voor beide betekenissen. Er is zelfs geen verschil in klemtoonpatroon hoorbaar.

(18) is echter het transformationele resultaat van twee dieptestructuren. Deze kunnen — iets vereenvoudigd — als volgt worden weergegeven.

(19)



(20)



De transformaties projecteren zowel (19) als (20) in (18). De diepte-structuren (19) en (20) laten het verschil zien tussen de twee betekenissen van (14). De dieptestructuren vormen nu de input voor de semantische component. De semantische interpretatie, die „van onder naar boven” werkt verloopt aldus geheel verschillend voor (19) en (20). Ook de P-markers (16) en (17) zijn afgeleide P-markers. Zij stammen uit verschillende dieptestructuren: dat is wat het verschil in semantische interpretatie bepaalt. Er kunnen zich dus twee soorten syntactische dubbelzinnigheid voordoen: de zin kan twee verschillende diepte-structuren bij slechts één oppervlaktestructuur hebben (dat geldt voor *het roepen van kinderen vulde de lucht*), of de zin kan twee verschillende diepte- én oppervlaktestructuren hebben (dat geldt voor *verstandige linguïsten en psychologen werken samen*) (en soortgelijks voor meerzinnigheid).

Zowel (19) als (20) vertonen de recursieve herschrijfrègel  $NG \rightarrow Z^2$ ). Boven werd opgemerkt dat het type regels waarin  $Z$  rechts van de pijl voorkomt het generatieve hart van de grammatica vormt. Op die manier worden zinnen „in elkaar gehaakt”. Herhaalde toepassing van zulk soort regels levert ingebedde constructies als *het roepen van kinderen die buiten speelden vulde de lucht* (hier is de zin *kinderen speelden buiten* ingehaakt), *het roepen van kinderen die buiten waar de zon scheen speelden vulde de lucht* (inbedding van *buiten scheen de zon*), enz. Ook het geval (7) komt hieruit voort.

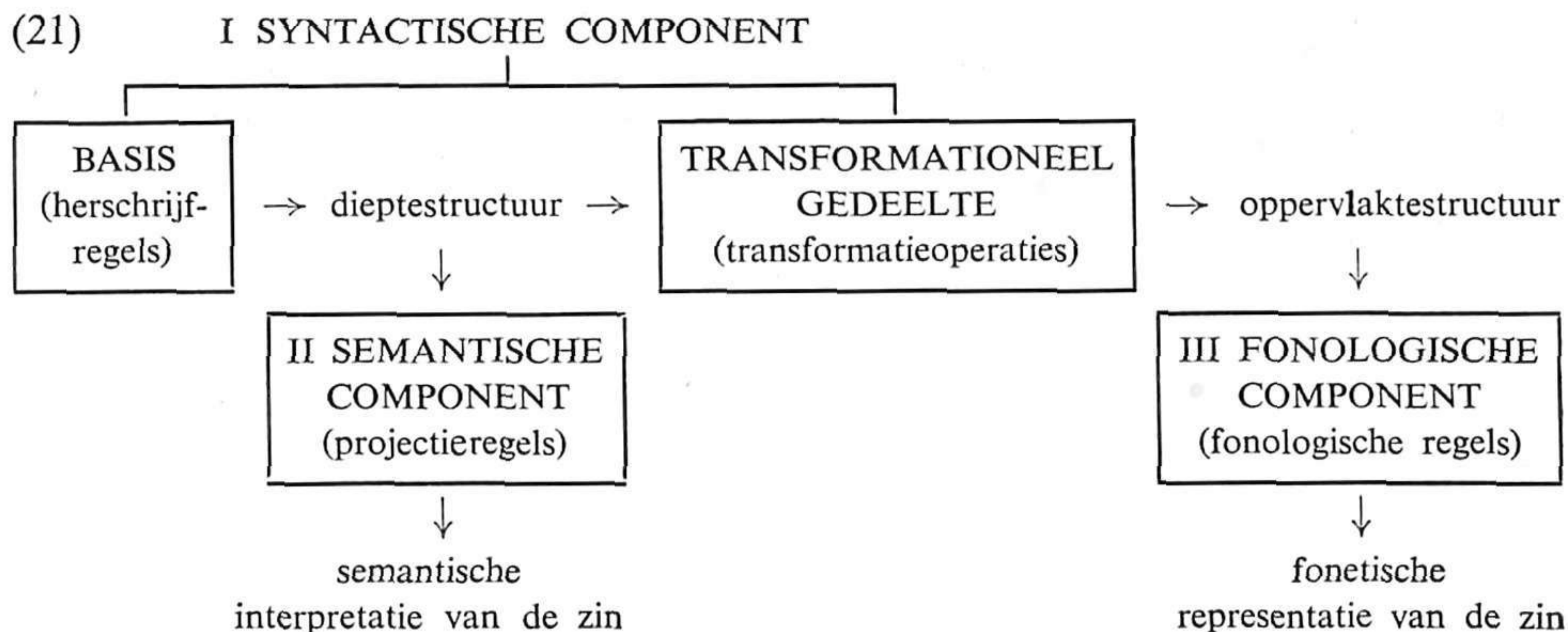
P-markers die het resultaat zijn van het type herschrijfregels waar  $Z$  rechts van de pijl voorkomt heten *gegeneraliseerde P-markers*. Zo zijn (19) en (20) *gegeneraliseerde P-markers*.

