V
W
N
I
L

CAHIER

12

Artificiele Intelligentie:
verwezenlijkingen en
verwachtingen

VWNIL Cahier 12

Artificiële Intelligentie:
Verwezenlijkingen en verwachtingen

1988

VWNIL-CAHIER WORDT GEPUBLICEERD DOOR DE VERENIGING VOOR WISKUNDIGEN, NATUURKUNDIGEN EN INFORMATICI LOVANIENSES

©1988

Eindredaktie:

A. Bultheel, Celestijnenlaan 200 A, B-3030 Leuven

Verantwoordelijke uitgever:

Y. Bruynseraede, Celestijnenlaan 200 D, B-3030 Leuven

Reproduction rights reserved.

Inhoudsopgave

Ten geleide **v**

De sprekers **vi**

Y. Willems

Kennisvoorstelling: wat, waarom, hoe? 1

H. Roelants

Mechanisering en denken 12

G. d'Ydewalle

Artificiële intelligentie menselijk bekeken 25

M. Van Hulle

Artificiële neurale netwerken: netwerken met verstand? 37

G. Adriaens

Linguistiek en artificiële intelligentie: Theorie en praktijk van een haat-liefde-verhouding 46

Het VWNIL-Colloquium, waarvan dit VWNIL-Cahier de blijvende neerslag is, werd ingericht in samenwerking met, en met de financiële steun van “Alumni Lovanienses”, de overkoepelende alumni-vereniging van de Katholieke Universiteit Leuven.

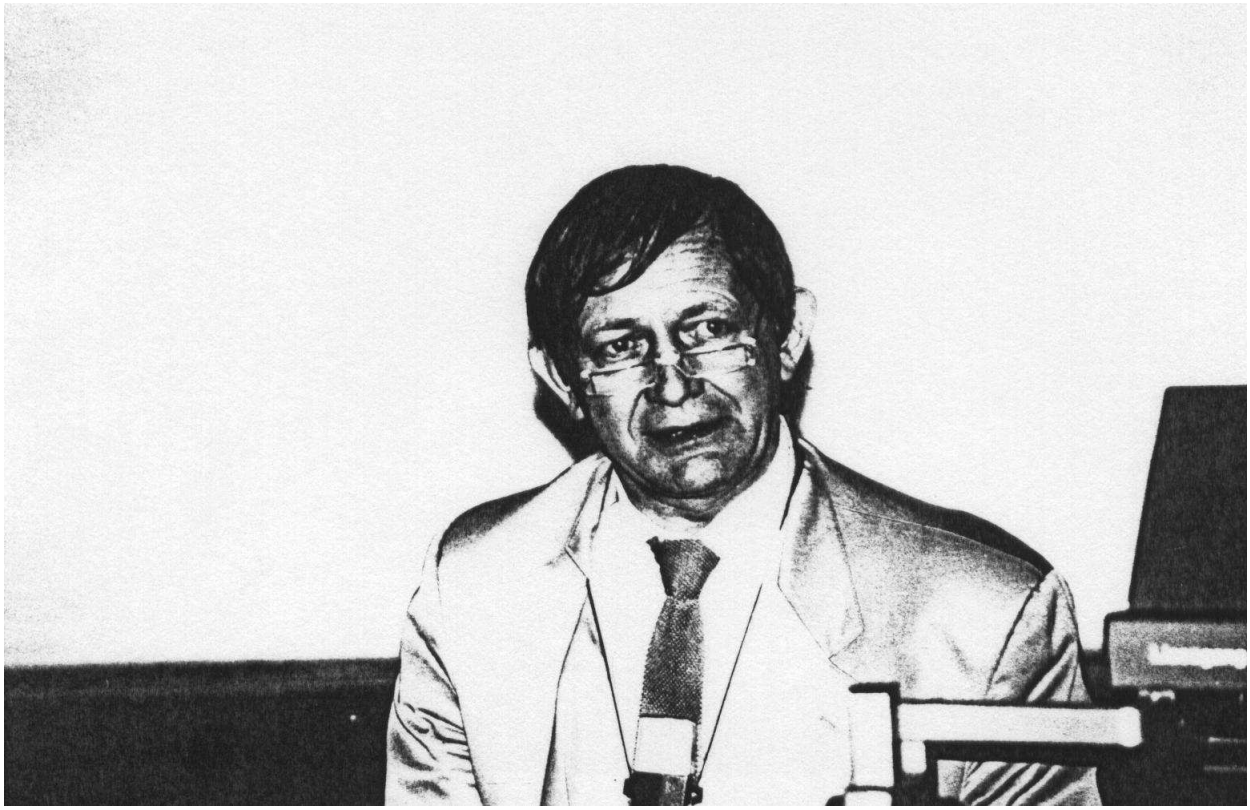
Ten geleide

Dit is reeds het twaalfde cahier in de reeks. Het eerste eerste sinds de naam van de vereniging van VWNL in VWNIL veranderd is. Inderdaad zijn er nu ook afgestudeerde informatici lid van onze vereniging. Voor een gedeelte ligt de aanwezigheid van deze specialiteit in onze vereniging aan de grondslag van de keuze van het onderwerp van het colloquium waarvan dit cahier de neerslag is. De artificiele intelligentie is weer in. De beginperiode, die zo ergens in de jaren 60 moet gesitueerd worden, werd gekenmerkt door een euforisch enthousiasme van de onderzoekers. Beloftes werden gedaan over resultaten die men nooit heeft kunnen inlossen ondanks de enorme investeringen van geld en brains in deze sector. Vandaar dat men in de tweede helft van de jaren 70 teruggevallen is op een zekere reflectie over het onderwerp. Men stelde namelijk een paradox vast: Dingen die voor een mens met gezond verstand als uiterst eenvoudig ervaren worden, blijken voor computers verschrikkelijk moeilijk te zijn terwijl er anderzijds experte systemen bestaan die betere diagnoses van ingewikkelde problemen kunnen maken dan het door menselijke experten mogelijk is. Na een periode van bezinning is de artificiële intelligentie de laatste jaren weer volop in de belangstelling. Met dit colloquium hebben we geprobeerd te benadrukken dat de artificiele intelligentie niet het privé-domein is van de informaticus. Het is een jachtterrein waar men allerlei soorten jagers, en ook allerlei soorten wild kan vinden. De vijf sprekers die op het colloquium hun visie op de artificiele intelligentie zijn komen voorstellen hebben dan ook een zeer verschillende achtergrond. Prof. Y. Willems geeft aan hoe een informaticus kennis op een formele manier probeert voor te stellen. Prof. H. Roelants geeft een filosofische kijk op het menselijk denken en hoe men reeds van oudsher probeert dit te mechaniseren. De visie van psychologen op het overdragen van experte kennis wordt belicht door Prof. G. d'Ydewalle. Neurobiologen bestuderen de werking van menselijke hersenen. Ir. M. Van Hulle zal uiteenzetten hoe men vereenvoudigde modellen van deze hersenen kan inzetten in artificiele intelligentie. Ten slotte behandelt Prof. G. Adriaens de onderzoeksaktiviteiten van linguisten in dit gebied. Ik hoop dat dit cahier een blijvende herinnering mag zijn aan het colloquium van 16 november 1988. Alvast veel leesplezier.

A. Bultheel



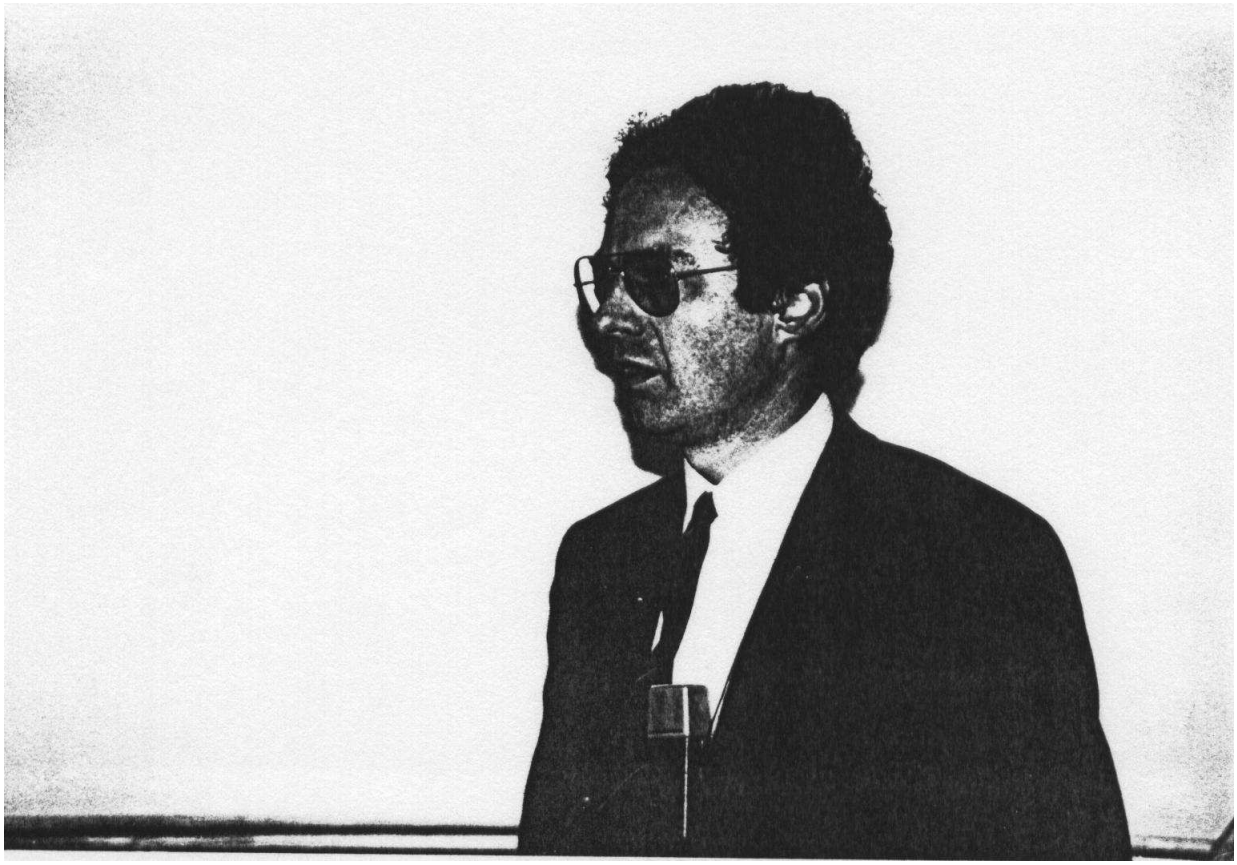
Yves D. Willems werd geboren te Brugge in 1942. Hij behaalde het diploma van burgerlijk werktuigkundig-elektrotechnisch ingenieur aan de Rijksuniversiteit Gent. Hij werkte van 1965 tot 1972 aan het Massachusetts Institute of Technology (USA) waar hij in 1972 een doctoraatsthesis maakte over automatisch herkennen van gesproken taal. Sinds 1972 is hij verbonden aan de Katholieke Universiteit te Leuven, Fakulteit Toegepaste Wetenschappen, waar hij vakken doceert in de studierichtingen Computerwetenschappen en Informatica en de leiding heeft van een twintigtal onderzoekers rond de thema's "programmeertalen gebaseerd op logica" en "experte systemen en kunstmatige intelligentie". Zijn onderzoeksgebieden zijn: methodologie en ontwerp van programmatuur, programmeertalen gebaseerd op logica, experte systemen en kunstmatige intelligentie en het voorstellen van kennis. Hij is nu gewoon hoogleraar aan de K.U.-Leuven en sinds 1981 ook gasthoogleraar aan de "Université Catholique de Louvain".



