

WAT IS EN DOET EEN VERTAALMACHINE?

door

Ir. IJ. Boxma

WAT IS EN DOET EEN VERTAALMACHINE?

door

Ir. IJ. Boxma

1. *Inleiding*

De omstandigheid dat het onmogelijk is alle wetenschappelijke literatuur, die in vele talen verschijnt, te lezen werkt sterk belemmerend op de vooruitgang van de wetenschap. Dit is natuurlijk vooral duidelijk voor Russische, Japanse en andere niet-West-europese talen. Maar in Frankrijk, Engeland en Amerika is ook niet altijd de literatuur in het Engels, Frans of Duits direct toegankelijk.

Het is vrijwel ondoenlijk alles wat daarvoor in aanmerking komt te vertalen. Men weet bovendien dikwijls pas of een artikel voor vertaling in aanmerking komt nadat deze is uitgevoerd, zodat men er meer en meer toe overgaat tijdschriften volledig te vertalen en opnieuw uit te geven. Er verschijnen tegenwoordig enorme aantallen tijdschriften (ongeveer 50.000), zodat er eenvoudig geen vertalers genoeg beschikbaar zijn. Deze moeten bovendien naast kennis van de beide talen ook een goed begrip hebben van het betreffende vakgebied. Het is dus vanzelfsprekend dat men enige jaren geleden begon te dromen van de een of andere machine, die dit vertaalwerk kon uitvoeren.

De eerste suggestie het eens met een elektronische rekenmachine te proberen kwam omstreeks 1947. Er had een correspondentie over dit onderwerp plaats tussen Warren Weaver van het Rockefeller Foundation en professor Norbert Wiener van het Massachusetts Institute of Technology, waarbij Weaver onder andere schreef: "When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode". "Aan diverse instellingen begon een onderzoek naar de mogelijkheid elektronische rekenmachines voor het vertalen te gebruiken, zoals aan het Massachusetts Institute of Technology, University of California, University of Washington en het Princeton Institute of Advanced Study. In 1952 werd over de vertaalmachine een congres gehouden. Behalve in Engeland en Amerika was er toen ook reeds belangstelling voor dit onderwerp in Rusland.

De moeilijkheden, die moeten worden overwonnen voordat een vertaalmachine kan worden geconstrueerd, zijn zeer groot. Een soort van mechanisch woordenboek moet in de machine worden

Naar een voordracht gehouden voor de Koninklijke Maatschappij voor Natuurkunde "Diligentia" te 's-Gravenhage op 6 maart 1959.

ingebouwd. Gegevens over grammatica van beide talen moeten ook aan de machine worden toegevoerd, maar daartoe heeft men eerst een uitvoerige studie van de opbouw van de taal moeten maken, zoals van woordfrequentie, woordvolgorde, enz. In beperkte mate zijn successen bereikt door toepassing van rekenmachines als de IBM 701, de Whirlwind en de Russische B.E.S.M. Momenteel wordt in Amerika in opdracht van het Rome Air Development Center van de Amerikaanse luchtmacht gewerkt aan de constructie van een machine, die uitsluitend bestemd is voor vertalen uit het Russisch in het Engels.

In de rest van deze voordracht zal ik spreken over de redenen, die de elektronische rekenmachine min of meer geschikt maken vertaalwerk uit te voeren, over de vele problemen, die men ontmoet bij het opzetten van een programma voor het automatisch vertalen en over toekomstmogelijkheden.

2. *De elektronische rekenmachine als vertaalmachine*

2.1 Overzicht van de werking van de elektronische rekenmachine

Nadat gedurende de tweede wereldoorlog enkele elektronische rekenmachines waren gebouwd, begon men in 1946 meer rekenmachines te construeren, die sneller konden rekenen, meer getallen tegelijk konden onthouden en gemakkelijker van het ene op het andere probleem konden worden omgeschakeld. Dit bereikte men door het toepassen van nieuwe technieken. De veelzijdigheid van deze machines werd aan allerlei mathematische en niet-mathematische problemen gedemonstreerd, zoals het oplossen van 30 vergelijkingen met 30 onbekenden, het spelen van een of ander spel tegen een menselijke tegenstander en ook het vertalen van een tekst in een andere taal.

De rekenmachines, waaraan wij hier uitsluitend zullen denken, zijn digitale of numerieke rekenmachines. In tegenstelling tot de analoge rekenmachines, die de grootheden uit het probleem vervangen door hanteerbare fysische grootheden, werkt een digitale rekenmachine met getallen, die op fysische wijze worden vastgelegd. Een rekenliniaal is een zeer eenvoudig analogo werkend rekenhulpmiddel, waarbij de analogie bestaat in de lengte van een lijnstuk tegenover de logaritme van een getal. Een telraam of een tafelrekenmachine met toetsenbord zijn voorbeelden van digitaal werkende rekenhulpmiddelen. Een voordeel van de digitale apparatuur is dat de nauwkeurigheid wordt bepaald door het aantal onderdelen en niet in de eerste plaats door de nauwkeurigheid, waarmee deze onderdelen zijn afgewerkt.