Het is duidelijk dat er nog heel wat transformaties voor nodig zijn om (19) en (20) in de afgeleide P-marker (18) om te zetten. Er moeten formatieven verschijnen (*van*) en verdwijnen („*iemand*”) en de volgorde moet worden gewijzigd. Er wordt hier geen poging gedaan om deze transformaties te beschrijven. Het zij slechts opgemerkt dat deze steeds in abstracte vorm kunnen worden gesteld, dwz. onafhankelijk van de toevallige formatieven van de zin. Zij gelden voor elke P-marker van dezelfde structuur.

Samenvattend: De *basis van een grammatica*, dat is dat deel van de grammatica dat de dieptestructuren van de zinnen genereert, is een constituenten-grammatica, bestaande uit contextvrije herschrijfregels. Sommige van deze herschrijfregels zijn recursief: zij introduceren opnieuw het symbool  $Z$ . Dit leidt tot het ontstaan van *gegeneraliseerde P-markers*. De dieptestructuur van een zin is bepalend voor de semantische interpretatie ervan; zij is de input voor de semantische component van de grammatica. De dieptestructuur wordt in de oppervlaktestructuur omgezet door transformatieoperaties. Transformatieoperaties veranderen P-markers in nieuwe P-markers door toevoegen, weglaten of verwisselen van elementen. De oppervlaktestructuur van de zin vormt de input voor de fonologische component van de grammatica. Deze regelt de fonetische vorm van de zin. Zo bewerkt de grammatica dat een semantische interpretatie wordt gepaard aan een fonetische representatie. De samenstelling van de transformationele grammatica is schematisch weergegeven in (21).

<sup>2)</sup> We geven hier een enigszins vereenvoudigde voorstelling van zaken. In werkelijkheid heeft de regel een iets ingewikkelder vorm. Deze verschillen zijn voor ons betoog niet van belang.





