

Samenvatting van “Complexiteit en leerbaarheid van fonologische patronen: simulaties, experimenten, typologie”

Je kan taalwetenschappers op veel verschillende manieren onderverdelen. Er is bijvoorbeeld een groep taalwetenschappers die wil begrijpen hoe taal wordt opgeslagen en verwerkt in de hersenen: dat zijn psycholinguïsten. Er is ook een groep taalwetenschappers die onderzoekt wat voor eigenschappen er vaak – of juist niet vaak – voorkomen in de talen van de wereld (zover die beschreven zijn), en waarom: die noemen we typologen. Er zijn taalwetenschappers die geïnteresseerd zijn in hoe klanken of gebaren worden gearticuleerd en waargenomen, en hoe ze bijvoorbeeld betekenissen van elkaar onderscheiden: zulke mensen heten fonologen en/of fonetici. En zo zijn er nog veel meer groepen taalwetenschappers — degenen die kijken hoe taal als communicatiemiddel wordt gebruikt, degenen die computermodellen bouwen, degenen die onderzoeken hoe kinderen taal verwerven, enzovoort. Alle taalwetenschappers behoren tot meer dan één groep: de onderverdelingen zijn gedefinieerd in meerdere, onafhankelijke dimensies. Voor het onderzoek dat ik in dit proefschrift beschrijf, heb ik bijvoorbeeld computersimulaties gedaan, experimenten uitgevoerd, en naar een database gekeken met informatie over echte talen. Heel algemeen geformuleerd heb ik de vraag willen beantwoorden: *komen dingen die mensen makkelijker kunnen leren ook vaker voor in talen?*

Het woord “dingen” is natuurlijk niet heel specifiek. In hoofdstuk 1 van dit proefschrift leg ik de achtergrond van mijn promotieonderzoek nader uit. Voor dit onderzoek heb ik naar foneeminventarissen gekeken, de verzamelingen klanken die gesproken talen inzetten om verschillende betekenissen van elkaar te onderscheiden. In het Nederlands verschillen de woorden *dal* en *bal* bijvoorbeeld maar in één klank van elkaar, maar ze verwijzen naar heel andere dingen. Deze observatie vertelt ons iets over het Nederlands: [d] en [b] zijn FONEMEN van het Nederlands.^a We kunnen fonemen, net als taalwetenschappers, ook in verschillende groepen indelen. Sommige Nederlandse fonemen, zoals {b f m w}, worden uitgesproken met één of beide lippen: zulke klanken heten labialen of lipklanken. Sommige Nederlandse fonemen, zoals {b d z l}, spreek je uit met trillende stembanden: zulke klanken noemen we stemhebbend. Deze eigenschappen heten FONOLOGISCHE KENMERKEN, en de namen van die kenmerken zijn doorgaans gebaseerd op de manier waarop we ze uitspreken of waarnemen. De [b] komt voor in het rijtje labialen, maar ook in het rijtje stemhebbende klanken: een foneem heeft meerdere kenmerken tegelijk. Maar hoe leren taalgebruikers die kenmerken? Er bestaat een redelijk populaire theorie, die

^a Fonemen noteer ik in [sluistekens], groepen klanken in {accolades}, en realisaties van klanken in [rechte haken].

van de Universele Grammatica, die stelt dat we worden geboren met kennis over die kenmerken: die kennis zou in ons DNA zitten. Die theorie is gebaseerd op de observaties dat kinderen schijnbaar makkelijk taal verwerven, en dat de talen van de wereld veel op elkaar lijken. Als deze theorie klopt zou het echter wel betekenen dat die kenmerken zich in alle talen precies hetzelfde moeten gedragen, en we weten dat dat niet zo is; sterker nog, hoe meer talen er door typologen beschreven worden, hoe meer onverwachte dingen we vinden, dingen die als het ware de poten onder de stoel van de universaliteit wegzagen.