Herman Roelants werd geboren in 1937. Hij behaalde het diploma van Licentiaat in de wiskunde in 1960 en behaalde in 1963 een doctoraat in de wijsbegeerte aan het Hoger Instutuut voor wijsbegeerte aan de Katholieke Universiteil Leuven.

Sedert 1974 is hij gewoon hoogleraar aan de fakulteit der wetenschappen en aan het Hoger Instituut voor Wijsbegeerte aan de K.U. Leuven, met als onderwijsopdrachten o.m. Logica, Wetenschapsleer, filosofie van de natuur, Grondige vraagstukken uit de wetenschapsfilosofie.

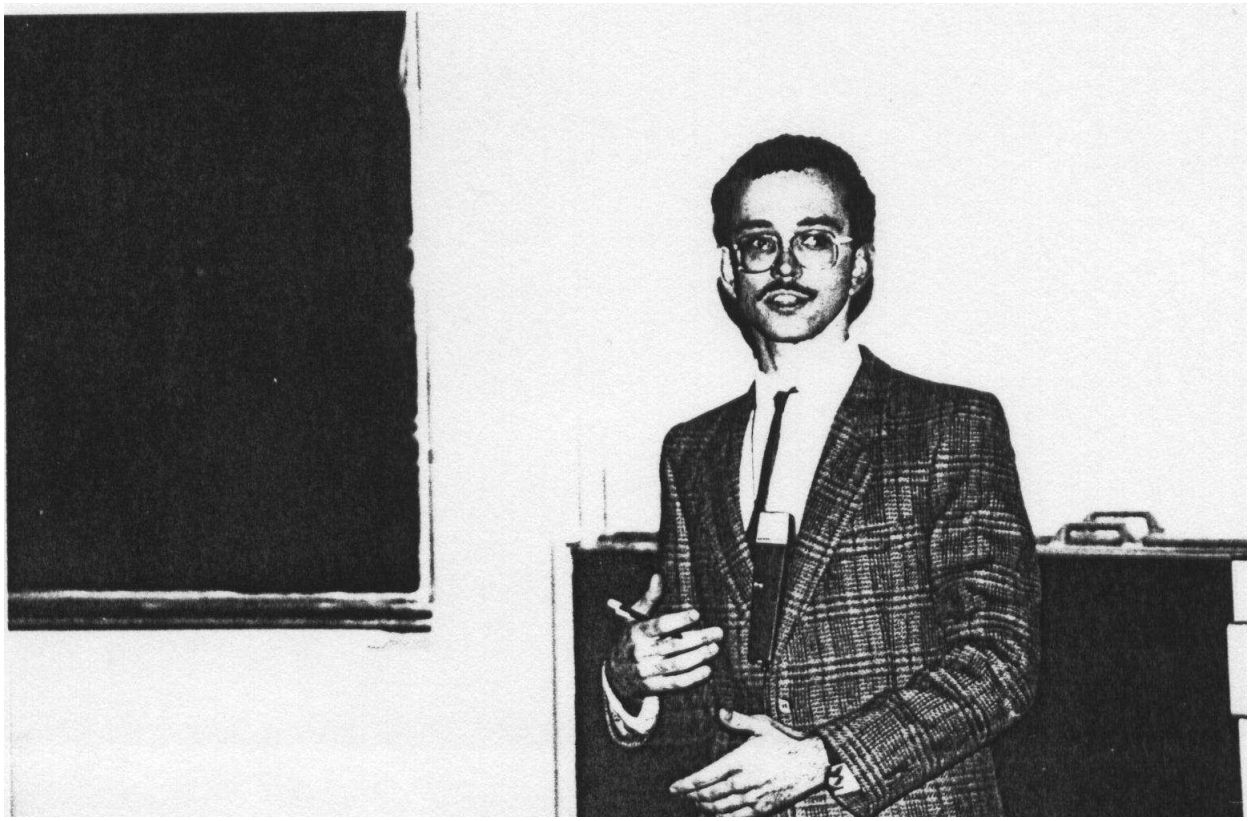
Publicaties heeft hij voornamelijk in het domein van de filosofie van de wetenschappen.



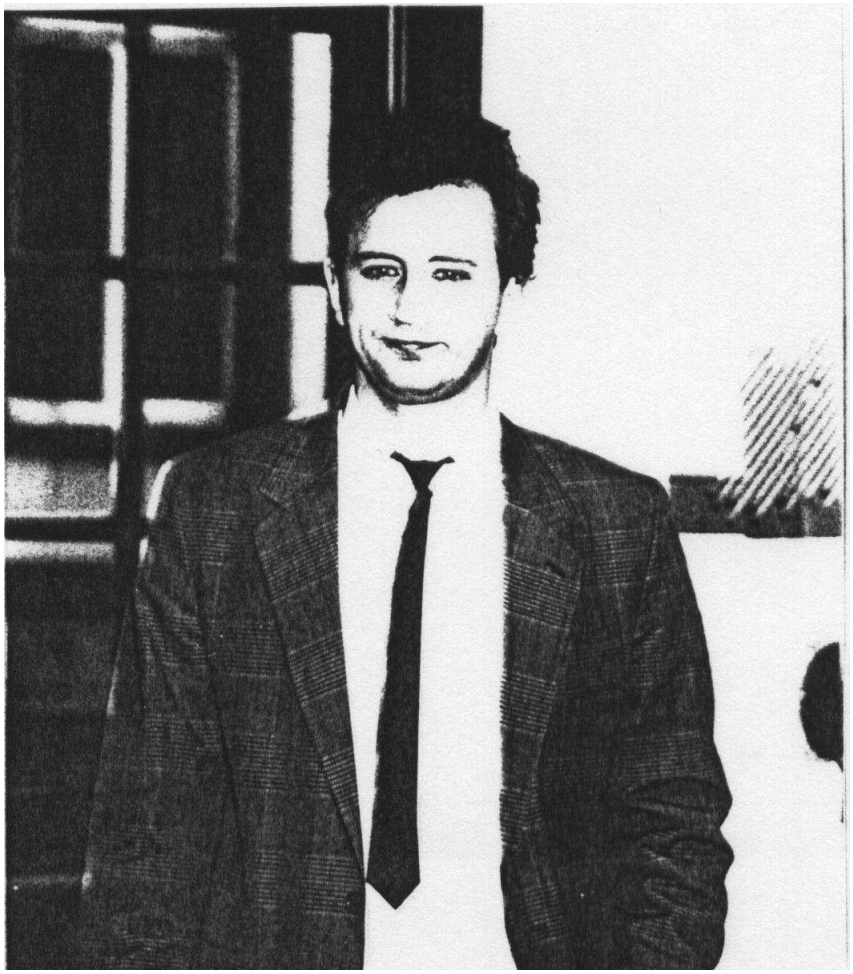
Géry van Outryve d'Ydewalle werd geboren in 1946. Hij behaalde een doctoraat in de psychologie aan de K.U.-Leuven in 1974. Hij is gewoon hoogleraar aan de fakulteit psychologie.

Op dit ogenblik is hij vice voorzitter van de Belgische Psychologische Vereniging en leidt het Laboratorium voor Experimentele Psychologie aan de K.U.-Leuven.

Daarnaast is hij lid van meerdere nationale en internationale psychologische verenigingen en zit in de "editorial board" van verscheidene wetenschappelijke tijdschriften.



Marc Van Hulle werd in 1960 geboren te Gent. Hij studeerde burgerlijk ingenieur electronica aan de K.U.-Leuven. Sinds 1985 is hij aspirant van het Nationaal Fonds voor Wetenschappelijk Onderzoek (NFWO). Hij is werkzaam aan het Laboratorium voor Neuro- en Psychofysiologie en aan de afdeling E.S.A.T. van de K.U.-Leuven waar hij een doctoraat in de toegepaste wetenschappen voorbereidt.



Geert Adriaens heeft linguistiek en computerwetenschappen gestudeerd aan de universiteiten van Leuven, Amsterdam en Rochester (NY). Hij is afgestudeerd als Germanist in 1981, behaalde in 1984 een aanvullende licentie in de informatica, en combineerde beide studies in een doctoraat rond natuurlijke-taalverwerking in 1986.

Momenteel is hij werkzaam bij Siemens en bij de K.U.-Leuven. Voor Siemens is hij verantwoordelijk voor de software-ontwikkeling op het automatisch-vertaalproject METAL; aan de universiteit doceert hij computerlinguistiek. Huidige NTV-interessedomeinen zijn automatische vertaling en parallelle analysesystemen, en dat alles in een ruim cognitief-wetenschappelijk kader.

Kennisvoorstelling: wat, waarom, hoe?

Y.D. Willems, Dept. Computyerwetenschappen, K.U.Leuven
Verslaggever: A. Bultheel, Dept. Computerwetenschappen, K.U.Leuven

De manier waarop men kennis voorstelt is voor een informaticus het kernprobleem van de informatica in het algemeen en van de artificiële intelligentie in het bijzonder. Alvorens we daar dieper op in gaan zullen we eerst het belang van deze kennisvoorstelling illustreren.

Waarom is kennis belangrijk?

Indien men aan informatica doet is men bezig met het formeel voorstellen van kennis zodanig dat deze automatisch kan verwerkt worden. Voor het domein van de A.I. is dit meer dan ooit waar. Het onderzoek in de A.I. is gestart rond 1960. Men heeft gedurende de eerste 10 jaar van onderzoek vastgesteld dat men over ontzaglijk veel kennis over een bepaald gebied moet beschikken om een A.I.-probleem uit dat gebied op te lossen. In deze beginjaren van de A.I. hield men zich bezig met onderwerpen waar voornamelijk een algemene kennis voor vereist werd. Hoe eenvoudig het probleem op het eerste gezicht ook leek, er ontbrak steeds zeer veel kennis waarover een mens met gezond verstand als vanzelfsprekend beschikt. Enkele voorbeelden hiervan:

- *Het spelen van allerlei spelletjes:* Uit die tijd dateren de eerste schaak- en dam-programma's. Deze programma's zijn gebaseerd op het vooraf onderzoeken van een aantal zetten waarna dan de zet gekozen wordt die de minst ongunstige gevolgen kan hebben. Men kan wel een goed schaakprogramma maken dat bekwame spelers aan kan maar het mist toch steeds het soort intelligentie waarmee een menselijke schaakspeler zijn strategie uitwerkt. Zijn dergelijke programma's dan wel een vorm van (artificiële) intelligentie?
- *Beeldverwerking:* Men wil bijvoorbeeld robotten uitrusten met een autonoom visiesysteem. Ook dit vereist zeer veel kennis die elke mens met wat gezond verstand als vanzelfsprekend heeft. Bepaalde voorwerpen kunnen delen van andere voorwerpen verbergen. Neem bijvoorbeeld een man die achter een paal staat. Door de paal wordt het beeld van de man in twee helften verdeeld. Een menselijke waarnemer zal er geen enkele moeite mee hebben om de twee halve mannen als een geheel te beschouwen: de paal waarnemen als iets wat zich vóór de man bevindt wordt als vanzelfsprekend ervaren. Voor een programma is het helemaal niet a priori duidelijk dat de twee halve mannen samenhoren. De programma's die uitzoeken welke voorwerpen er in het beeld voorkomen vereisen daarom algauw zeer ingewikkelde herkenningsalgoritmes.
- *Taalverwerking:* Het begrijpen van geschreven (en gesproken) natuurlijke taal. Neem het volgende voorbeeldje:

> Linda bracht An een roos voor *haar* verjaardag omdat *zij* verteld had dat *zij* van bloemen hield.

