

De vertaalmachine

Battus

Battus was een van de minstens 60 schuilnamen van Hugo Brandt Corstius (1935-2014), die wiskunde en informatica studeerde en promoveerde op een proefschrift over computertaalkunde. Hij was hoogleraar automatische informatieverwerking aan de Erasmus Universiteit in Rotterdam en docent semantiek en computerlinguïstiek aan de Universiteit van Amsterdam. Na zijn emeritaat doceerde hij nog neerlandistiek aan de Sorbonne. Onder zijn vele pseudoniemen (zoals Piet Grijs, Stoker, Raoul Chapkis, Victor Baarn, Jan Eter, Maaïke Helder) schreef hij columns en kritieken (onder meer in Vrij Nederland, de Volkskrant en NRC Handelsblad). Naar eigen zeggen waren deze schuilnamen elk een deel van zijn karakter. Hij begon zijn schrijversloopbaan als redacteur van het studentenblad Propria Cures. Hij werd bekroond met de Anne Frank-prijs (Raoul Chapkis, Ik sta op mijn hoofd), Cestoda-prijs (voor de moeiteloze beoefening van de Nederlandse taal in al haar genres), Burgemeester van Grunsven-prijs, Busken Huetprijs (Rekenen op taal). Toen hij in 1984 voor de P.C. Hooft-prijs werd voorgedragen, weigerde de verantwoordelijke minister hem deze uit te reiken omdat hij 'het kwetsen tot instrument had verheven'. Uit protest trad de jury af en werd de prijs twee jaar niet uitgereikt. Toen de Staatsprijs daarna tot 'gewone' prijs werd gedegradeerd, kreeg hij de vernieuwde P.C. Hooft-prijs in 1988 alsnog uitgereikt. Als Battus schreef hij onder meer zijn befaamde Opperlandse taal- & letterkunde (Multatuli-prijs), twintig jaar later nog gevolgd door Opperlans!. 'De vertaalmachine' verscheen oorspronkelijk in 1973 in Hollands Maandblad. VertaalVerhaal dankt Ina Rilke voor de toestemming het hier te mogen publiceren.

De vertaalmachine

Een jaar of zes geleden stuurde een grote encyclopedieënuutgever me een formulier om het artikel te schrijven bij het trefwoord 'vertaalmachine'. Ik retourneerde dat met de twee woorden 'bestaat niet'.

Het niet-bestaan is de voornaamste eigenschap van de vertaalmachine. Net als bij discussies over God mag die belangrijke eigenschap nooit uit het oog worden verloren.

Toch waren er jaren dat in allerlei kranten en tijdschriften over vertaalmachines werd geschreven alsof ze er wel waren. De publieke opinie is er dan ook vast van overtuigd dat ze bestaan. Wel is de aandacht de laatste jaren wat afgenomen. In een encyclopedie die thans wordt voorbereid komt het trefwoord 'vertaalmachine' niet voor. Hoe zit het met die vertaalmachine? Is het ding even onmogelijk als het perpetuum mobile? Wordt er nog ergens serieus aan gewerkt? Zijn de mensen, die er tussen 1950 en 1965 in grote aantallen aan werkten, bedriegers of bedrogenen? Wat voor conclusie valt er uit het mislukken van de zoek naar de vertaalmachine te trekken over de verdeling van geld voor wetenschappelijk onderzoek? Die vragen wil ik beantwoorden.¹⁾

Historie

De geschiedenis van de vertaalmachine moet nog geschreven worden. Het zal moeilijk zijn er iemand voor te vinden. Want wie er wat van af weet is meestal gecompromitteerd. In 1964 verscheen van George Mounin *La machine à traduire*. Hoewel toen het failliet al vast stond zwijgt hij daarover als het graf. Vorig jaar nog verscheen van Paul Garvin een collectie oude stukken²⁾, waaruit blijkt dat hij de hoop niet heeft opgegeven. Toch is die geschiedenis best te overzien: er zijn ongeveer 150 namen van personen, het geheel speelt zich af in de genoemde vijftien jaar, er waren op het hoogtepunt 30 centra van onderzoek, de totale literatuur kan op drie meter plank.