Daar de meeste mathematische problemen een groot aantal herhalingen bevatten, heeft het zin zeer snel werkende, en dus elektronische, rekenmachines te construeren, die echter in staat moeten zijn een in detail meegedeeld proces te herhalen gedurende een vast aantal malen, of tot een of ander criterium is bereikt.

In dergelijke elektronische rekenmachines worden onderdelen gebruikt, die twee evenwichtstoestanden bezitten, zoals zogenaamde trigger- of flip-flop-schakelingen. Dit zijn een soort van elektronische relais, bestaande uit twee radiobuizen of twee transistoren (een moderne vervanging van een radiobuis), waarbij altijd een van de beide buizen of transistoren de elektrische stroom geleidt. Het is mogelijk de overgang van de ene naar de andere evenwichtstoestand te doen plaatshebben door aan de schakeling een korte elektrische spanningsstoot toe te voeren.

Hoewel men oorspronkelijk, gedurende de tweede wereldoorlog, nog trachtte met deze onderdelen rekenmachines te construeren, die in het ons vertrouwde tientallige stelsel rekenen, heeft men bij latere ontwerpen overwogen dat het gebruik van dit talstelsel niet noodzakelijk is. Het tweetallige stelsel is eenvoudiger uit te voeren, omdat flip-flop-schakelingen, zoals gezegd, twee evenwichtstoestanden bezitten, zodat aan deze beide toestanden de cijfers 0 en 1 kunnen worden toegekend. Dit geeft een aanzienlijke materiaalbesparing (ca 70 %) ten opzichte van het tientallige stelsel. Bovendien kunnen andere elementen, die volgens het ja/nee principe werken, worden gebruikt, bijv. als geheugen.

Ook de optel- en vermenigvuldigregels zijn in het tweetallige stelsel aanzienlijk eenvoudiger dan in het tientallige. Hierop zullen we echter niet nader ingaan. Wel is het voor een goed begrip van de vertaalmachine nuttig te weten hoe een getal in het tweetallige stelsel moet worden gelezen.

Zoals in het tientallige stelsel de betekenis van een getal de volgende is: $935 = 9 \times 10^2 + 3 \times 10^1 + 5 \times 10^0$, zo is in het tweetallige stelsel de betekenis van een getal:

$$10110 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0.$$

Weer uitgedrukt in het tientallige stelsel, dus in begrijpelijke taal, is dit $16 + 4 + 2 = 22$.

In de volgende tabel zijn alle getallen van 0 tot 31 in het tweetallige stelsel weergegeven:

0 = 00000	16 = 10000
1 = 00001	17 = 10001
2 = 00010	18 = 10010
3 = 00011	19 = 10011
4 = 00100	20 = 10100
5 = 00101	21 = 10101
6 = 00110	22 = 10110
7 = 00111	23 = 10111
8 = 01000	24 = 11000
9 = 01001	25 = 11001
10 = 01010	26 = 11010
11 = 01011	27 = 11011
12 = 01100	28 = 11100
13 = 01101	29 = 11101
14 = 01110	30 = 11110
15 = 01111	31 = 11111

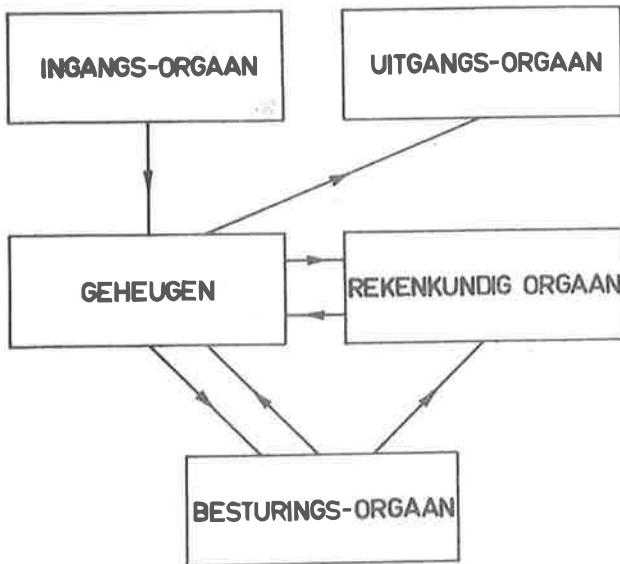
Figuur 1. De getallen van 0 tot 31 in het tweetallige stelsel.

2.2 De opbouw van de elektronische rekenmachine

In elke digitale rekenmachine - dus zowel de universeel bruikbare als de machine voor speciale doeleinden - kan men vijf hoofdorganen onderscheiden. Deze zijn:

- het rekenkundig orgaan;
- het geheugenorgaan;
- het besturingsorgaan;
- het ingangsgaan;
- het uitgangsgaan.