In deze zin verwijzen de drie schuin gedrukte woordjes allemaal naar An. Neem het volgende voorbeeld daarentegen:

> Linda bracht An een roos uit *haar* tuin omdat *zij* wist dat *zij* van bloemen hield.

Hier verwijzen *haar* en de eerste *zij* naar Linda en de tweede *zij* naar An. Een laatste voorbeeld:

> Linda toonde An een van *haar* creaties omdat *zij* hoopte indruk op *haar* te maken met *haar* werk.

De eerste en de laatste *haar* en *zij* verwijzen naar Linda, de voorlaatste *haar* naar An. Deze voorbeelden van menselijk onderscheidingsvermogen om dezelfde woorden in een gelijkaardige zin anders te begrijpen illustreren nog eens hoe belangrijk het is dat men over zeer veel kennis beschikt om op een verstandige manier te kunnen reageren op dergelijke zinnen. Men realiseerde zich dus in de jaren '70 dat men op het gebied van opslag en *formeel* voorstelling van kennis eigenlijk nog nergens stond. Zelfs nu zijn de verwezenlijkingen nog maar beperkt.

Hoe moet men kennis voorstellen?

Rond 1970 is men zich gaan toeleggen op het probleem hoe die kennis moest voorgesteld worden zodanig dat ze effectief kon gebruikt worden in allerhande automatische systemen. Dit onderzoek was voornamelijk gericht op het *formeel* voorstellen van kennis. De fundamentele opvatting die tot nog toe aan de grondslag ligt van de A.I. is dat elk artificieel intelligent systeem een gesofisticeerde symbolen-manipulator is. De computer is hiervan het beste voorbeeld. Er zijn echter onderzoekers die heel andere paradigma's vooropstellen. Bij artificiële neurale netwerken bijvoorbeeld ziet men volledig af van deze opvatting, maar hiervoor verwijzen we naar de desbetreffende bijdrage in dit cahier.

We zullen eerst de eisen vermelden waaraan de voorstellingswijze van kennis best zou voldoen. Ze moet *expliciet* zijn, *declaratief* en *vraag-onafhankelijk*. Hiermee wordt het volgende bedoeld: De voorstelling moet expliciet zijn en gebeurt daarom best in een soort declaraties. Dit in tegenstelling met een impliciete voorstelling van de kennis die aanwezig is in een klassiek computer-programma. In zo'n programma heeft men duidelijk kennis gebruikt over een bepaald gebied om een probleem uit dat gebied op te lossen. Deze kennis was immers nodig om stap voor stap te beschrijven hoe het probleem moet opgelost worden. Het nadeel van deze zogenaamde *procedurale* beschrijving is echter dat om een analoog probleem uit hetzelfde gebied op te lossen (waarvoor dus grotendeels dezelfde kennis vereist is), men meestal een ander programma moet schrijven. De kennis wordt daarom beter expliciet voorgesteld zodat ze ook bruikbaar

blijft om automatisch gelijkaardige problemen op te lossen. Deze voorstelling is dan probleem- of vraag-onafhankelijk geworden.

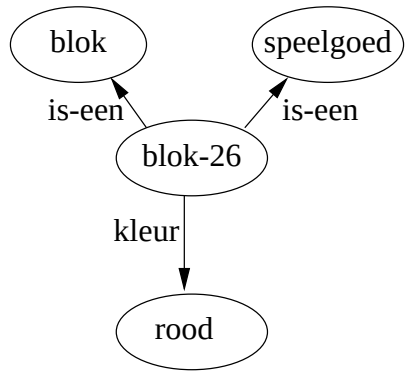
Deze eigenschappen van kennisvoorstelling hebben duidelijk een aantal voordelen: Ze geven ons een *polyvalente* voorstelling. Dit wil zeggen bruikbaar om verschillende problemen uit eenzelfde gebied op te lossen. Met een dergelijke voorstelling wordt het ook mogelijk *verklaringen* te geven over de werking van een programma: het programma zou kunnen zeggen *waarom* en *hoe* het tot een antwoord is gekomen. Men zegt soms dat deze programma's een zekere vorm van *introspectoe* vertonen. Men krijgt ook *begrijpelijker* programma's omdat de kennis op een voor de mens meer begrijpelijke manier is voorgesteld. Dit is bijzonder nuttig als men programma's moet ontluizen (= fouten verbeteren). Men verkrijgt ook een grotere *flexibiliteit* want de programma's worden gemakkelijker aanpasbaar. Als de kennis immers expliciet aanwezig is kan men economischer, sneller en eenvoudiger die kennis wijzigen of aanpassen.

Men kan uit het voorgaande besluiten dat niet alleen de kennis op zichzelf, maar ook de manier van voorstellen van deze kennis zeer belangrijk is als men aan "software engineering" wil gaan doen. De beste illustratie hiervan is de rol die *experte systemen* de laatste tijd toebedeeld krijgen. In feite zijn experte systemen een nieuwe vorm van programma ontwikkeling waarbij men gebruik maakt van een expliciete voorstelling van kennis. Hierdoor wordt het mogelijk om sneller programma's te ontwikkelen en kan men ook de commerciële waarde van de know-how van experten beter te gelde maken: men kan gemakkelijker expertsystemen over de wereld verdelen dan experten. Het succes van deze *experte systemen* is eigenlijk een paradox van de artificiële intelligentie. Enerzijds staat men nog bijna nergens voor het automatisch uitvoeren van heel wat eenvoudige taken waartoe de man in de straat wel in staat is (bv. een eenvoudig gesprek voeren). Anderzijds is men in het modeleren van de expertise van menselijke experten in zekere gebieden reeds heel ver geraakt.

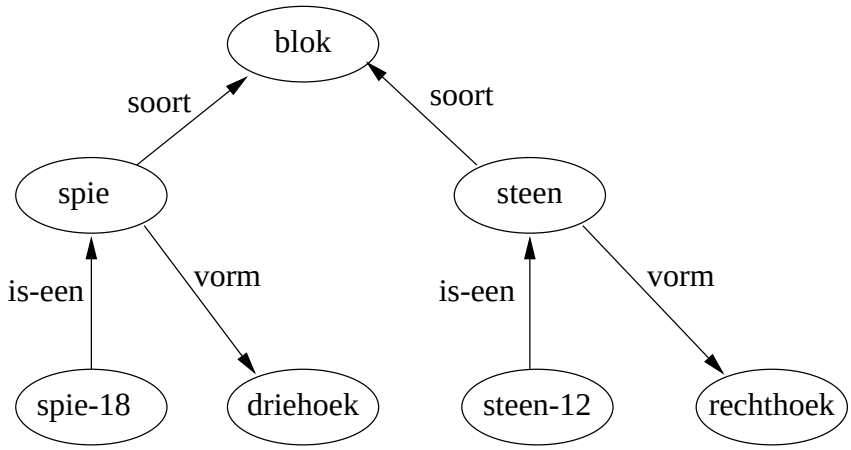
Enkele praktische voorstellingswijzen van kennis

Er zijn vier grote families van kennisrepresentatie-formalismen. We kunnen ze groeperen in twee grote groepen: De *netwerken* en *frames* (objecten) enerzijds en de systemen beschreven door *regels* en door *wiskundige logica* anderzijds. We zullen deze vier families wat van dichterbij bekijken.

1. Netwerken: Een eerste vorm is een zogenaamd semantisch netwerk. Een voorbeeldje van een semantisch netwerk staat in Fig.1. Het bestaat uit objecten, die omcirkeld zijn en relaties tussen deze objecten, die weergegeven zijn met pijlen. Op de figuur zijn 3 relaties voor blok-26 weergegeven. Men leest dit als "blok-26 is een blok", "blok-26 is (een) speelgoed" en "blok-26 heeft een rode kleur". Een wat uitgebreider voorbeeld vindt men in Fig.2. Dit voorbeeld illustreert hoe men niet expliciet vermelde relaties toch kan terugvinden in een dergelijk netwerk. Zo ziet men dat spie-18 een speciaal geval van een spie is en dat elke spie de eigenschap heeft dat ze driehoekig is. Een gevolgtrekking is dan dat ook spie-18 een driehoekige vorm heeft. De eigenschap van een driehoekige vorm te hebben wordt door spie-18 *overgeërfd* van de spieën. Op



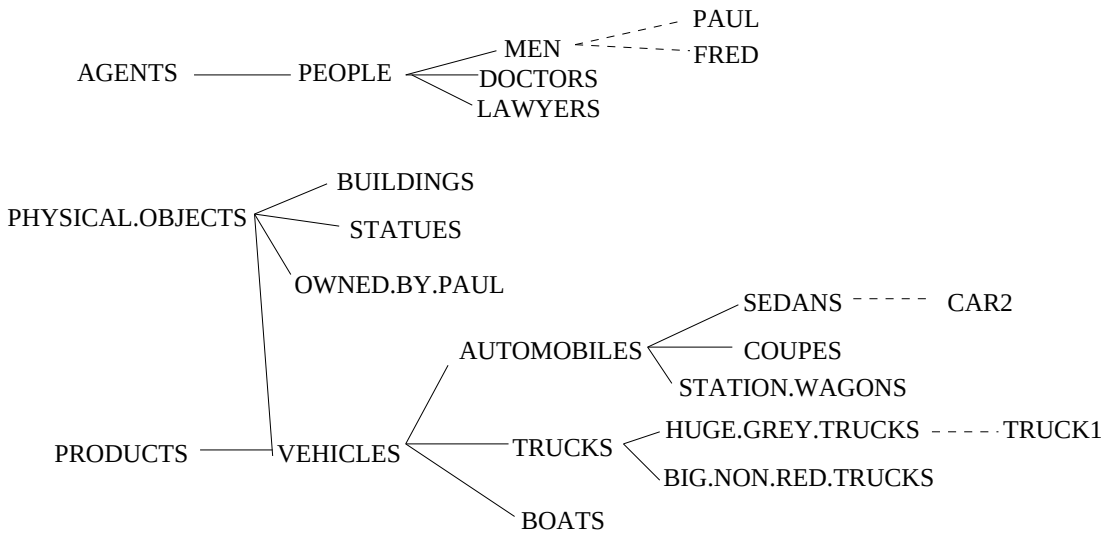
Figuur 1:



Figuur 2:

dezelfde manier kan men afleiden dat steen-12 een rechthoekige vorm heeft. Men kan eens en vooral een algemeen programma schrijven dat eigenschappen in een dergelijke grafe gaat opzoeken.