De gezochte vertaalmachine is niet een ding, zoals een was- of schrijfmachine. Bedoeld wordt een programma waarmee een computer (alle computers zijn gelijk, net als mensen) in staat wordt gesteld om een willekeurige (dat wil zeggen bij de constructie van het programma nog onbekende) tekst uit de ene taal in de andere te vertalen. Daarbij mag de te vertalen tekst niet tevoren door een mens zijn behandeld, terwijl ook het eindproduct zonder menselijk ingrijpen bruikbaar moet zijn. Dat ligt wel erg voor de hand, maar het is juist op dit punt dat het bedrog hoog oploopt. Vele bekende 'vertaalmachines' lijken op de beroemde schaakmachine die in de 18e eeuw van hof naar hof werd gedragen, en die vele spelletjes won tot men ontdekte dat er een kleine schaker in opgeborgen zat.

Men zou de geschiedenis van de vertaalmachine kunnen schrijven aan de hand van historische voorbeeldzinnen met hun al dan niet bedrieglijke computervertaling. De eerste zin uit die geschiedenis staat in een boek uit 1958³): ‘Disease comes thus very rapidly up and has in many cases a total amiss crop then follow.’

Uit welke taal denkt u dat dit vertaald is? Het is volgens Andrew Booth een woord-voor-woord-vertaling die met de hand is uitgevoerd maar op geheel formele wijze, dus zoals een computer met een woordenboek het zou doen: zoek het woord op en vervang het door de eerste vertaling. De vertaling is, raadde u het? uit het Nederlands. Maar in geen Ned.-Engels woordenboek zal als vertaling van *ten then* staan, en hoe *misoogst* in twee delen werd gehakt is ook niet duidelijk. Bedrog is hier een te groot woord; Booth bedroog zichzelf als hij denkt echt formeel uit het Nederlands vertaald te hebben. Maar hij, en vele anderen, bedrogen zich nog meer wanneer ze meenden dat zulke betrekkelijk begrijpelijke vertalingen hoop opleverden voor het vinden van de vertaalmachine.

De voorlopig laatste zin uit die historie zou luiden: ‘It may be obtained after bombarding the element with high-energy particles.’

Dit is de Engelse vertaling van een Chinese zin die is te vinden in de laatste aflevering van de *Scientific American*. William Wang schrijft daar een zeer lezenswaard artikel⁴) over het Chinees. Wang heeft zich met computervertaling van Chinees naar Engels bezig gehouden, maar gebruikt, terecht, zijn computer nu voor dialectstudies. Toch kan hij het niet laten in het onderschrift bij de analyse van de boven genoemde Chinese zin de indruk te wekken alsof de computer voor de Engelse vertaling zorgt. Geloof er niets van.

In die historie aan de hand van zinnen zou dan ook het apocriefe verhaal niet mogen ontbreken dat een computer de tekst Marcus 14:38 vertaalde tot: ‘De whisky is wel smakelijk, maar de biefstuk is slap.’

Zo'n geschiedschrijving zou de zaak onrecht doen. In het begin was er eerlijk enthousiasme. Over dat eerste begin is een onsmakelijke prioriteitsstrijd ontbrand, die des te gekker is omdat men elkaar de prioriteit op een slecht idee betwist. Vast staat dat de Amerikaan Warren Weaver en de Engelsman Booth er in 1946 in New York over gesproken hebben. Zij eisen dus ieder de prioriteit op. Later kwamen de Russen met de onvermijdelijke eerdere uitvinder: ene P.P. Smirnov-Trojanskij die inderdaad in 1933 zoiets als computervertaling voorstelde (de eerste computer zag in 1943 het daglicht) zonder dat iemand er op lette⁵). De bovengenoemde Franse geschiedenis komt met René Descartes aan die in 1629 al een cijfertaal voorstelde, maar dan kun je ook wel op Ramon Lull teruggaan die omstreeks 1300 dergelijke ideeën had. Het was in ieder geval een rondschrijven van Weaver in 1949 dat het idee verspreidde. Het sloeg aan vanwege deze redenen:

Vijf oorzaken

1. De woord-voor-woord-vertaling is bij verwante talen (en Engels en Russisch waar het in die eerste jaren steeds om ging, zijn allebei Indo-Europees) al zo goed dat het gevoel ontstond: nog even en we zijn er.
2. De computer leek een wonder. Men schreef hem in die dagen ook het vermogen toe de wereldvrede te vestigen. Op het ogenblik is het meer mode hem domme fouten aan te wrijven.
3. In Amerika leefde het angstige gevoel (na de Sputnik helemaal) dat de Russen ze gingen inhalen. Ze wilden, dat was een serieus voorstel, alles wat in Rusland gedrukt werd, direct vertaald hebben. Daarvoor leek mechanische vertaling de enige oplossing.
4. Na 1957 kwam er een stroming in de taalkunde die de grammatica formeler maakte. Hoewel de leider, Noam Chomsky, nooit iets in vertaalmachines heeft gezien, ja zelfs ten onrechte nooit iets in het nut van computers heeft gezien, hebben ongetwijfeld veel van zijn volgelingen onder het mom van automatische vertaling geld gekregen om onderzoek te kunnen doen. In de Sovjet-Unie was de enige manier om serieuze taalkunde te doen de smoes van de vertaalmachine. Uit dat land bereiken ons dan ook berichten over computers die in en uit het Birmees, Hindoestaans, Noors, Indonesisch, Georgisch, Armeens, Tsjechisch en Hongaars vertalen. Andere in de loop der tijd behandelde talen zijn Frans, Engels, Duits, Russisch, Arabisch, Japans, Vietnamees, Spaans, Chinees, Hebreeuws. Geen Nederlands.
5. Het belang voor de computermaatschappijen. Het eerste, goed geadverteerde, publieke experiment (een doorgestoken kaart) vond onder de hoede van I.B.M. en Leon Dostert in 1954 plaats.

Het hoogste punt wordt bereikt in 1962. De *Times Literary Supplement* publiceert een serie artikelen⁶⁾ over de mogelijkheid van mechanisch vertalen. Maar in Venetië vertelt Yehoshua Bar-Hillel haarfijn waarom het allemaal een utopie is⁷⁾. Wie daarna toch nog aan mechanisch vertalen doet moet dus eerst zeggen waarom hij Bar-Hillel niet gelooft.

Zijn argumenten komen in het kort hier op neer dat voor formeel vertalen behalve een woordenboek ook een encyclopedie nodig is, en wel een encyclopedie waarin alle kennis over de wereld te vinden is. In de zin: 'Ik kocht een piano en een viool die ik onder de arm mee naar huis nam' moet je bij vertaling in bepaalde talen weten of dat die op piano en viool, of alleen op viool slaat. Bij elk substantief moet dus vermeld zijn of het onder-de-arm-mee-naar-huis-draagbaar is.

Misschien vindt u, als u ooit iets vertaalde, het argument voor de hand liggend. Maar vanzelfsprekend is het niet. Voor vertaling tussen dialecten van één taal heb je die encyclopedische kennis niet nodig. Bar-Hillel was niet zomaar een filosoof die iets beweerde: hij was in 1951 en 1952 de eerste fulltime werker op vertaalmachinegebied. Er breekt nu een periode aan waar het bedrog de overhand heeft. Zo beweert Silvio Ceccato op dezelfde Venetiaanse bijeenkomst: 'IBM heeft een machine die opmerkelijk goede resultaten oplevert. Het enige dat nu nog de productie van vertalingen vertraagt is het ontbreken van een machine die automatisch letters kan lezen.'⁸⁾ Dat is gewoon gelogen. Edmundson zegt in datzelfde jaar: 'We zullen een

versnelling in mechanische vertaling zien. Terwijl de afgelopen 10 jaar tamelijk weinig is gebeurd, voorspel ik dat de komende 10 jaar doorslaggevend succes te zien zullen geven⁹⁾. Het heeft niet zo mogen zijn.