Ze zijn onderling verbonden. In figuur 2 zijn de belangrijkste koppelingen weergegeven.



Figuur 2. De opbouw van een elektronische rekenmachine.

Het *rekenkundig orgaan* heeft tot taak alle elementaire bewerkingen, d.w.z. optellen, aftrekken, vermenigvuldigen en delen, uit te voeren. Dit betekent dat alle bewerkingen, waaruit een bepaald wiskundig probleem bestaat, na elkaar moeten plaats vinden. Hierbij zullen tussenresultaten ontstaan, die later weer moeten worden gebruikt. Het *geheugenorgaan* onthoudt deze getallen zolang dit nodig is, evenals alle getallen en instructies, die bij het begin van het probleem aan de machine zijn toegevoerd. Het bergt deze getallen op in registers, "vakjes", die op de een of andere wijze genummerd zijn.

Alvorens het rekenkundig orgaan een bewerking kan uitvoeren, dient het een opdracht te ontvangen, die inhoudt welke getallen uit

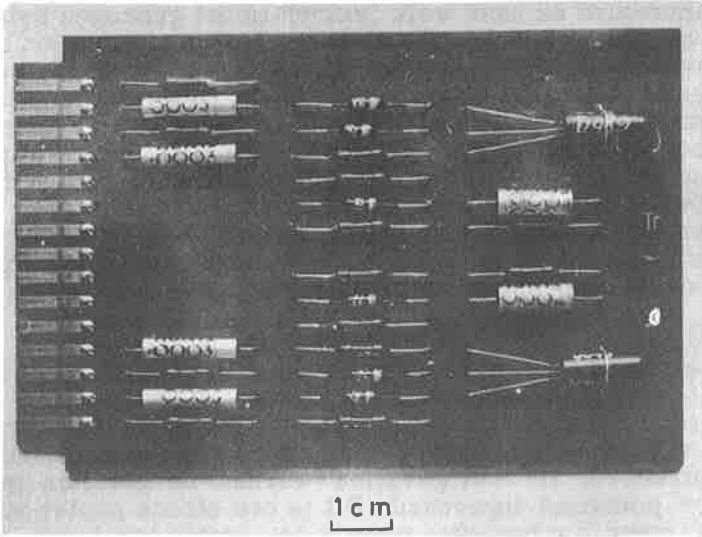
het geheugen moeten worden gehaald, welke bewerking moet worden uitgevoerd en naar welk "vakje" in het geheugen het resultaat moet worden gebracht. Deze opdrachten worden door het *bestu- ringsorgaan* gegeven, dat de orders in een bepaalde volgorde in gecodeerde vorm uit het geheugen haalt. Deze instructies zijn van buitenaf aan het geheugen toegevoerd, of kunnen tijdens het rekenproces ontstaan. Een instructie kan bijv. afhankelijk gesteld worden van het positief of negatief zijn van een vorige rekenkundige bewerking. Het *ingangs-* en het *uitgangsgorgaan* tenslotte brengen de machine met de buitenwereld in contact. Hierdoor worden namelijk de getallen en instructies aan de machine toegevoerd en de eindresultaten afgeleverd.

2.3 Technische uitvoering van het geheugenorgaan

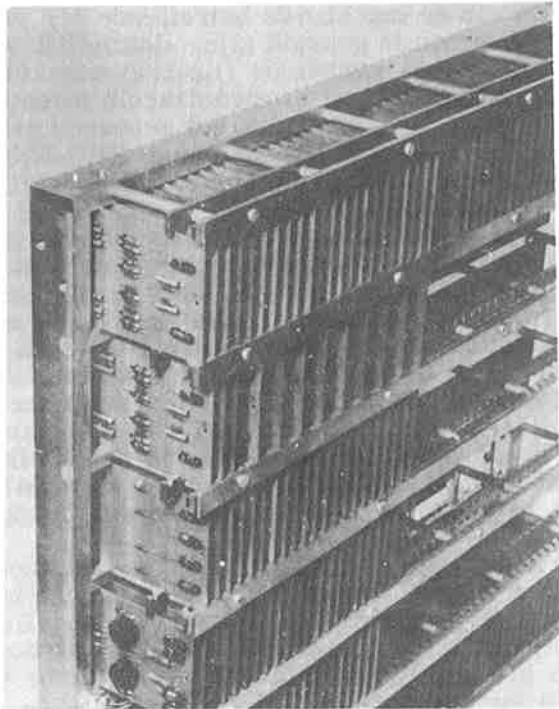
De vijf genoemde organen kunnen op allerlei manieren worden geconstrueerd. In veel gevallen worden de getallen met behulp van een ponsband ingevoerd. Dit is een strook papier van anderhalf of meer cm breedte, waarin een gaatje een 1 voorstelt. Een aantal gaatjes naast en/of achter elkaar vormen een getal; dit getal kan bijv. met een aantal pennetjes worden afgetast. Een gaatje laat toe dat het pennetje een stroomkring sluit. Hiermee wordt in het geheugen van de machine de betreffende één vastgelegd, waarvoor vele systemen in gebruik zijn, afhankelijk van de eisen, die men stelt. De reeds genoemde flip-flop-schakelingen zijn zeer snel, maar vrij groot; zij worden daarom meestal wel in het rekenkundig orgaan, maar niet in het geheugen gebruikt. Figuur 3 geeft een op het Fysisch Laboratorium RVO-TNO geconstrueerde flip-flop, terwijl figuur 4 een rek met dergelijke schakelingen toont.