In hoofdstuk 2 van dit proefschrift gebruik ik computersimulaties om te kijken hoe fonologische kenmerken ontstaan in een NEURAAL NETWERK, een model van een taalleerder dat gebaseerd is op hoe de hersenen werken. Zo'n netwerk bestaat uit zogeheten knopen, die met elkaar verbonden zijn: die knopen kunnen geactiveerd worden, en door de verbindingen andere knopen ook actief maken ("activeren") of juist minder actief maken ("inhiberen"), net zoals synapsen in onze hersenen elkaar activeren of inhiberen. Zo kan activiteit door het model stromen. In het model dat ik gebruik zijn de knopen ingedeeld in drie lagen. Twee daarvan stellen in zekere zin dingen buiten de leerder voor: een van de lagen is een representatie van verschillende betekenissen (zoals objecten die de leerder kan waarnemen), en een andere is een representatie van geluiden die de leerder kan horen, zoals frequenties (met oneindig veel mogelijke waarden). Daartussenin zit nog een derde laag, die een representatie in de hersenen van de leerder voorstelt. Door het computermodel combinaties van objecten en geluiden aan te bieden, en de activaties te laten stromen, ontstaat er gaandeweg op die tussenliggende laag een abstract representatieniveau: de oneindig vele mogelijke geluiden worden in de tussenlaag teruggebracht tot een beperkt aantal activatiepatronen, en dat zijn de kenmerken waar ik het hierboven over had. Die kenmerken kan het model dus leren uit zijn omgeving; we hoeven helemaal niet aan te nemen dat het ze meteen al kent, zoals de Universele Grammatica aanneemt.

Het netwerk leert door alleen te luisteren, maar in hoofdstuk 3 laat ik het ook spreken: daarvoor is nog een vierde laag nodig in het netwerk, een articulatorische laag. Ik laat netwerken ook van elkaar leren, een beetje zoals in een doorfluister-spelletje — een netwerk leert eerst van zijn omgeving en spreekt daarna, waarna een nieuw netwerk van de output van het vorige leert. Zo ontstaan ketens van opeenvolgende "generaties" netwerken die van elkaar leren, zoals echte talen ook van generatie op generatie worden overgedragen. In dit proces van overdracht spelen twee effecten een belangrijke rol: het PROTOTYPE-EFFECT en het ARTICULATORISCH EFFECT. Het prototype-effect houdt in dat je als luisteraar sommige realisaties van fonemen prototypischer, eenvoudig gezegd misschien "beter", vindt dan andere: een [s] die goed sist, met een hoge frequentie, is een duidelijke realisatie van dat foneem, die je eigenlijk onmogelijk kunt verwarren met een andere klank. Zulke prototypische realisaties van klanken kosten echter meestal ook meer moeite om uit te

spreken, en als spreker wil je niet meer moeite doen dan strikt noodzakelijk: de twee effecten zijn dus tegengesteld aan elkaar. In het netwerk is dat geïmplementeerd door geluiden die meer articulatorische moeite kosten sterker te inhiberen dan geluiden die minder moeite kosten. De simulaties in Hoofdstuk 3 laten zien dat de twee effecten gedurende de overdracht tussen meerdere generaties een optimale balans vinden: als je bijvoorbeeld begint met een taal waar sprekers heel veel moeite moeten doen om betekenissen uit elkaar te houden, dan verandert zo'n taal totdat het onderscheid zo klein is als nodig, en dat dat eenmaal is gebeurd, verandert hij niet noemenswaardig meer. Dat talen in sommige opzichten vaak op elkaar lijken kan dus komen doordat ze zo evolueren, zonder dat daar iets doelmatig voor hoeft te gebeuren, en zonder dat er iets aangeboren hoeft te zijn. Het idee dat mensen – en eigenlijk niet alleen mensen – zo min mogelijk moeite doen om hun doel te bereiken zie je trouwens overal om je heen: denk bijvoorbeeld maar aan de “olifantenpaadjes” die langzaam uitslijten als mensen ergens een stukje willen afsteken.