Op basis van het onderscheid tussen diepte- en oppervlaktestructuur kan nu ook het intuïtieve begrip *kernzin* worden verduidelijkt. De zinnen (2), (3) en (10) zijn kernzinnen. Voor deze zinnen is de overeenkomst tussen oppervlakte- en dieptestructuur maximaal. Dat wil zeggen, er heeft een minimum aan transformaties plaats gevonden in het voortbrengen van deze zinnen (de vervoeging van het werkwoord). Vergelijk hiermee een zin als *wordt de bal door Jan geschopt?* In deze zin zien we onder andere het resultaat van een  $T_{pas}$ . Zonder deze transformatie hadden we *schopt Jan de bal?* gekregen. Voor deze laatste zin is daarom de overeenkomst tussen diepte- en oppervlaktestructuur groter dan voor de eerste. Er is echter in (10) *Jan schopt de bal* een nog grotere overeenkomst. De  $T_{vr}$  die in de generatie van *schopt Jan de bal?* nog was opgetreden heeft hier niet plaats gevonden. Het is nu echter onmogelijk nog meer transformaties achterwege te laten. Er is geen eenvoudiger vorm dan *Jan schopt de bal*, daarom is deze zin een kernzin. In de generatie van kernzinnen treden echter ook steeds transformaties op, met name de vervoeging van het werkwoord.

Transformaties als  $T_{pas}$ ,  $T_{vr}$ ,  $T_{neg}$ , die in een taal vaak gemakkelijk zijn te onderkennen, kunnen de mogelijkheid geven de regels van de basis te achterhalen. Met name de afbakening van de constituenten (NG, WG, N, enz.) wordt vergemakkelijkt door op te sporen wat zich onder transformaties als eenheid gedraagt.

## 6. UNIVERSALIA VAN DE TAAL EN TAALVERWERVING

Het is een oud probleem in de algemene taalwetenschap hoe allerlei trekken van individuele talen terug te voeren zijn op universele eigen-

schappen van taal als zodanig. Een recent overzicht over de stand van zaken op dit gebied is te vinden in GREENBERG (1963). De grammaticale beschrijving van een taal kan enorm worden vereenvoudigd door *substantieve* en *formele universalia* niet meer in de grammatica's van de afzonderlijke talen op te nemen. Substantieve universalia zijn „elementen” die in elke taal voorkomen, formele universalia zijn abstracter, zij zeggen iets over de structuur van de grammaticale regels. Formele universalia van de fonologische component bijvoorbeeld, zijn mogelijk te vinden in het cyclische karakter van sommige fonologische regels (met name de klemtoonregels). Het schijnt algemeen voor te komen dat een geordende verzameling van fonologische regels steeds weer wordt herhaald, beginnend „onderaan” in de oppervlaktestructuur en eindigend bij Z. Formele universalia van de syntactische component zijn waarschijnlijk gelegen in de transformationele structuur ervan, dwz. het fundamentele onderscheid tussen diepte- en oppervlaktestructuur. Het lijkt de moeite waard te onderzoeken of talen die in oppervlaktestructuur sterk verschillen in de basis van de grammatica soms diepgaande overeenkomsten vertonen, die alleen door een verschil in transformationele structuur verdoezeld worden. Voorbeeld van een algemene formele eigenschap in het semantische vlak is dat in elke taal de kleurwoorden het kleurenspectrum in aaneensluitende segmenten verdelen (dit is geenszins noodzakelijk!).

Als substantieve universalia in de fonologische component worden de distinctive features voorgesteld; in de syntactische component sommige constituenten zoals N, W, WG. In de semantische component schijnen trekken als „persoon” een universele rol te spelen.

Het ligt nu voor de hand om deze universalia, die geen enkele logische noodzaak hebben, te beschrijven als eigenschappen van het menselijke taalverwervingssysteem. Het zouden dan aangeboren structuren zijn, die worden verwerkt indien de adekwate uitwendige stimulus aan het kind wordt gepresenteerd. In beginsel wordt hier hetzelfde mechanisme gesuggereerd als wat ethologen „inprenting” noemen. Dit mechanisme is van totaal andere aard dan een leer- of conditioneringsproces (zie LENNEBERG, ter perse).