Tegenwoordig past men dikwijls magnetische geheugenelementen toe. Een uitvoeringsvorm is een trommel, waarvan de mantel bedekt is met een laagje nikkel. Terwijl de trommelsnel ronddraait wordt door een zogenaamde schrijfkop een klein oppervlakte-elementje van de mantel gemagnetiseerd. Door heel korte stroomstoten toe te passen kan dit gemagnetiseerde oppervlak zeer klein worden gehouden, zodat langs de omtrek van een trommel van bijv. 20 cm diameter wel 1000 enen kunnen worden geschreven. Door bijv. 30 schrijfkoppen naast elkaar te plaatsen zou men 30.000 tweetallige cijfers op de trommel kunnen onthouden. Voor het weer aflezen kunnen de schrijfkoppen worden gebruikt.

Bij een ander type magnetisch geheugen is voor ieder tweetallig cijfer een ringetje aanwezig, dat gemaakt is van materiaal met een vrijwel rechthoekige hysteresis-lus. De diameter van zo'n ringetje is bijv. 2 mm. Na een korte stroomstoot blijft het ringetje in een goed gedefinieerde magnetische toestand achter, die we één kunnen noemen. Magnetisatie in de andere richting noemen we dan nul. Een stroomstoot in tegengestelde richting



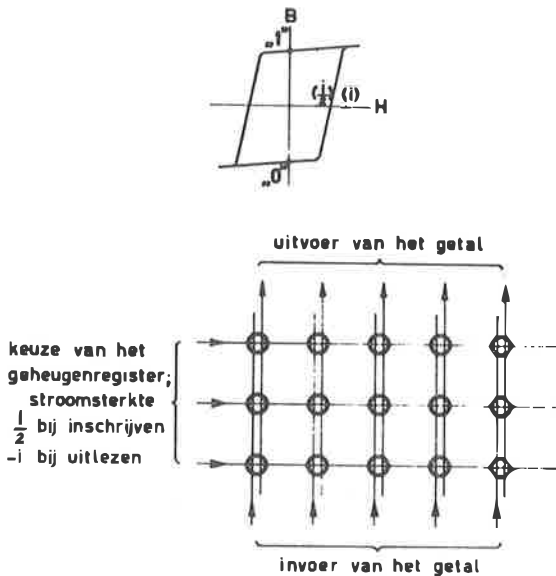
Figuur 3. transistor flip-flop.



Figuur 4. rek met transistor flip-flop schakelingen.

door dezelfde wikkeling of door een tegengesteld gewikkelde speciale afleeswikkeling doet de magnetisatie-richting omklappen als er een één in werd vastgehouden. Aan de uitgangswikkeling verschijnt dan een spanningsstoot, die verder kan worden gebruikt.

Bij een veel gebruikte uitvoeringsvorm worden draden door de ringetjes gevlochten, waardoor geheugen "matjes" ontstaan. Het is mogelijk hierin een bepaald ringetje te magnetiseren door door de twee draden, die door het ringetje gaan, de halve magnetiseringsstroom te voeren. De ringen in dezelfde horizontale en verticale rij krijgen dan de halve stroom, waarop ze niet reageren door de rechthoekige vorm van de hysteresislus. Het uitgekozen ringetje krijgt echter de volle benodigde stroom. Figuur 5 geeft het principe weer.



Figuur 5. Magnetisch geheugen met ringkernen.

Een fotografische film of plaat is ook bruikbaar als geheugen, waarbij de enen en nullen door doorschijnende en ondoorschijnende blokjes worden voorgesteld. Een nadeel is dat het moeilijk is om getallen in te schrijven, terwijl het schoonvegen en opnieuw gebruiken onmogelijk is. Evenals bij een magnetische trommel kan het opzoeken van een getal lang duren wanneer het afleessysteem op een ander deel van de film of plaat staat. Een voordeel is de grote capaciteit; met verfijnde fotografische technieken kunnen wel 10.000 bits (samentrekking van binary digits = tweetallige cijfers) per mm^2 worden vastgelegd. Om deze redenen is het fotografisch geheugen zeer geschikt voor het vastleggen van tabellen en dergelijke, hoewel het in universele rekenmachines vrijwel

nooit wordt toegepast. Figuur 6 toont een deel van een fotografische schijf, die snel onder een aantal fotocellen, die de getallen kunnen aflezen, kan ronddraaien.