Terug naar de vraag die centraal staat in dit proefschrift. Om die te beantwoorden moest ik uitzoeken welke dingen mensen makkelijk kunnen leren, en dat heb ik gedaan in hoofdstukken 4 en 5. Ik schreef eerder dat foneeminventarissen eigenlijk altijd worden geanalyseerd aan de hand van de kenmerken die relevant zijn. Als voorbeeld zie je hieronder een tabel met een inventaris van plosieven, klanken die je maakt door heel kort een volledige afsluiting in het spraakkanaal te maken en die vervolgens weer los te laten, zoals [p], of [t], of [d]. Om deze inventaris te beschrijven hebben we twee kenmerken nodig: stemhebbendheid (trillen de stembanden of niet?) en plaats van articulatie (waar wordt de afsluiting in het spraakkanaal gemaakt?). De stembanden kunnen niet of wel trillen (respectievelijk “stemloos” en “stemhebbend”), en de plaats van articulatie kan “labiaal” zijn (bij de lippen), “coronaal” (bij het harde verhemelte), of “dorsaal” (met de achterkant van de tong). Dat levert de zes mogelijkheden in Tabel 1 op (het foneem [g] is de eerste klank in het woord *goal*, niet die in *goed*). Het Engels heeft deze inventaris, bijvoorbeeld.

Tabel 1. *Plosiefinventaris 1.*

	<i>labiaal</i>	<i>coronaal</i>	<i>dorsaal</i>
<i>stemloos</i>	p	t	k
<i>stemhebbend</i>	b	d	g

Dit is een overzichtelijke inventaris: alle combinaties van kenmerken bestaan. Nu een andere mogelijke inventaris (“patroon”):

Tabel 2. *Plosiefinventaris 2.*

	<i>labiaal</i>	<i>coronaal</i>	<i>dorsaal</i>
<i>stemloos</i>	p	t	k
<i>stemhebbend</i>			

Dit is ook overzichtelijk: er zijn alleen maar stemloze plosieven, geen stemhebbende (daarom is dat woord grijs gemaakt). Maar neem nu een taal als het Efik, gesproken in Nigeria:

Tabel 3. *Plosiefinventaris 3.*

	<i>labiaal</i>	<i>coronaal</i>	<i>dorsaal</i>
<i>stemloos</i>		t	k
<i>stemhebbend</i>	b	d	

Dit is al een stuk ingewikkelder: deze taal heeft allebei de coronalen, de stemloze dorsaal, en de stemhebbende labiaal. De notie “ingewikkeld”, of zoals we liever zeggen: “complex”, kunnen we op verschillende manieren preciezer maken: bijvoorbeeld door te kijken hoeveel fonemen, ofwel combinaties van kenmerken, een taal gebruikt, en dat aantal te delen door het aantal combinaties dat *in principe* mogelijk is. Dat getal heet KENMERKZUINIGHEID. In deze laatste taal, bijvoorbeeld, worden alle waarden die de kenmerken kunnen aannemen gebruikt (er zijn stemloze klanken, stemhebbende, labiale, coronale, én dorsale): er zijn dus $3 \times 2 = 6$ mogelijke combinaties. Daarvan worden er maar 4 gebruikt, dus de kenmerkzuinigheid is $4/6 = 2/3$. Inventarissen 1 en 2 zijn maximaal zuinig, want ze gebruiken alle mogelijke combinaties, dus hun kenmerkzuinigheid is 1.

Een andere mogelijkheid om complexiteit te definiëren is om het aantal kenmerken te tellen dat je nodig hebt om een inventaris te beschrijven: voor inventaris 2 is dat er maar eentje (“stemloos”), voor inventaris 3 zijn het er wel vijf (“coronaal”, “stemloos en dorsaal”, “stemhebbend en labiaal”). Deze definitie heet LOGISCHE COMPLEXITEIT. In het algemeen kun je zeggen dat een taal zonder gaten in het systeem, dus een taal waarin geen mogelijke combinaties van kenmerken ontbreken, minimaal complex zijn; zulke inventarissen heten ook wel “regelmatig”. Inventarissen 1 en 2 zijn voorbeelden van zulke inventarissen. In het algemeen geldt dat hoe meer gaten er zijn, hoe hoger de complexiteit is, zoals in Tabel 3.