2. Frames of objecten: In plaats van (zoals in het semantisch netwerk) cirkels en pijlen te gebruiken, zou men dezelfde kennis ook op een wat meer computermatige manier kunnen voorstellen. In het voorbeeld van Fig.3a ziet men dat bijvoorbeeld de auto's een deelverzameling (subklasse) vormen van de voertuigen, die op hun beurt dan weer deelverzamelingen zijn van de fysieke voorwerpen én van de industriële produkten. Dit betekent dat auto's bepaalde eigenschappen kunnen overerven van de fysieke voorwerpen en andere van industriële produkten. In de eenheid auto's van Fig. 3b wordt deze informatie ook aangegeven. Zo ziet men dat de verzameling voertuigen de verzameling auto's omvat, dus een superklasse is van de auto's en dat allerlei soorten auto's (stationwagons, coupes, sedans) deelverzamelingen (subklassen) vormen van de auto's. Er wordt ook aangegeven dat de eigenschap "kleur" een eigenschap is die wordt overgeërfd van de fysieke objecten enz. Het gebeurt dat er bepaalde eigenschappen zijn



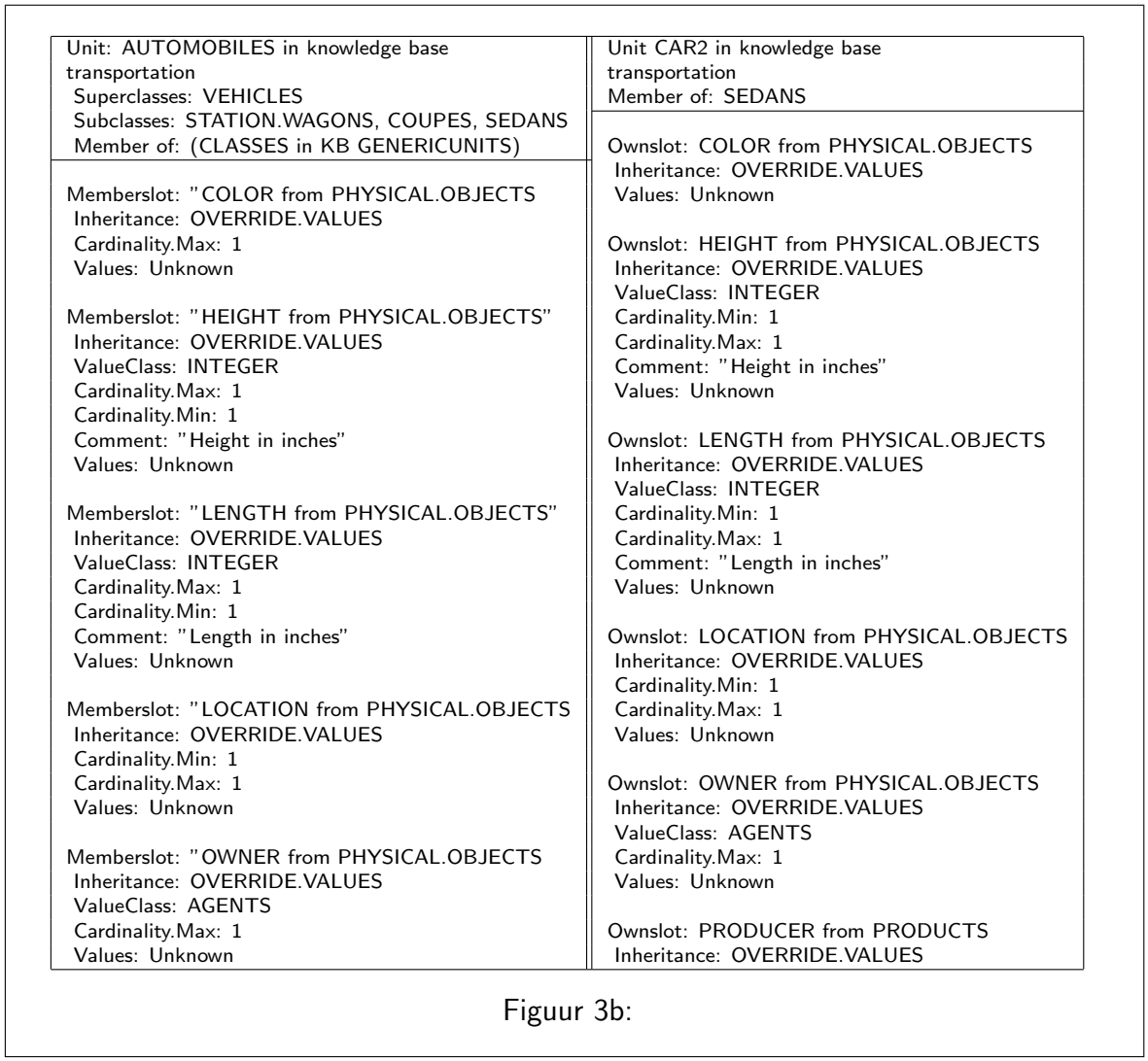
Figuur 3a:

van een klasse die niet overgeërfd worden door de elementen van die klasse. Bijvoorbeeld "het gewicht van de zwaarste vrachtwagen" is een eigenschap van de verzameling vrachtwagens in haar geheel, die echter niet wordt overgeërfd.

Eigenschappen hoeven ook niet altijd expliciet vermeld te worden. Het is voldoende dat men met een procedure beschrijft *hoe* men bepaalde eigenschappen kan berekenen. Een voorbeeld is het geval waarbij men de leeftijd van mensen nodig heeft als eigenschap. Men zal dan niet de leeftijd stockeren maar wel de geboortedatum. Het is dan eenvoudig om uit de huidige datum met een aftrekking de leeftijd te berekenen. Dit is natuurlijk veel ekonomischer dan elke dag na te kijken of de leeftijd niet moet aangepast worden omdat de persoon misschien verjaard is. Dergelijke procedures noemt men dikwijls "demons" (duiveltjes) die ervoor zorgen dat de informatie steeds up-to-date blijft; de informatie zelf noemt men soms "actieve waarde".

Als besluit kunnen we voor deze eerste twee families van kennisvoorstelling (netwerken en frames), de volgende eigenschappen opsommen:

- *Hiërarchie:* Doordat men een hiërarchie heeft ingebouwd, nl. het "element zijn van" weergegeven in de "is-een" relatie, heeft men een zekere vorm van abstractie ingevoerd.
- *Structuur:* Men kan relatief ingewikkelde structuren gaan beschrijven doordat men de relatie "deelverzameling-van" heeft ingevoerd en omdat een verzameling deel kan zijn van verschillende superklassen.
- *Erfenis:* De elementen in een zekere klasse kunnen eigenschappen over-erven van een of meerdere van de superklassen.



Figuur 3b:

- *Verzuimwaarde:* Men kan een algemene eigenschap geven voor de elementen van een klasse en slechts voor die elementen waar deze eigenschap afwijkt van de algemene eigenschap dit expliciet vermelden. Voor alle andere elementen vermeldt men dit niet expliciet en wordt een verzuimwaarde (default) door over-erving teruggevonden.

- *Demons:* Dit zijn procedurale beschrijvingen van kennis die ook kunnen weggeborgen worden. Ze worden gebruikt om actieve waarden (active values) te berekenen. We zullen nu de tweede groep van kennisvoorstellingen behandelen.

3. Regels: Zeer dikwijls is kennis voorstelbaar onder de vorm van productie-regels. In Fig. 4 ziet men een voorbeeld van een regel gebruikt in het expert systeem van een Amerikaanse verzekeringsmaatschappij om eisen tot schadevergoeding te behandelen. Het systeem omvat een zeer groot aantal regels zoals de regel in Fig.4. Elke regel stelt een klein stukje informatie voor. Voor een specifiek geval van schadevergoeding zal

```
If the plaintiv did receive an eye injury
  and there was just one eye that was injured
  and the treatment for the eye did require surgery
  and the recovery from the injury was almost complete
  and visual acuity was slightly reduced by the injury
  and the condition is fixed
increase the injury trauma factor by $10,000.
```

Figuur 4:

men nagaan voor welke regel(s) de premissen van toepassing zijn en men voert dan de overeenkomstige aktie(s) uit, anders niet. Een ander voorbeeld vindt men in Fig. 5. Hier gaat het om een systeem waarbij men de mogelijke oorzaak van een autodefekt wil opsporen. In het voorbeeldje van Fig. 5. ziet men dat het ook mogelijk is om

```
C is a car.
If:
  the pattern observed by attaching an
  oscilloscope to the charging circuit of
  the car C is fluctuating.arches, and the
  alternator of car C responds properly to
  different loads,
Then:
  there is strongly suggestive evidence <.9>
  that the cause of the problem with the car
  is voltage.regulator.bad.
```

Figuur 5:

bepaalde conclusies te kwantificeren. Het getal 0.9 (in het algemeen een getal tussen 0 en 1) geeft hier een waarschijnlijkheid aan, een betrouwbaarheid voor de juistheid van de conclusie. Praktisch zal men dan eerst de meest waarschijnlijke oorzaken van het defekt onderzoeken en indien dit niet helpt, zal men de minder waarschijnlijke mogelijkheden nagaan.

Dikwijls is het ook interessant om niet alleen in de conclusies waarschijnlijkheden toe te laten, maar ook in de premissen. Een expert systeem zou aan een niet experte gebruiker allerlei vragen kunnen stellen waarop die niet met zekerheid kan antwoorden. Bijvoorbeeld een landbouwer die allerlei vragen gesteld krijgt over de scheikundige samenstelling van de bemestingsprodukten die hij gebruikt. In een dergelijk geval is het ook nuttig om waarschijnlijkheden of betrouwbaarheden bij de premissen in te lassen, en een of andere methode te bedenken om uit de betrouwbaarheden van de verschillende premissen, en rekening houdend met de betrouwbaarheid van de regel zelf, de betrouwbaarheid te berekenen van de conclusie.

Alvorens we de eigenschappen van deze voorstellingswijze door regels nog eens samenvatten, kunnen we opmerken dat het natuurlijk ook mogelijk is om verschillende systemen met elkaar te mengen. Zo kan men zich best indenken dat men in een frame systeem zekere kennis kan inbrengen onder de vorm van produktie-regels. De voordelen van de voorstelling met regels zouden we als volgt kunnen samenvatten:

- *Homogeen:* Alle kennis wordt op dezelfde manier voorgesteld en op dezelfde manier behandeld.
- *Menselijk model:* Er bestaat psychologische evidentie om aan te nemen dat het redeneren via de produktie-regels een weerspiegeling is van de menselijke redeneringswijze.
- *Declaratief:* De regels zijn een declaratieve weergave van kennis en dus probleemonafhankelijk.
- *Modulair:* Omdat elke regel een afzonderlijk stukje kennis voorstelt is de opbouw zeer modulair.
- *Incrementeel:* Het is eenvoudig om de regels uit te breiden met nieuwe regels. Dit heeft wel het nadeel dat men na verloop van tijd met een ontstellend groot aantal regels zit waarvoor consistentie enz. nog moeilijk controleerbaar is.
- *Verklarend:* Men kan eenvoudig de gevolgde weg bij de besluitvorming nagaan.
- *Onderhoudbaar:* Het systeem is zeer flexibel. Aanpassingen zijn eenvoudig te doen door regels te wijzigen of nieuwe regels bij te voegen. Dit kan echter leiden tot het hiervoor vermelde nadeel als het aantal regels te groot wordt.
- *Synthese-analyse:* Zowel synthese als analyse-problemen kunnen opgelost worden. Dit door de weg die tot de conclusie leidt voor of achterwaarts te volgen.
- *Onzekerheid:* Zoals gezegd heeft men de mogelijkheid om bij premissen en conclusies waarschijnlijkheidsfactoren in te bouwen. Dergelijke systemen gebruiken meestal een ad-hoc strategie om deze factoren te combineren.
- *Experte systemen:* Zeer vele experte systemen in de praktijk zijn op deze voorstellingswijze van kennis gebaseerd.