Figuur 6. Deel van een fotografisch geheugen.

Figuur 7 geeft een schatting van de snelheid en het volume van de vier genoemde geheugensystemen, compleet met de bijbehorende hulpapparatuur. Hoewel de gegeven getallen slechts een ruwe aanduiding zijn, komt toch duidelijk uit dat een toename van de snelheid gepaard gaat met een toename van het volume.

Type geheugen	tijd in sec. nodig voor		volume in cm^3/bit
	schrijven	uitlezen	
Transistor flip-flop	10^{-6}	10^{-6}	200
Magnetische ringkernen	10^{-5}	10^{-5}	2
Magnetische trommel	10^{-5}	$10^5 \cdot 10^{-2}$	0,2
Fotografische schijf	—	$10^4 \cdot 10^{-1}$	0,04

Figuur 7. Schatting van snelheid en volume (bij 50.000 bits) van enkele geheugensystemen.

2.4 Technische problemen bij het gebruik van de rekenmachine als vertaalmachine

Om een elektronische rekenmachine te kunnen gebruiken als vertaalmachine moet het mogelijk zijn woorden in te voeren in plaats van getallen. Dit gaat zeer eenvoudig door iedere letter te vervangen door een getal onder 32. In tweetallige vorm kunnen deze getallen dan in de machine worden vastgelegd. In feite ge-

beurt iets dergelijks reeds bij de telex. Op deze wijze zal een woord van 10 letters in de machine worden ingevoerd als een getal met $10 \times 5 = 50$ enen en nullen.

Zou bijv. een codering worden gebruikt waarin $a = 1$, $b = 2$, enz., dan zou het woord "tafel" worden gecodeerd als 20, 1, 6, 5, 12 en dus in de machine worden ingevoerd als 10100 00001 00110 00101 01100.

Op deze wijze kunnen in het geheugen van de machine vele woorden worden vastgelegd. Wordt achter ieder woord het ermee overeenkomende woord in de andere taal geschreven, natuurlijk ook weer in de vorm van enen en nullen, dan ontstaat een woordenboek. Komt een woord, dat vertaald moet worden, de machine binnen, bijv. op een ponsband, dus als enen en nullen, dan kan dit woord - of zo men wil dit getal - achtereenvolgens worden afgetrokken van alle woorden in het geheugen. Is de uitkomst nul dan is het juiste woord gevonden. Naar de uitgang van de machine kan dan het overeenkomstige woord in de andere taal worden gevoerd, waarna het met behulp van bijvoorbeeld een telex-machine kan worden getikt.

Het is nuttig hierbij nog te wijzen op twee technische problemen, die zich voordoen wanneer we een woordenboek in het geheugen willen vastleggen. Het ene probleem betreft het benodigde volume. Stel dat men 60.000 woorden wil vastleggen en hierbij ook 60.000 woorden in een andere taal. In werkelijkheid is het door meer betekenissen en andere redenen nog aanzienlijk meer. Deze 120.000 woorden hebben een gemiddelde lengte van 5 letters, dus 25 bits. In totaal moeten dus 3.000.000 bits worden onthouden. Dit is de reden dat men aan de ene kant op allerlei manieren probeert het woordenboek te beperken en aan de andere kant zoekt naar een ander type geheugen, dat speciaal voor het vertaalproces geschikt is. Het fotografische geheugen lijkt hiervoor zeer geschikt te zijn. Het niet-uitwisbaar zijn van het geheugen is hierbij juist een voordeel, terwijl het om deze reden in universele rekenmachines vrijwel nooit wordt toegepast.

Het andere technische probleem betreft de snelheid. Komt met de ponsband via het ingangsgaas een woord in de machine binnen, dan moet dit woord worden vergeleken met de woorden uit het geheugen. Wordt het binnenkomende "woordgetal" achtereenvolgens afgetrokken van de "woordgetallen" (bijv. 60.000) in het geheugen, dan zullen gemiddeld 30.000 aftrekkingen moeten worden uitgevoerd voordat de uitkomst nul is, dus voor het juiste woord wordt gevonden. Stel dat 0,001 seconde nodig is voor het uitlezen van het woord en het uitvoeren van de aftrekking, dan kost het opzoeken van elk woord gemiddeld 30 sec. Een betere methode is de woorden in volgorde te plaatsen, d.w.z. in een opklimmende getallenreeks. Trekken we nu het binnenkomende woord af van het middelste woord uit het geheugen, dan kan de uitkomst positief of negatief zijn. Het gezochte woord zal zich dan dus in de eerste of tweede helft van het geheugen bevinden. Het proces wordt herhaald voor het midden van die helft, waarmee het juiste kwart vastligt. Zo doorgaande geven 16 aftrekkin-

gen de plaats aan tot op 1: 65.536 nauwkeurig. Dit is dus voldoende voor 60.000 woorden. Op deze wijze is het woord gevonden in $16 \times 0,001 \text{ sec} = 0,016 \text{ seconde}$. Met een gewoon woordenboek kan men dit eenvoudig nagaan.