Uit eerder onderzoek weten we dat mensen meer moeite hebben om complexere systemen te leren: ze hebben bijvoorbeeld meer moeite om te beslissen of een combinatie van kenmerken in de inventaris zit, of ze hebben meer moeite om de regelmaat in de inventaris te beschrijven. Ik heb experimenten gedaan waarin mensen inventarissen van verschillende complexiteit moesten leren, en gekeken hoe goed ze dat deden. Aan mijn onderzoek deden alleen volwassenen mee, die in hun moedertaal of -talen al een heel systeem van kenmerken hadden geleerd; om te

zorgen dat ze niet op die kennis konden leunen moest ik dus iets anders dan spraak gebruiken, en op aanraden van collega's heb ik daarvoor een set heel simpele gebaren gebruikt, die dus andere kenmerken gebruikte dan stemhebbendheid en plaats van articulatie, maar wel op dezelfde manier kon worden gerepresenteerd (een foto van alle handvormen in de taal is te zien op pagina 106).

De resultaten van mijn experimenten laten zien dat complexiteit doorgaans inderdaad goed voorspelt hoe mensen het doen (logische complexiteit weliswaar nog iets beter dan kenmerkzuinigheid): hoe hoger de complexiteit, hoe slechter het resultaat. En als mensen de complexiteit blijkbaar te hoog vinden, doen ze er iets aan, door de inventaris dusdanig aan te passen dat hij minder complex wordt: meestal voegen ze een of meerdere fonemen toe die niet in hun input zaten, en vaak resulteert dat in een regelmatig systeem. Een leerder van plosiefinventaris 3, bijvoorbeeld, zal dus geneigd zijn om een [p] toe te voegen, of een [g], of allebei; en hij of zij zal zich daar niet eens bewust van zijn.

Het is mooi dat mensen in een experiment hun input minder complex maken, maar bij echte taalverwerving komt natuurlijk nog veel meer kijken: dat duurt veel langer, is interactief, enzovoort — en, heel belangrijk: in foneeminventarissen zijn het prototype-effect en het articulatorisch effect ook relevant, waar ze dat niet echt waren in de experimenten. Van de drie paren stemloze en stemhebbende klanken in Tabel 1, bijvoorbeeld, lijken de [k] en de [g] in zekere zin meer op elkaar dan de andere twee paren, dus verwarren luisteraars ze sneller met elkaar en moeten sprekers relatief veel moeite doen om het verschil duidelijk te maken: dat zijn geen nuttige eigenschappen, en het is ongetwijfeld de reden dat het Nederlands geen [g] heeft (alleen in een paar leenwoorden, waaronder dus *goal*). Daarom heb ik ook gekeken naar klankveranderingen in echte talen (hoofdstuk 6), om te kijken of die het systeem minder complex maakten, en naar plosiefinventarissen van echte talen (hoofdstuk 7), om te kijken hoe complex die zijn. En zo overtuigend als mijn proefpersonen in het lab de complexiteit van hun input reduceerden, zo vaak vinden we uitzonderingen in natuurlijke talen: klankveranderingen vullen lang niet altijd gaten in het systeem op, en hoewel een krappe meerderheid van de 317 plosiefinventarissen waar ik naar heb gekeken regelmatig is, zijn er ook genoeg inventarissen vol gaten. Dat is niet per se wat we misschien zouden verwachten, maar het is wel heel interessant: het laat namelijk zien wat mensen blijkbaar toch kunnen leren, ondanks dat het niet ideaal is voor ze, en het laat ook zien hoe verschillende factoren (zoals leerbaarheid, perceptie en articulatie) in elkaar grijpen.