Nu komen we tot een laatste voorstellingswijze van kennis:

4. Wiskundige logica: Meer bepaald maakt men hier gebruik van de predikaten logica. In feite zijn de hoger besproken produktie-regels niets anders dan logische implicaties: deze implicaties zijn een bepaalde vorm van logische formules. In de predikaten-logica kan men nog andere logische formules construeren. Hierbij maakt men gebruik van verschillende symbolen: $\vee$ ("of"), & ("en"), $\Rightarrow$ ("impliceert"), $\neg$ ("niet"). In de logica worden ook verschillende *geldige afleidingsregels* bestudeerd.

Twee ervan die men zich vermoedelijk nog herinnert zijn de *Modus Ponens* en de *Modus Tollens*. Zij worden hieronder formeel weergegeven:

uit:	$A \Rightarrow B$	uit:	$A \Rightarrow B$
en:	A	en:	$\neg\, B$
volgt:	B	volgt:	$\neg\, A$
	(MP)		(MT)

Men kan zo allerlei afleidingsregels gaan bedenken, maar in 1965 ontdekte A. Robinson dat al de verschillende geldige afleidingsregels kunnen vervangen worden door één enkele regel: de Symbolisch kan men de resolutieregel als volgt weergeven:

uit:	$A \vee B \vee C \vee \cdots \vee K$
en:	$(\neg A) \vee P \vee Q \vee \cdots \vee Z$
volgt:	$B \vee C \vee ... \vee K \vee P \vee Q \vee ... \vee Z$

Uit de wiskundige logica weten we ook dat een implicatie $A \Rightarrow B$ kan geschreven worden als $\neg A \vee B$. Daaruit kunnen we onmiddellijk aantonen dat bijvoorbeeld de Modus Ponens afleidingsregel inderdaad een speciaal geval is van de resolutieregel:

uit:	$\neg A \vee B$
en:	A
volgt:	B

Bij deze illustratie van het gebruik van wiskundige logica voor kennisrepresentatie, kunnen wij de notatie van de programmeertaal PROLOG gebruiken. In deze taal worden de implicaties genoteerd als: $\Leftarrow$ (de conclusies staan dan links en de premissen of condities staan rechts). Een prolog uitdrukking zoals:
$A \Leftarrow (B_1 \,\&\, B_2 \,\&\, \cdots \,\&\, B_n)$
kan ook geschreven worden als:
$A \vee \neg B_1 \vee \neg B_2 \vee \cdots \vee \neg B_n$.
Hieruit volgt dat de resolutieregel ook op de volgende manier kan geschreven worden:

uit:	$A \Leftarrow (B_1 \,\&\, B_2 \,\&\, \cdots \,\&\, B_n)$
en:	$B_1 \Leftarrow (C_1 \,\&\, C_2 \,\&\, \cdots \,\&\, C_m)$
volgt:	$A \Leftarrow (C_1 \,\&\, C_2 \,\&\, \cdots\, C_m \,\&\, B_2 \,\&\, \cdots \,\&\, B_n)$

Rond 1971 kwam een onderzoeker in A.I., R. Kowalski, tot het inzicht dat bovenstaande schrijfwijze van de resolutieregel, naast zijn "declaratieve" (of logische) betekenis, ook nog een "procedurale" interpretatie toelaat, namelijk:

eerste regel: als men de relatie A wil bewijzen, dan kan dat gebeuren door de relaties B_1 B_2 ... B_n te bewijzen.

tweede regel: als men de relatie B_1 wil bewijzen, dan kan dat gebeuren door de relaties C_1 t.e.m. C_m te bewijzen.

Toepassing van resolutie komt er dan op neer dat men de relatie A kan bewijzen, door de relaties C_1 t.e.m. C_m B_2 t.e.m. B_n te bewijzen. Een typische redenering in een dergelijk systeem is bijvoorbeeld als volgt: Zij gegeven:

```
grootvader(X,Y)~\Leftarrow~vader(X,Z)  & vader(Z,Y)
vader(tom,bob)
vader(bob,pat)
```

De eerste regel kunnen we lezen als: X is de grootvader van Y indien X de vader is van Z en Z de vader van Y. De tweede regel zegt dat tom de vader is van bob en de derde regel zegt dat bob de vader is van pat. Om nu te bewijzen dat men ook heeft:

```
grootvader(tom,pat)
```

dit wil zeggen dat tom de grootvader is van pat, gaat men in twee stappen te werk: eerst stelt men X =tom en Y =pat in de eerste regel:

```
grootvader(tom,pat) ⇐ vader(tom,Z) & vader(Z,pat)
```

In een tweede stap zal men Z door bob vervangen en regels twee en drie gebruiken om het bewijs af te leiden.

Dit eerste voorbeeld toont aan hoe een Prolog systeem het antwoord kan vinden op een "ja-of-nee" vraag (namelijk: "is tom de grootvader van pat?"). Het resolutieprincipe kan ook gebruikt worden om andere vragen te beantwoorden, waarin onbekenden voorkomen, zoals: "wie is de vader van pat?". Deze vraag wordt geschreven als:

```
vader(X,pat)
```

De methode die hierbij gevolgd wordt, is het klassieke "bewijs uit het ongerijmde". Men veronderstelt eerst dat de gestelde vraag *geen antwoord* heeft (m.a.w. "er bestaat geen waarde voor de veranderlijke X waarvoor geldt: `vader(X,pat)`"). Het resolutieprincipe wordt gebruikt om te bewijzen dat deze veronderstelling aanleiding geeft tot een contradictie: gedurende dit bewijs wordt de variabele X echter vervangen door een bepaalde waarde, namelijk de gezochte waarde! In ons voorbeeld zou dat "bob" zijn.

Bovenstaande voorbeeld illustreert hoe de notatie en de technieken van de wiskundige logica kunnen gebruikt worden om kennis voor te stellen, en deze kennis effektief te gebruiken om geldige deducties te maken.

We vatten tenslotte enkele belangrijke voordelen van deze laatste kennisvoorstelling samen:

- *Gekend formalisme:* De wiskundige logica is goed gekend en wordt reeds sinds zeer lang bestudeerd. Men kent allerlei eigenschappen i.v.m. volledigheid, beslisbaarheid, bewijsbaarheid enz.
- *Homogeen kader:* De voorstelling en de verwerking van de kennis is zoals bij het gebruik van produktie-regels zeer homogeen.
- *Semantisch model:* De wiskundige logica biedt een formeel model waarin de correctheid van de deductieprocessen kan nagegaan en bewezen worden.

- *Implementatie:* De taal PROLOG is een computertaal die ontworpen is om precies met deze kennisvoorstelling te werken.

Toch blijven er ook een aantal problemen bij deze manier om kennis voor te stellen. Vermelden we bijvoorbeeld:

- de *monotoniciteit* van de logica: het toevoegen van axioma's heeft nooit tot gevolg dat een vroegere deductie ongeldig wordt; dit in tegenstelling met onze dagelijkse ervaring, waarbij de beschikbaarheid van bijkomende kennis kan leiden tot het verwerpen van eerder gemaakte "deducties".
- het werken met *onzekerheden* valt buiten het kader van de formele logica; om dit te behandelen moeten uitbreidingen voorgesteld worden.
- ook om het begrip *"tijd"* correct te kunnen behandelen is het aangewezen nieuwe uitbreidingen ("temporal logic") in te voeren.

Besluit

Gedurende de laatste tien jaar hebben meer en meer onderzoekers uit de A.I. zich toegelegd op het probleem van kennisvoorstelling. Verschillende formalismen werden hierbij voorgesteld, die elk voor- en nadelen vertonen. Nieuwe programmeertalen (zoals PROLOG) en zeer complexe programmeeromgevingen voor de ontwikkeling van experte systemen (zoals "Knowledge Engineering Environment" (KEE), "Knowledge Craft", S1,...) werden hierbij ontwikkeld. Het commerciële succes van enkele experte systemen toont aan dat er reeds betrouwbare resultaten werden bereikt. Toch zijn de onderzoekers nog verdeeld over de vraag welk formalisme nu het best geschikt is voor een bepaald probleemgebied, en hoe men de verschillende formalismen tot een homogeen geheel kan combineren. Er is dus nog altijd plaats voor interessante research!

Mechanisering en denken

H. Roelants, Hoger Instituut voor Wijsbegeerte, K.U.-Leuven

Mijn uiteenzetting betreft een problematiek waarvan de wortels ver terugreiken in de geschiedenis van het denken, de vraag nl. naar de verhouding mens-machine; een problematiek die door recente ontwikkelingen in verschillende wetenschappen en in het bijzonder door de filosofische discussie die is ontstaan rond Artificial Intelligence een nieuwe actualiteit heeft gekregen[1]. Het gaat om de bespreking van een belangrijke etappe in de evolutie van de opvattingen over de verhouding natuurzijnde - artefact, (phusis - techne in de oude Griekse terminologie). Met de wetenschapstheoreticus Dijksterhuis kan de overgang van een organische natuurvisie (waarin de natuur gezien wordt in analogie met een levend organisme) naar een mechanistische natuuropvatting (waarin de natuur gezien wordt in analogie met een machine) gekarakteriseerd worden als "de mechanisering van het wereldbeeld". Deze overgang hangt, zoals bekend, samen met de opkomst van de moderne natuurwetenschap sedert de 17e eeuw. Mij zal het op de eerste plaats gaan om een *betrekkelijk recent aspect* hiervan, nl. de pogingen tot mechanisering van het creatief menselijk denken. Ik zal o.m. argumenteren dat de mogelijkheid om een band te leggen tussen "denken" enerzijds en "automaat" anderzijds slechts echt zinvol kon worden nadat het begrip "machine" een belangrijke inhoudelijke wijziging had ondergaan. Dit is een zaak geweest van de laatste vijftig jaar: bepaalde ontwikkelingen in o.m. de symbolische logica, de zogenaamde algoritmentheorie en de computerwetenschappen hebben geleid tot het leggen van zo'n band tussen een nieuwe inhoud van "automaat" (het duidelijkst belichaamd in het begrip van een universele digitale computer) en bepaalde creatief genoemde aspecten van het menselijk denken en handelen. Het leggen van zo'n band heeft uiteraard ver-reikende gevolgen en is ook niet zonder kritiek gebleven. De duidelijkste en meest geprononceerde belichaming van die band vindt men in het domein van het AI-(Artificial Intelligence) onderzoek. Ik zal nader ingaan op een aantal achtergronden van die evolutie en op enkele kritische beschouwingen. Eerst wil ik echter nog een mogelijk misverstand elimineren. Met de tegenwoordig erg courante term 'computer-revolutie' kunnen allerlei soorten problemen in verband gebracht worden, en voornamelijk twee belangrijke probleemgebieden:

- Vragen omtrent de maatschappelijk-ethische aspecten van die revolutie. Deze problematiek van het klassieke sprookje van de leerling-tovenaar maakt deel uit van een algemene ondervraging over de techniek en over de technocratie. Hoe belangrijk deze problematiek ook is, ik zal ze vandaag niet aan de orde stellen.