Behalve het woordenboek moeten allerlei andere regels voor het vertalen, zoals grammaticale gegevens, in de machine worden vastgelegd. Ook deze worden in het geheugenorgaan opgeslagen en beïnvloeden het besturingsorgaan. Het resultaat van verschillende tests, die op de ingevoerde tekst worden uitgevoerd, vullen op deze wijze de woord voor woord vertaling aan.

3. *Taalkundige problemen bij het ontwerpen van een vertaal-machine*

3.1 Algemeen overzicht

Vertalen is eigenlijk niets anders dan het vervangen van de ene taal door de andere met de bedoeling dezelfde ideeën uit te drukken. Zowel wanneer dit vertalen door de mens als wanneer het door een machine gebeurt, is het woordenboek hiervoor de basis. Met de woord voor woord vertaling, die hiermee mogelijk is, bereikt men echter vrijwel nooit volledig het doel, onder andere doordat vele woorden meer dan één betekenis hebben, die met verschillende woorden in de andere taal overeenkomen. Er is echter nog een andere reden waarom woord voor woord vertaling niet voldoende is; de grammatica bevat namelijk ook een aanzienlijk deel van de over te brengen informatie, wat onder anderen tot uiting komt in een min of meer streng voorgeschreven zinsbouw en in de verbuigingen. Men drukt het wel als volgt uit: de woorden geven aan waarover wordt gesproken, de grammatica wat er over wordt gezegd.

Uit het bovenstaande volgt dat het noodzakelijk is naast een woordenboek grammaticale regels voor het vertalen in de vertaal-machine in te voeren. Deze regels moeten evenals het woordenboek in het geheugen worden vastgelegd en dienen op een zodanig logische wijze te zijn opgesteld dat ze door een machine kunnen worden opgevolgd. Dit probleem is zeker niet eenvoudig en heeft een nauwkeurige bestudering van de opbouw van de taal nodig gemaakt.

In de rest van dit hoofdstuk zal een aanduiding worden gegeven van de richting waarin oplossingen van deze moeilijkheden bijvoorbeeld kunnen worden gezocht, teneinde een vertaling met een machine van redelijke omvang en gecompliceerdheid mogelijk te maken.

3.2 Verbuigingen

Het is aan de ene kant nodig de verbuigingen in de tekst te herkennen om er conclusies voor de vertaling uit te kunnen trekken.

Aan de andere kant is het nodig de woorden van de vertaling op de juiste wijze te verbuigen. Deze beide zijden van het verbuigingsprobleem dekken elkaar slechts ten dele. Zo kunnen verbuigingen in sommige talen door woorden worden vervangen, zoals voorzetsels of lidwoorden.

Voorbeeld 1: Duits "der" kan "van de" betekenen.

Voorbeeld 2: Noors kake = koekje
 kaken = het koekje
 kaker = koekjes
 kakene = de koekjes.

Bij het gebruiken van elektronische rekenmachines voor het vertalen dient men te woekeren met de geheugencapaciteit, hoe groot deze geheugens ook lijken te zijn. Daarom zijn er vele pogingen gedaan voor- en achtervoegsel af te snijden. De machine kan bijvoorbeeld proberen of de laatste twee letters "en" zijn en dan in het woordenboek nazoeken of de rest voorkomt. Deze rest draagt dan in de een of andere code een aanduiding voor de woordsoort, terwijl de machine in een achtervoegselwoordenboek vindt wat in de vertaling het gevolg van deze uitgang moet zijn.

Waarschijnlijk zal echter het geheugen van een echte vertaalmachine zo groot worden gemaakt dat alle verbuigingen, die kunnen voorkomen, in het woordenboek kunnen worden opgenomen.

3.3 Zinsbouw

Ook het probleem van de zinsbouw heeft twee kanten. In de eerste plaats kan de zinsbouw van de ingevoerde tekst van groot belang zijn, bijv. bij het bepalen van de juiste betekenis van woorden met meer dan één betekenis. In de tweede plaats kan de zinsbouw van de vertaling voor het goede begrip van groot belang zijn.

Het meest voor de hand ligt de machine een analyse te doen uitvoeren van de zinsbouw van de tekst. Hiervoor zal elk woord in het geheugen een aanduiding moeten hebben, zoals "lidwoord", "werkwoord", enz., voor zover dit mogelijk is. Voor die woorden, waarvoor dit niet mogelijk is, kan toch meestal de combinatie met de bekende woorden enig gegeven opleveren. Daarna zal met de opbouwregels van de tweede taal de zinsbouw van de vertaling moeten worden opgezet.