Het einde

Tot 1966 suddert het door. Dan wordt op typisch Amerikaanse wijze een hardvochtig einde aan het avontuur gemaakt. Een commissie stelt aan de Academie van wetenschappen voor om geen geld meer ter beschikking te stellen voor de vertaalmachine.¹⁰⁾ ‘There has been no Mechanical Translation of general scientific text, and none is in immediate prospect’. Het rapport schat dat in tien jaar door de overheid 20 miljoen dollar aan de vertaalmachine is uitgegeven. Voor de hele wereld en de hele periode schat ik het uitgegeven geld op 60 miljoen dollar. Is dat geld weggegooid? Ik heb een lijstje gemaakt van nuttige bijeffecten van het onderzoek. Het lijkt erg op soortgelijke lijstjes wanneer je vraagt waar het Apollo-project eigenlijk goed voor was. Overigens is 60 miljoen dollar daarbij vergeleken niets. Het belangrijkste algemene nut is dat taalkundigen gelegenheid kregen zich met hun vak bezig te houden, en dat de computer voor niet-wiskundige toepassingen toegankelijk werd.

Wat is er in 1973 nog over? In Amerika is nog een universiteit waar men aan Duitse vertaling werkt (merkwaardig genoeg was de eerste taalkundige publicatie uit 1951 ook Duits); en er zijn een aantal militaire centra. Die zijn natuurlijk geheim, maar uit een artikel van Wallace Sinaiko¹¹⁾ blijkt dat de door de Luchtmacht gebruikte machine, die Engels in Vietnamees vertaalt, onbruikbare producten aflevert. In Rusland is volgens een mededeling van Mel’cuk aan Frits Droste in 1968 ‘de laatste tien jaar, bij zijn weten, nergens in Rusland feitelijk met de computer vertaald’.¹²⁾ Dit is notabene dezelfde Mel’cuk die in een in 1963 in het Engels vertaald boek¹³⁾ schreef: ‘De door deze experimenten verzamelde ervaring stelt ons in staat serieus over te gaan tot massale mechanische vertaling.’ Ik moet nu denken aan een aardige Rus die me in 1966 in Moskou een computerprogramma liet zien dat, getuige de output, een interessant probleem had opgelost. Ik wees op een foutje, hij verbeterde het, en ineens begreep ik dat er nooit een computer aan te pas was gekomen. Dat was geen bedrog, maar de enige manier waarop deze man kon werken.

In Frankrijk waar het wat later begon is de vertaalmachine het langst gebleven. Emile Delavenay, auteur van een *Que sais-je* uit 1959 over de vertaalmachine¹⁴⁾, beweert in 1969 in het orgaan van de ATALA¹⁵⁾, waarin de T dan niet meer voor *translation* maar voor *traitement* staat: ‘Het perfectionisme dat een machine wil maken die alle theoretisch mogelijke zinnen aan kan is de *ennemi du bien*... Wij moeten ons beperken tot de waarschijnlijkste zinnen.’ Daarentegen schrijft zijn landgenoot Maurice Gross in december 1972: ‘Het mechanisch vertalen misbruikt zichzelf, misbruikt de taalkundigen door hun modieuze terminologie... en misbruikt ten slotte het publiek en de geldverschaffende instanties door de belofte alle

problemen op te lossen¹⁶). In Italië, ten slotte, draait in Ispra nog steeds een vertaalmachine. Hij produceert Engelse vertalingen van Russische artikelen, waaraan, aldus werd me verteld, je kon zien waar ze over gaan.

Misverstanden

Bepaalde ideeën zie je steeds weer opduiken. Zo de hoop dat de statistiek je kan helpen. Het blijkt van niet. Hardnekkig is het idee van de tussentaal, het *machinees*. In plaats van om tien talen in elkaar te kunnen vertalen 90 programma's te maken (van iedere taal naar elk der 9 andere), zou men dan kunnen volstaan met 20 vertaalprogramma's (van iedere taal naar het machinees, en van het machinees naar iedere taal). Hiertegen kan ingebracht worden dat wanneer je één van de tien talen die scharnierfunctie geeft, zelfs 18 programma's genoeg zijn. Maar ernstiger is dat je om b.v. van het Russisch naar het Chinees te vertalen ook alle onderscheidingen moet maken waarmee je bij een vertaling naar het Engels de juiste lidwoorden kan plaatsen. Het Russisch kent geen lidwoorden; bij de Engelse vertaling moet je kiezen tussen *the*, *a* of niets. 20 jaar onderzoek heeft nog geen begin van een oplossing van dit probleem gebracht. Het hele idee is natuurlijk afkomstig uit de metafysische taalopvatting waar ik in het decembernummer mijn hart over luchtte. Als werkbare tactiek beperkt men zich toch steeds tot één talenpaar.