- Anderzijds vragen met een meer kennistheoretische gerichtheid, belichaamd in zulke prototypische (hoewel niet noodzakelijk klare) voorbeelden als: is de mens een machine? Kunnen machines denken? Ook hier zou men de problematiek kunnen illustreren met een klassiek sprookje als bv. dat van de keizer van China en de nachtegaal.

Het is tot enkel aspecten van *deze* problematiek dat ik mij verder beperk en ik hoop o.m. te kunnen duidelijk maken, dat bepaalde traditionele filosofische problemen en argumenten in een nieuw en intrigerend kleedje terug geactualiseerd worden.

I. Enkele achtergronden

1. De Aristotelische en middeleeuwse fysica is o.m. gekenmerkt door een tegenstelling tussen phusis en techne, tussen het natuurlijke en het kunstmatige, tussen 'fysica' en 'mechanica'. De natuurdingen (de dingen die 'van nature zijn') worden geacht een intern ontwikkelingsprincipe in zichzelf te hebben, de technische produkten niet. Een Aristotelische boutade stelt het kort en frappant: 'Een mens brengt een mens voort, maar een bed geen bed'. De natuur is gekenmerkt door zelfontplooiing en daartegenover staan de produkten van de kunde als datgene wat door de mens met behulp van vakkennis (mechanica) wordt voortgebracht: dit laatste heeft geen natuurlijk wezen, maar enkel een kunstmatige vorm. De technische kunde kan wel natuurlijke dingen imiteren, maar niet tot stand brengen, want ze bezit het daartoe nodige voortbrengingsvermogen niet. Imiteren is nooit evenaren, laat staan overtreffen. Waar de natuur voortbrengt, kan de kunde slechts samenvoegen.

2. Na de middeleeuwen ziet men een revolutie starten die als vaandeldragers o.m. Francis Bacon en Descartes heeft, en die als een centrale leidraad stelt dat de werkingen van de kunde niet wezenlijk verschillend zijn van deze van de natuur en dat hetzelfde dus ook moet gelden voor de voortbrengselen van beide. De meest extreme formulering vinden we bij Descartes: 'Alle regels van de mechanica gelden ook in de fysica, zodat alle dingen die kunstmatig zijn, tegelijk natuurlijk zijn. Want wanneer een uurwerk de tijd aangeeft met behulp van de raderen waarmee het gemaakt is, dan is dit voor het uurwerk niet minder natuurlijk dan het voor een boom is vruchten te dragen'. Ik zal verder nog terugkeren op de rol die automaten hebben gespeeld in het cartesiaans denken maar ik wil eerst kort een aanwijzen tussen de manier waarop ten tijde van Francis Bacon (begin 17de eeuw) het denken van de verhouding 'natuurlijk' - 'kunstmatig' een wijziging onderging, en de manier waarop hedentendage door sommigen de verhouding 'creatief denken' - 'mechaniseerbaar denken' wordt gezien. Nog ten tijde van Bacon gelukte het de scheikunde (een vorm van technè) totdantoe natuurlijk geachte verbindingen 'artificieel' te produceren. Van oudsher had men een tegenstelling gesteld tussen 'natuurlijke verbindingen', waarin de bestanddelen 'opgelost' zijn en waarbij er nog slechts een door een natuurlijk voortbrengingsproces ontstaan homogeen produkt voorhanden is; en anderzijds kunstmatige schijnverbindingen (zoals glas bv.) waar men ogenschijnlijk een homogeen produkt had, maar in werkelijkheid een samenstelling van uiterst fijn verdeelde heterogene bestanddelen. Welnu, meer en meer tevoren als natuurlijk aangeziene verbindingen worden door de scheikundige kunde van de 17de eeuw kunstmatig samengesteld. En de grens tussen natuur en kunst wordt daarmee verlegd naar de grens tussen enerzijds datgene wat kunstmatig kan uiteengehaald en weer samengesteld worden en anderzijds datgene waarbij dit onmogelijk is. Men haalt op die manier natuurlijk het Trojaanse paard binnen, want op die wijze beslist om

zo te zeggen het laboratorium over wat 'natuurlijk' en wat 'mechanisch' is. En het is niet verwonderlijk dat dit er uiteindelijk toe zal leiden dat de 'naturen' geleidelijk zullen geëlimineerd worden uit de scheikunde en dat in zekere zin *alle* (anorganische) verbindingen als mechanisch zullen worden aangezien.

Ik meen dat er een frappante analogie is tussen de vermelde episode en de wijze waarop bepaalde AI-protagonisten heden ten dage de menselijke creativiteit benaderen. Ik citeer ter illustratie een tekst van Marvin Minsky uit 1959: 'In onze poging om intelligente machines te ontwerpen zijn we inderdaad bezig met de problemen van "creativiteit". Veel mensen staan vijandig tegenover zulk onderzoek en stellen dat creativiteit (of intelligentie) een soort "gave" is die gewoonweg niet kan begrepen of gemechaniseerd worden. Deze houding kan slechts volgehouden worden door een voortdurende verschuiving in definitie. Zo gauw een of ander proces of handeling gemechaniseerd wordt. . . , wordt het afgevoerd. . . van de lijst van creatieve handelingen. Dat is volkomen redelijk: eens dat een proces wordt gemechaniseerd, zijn termen als "creatief" niet langer nodig voor de beschrijving ervan. . . De zwakheid van de voorstander van een onverklaarbare creativiteit ligt in de niet gesteunde overtuiging dat er, nadat *alle* mogelijke machines onder ogen zijn genomen, nog steeds een aantal items op de lijst zullen blijven staan'.

3. Mens en automaat

3a) Machine en leven De mythe van Narcissus (de jongeling die verliefd werd op zijn eigen spiegelbeeld) resoneert doorheen de menselijke geschiedenis. Het thema van het op een min of meer artificiële wijze van creëren van wezens 'naar zijn beeld en gelijkenis' heeft de mens steeds geïntrigeerd. Reeds de Griekse en andere mythen getuigen daarvan en dichter bij ons de vele varianten op de thema's van de Golem, de homunculus, Frankenstein e.d.

Daarnaast zijn er natuurlijk de 'echte' automaten geweest en deze hebben zeker mee een rol gespeeld in de 'mechanisering van het wereldbeeld' meer in het bijzonder in de ontwikkeling van *mechanistische visies op de levensverschijnselen*. Ik wil daar zeer kort even bij blijven stilstaan om nadien op de geschiedenis van de mechanisering van het *denken*terug te keren, want er zijn interessante parallellen te trekken.

In de 17de en 18de eeuw heeft de automaat model gestaan voor de levensverschijnselen. en dit is op de meest uitdrukkelijke wijze gebeurd in het cartesianisme en de uitlopers ervan. Niet alleen heeft de automaat en de terminologie die ermee verband hield, een reeks belangrijke *metaforen* geleverd om te spreken over de levensverschijnselen, maar hij heeft ook een verklarende rol gespeeld t.o.v. die verschijnselen. In dat proces is de invloed van het werk van Descartes centraal: hij is de eerste die van het automaatmodel een *theoretisch denkinstrument* bij uitstek heeft gemaakt: het is een Leitmotiv van zijn fysica, van zijn fysiologie en van zijn theorie der diermachines. De automaat dient er als een soort *speculatieve of theoretische microscoop* die een 'visie in de diepte' toelaat: de interne anatomie en de bewegingen worden er door 'gevisu-

aliseerd'; men treedt in het binnenste van de doos van het uurwerk. En men 'ziet': de aderen die als orgelpijpen zijn, de zenuwen die gelijken op koorden of buizen, de hersenen die een zeil vormen, het hart dat een grote veer is... En nog in de *Principia* kunnen we lezen dat het enige verschil tussen de machines die de mensen maken en de natuurlijke lichamen hierin ligt: 'dat de werking van de machines afhangt van de opstellingen van de buizen, veren en andere instrumenten die, omdat ze enige proportie moeten hebben met de handen van hun makers, altijd groot genoeg zijn, zodat hun figuren en bewegingen zichtbaar zijn, terwijl de buizen en veren die de oorzaak zijn van de werkingen der natuurlijke lichamen gewoonlijk te klein zijn om door onze zintuigen waargenomen te worden...'; en de *Traité de l'homme* specifieert dat de functies van het menselijk lichaam volgen uit de opstelling van de organen 'zoals de bewegingen van een uurwerk of een andere automaat volgen uit de opstelling van zijn tegengewichten en van zijn raderen'. De automaat levert het model voor een klaar en onderscheiden inzicht, want hij is een constructie die doet begrijpen 'par figures et mouvements'. Het inwendig lichaam wordt opengelegd voor de ogen als een geheel van vertrouwde werktuigen en bekende technische werkingen (verwarming, distillatie, scheiding*ldots*).

Ik kom nu tot de geschiedenis van de mechanisering van het denken.

3b) Mechanismen en denken. De idee van de mogelijkheid van mechanisering van bepaalde vormen van gaat ver terug. Men denkt hierbij wellicht op de eerste plaats aan het proces van *rekenen*: aan de abacus, de rekenmachines van Kepler, Pascal, Leibniz, Babbage, enz. Maar ook de idee van mechanisering van meer creatieve vormen van denken (dan rekenen) heeft een voorgeschiedenis lang vóór het silocoonparadijs. Een groep Arabische astrologen wordt gekrediteerd met de constructie (ergens tussen de 8ste en de 12de eeuw na Kristus) van een soort denkmachine, *zairja* genaamd, en die gebaseerd was op een schema waarbij de letters van het Arabisch alfabet ideeënklassen voorstelden. Wellicht haalde de 13de eeuwse Raymondus Lullus daar een deel van de inspiratie voor zijn concept van *Ars magna*, het concept nl. van een schema waarin een reeks kenniskategorieen werden geplaatst in segmenten van concentrische cirkels, die dan in verschillende standen t.o.v. elkaar konden worden gezet, de idee dus van een systeem dat op een mechanische wijze begrippen kan combineren zodat alle mogelijke alternatieven exhaustief aan bod komen.

Lang voor de huidige critici van AI is er al patent genomen op een bepaalde vorm van zo'n kritiek, nl. door Jonathan Swift, waar hij in *Gulliver's reizen* (1727) een burleske parodie op Lullus' *Ars Magna* brengt: Gulliver bezoekt gedurende zijn reis naar Laputa de Akademie van Lagado (Part III, Hfdst. 5) waar hij o.m. een professor vindt die een 'denkmachine' aan het demonstreren is: 'a project for improving speculative knowledge by practical and mechanical operations' noemt hij dat. En de uitvinder gaat verder: 'Iedereen weet hoe moeizaam de gewone methode is om zich kunden en wetenschappen eigen te maken; welnu deze uitvinding zou de domste persoon met weinig kosten en met een kleine lichamelijke inspanning in staat stellen boeken te schrijven over filosofie, poëzie, politiek, recht, wiskunde en theologie zonder enige noodzak begaafd te zijn of gestudeerd te hebben'. Een aantrekkelijk perspectief uiteraard!