3.4 Woorden met meer dan één betekenis

Het vinden van de juiste vertaling van de woorden waarvoor meer dan één vertaling mogelijk is, is wel het moeilijkste probleem voor de constructeurs van een vertaalmachine. Het is door verschillende onderzoekers op vele manieren aangepakt, zodat het onmogelijk is in het kader van dit artikel meer dan een indruk te geven van het probleem.

Het aantal woorden dat meer dan één betekenis heeft is zeer

groot, zelfs wanneer men er genoeg mee neemt dat synoniemen niet worden gebruikt. Vermoedelijk heeft ieder woord gemiddeld twee betekenissen. Dit vermindert wel aanzienlijk wanneer de vertaalmachine slechts bestemd is voor een bepaald vakgebied, zoals wiskunde.

Het heeft zin een verdeling te maken tussen grammaticale en niet-grammaticale meervoudige betekenissen. Het onderscheid maken tussen de betekenissen is in de eerste groep (weer - het weer) eenvoudig, indien de woordsoort reeds bepaald is. Het onderscheid tussen de betekenissen in de tweede groep (as ; bloem) moet blijken uit de rest van de tekst. Het is gewoonlijk wel voldoende twee woorden voor en achter het betreffende woord hiervoor te onderzoeken, hoewel in sommige gevallen de hele zin of zelfs meer zinnen moeten worden afgezocht naar een aanduiding waaruit de betekenis kan worden bepaald. Een dergelijke aanduiding kan zijn het behoren tot een bepaald vakgebied. Als deze met nummers worden aangeduid kan worden nagegaan welk nummer het meeste voorkomt in de tekst. De kans is groot dat de vertaling van het betreffende woord ook dit nummer moet dragen.

Het is mogelijk de woorden te verdelen over twee groepen. De ene groep geeft de wezenlijke inhoud van de tekst; het zijn woordsoorten als werkwoorden, zelfstandige naamwoorden, bijvoeglijke naamwoorden. Ze hebben meestal weinig verschillende betekenissen. De andere groep woorden wordt wel aangeduid als "cementwoorden" omdat ze de woorden uit de eerste groep aan elkaar lijmen; voorzetsels behoren hiertoe. Ze hebben gewoonlijk vele betekenissen en komen veel vaker voor dan de woorden uit de eerste groep (het Engelse "to" kan onder anderen betekenen: te, tot, ten, aan, naar, tegen, in, tot aan, voor, bij, in vergelijking met). Opgemerkt kan worden dat ook bij het opstellen van een woordenboek ermee rekening moet worden gehouden dat de woorden, die het meeste voorkomen, niet altijd het belangrijkste zijn en dat het weglaten van een weinig frequent voorkomend woord wel eens ernstiger gevolgen kan hebben dan het weglaten van een "cementwoord".

Ook de woorden, die in combinatie met een ander woord een speciale vertaling hebben, kan men tot de woorden met meer dan een betekenis rekenen. Zo kan de combinatie "private person" worden vertaald met "particulier". Het woord "private" zal in het woordenboek een aanwijzing moeten bevatten om na te gaan of het wordt gevolgd door "person".

Eigennamen vormen in deze groep van woorden met meer dan een betekenis een eigenaardig probleem. Het zal dikwijls niet vermeden kunnen worden dat de machine tracht zo'n woord in zijn geheugen te vinden. Als hij daarin slaagt zal de vertaling ervan worden afgegeven. Als het niet in het geheugen voorkomt wordt het zonder meer naar de uitgang gevoerd, zoals trouwens elk woord waarvoor geen vertaling aanwezig is. Een hoofdletter kan soms een indicatie voor een eigennaam zijn, maar geeft niet altijd zekerheid, bijvoorbeeld als het het eerste woord van een zin is.

4. Enige beschouwingen over vertaalmachine-mogelijkheden

4.1 De huidige situatie

De grootste belangstelling voor vertaalmachines komt van de zijde van mathematici en fysici, gevolgd door beoefenaars van andere wetenschappen en door militaire instanties. De verklaring hiervoor ligt enerzijds in het feit dat dezen het meest vertrouwd zijn met de mogelijkheden, die universele elektronische rekenmachines bieden, en anderzijds het meest de behoefte gevoelen kennis te nemen van wetenschappelijke publikaties in andere talen. Hiermee hangt samen dat de woordenlijsten, die zijn ontworpen om een elektronische rekenmachine als vertaalmachine te gebruiken, gebaseerd zijn op een gebruik in deze speciale vakgebieden. Het aantal woorden vermindert daardoor aanzienlijk, terwijl vooral de meervoudige betekenissen sterk afnemen. Een niet geheel correcte zinsbouw wordt wel geaccepteerd, als de inhoud van het artikel maar begrepen kan worden. Een goede vertaler daarentegen zal bij het zien van het resultaat waarschijnlijk opmerken dat hij het sneller en beter zelf kan doen. Het gebrek aan goede vertalers geeft echter de vertaalmachines zeker een bestaansrecht.