Binnen de generatieve taaltheorie wordt een belangrijke plaats ingeruimd voor een abstracte studie van de eigenschappen van het taalverwervingssysteem. Men vraagt zich af hoe de formele structuur van zo'n systeem is geconstitueerd, uitgaande van de overweging dat de input van het systeem bestaat uit een eindige verzameling min of meer wel-

gevormde zinnen en de output uit een functionele grammatica van de betreffende taal. Daarbij kan men enkele condities toevoegen. Ten eerste een zekere terugkoppeling (zoals beloning voor welgevormde zinnen), ten tweede een tijdsbeperking (naar analogie van het menselijke taalverwervingssysteem, dat zijn werk grotendeels in een periode van 24 maanden voltooit). Vanuit formele overwegingen van deze aard is het vrijwel zeker dat zo'n verwervingssysteem over andere dan inductieve mogelijkheden moet beschikken. Op de basis van inductieve processen als stimulusgeneralisatie (die tot het onderkennen van de eenheden in het binnenkomende materiaal kan leiden) en associatie (die de eenheden tot groepen samenstelt) zou er een grammatica ontstaan die fundamenteel een verkeerde formele structuur heeft (zie o.a. CHOMSKY, 1959 en KATZ & POSTAL, 1964).

In ieder geval is er een *a priori* onder- en bovengrens aan de structuur van het menselijke verwervingssysteem. De ondergrens is niet de *tabula rasa*: er moet minstens de structurele mogelijkheid zijn om iets te doen met de binnenkomende zinnen zodat ze uiteindelijk de grammatica opleveren, er moet een mediatiemogelijkheid zijn. De bovengrens wordt gesteld door de verscheidenheid van talen. Het kan niet worden verondersteld dat ons verwervingssysteem zo rijk gestructureerd is dat het ons determineert tot het spreken van Nederlands. Johannes Goropius Becanus uit Antwerpen schijnt echter ook voor deze stelling niet te zijn teruggeschrokken (zie Bloomfield's Language).

#### SAMENVATTING

Dit artikel is een terminologische inleiding in de generatieve grammatica. Na behandeling van de generatieve regels die ten grondslag liggen aan het finite state model, het phrase structure model en het transformationele model wordt een korte beschrijving gegeven van de syntactische, de fonologische en de semantische component van de transformationele grammatica. Speciaal wordt de nadruk gelegd op het verschil tussen basis en transformationeel gedeelte van de syntactische component. Hierbij wordt uitgegaan van de theorie van gegeneraliseerde P-markers. Tenslotte volgen enkele opmerkingen over universalia van de taal en taalverwerving.

#### SUMMARY

This paper is a terminological introduction to generative grammar. After a discussion of the generative rules underlying the finite state model, the phrase structure model, and the transformational model, a short description is presented of the syntactic, the phonological, and the semantic components of transformational grammar. Special stress is laid on the difference between base and transformational part of the syntactic component. Herein the theory of generalized P-markers is adopted. Finally some remarks are given on language universals and language acquisition.

## DANKBETUIGING

Gaarne dank ik Drs. R. P. G. DE RIJK, die het eerste concept van dit artikel van veel waardevolle op- en aanmerkingen voorzag. Fouten komen echter geheel voor mijn rekening.

## ACKNOWLEDGEMENT

This investigation was supported by Grant # B-3233 from the Carnegie Corporation of New York to Harvard University, Center for Cognitive Studies.