Wij vinden de discussie rond de mogelijkheid van mechanisering van bepaalde vormen van creatief denken op een meer gesofisticeerde wijze terug bij enkele grote denkers van de 17de eeuw die zich met het probleem van de *ontdekking van nieuwe kennis* (en dus eigenlijk met de problematiek van een ontdekkingslogica) hebben beziggehouden: ik bedoel hier in het bijzonder Descartes, Leibniz en Tschirnhaus. Hun gaat het in dat verband niet zozeer om de constructie van materiële machines als om het ontwerpen van een denkinstrument dat de beregeling van bepaalde vormen van het creatief denken zelf naar voor zou brengen. In een hedendaagse terminologie gesteld en met bepaalde reserves: het gaat in zekere zin om een project van *algoritmisering* van het creatief denken, om het zoeken naar een algemene heuristiek, een 'ars inveniendi', een algemene systematische 'methode' van 'problem solving'.

En, en passant gezegd, het zou bijzonder belangwekkend zijn de *Regulae* van Descartes (1628) en de *Medicina mentis sive artis inveniendi praecepta generalia* van Tschirnhaus (1687) vanuit een hedendaags perspectief van 'problem solving' te bestuderen. Aan de peripetieen die de idee van een ontdekkingslogica in de geschiedenis heeft gekend, zou een aparte studie kunnen gewijd worden. Ik wil in dit verband enkel twee dingen beklemtonen:

- Het is bijzonder frappant dat iemand als Descartes met zijn fundamentele interesse in automaten het onmogelijk achtte een intelligente automaat te construeren (en ik kom verder terug op zijn 'onmogelijkheidsargumenten') *maar* terzelfdertijd een regelsysteem zocht voor het creatieve denken, zonder de band te kunnen of willen leggen tussen beide.
- De hedendaagse ontwikkeling van AI-onderzoek is meen ik *ook* o.m. de meest recente figuur van de idee van een ontdekkingslogica.

De tijd ontbreekt om in enig detail de meer recente voorgeschiedenis van AI in de 20ste eeuw te behandelen. In zo'n behandeling zou aandacht moeten gegeven worden aan de preciseringen van het algoritmebegrip (o.m. met behulp van een 'machine van Turing'), aan de onderzoekingen van Polya in verband met de heuristiek van de wiskune (rond 1950), an de ontwikkeling van de informatietheorie (Shannon) en van de cybernetica (Wiener) en aan heel wat andere dingen. Een centraal belangrijke zaak echter, die afzonderlijke vermelding verdient, betreft de ontwikkeling van de *digitale computer* (zo genoemd omdat ze hun informatie opslaan onder de vorm van bitstrengen).

Het concept van AI zal een preciesere inhoud kunnen krijgen op het ogenblik dat de digitale computer zal gezien worden als een *symboolmanipuator*. Met die evolutie vindt naar mijn mening een nieuwe fase in de 'mechanisering van het wereldbeeld' haar ontstaan. Langzaamaan realiseert men zich dat een bijzonder belangrijk aspect van 'mechanisme' niet op de eerste plaats te maken heeft met overdracht van materiële kracht of energie, maar met het omzetten van (symbolische) informatie. En zo komt langzaamaan een verbreding van de inhoud van de begrippen 'machine' en 'automaat' tot stand: van 'transformator' van kracht of energie naar 'symbolenmanipulator'.

Een mijlpaal in de ontwikkeling van de idee an AI blijkt de zogenaamde Dartmouth

Conference van 1956 geweest te zijn. Deze bracht een aantal onderzoekers uit zeer verschillende domeinen samen die het gemeenschappelijk geloof deelden dat wat 'denken' genoemd wordt ook buiten het menselijk hoofd kan gebeuren, dat het begrepen moest worden op een formele wijze en dat het beste niet-menselijk instrument om het te verwezenlijken de digitale computer was. Tot die groep behoorden John McCarthy (die de term AI toen introduceerde, hoewel die heel wat tegenstand opriep), Marvin Minsky, Claude Shannon, A. Samuel, Allen Newell, Herbert Simon, e.a.

A posteriori gezien was de bijdrage van Newell en Simon ongetwijfeld de belangrijkste, omdat zij op de proppen kwamen met een nieuw *paradigma* nl. het paradigma van een *'informatie-verwerkend systeem'*, een begrip dat later en tot zeer recent het onderzoek in AI zal domineren. Het programma dat Newell en Simon presenteerden was de *Logic Theorist* een programma dat in staat was theorema's uit de *Principia Mathematica* te bewijzen en dat daarbij 'menselijke werkwijzen' simuleerde. Wat Newell en Simon zich in feite realiseerden was dat de digitale computer kon beschouwd worden als een symbolenverwerker, m.a.w. als een instrument met immense niet-numerische mogelijkheden. Dat standpunt was–op Turings speculatieve positie na–nieuw en erg vruchtbaar. Het liet hun o.m. toe de menselijke geest te vergelijken met een computer in de zin dat de menselijke geest kon beschouwd worden als iets dat, met een bepaalde programma-input en data, zekere bewerkingen op die data uitvoerde en een output produceerde. Volgens Newell en Simon kan een algemene band gelegd worden tussen een systeem van symbolen die op bepaalde wijzen getransformeerd worden en zogenaamd 'intelligent gedrag'. Zo'n opvatting is zelfs volgens hen even centraal voor het begrijpen van de menselijke geest in de 20ste eeuw als Darwins principe van natuurlijke selectie geweest is voor het begrijpen van bepaalde biologische verschijnselen in de 19de eeuw. De opvatting van Newell en Simon impliceerde het inzicht dat *verzamelingen van eenvoudige processen in interactie een systeem konden produceren dat complex (nl. intelligent) gedrag manifesteert*[2]. In die zin kunnen ze met een boutade stellen dat de mens een ondersoort is van het geslacht der informatieverwerkende en symboolmanipulerende systemen. Symbolen liggen aan de basis van intelligent gedrag en de intelligentie van een systeem kan afgewogen worden aan zijn geschiktheid om gestelde doelen te bereiken ondanks variaties, moeilijkheden, complexiteiten, gesteld door de taakomgeving.

Het onderzoek van de vereisten waaraan een systeem moet voldoen om intelligent gedrag te produceren toont volgens Newell en Simon aan dat *er geen uniek elementair iets is* dat rekenschap geeft van intelligentie in al haar manifestaties. Er is m.a.w. evenmin een 'intelligentie-principe' als er een 'levensprincipe' is dat de essentie van het levende zou uitmaken. Maar de afwezigheid van een eenvoudige deus ex machina betekent niet dat er geen *structurele vereisten* zouden zijn *voor intelligentie*. En één van die basisvereisten is *de geschiktheid om symbolen op te slaan en te manipuleren*.

Samengevat brengt een opvatting als die van Newell en Simon mee dat de mens wordt gezien als een bepaalde verwerker van symbolische informatie, die gesofisticeerde cognitieve (creatieve) taken volbrengt als zoektochten tot het bereiken van bepaalde doelen, geleid door een groot aantal informele vuistregels, *heuristische regels*, die constraints leggen op dat zoeken. De bron van wat courant 'intelligentie' en 'creativiteit'

wordt genoemd wordt dan gezien als een zeer efficiënte exploratie van een a priori immense zoekruimte.

4. Situering van de AI-benadering in het zoeken naar een 'naturalistische' verklaring van het mentale.
Het lijkt mij goed er op te wijzen dat de ontwikkeling van de AI-benadering van het mentale kan geplaatst worden in de ruimere context van wat ik zou noemen het zoeken van een *naturalistische verklaring van (o.m.) menselijke ervaring en gedrag*. In die zin lijkt er mij een duidelijke verwantschap te bestaan tussen bepaalde filosofische problemen opgeroepen door zekere ontwikkelingen in de volutiebiologie (ik denk meer in het bijzonder aan de opkomst van de sociobiologie sedert een 10-tal jaren) en in het domein van AI. Ik licht dat parallelisme kort toe:

- Vanaf de 19de-eeuwse tot bepaalde hedendaagse sociobiologen, zijn er denkers geweest die op een nieuwe wetenschap van de mens aanstuurden gebaseerd op een naturalistische visie op de *homo sapiesn* als een dier(soort) onderworpen aan evolutie doorheen processen van de soort die Darwin in zijn *Order of Species* benadrukte: evolutie door natuurlijke selectie. De hoop is er dan o.m. bij die denkers dat men de 'geest' (mind, het mentale) zal kunnen verklaren als een produkt van evolutie. In elk geval gaat het hier principieel om een zoeken naar de plaats van 'mind' in 'nature'.

- Vanaf de 50-er jaren van deze eeuw is zoals gezegd de filosofische speculatie over de mind ook gericht geworden *door een ander naturalistische model* nl. dat van de computer. In verscheidene domeinen (cognitieve psychologie, AI) vindt men de aanname dat computermodellen van onze menselijke vermogens licht werpen op die vermogens. In het licht van een bepaalde visie op wat en machine is en de computer aanziend als een bepaald soort machine steunt dat model eveneens een naturalistische opvatting: als een machine onze mentale activiteit kan stimuleren, aldus enigszins extreem gesteld het argument, dan kan onze mentale activiteit niet als aanwijzing of bewijs dienen dat we meer zijn dan uiterst gecompliceerde fysische objecten. *In die zin* lijkt mij die ontwikkeling complementair aan bepaalde ontwikkelingen voortkomend uit de theorie der natuurlijke selectie.

II. Enkele fundamentele aspecten voor AI en discussie van bepaalde klassieke cartesiaanse arumenten

1. Het representatieprobleem
De keuze van de representatie betreft de wijze waarop een bijzondere taak vertaald wordt in informatiestructuren en informatieprocessen binnen de computer. Zo'n representatie kan zowel de oplossing efficient mogelijk maken als ze *praktisch* onmogelijk maken. Eén van de lange-termijnobjectieven van AI is de automatisering van de processen zelf die de beslissing nemen omtrent een bepaalde bijzondere representatie na het specificeren van de taak (eventueel in natuurlijke taal). Op dit ogenblik is dit nog

goeddeels een vrome gedachte en worden de informatiestructuren en informatieprocessen *ad hoc* gekozen voor de gegeven taken.

Het probleem van algemeenheid versus efficiëntie
Menselijke intelligentie is wezenlijk 'multipurpose', veelzijdig. De meeste projecten in AI hebben daarentegen een *betrekkelijk eng* programma met een *redelijk grote efficiëntie* tot resultaat gehad.