Een machine, die wordt ontworpen om te vertalen, dus geen geprogrammeerde rekenmachine, zal men aan zijn taak geheel kunnen aanpassen en de resultaten, die zo'n machine levert zullen dan ook bevredigend kunnen zijn. De vertaalmachine, die momenteel wordt gebouwd in de Verenigde Staten van Amerika voor het Rome Air Development Center is voorzien van een fotografisch geheugen voor 30 miljoen bits, waarmee ongeveer een half miljoen woorden worden vastgelegd. Hij is bestemd voor vertalingen van het Russisch in het Engels. Voor zover bekend zal dit de eerste echte vertaalmachine zijn.

Zowel uit Amerikaanse als uit Russische literatuur blijkt dat men een onderzoek instelt naar de mogelijkheid een vertaalmachine te construeren, die niet slechts geschikt is voor het vertalen van één taal in een andere, maar die meer talen kan verwerken in beide richtingen. Bij N talen heeft men dan $N \times (N - 1)$ vertaalmogelijkheden. Indien men één taal als een centrale taal behandelt, moet men alle vertalingen via deze taal leiden en heeft men dus $2 \times (N - 1)$ vertaalmogelijkheden. De $(N - 2) \times (N - 1)$ mogelijkheden, die hier niet onder vallen, vinden dus in twee stappen plaats, waaraan uiteraard nadelen verbonden zijn. In Rusland overweegt men dit systeem voor de talen Duits, Chinees, Japans en Russisch, waarbij het Russisch de centrale taal is, zodat het meesté vertaalwerk toch in één stap kan plaatsvinden. In Amerika is wel voorgesteld een logisch opgebouwde taal zonder woorden met meer betekenissen als tussentaal te gebruiken, waarvoor het Chinees in aanmerking zou kunnen komen. Omdat dan toch vrijwel al het vertaalwerk in twee stappen plaatsvindt lijkt het voor de hand te liggen voor dit doel een kunstmatige machinetaal op te stellen, waarvoor men wel de naam "Machinees"

heeft gebruikt. Hiermee zou het aantal in te bouwen vertaalmogelijkheden $2N$ moeten worden. Hierbij kan worden opgemerkt dat het niet nodig is dat een dergelijke kunstmatige machinetaalkan worden uitgesproken. Het is voldoende wanneer elk begrip in een code in de vertaalmachine wordt vastgelegd. Hoewel we hierbij zelfs niet aan letters behoeven te denken, kan toch hiermee het voordeel van het niet uitspreken van deze taal worden geïllustreerd. Wanneer we namelijk alle lettercombinaties toestaan kunnen bijna een half miljoen "woorden" van vier letters worden gemaakt, waarmee dus een even groot aantal begrippen kan worden vastgelegd.

4.2 Toekomstmogelijkheden

De omstandigheid dat men de te vertalen tekst altijd op een soort telexmachine moet overtikken, teneinde hem op een ponsband in de vertaalmachine te kunnen invoeren, zal ongetwijfeld aanleiding zijn het invoerprobleem nader te onderzoeken.

Er bestaat reeds apparatuur waarmee gedrukte cijfers kunnen worden herkend. Bankinstellingen hebben hiervoor belangstelling. Het aftasten van letters zal echter veel moeilijker zijn, doordat letters uit vele talen moeten kunnen worden onderscheiden en doordat er geen eenheid bestaat in lettertypes. Waarschijnlijk zou internationale standaardisatie van literatuur een eerste vereiste zijn. Dit zou dan niet alleen het te gebruiken lettertype betreffen, maar ook de wijze waarop formules worden aangeduid, enz.

Een andere mogelijkheid, die niet moet worden uitgesloten voor het invoeren van de te vertalen tekst, is de gesprokeninvoer. Bij sommige communicatiesystemen, die zich in het researchstadium bevinden, worden de klanken geanalyseerd en later weer samengesteld. Indien een op een dergelijke wijze geanalyseerde tekst wordt vastgelegd en vergeleken met een hieraan aangepast woordenboek, dan lijkt een vertaling mogelijk. Moeilijkheden kunnen worden veroorzaakt door slordig spreken en dialecten, terwijl ook een aantal woorden moeilijk zal kunnen worden onderscheiden. Zo zullen de Engelse woorden "to", "too" en "two", evenals de Duitse woorden "her", "hehr" en "heer" door een dergelijke machine nauwelijks verschillend worden beoordeeld. Een dergelijke machine lijkt nog verre toekomst te zijn, vooral wanneer men een gesproken vertaling wenst.