



Een vertaalcomputer met een zes op zijn rapport

De tv-serie *Star Trek* was achteraf een inspiratiebron voor de huidige Tilburgse hoogleraar Antal van den Bosch (38). Toen hij als kind die serie zag, waarin een “universele vertaler” voorkwam, die alles kon vertalen van iedere taal naar de andere, wist hij nog niet dat hij later zou gaan meehelpen werkelijkheid te maken van die sciencefiction.

De droom zelf is al ouder. Vrijwel zo lang als er computers bestaan, proberen mensen een vertaalmachine te maken. Op het hoogtepunt van de Koude Oorlog werkten Amerikaanse geleerden en ingenieurs aan een computer die snel van het Russisch naar het Engels zou kunnen vertalen. Volgens een anekdote voerde men aan een prototype eens de bijbeltekst “De geest is wel gewillig, maar het vlees is zwak” (Matteüs 26:41). Men liet die zin naar het Russisch vertalen, en die vertaling vervolgens ook weer door de computer omzetten naar het Engels. Het resultaat: “The wodka is good, but the beef is rotten” (‘De wodka is goed, maar het rundvlees is bedorven’).

■ GELD VERDIENEN

Inmiddels zijn vertaalcomputers wel iets beter geworden, maar nog steeds is een machine niet in staat om een doorsneetekst zó van de ene taal naar de andere over te zetten dat je die vertaling ook kunt begrijpen zonder het origineel te raadplegen. “De beste algemene vertaalcomputers halen tegenwoordig een vier of een vijf op hun rapport”, zegt Van den Bosch. “Er zijn er wel een paar die een acht halen, maar die vertalen dan alleen heel specifieke teksten, zoals weerberichten. Binnen een paar jaar moet dat een vijf of een zes worden. Een menselijke vertaler zal er dan altijd nog naar moeten kijken als je de tekst wilt gebruiken, maar die hoeft dan in ieder geval een heleboel routinewerk niet meer te doen.”

Van den Bosch heeft er vertrouwen in, omdat er de laatste jaren nieuwe technieken worden beproefd. “De zoekmachine Google heeft sinds kort een dienst waarmee je webpagina’s onder andere van het Nederlands naar het Engels kunt vertalen en omgekeerd. Die werkt met een nieuwe technologie die ik ook aan het uitproberen ben.” Anders dan voor het commerciële Google is het voor Van den Bosch niet per se de bedoeling om geld te gaan verdienen met

Foto: Ben Bergmans



Antal van den Bosch: “Mijn uitgangspunt is om een werkend computermodel te maken.”

zijn vertaalcomputer. De geleerde wil vooral achterhalen hoe een dergelijk systeem in elkaar kan zitten.

■ MENEER DE VOORZITTER

Hoe werken die nieuwe technieken? “De grondslag wordt gevormd door grote databases met vertalingen die door mensen gemaakt zijn. Google maakt daarvoor gebruik van alle internetpagina’s waar ergens ‘klik hier voor een vertaling’ staat. Wij in Tilburg gebruiken een grote database met verslagen van vergaderingen in het Europees Parlement die in allerlei talen vertaald zijn. Als de computer nu een nieuwe zin moet omzetten, zoekt hij in de database alle zinnen op die een beetje op de nieuwe zin lijken. Dan probeert hij de vertalingen van die zinnen zó in elkaar te passen dat er een nieuwe vertaling ontstaat.”

De techniek is nieuw omdat er tien, twintig jaar geleden nog geen mogelijkheden voor waren. Zulke grote databases met vertalingen bestonden nog niet, en computers zouden ook niet het vermogen hebben gehad om er snel genoeg in te zoeken. Er zijn ook wel pro-

blemen die specifiek zijn voor deze methode. Van den Bosch: “Omdat onze computer zich baseert op verslagen van parlementaire vergaderingen, weet hij niet zo goed raad met gewone alledaagse taal. Hij laat een stukje tekst nu nog regelmatig beginnen met de woorden *Meneer de voorzitter*, omdat hij denkt dat dit zo hoort.” Door de technologie te verfijnen en de databases groter te maken, worden dit soort kwesties hopelijk opgelost.

■ CREATIVITEIT

Van den Bosch hoopt ook andere toepassingen te kunnen maken met dezelfde technieken. Voor een verbeterde spellingcontrole die bijvoorbeeld problemen met de spelling van werkwoorden – *bestuurd* of *bestuurt*? – aankan, kun je ook kijken hoe het woord in vergelijkbare zinnen meestal gebruikt wordt. En uiteindelijk denkt Van den Bosch dat de techniek ook kan worden toegepast in vraag-en-antwoordsystemen. Iemand belt naar een informatielijn om bijvoorbeeld inlichtingen te krijgen over een vliegreis, theaterkaartjes of het weer, de computer zoekt de vraag op en formuleert het antwoord dat op soortgelijke vragen gegeven werd.

Is dit alles uitsluitend een handigheidje voor de computer, of zegt Van den Bosch’ onderzoek ook iets over de manier waarop wij mensen met taal omgaan? Zouden wij ook bij ieder gesprek het voorafgaande in ons hoofd razendsnel vergelijken met alle eerdere gesprekken die we hebben gehoord, om op basis daarvan uit te rekenen wat nu het gebruikelijkste antwoord is? Hoeveel ruimte laat dat voor creativiteit? Van den Bosch: “Mijn uitgangspunt is om een werkend computermodel te maken. Maar ik lees de laatste tijd boeken en artikelen van psychologen en taalkundigen die inderdaad op dat soort gedachten gekomen zijn. Mensen hebben wel het gevoel dat we telkens nieuwe dingen bedenken, maar misschien is dat wel een illusie. We hebben er eigenlijk nog steeds geen idee van hoe wij komen tot de zinnen die we zeggen.” ■