Moraal

Hoe is het mogelijk dat geldgevende instanties lange tijd dit onderzoek subsidieerden? Een paar redenen heb ik al gegeven. Dat Weaver hoog bij de Rockefeller Foundation was hielp natuurlijk. Voor de taalkundigen was het een uitkomst: hun vak had ineens nut gekregen! Maar tegelijkertijd was er ook een operationele definitie van succes: die machine moest gaan werken en goede vertalingen afleveren. Dat deed hij niet. Alle genoemde redenen zouden in de Nederlandse situatie, waar taalkundige research alleen aan de universiteiten plaats vindt, ook opgaan. Maar in de Amerikaanse situatie zat en zit een belangrijk verschil. Daar moet je om geld voor onderzoek te krijgen ambtenaren overtuigen dat je op de goede weg bent. Dan worden wetenschappers, zo blijkt uit het avontuur van de vertaalmachine, gedwongen om de grens tussen enthousiasme en bedrog te overschrijden. De vele voorstellen die op het ogenblik de ronde doen om het wetenschappelijk onderzoek los te maken uit de universiteiten kunnen maar op twee dingen uitdraaien:

- 1) òf de fundamentele niet-technische research verdwijnt helemaal. Ik zou dat betreuren maar als het parlement het geld liever aan iets anders besteedt dan moet het.

2) òf de wetenschapper wordt een middenstander (de term is van Professor Tennekes, Nederlander in Pennsylvania¹⁷) die zichzelf moet verkopen. De gevolgen zijn in Amerika waar te nemen: overbodige artikelen die het geld voor de voortzetting van het werk moeten losmaken. Ik zou niet willen beweren dat in de Nederlandse wetenschap opschepperij en bedrog helemaal niet voorkomen, maar het zijn nu nog betreurenswaardige karakterfouten, niet onvermijdelijkheden van het systeem.

Hoe het dan wel moet? Alleen achteraf is vast te stellen of iemand goed werk heeft geleverd. Geef dus wie een doctoraalexamen heeft gedaan en dat wil, een krediet van drie jaar. Blijkt na drie jaar dat er niet genoeg is uitgekomen dan moet hij wat anders gaan doen. Beter is het zulke kredieten en straffen per instituut te geven. Zulke instituten kunnen beter aan universiteiten zijn verbonden zodat ze ook voor onderwijs moeten zorgen, dan aan de centra voor uitzonderlijk hoogstaand onderzoek waar de ministers aan denken, zonder ons te vertellen hoe ze zullen vermijden dat daar 20 jaar lang naar een vertaalmachine wordt gezocht.

Onmogelijk?

De voorspelling dat er nooit een vertaalmachine zal zijn is zinloos. Er bestaat ook geen perfecte menselijke vertaling. De onmogelijkheid van mechanisch vertalen volgt niet uit bepaalde natuurwetten zoals de onmogelijkheid van het perpetuum mobile. Zijn komst is alleen verbonden aan die van andere droommachines, zoals de verstamachine (automatische schrijfmachine die je toe kan spreken) en de *information retrieval* (de computer die al je vragen over de inhoud van een bibliotheek beantwoordt), waarvan de historie een opvallende parallellie met de onze vertoont. Wij weten niet hoe we onze ervaringen zo opslaan dat we er later uit kunnen putten. We weten ook niet hoe we dat een computer moeten leren.

¹) Enkele oneffenheden in dit stuk zijn veroorzaakt doordat het met een speciaal programma uit het Nederlands in het machinees, en met een ander programma uit het machinees in het Nederlands werd vertaald. Binnen enkele maanden zijn beide programma's volmaakt.

²) t/m ¹⁷)

Heus, al die aanhalingen zijn echt. Wilt u de vindplaats weten, vraag me er dan naar.