Cambridge, Mass., februari 1965

## LITERATUUR

- BACH, E., 1964. An introduction to transformational grammars. Holt, Rinehart & Winston, New York.
- BOLINGER, D. L., 1965. The atomization of meaning. *Language*, 41, 555.
- CHOMSKY, N., 1956. Three models for the description of language. Herdrukt in: *Readings in mathematical psychology*, Vol. II, R. D. Luce, R. Bush & E. Galanter, eds. John Wiley & Sons, Inc., New York, 1965.
- , 1957. *Syntactic structures*. Mouton & Co., Den Haag.
- , 1959. A review of B. F. Skinner's *Verbal behavior*. Herdrukt in FODOR & KATZ, 1964.
- , 1961. On the notion "Rule of grammar". Herdrukt in FODOR & KATZ, 1964.
- , 1963. Formal properties of grammars. In: R. D. Luce, R. Bush & E. Galanter, eds., *Handbook of mathematical psychology*, Vol. II. John Wiley & Sons, Inc., New York.
- & G. A. MILLER, 1963. Introduction to the formal analysis of natural languages. In: R. D. Luce, R. Bush & E. Galanter, eds., *Handbook of mathematical psychology*, Vol. II. John Wiley & Sons, Inc., New York.
- , 1964. *Current issues in linguistic theory*. Mouton & Co., Den Haag.
- , 1965. *Aspects of the theory of syntax*. MIT-press, Cambridge, Mass.
- , *ter perse*. *Topics in the theory of generative grammar*. In: T. A. Sebeok, ed., *Current trends in linguistics*, Vol. III. *Linguistic theory*. Mouton & Co., Den Haag.
- DIEBOLD, A. R., 1965. A Survey of psycholinguistic research 1954-1964. In: C. E. Osgood & T. A. Sebeok, eds., *Psycholinguistics*. Indiana University Press, Bloomington.
- FODOR, J. A. & J. J. KATZ, eds., 1964. *The structure of language*. Readings in the philosophy of language. Prentice-Hall, Englewood Cliffs.
- GEER, J. P. VAN DE, 1957. Psychologische toepassingen van de informatietheorie I. *Ned. Tijdschr. Psychol.*, 12, 295.
- GREENBERG, J. H., ed., 1963. *Universals of language*. MIT-press, Cambridge, Mass.

- HALLE, M., 1962. Phonology in generative grammar. Herdrukt in FODOR & KATZ, 1964.
- , 1964. On the basis of phonology. In: FODOR & KATZ, 1964.
- HOCKETT, C. F., 1955. A manual of phonology. Indiana University Publications in Anthropology and Linguistics, Memoir 11, Baltimore.
- KATZ, J. J. & J. A. FODOR, 1964. The structure of a semantic theory. Herdrukt in: FODOR & KATZ, 1964.
- & P. POSTAL, 1964. An integrated theory of linguistic descriptions. MIT-press, Cambridge, Mass.
- , 1966. The philosophy of language. Harper & Row, New York.
- KRAAK, A., 1966. Negatieve zinnen. Een methodologische en grammatische analyse. W. de Haan, Hilversum.
- LEES, R. B., 1960. The grammar of English nominalizations. Mouton & Co., Den Haag.
- LENNEBERG, E., ter perse. The biological foundations of language. John Wiley & Sons, Inc., New York.
- MILLER, G. A. & N. CHOMSKY, 1963. Finitary models of language users. In: R. D. Luce, R. Bush & E. Galanter, eds., Handbook of mathematical psychology, Vol. II. John Wiley & Sons, Inc., New York.
- MORSBACH, L., 1896. Mittelenglische Grammatik. M. Niemeyer, Halle.
- OSGOOD, C. E. & T. A. SEBEOK, 1954. Psycholinguistics, a survey of theory and research problems. Herdrukt in C. E. Osgood & T. A. Sebeok, eds., Psycholinguistics. Indiana University Press, Bloomington, 1965.
- POSTAL, P., 1964. Limitations of phrase structure grammars. In: FODOR and KATZ, 1964.
- REICHLING, A. J. B. N., 1961. Principles and methods of syntax: cryptanalytical formalism. *Lingua*, **10**, 1. Een Nederlandse versie hiervan verscheen in: A. J. B. N. Reichling. Verzamelde studies over hedendaagse problemen der taalwetenschap, Hoofdstuk V. Tjeenk Willink, Zwolle, 1965 (3e druk).
- UHLENBECK, E. M., 1963. An appraisal of transformation theory. *Lingua*, **12**, 1.