

Hoofdstuk 6: Hoe werkt ChatGPT zelf

Ik

In het voorafgaande is ChatGPT naar voren gekomen als taalexpert, om vragen te beantwoorden als wat zijn de verschillen tussen grammatica's van verschillende talen en wat zijn de verschillen in woordvolgorde bij verschillende talen. Dat kan omdat ChatGPT heel veel weet over talen en het maken van vertalingen.

Nu kom ik tot de volgende vraag. Ik ga ervan uit dat ChatGPT uitgaat van een zekere oertaal. Is dit zo?

ChatGPT

Interessante vraag! ChatGPT gaat niet expliciet uit van een *oertaal* in de taalkundige zin van het woord, zoals bijvoorbeeld het Proto-Indo-Europees, dat als voorouder van veel Europese en Aziatische talen wordt gezien.

Wat ChatGPT wél doet, is werken met enorme hoeveelheden tekst in verschillende talen, waarbij patronen en structuren in taal worden geleerd. Hierdoor kan het verschillen in grammatica, woordvolgorde en syntaxis tussen talen herkennen en beschrijven. Het model zelf heeft geen ingebouwde oertaal van waaruit alle vertalingen of grammaticale analyses worden afgeleid.

Wat gebeurt er dan onder de motorkap?

- ChatGPT maakt gebruik van statistische patronen en neurale netwerken om verbanden tussen woorden en zinnen te begrijpen.
- Het vergelijkt automatisch de structuren van verschillende talen, maar zonder dat er een 'centrale basistaal' als uitgangspunt wordt genomen.
- Wel is het zo dat Engels vaak een prominente rol speelt, omdat veel AI-trainingen gebaseerd zijn op grote Engelstalige datasets.

Als je uitgaat van een hypothetische *oertaal*, zou je kunnen zeggen dat ChatGPT functioneert op basis van een conceptuele kern van taalkundige structuren, eerder dan een echte oertaal. Het herkent universele grammaticale eigenschappen en kan verbanden leggen tussen verwante en niet-verwante talen, maar zonder een vast referentiepunt als 'oertaal'.

Ik

Als er geen sprake is van een oertaal die gebruikt wordt om verschillende talen te vergelijken, hoe komt ChatGPT dan tot vergelijkingen?

ChatGPT

Een menselijke taalkundige onderzoekt taal door regels, structuren en historische ontwikkelingen te analyseren. Taalkundigen werken met primaire bronnen (zoals oude teksten

en gesproken taal), ontwikkelen theorieën over taalverwantschap en gebruiken logische redeneringen om bijvoorbeeld de ontwikkeling van grammaticale structuren te verklaren.

ChatGPT en andere AI-systemen daarentegen werken heel anders:

- Ze **leren** taal op basis van enorme hoeveelheden tekst, zonder expliciet geprogrammeerde grammaticaregels.
- Ze **voorspellen** welke woorden en zinnen waarschijnlijk correct zijn op basis van patronen uit de trainingsdata.
- Ze **maken geen bewuste analyses** zoals een menselijke taalkundige, maar werken puur op basis van kansberekening en patroonherkenning.

Een menselijke taalkundige kan proberen een *oertaal* te reconstrueren door verwante talen te vergelijken. ChatGPT doet dat niet. In plaats van uit te gaan van een enkele oertaal, werkt het met statistische verbanden tussen duizenden talen.

Toch zou je kunnen zeggen dat ChatGPT indirect een soort *statistische kern van taal* heeft, omdat het universele grammaticale structuren herkent en verbanden legt tussen talen zonder te weten dat die talen verwant zijn.

Voorbeeld:

- In veel talen staat het werkwoord op de tweede plaats (zoals in Nederlands en Duits).
- In andere talen, zoals Japans, komt het werkwoord altijd aan het einde van de zin.
- ChatGPT weet dit niet als ‘taalkundige regel’, maar *ervaart* het via patronen in de teksten waarop het is getraind.

Omdat ChatGPT niet zoals een mens logisch redeneert over taal, werkt het anders bij het vergelijken van grammatica:

Aspect	Menselijke taalkundige	ChatGPT / AI
Analyse-methode	Gebaseerd op historische en structurele analyse	Gebaseerd op patroonherkenning
Taalontwikkeling	Probeert evolutie te verklaren	Ziet alleen verbanden tussen bestaande teksten
Woordvolgorde	Kijkt naar taalfamilies en typologie	Herkent veelvoorkomende structuren
Betekenis van woorden	Plaatst woorden in culturele context	Leert betekenissen door veelvoorkomend gebruik
Vergelijken van talen	Gebruikt systematische regels	Berekent statistische gelijkenissen

Ik

Kan AI ooit een echte taalkundige vervangen?

ChatGPT

ChatGPT kan zeer snel grammaticale structuren herkennen en vergelijken, maar:

- Het **begrijpt** taal niet op de manier waarop een mens dat doet.
- Het kan **geen nieuwe taalkundige theorieën** ontwikkelen of diepgaand nadenken over taalgeschiedenis.
- Het kan **geen onbekende talen reconstrueren**, zoals een menselijke taalkundige dat kan.

Toch is AI een waardevolle aanvulling op traditioneel taalonderzoek. Bijvoorbeeld:

- AI kan snel patronen opsporen in grote datasets van talen.
- AI helpt bij automatische vertalingen en taalherkenning.
- AI kan taalkundigen ondersteunen bij het analyseren van zeldzame of verdwijnende talen.

Conclusie: AI als taalexpert

ChatGPT is een krachtige taaltool en kan verschillen tussen talen beschrijven en vertalingen maken. Maar een menselijke taalkundige denkt anders: die zoekt naar verklaringen en historische verbanden. AI werkt daarentegen puur op basis van patroonherkenning en kansberekening.

In de toekomst zou AI wellicht nieuwe inzichten kunnen bieden in taalstructuren, maar vooralsnog is het vooral een hulpmiddel – een taalexpert zonder een diepgaand begrip van taal zelf.

Ik

Ik snap wat je zegt over statische koppelingen in de zin dat woorden en reeksen woorden voorspellingen kunnen doen over het volgende woord. Maar dat geeft geen antwoord op de vraag waarom de modellen waar ChatGPT gebruik van maakt, zouden kunnen redeneren. Dan zou ik eerder denken aan een soort mindmapping. Klopt dat?

ChatGPT

Ja, je zit op het juiste spoor! Het idee van *mindmapping* als metafoor voor hoe ChatGPT redeneert, komt dichterbij dan pure statistische voorspelling. Hoewel ChatGPT niet echt *denkt* zoals een mens, maakt het wel verbindingen tussen concepten, net zoals een mindmap dat doet. Dit komt door de manier waarop neurale netwerken en *vectorruimtes* werken.

Mindmapping en hoe ChatGPT "redeneert"

Een traditionele mindmap ordent informatie door concepten te verbinden. Bijvoorbeeld:

- Je schrijft het woord "*taal*" in het midden en trekt lijnen naar gerelateerde begrippen, zoals "*grammatica*", "*spraak*", "*schrijven*", "*woorden*", etc.
- Binnen grammatica maak je nieuwe verbindingen: "*woordvolgorde*", "*naamvallen*", "*zinsstructuur*", enz.