Een ander lange-termijn-objectief van AI is daarom *het ontwerpen van hoog efficiënte programma's met min of meer 'universele' toepasbaarheid*. Ook hier is de wens nog steeds goeddeels de vader van de gedachte. En iemand als Weizenbaum, zelf een centrale figuur in AI als ontwerper van één der eerste zogenaamde conversatieprogramma's (bekend onder de naam Eliza) en een berucht criticus van bepaalde extreme pretenties ervan (o.m. in zijn controversieel werk *Computer Power and Human Reason* van 1976) is niet mals in zijn kritiek op de zogenaamde *General Problem Solver* van Newell en Simon. Weizenbaum lijkt de mogelijkheid zelf van het bereiken van dat doel te betwijfelen.

Het lijkt mij in dit verband interessant *even een sprong terug te maken naar Descartes* want ik herken in bepaalde kritieken van Weizenbaum en anderen een geluid dat al te vinden was in Descartes' poging tot demonstratie van een onderscheid in natuur tussen de mens als 'res cogitans' (denkend ding) en de machines (en ook de dieren) anderzijds. Descartes beklemtoont het *universaliteitskarakter van de menselijke rede*. Dat is volgens hem (naast het beschikken over taal) *één van de criteria voor een wezenonderscheid*. In zijn vergelijking tussen automaat en mens wijst Descartes erop dat 'hoewel machines meerdere zaken evengoed of zelfs beter dan elk van ons zouden kunnen doen, zouden ze zonder enige twijfel andere zaken niet kunnen doen, en daardoor zou men ontdekken dat ze niet handelen *vanuit kennis*, maar enkel door *de dispositie van hun organen*'. Descartes stelt m.a.w. tegenover de enge programmatie van de machine de universaliteit van de reacties die voortkomen uit een wezen dat met rede begaafd is: 'Want waar de rede een universeel instrument is, dat dienstig is in alle soorten situaties, hebben die organen nood aan een bepaalde bijzondere dispositie voor elke bijzondere actie; en daarom is het moreel onmogelijk. . . dat een machine in alle omstandigheden van het leven zou handelen zoals onze rede ons doet handelen'. Anders gesteld: de cartesiaanse automaten zijn *allemaal special purpose machines*.

Dit nog steeds inspirerende cartesiaans argument stelt duidelijk de *plasticiteit en universaliteit van de rationele antwoorden* tegenover *de instrumentele 'particulariteit' van de machine*.

Het zoekprobleem
Van een goed programma verwacht men dat het een sterke zoeksturing, dus een *sterke heuristiek* incorporeert. In praktisch alle AI-programma's is de heuristiek nog steeds gepreprogrammeerd. M.a.w. het is niet de machine die de heuristische regels ontdekt, selecteert en optimaliseert. Het performatieniveau van die programma's is dan ook

vooral bepaald door de ervaring, het inzicht en het geluk van de onderzoeker. Een wenselijker en 'menselijker' situatie zou deze zijn waarin de heuristische processen zelf geautomatiseerd zouden zijn, m.a.w. leerprogramma's die hun eigen heuristiek gradueel zouden verbeteren. Nodeloos te zeggen dat ook hier nog een lange weg moet worden afgelegd.

De voorstelling van verworven kennis
De vraag hoe verworven kennis en vaardigheden in een bepaald domein gestructureerd zijn en hoe ze kunnen gerepresenteerd worden, is een verder fundamenteel probleem voor AI en uiteraard een centrale zaak voor de constructie van zogenaamde 'expertsystemen'. Maar andere sprekers vandaag zullen op dit probleem meer in detail ingaan.

De communicatie mens-machine in natuurlijke taal
Het gaat hierbij in feite om een ganse klasse van problemen, en waarvan sommige volgens bepaalde prominente critici van AI niet gemakkelijk een oplossing zullen bekomen (Weizenbaum, Searle, Dreyfus, e.a.).

Weer herken ik in *bepaalde* van de hedendaagse kritieken op dit punt een echo van een interessante cartesiaanse argumentatie, die ik daarom ook presenteer: Descartes wijst als tweede onderscheidingskriterium tussen mens enerzijds en automaat en dier anderzijds op het bezit door de mens van een taal. Volgens Descartes kunnen machines 'geen woorden gebruiken of andere tekens, zoals wij doen, om onze gedachten aan anderen duidelijk te maken'. Uiteraard weet Descartes wel dat er automaten zijn die bepaalde klanken uitstoten, en dat er dieren zijn die bepaalde zinnen kunnen leren zeggen. 'Men kan wel denken dat een machine zodanig gemaakt is dat ze woorden uitspreekt en zelfs dat ze bepaalde woorden uitspreekt in verband met de acties die enige verandering in haar organen zullen teweegbrengen: bv. als men ze op een bepaalde plaats aanraakt, dat ze vraagt wat men wil zeggen; op een andere plaats, dat ze roept dat men haar pijn doet en zulke dingen meer'. M.a.w. Descartes beseft heel goed dat men het spreken kan *simuleren* met behulp van mechanische werkwijzen. Maar datgene waartoe de machine onbekwaam is, is 'de diversiteit om te reageren op de betekenis van alles wat in haar aanwezigheid zal gezegd worden, waartoe zelfs de meest idiote mensen in staat zijn'. In dit argument wordt dus een dubbele voorwaarde benadrukt:
– 'antwoorden op betekenis' (als het eigene van de menselijke taal)
– en dit slechts te kunnen door 'des arrangements' die zo verscheiden zijn dat geen enkel mechanisch proces daartoe in staat zou zijn.

Voor Descartes zal de automaat slechts kunnen worden begiftigd met een taal zonder 'arrangement ni diversité' en zonder 'réponse au sens' en dus met een schijntaal. In Descartes' argument lijkt het er niettemin op dat voor hem het 'antwoorden op betekenis' ten nauwste verbonden is met die 'arrangements divers' en dat dit ruimte laat voor de mogelijkheid dat een 'antwoorden op betekenis' de tegenhanger zou kunnen zijn van een immens grotere capaciteit van de 'arrangementen' (dan die waartoe de automaten van zijn tijd in staat waren). Vanuit AI-standpunt een aantrekkelijke gedachte. Een nutteloze maar niet zinloze vraag zou hier kunnen zijn: hoe zou Descartes hebben

gereageerd op de nieuwe inhoud van 'automaat' belichaamd in het concept van een universele digitale computer en gekarakteriseerd door een capaciteit van 'arrangements divers' zonder precedent?

III Searle's kritiek en een intrigerend gedachtenexperiment[3]

In dit deel van mijn uiteenzetting zal ik, tussen de veelheid en de diversiteit van actuele kritische stemmen t.o.v. AI kort het skelet schetsen van en enig commentaar geven bij één der meest interessante kritieken, nl. die van J. Searle. Searle's kritiek is niet gericht tegen wat hij 'weak AI' noemt: de idee nl. dat de digitale computer een uitzonderlijk nuttig hulpmiddel is bij de studie van de 'human mind'. Wel *tegen 1strong AI'* waarvan volgens Searle aldus kan geformuleerd worden: de geschikt geprogrammeerde computer heeft letterlijk een 'mind' in de zin dat van computers met de geschikte programma's letterlijk kan gezegd worden dat zij *verstaan* en andere zogenaamde mentale toestanden hebben die door Searle onder de generische term intentionaliteit of intentioneel gedrag worden gerangschikt ('intentionaliteit' is voor Searle een naam voor een ganse familie van mentale toestanden, nl. die toestanden die gericht zijn naar of betrekking hebben op objecten en standen van zaken in de wereld: toestanden als verstaan, geloven, vrezen, begeren, bedoelen, enz.). Eigenlijk – aldus Searle – stelt 'strong AI' dat wat men 'geest' noemt niets anders is dan zeer complexe digitale computerprogramma's en dat mentale toestanden eenvoudigweg computertoestanden zijn en mentale processen computatieprocessen.

Met die eerste centrale thesis van 'strong AI' verbindt Searle nog twee andere thesissen die hij eraan toeschrijft en waarvan hij meent dat zijn in 'cognitieve science' wijd onderschreven worden. Thesis 2 betreft de *irrelevantie van de specifieke neurofysiologie van de hersenen*[4]. Searle citeert in dit verband 'the emerging view of the mind as software or program, as an abstract sort of thing whose identity is independent of any particular physical embodiment'. Thesis 3 stelt de zogenaamde Turing-test als kriterium voor het mentale[5]: de idee dat het conclusief bewijs voor de aanwezigheid van mentale toestanden en capaciteiten ligt in de geschiktheid van een systeem een competente expert er te kunnen van overtuigen dat het systeem mentale toestanden bezit.

Nu argumenteert Searle als volgt tegen de stellingen:

1. *Intentionaliteit* in menselijke wezens (en in dieren) *is een produkt van causale aspecten van de hersenen.* Dit is volgens Searle een empirisch gegeven omtrent de feitelijk causale relaties tussen mentale processen en hersenen, m.a.w. een empirisch feit over hoe een deel van de wereld ineen zit en werkt. Kortweg: 'mind' is een konkreet biologisch verschijnsel. En intentionele toestanden worden veroorzaakt door en gerealiseerd in de structuur van de hersenen.

2. *Instantiëring (afwikkeling) van een computerprogramma is uit zichzelf nooit een oldoende voorwaarde voor intentionaliteit.* Ter staving van deze centrale premisse werkt Searle het gedachtenexperiment uit van de 'Chinese kamer', dat moet

aantonen hoe een mens een programma kan instantiëren zonder de relevante intentionaliteit te hebben.

3. Uit 1 en 2 volgt dan dat de verklaring van hoe hersenen intentionaliteit produceren niet kan zijn dat dit gebeurt door het instantiëren van een computerprogramma.

4. Gelijk welk mechanisme dat in staat zou zijn intentionaliteit te produceren moet causale vermogens hebben die minstens gelijk zijn aan die van de hersenen.

5. Pogingen om kunstmatig intentionaliteit te produceren kunnen niet slagen enkel door het produceren van programma's maar zouden moeten de causale vermogens van de menselijke hersenen dupliceren (een gevolg van 2 en 4).

Bijzonder belangwekkend is *het gedachtenexperiment ter staving van 2*: Onderstel dat een programma opgesteld wordt om het verstaan van Chinees (een bepaalde vorm van intentioneel gedrag) te simuleren zodat bv. als aan de computer vragen in het Chinees worden gesteld, het programma antwoorden kan geven in het Chinees, enz. Searle argumenteert nu dat geen (formeel) programma uit zichzelf enige intentionele inhoud heeft en de instantiëring ervan dus geen intentionele toestand kan produceren. Onderstel nl. dat ik, die hoegenaamd geen Chinees begrijp en zelfs geen Chinese symbolen kan onderscheiden van bv. Japanse, opgesloten wordt in een kamer met een aantal fiches vol met Chinese symbolen. Veronderstel dat men mij een regelboek in het Nederlands meegeeft dat mij leert hoe zulke Chinese symbolen aan elkaar te koppelen. De regels bv. dat een 'zus en zus' teken moet gevolgd worden door een 'zo en zo' teken. Onderstel dat mensen buiten de kamer mij bepaalde reeksen van Chinese symbolen doorgeven en dat ik, de instructies volgend, bepaalde strengen van Chinese symbolen terug aan hen aflever. Zonder dat dit mij bekend is, zouden de mensen buiten de kamer de eerste reeks symbolen een 'vraag' noemen, het regelboek 'het programma', de tweede reeks symbolen een 'antwoord op de vraag' en mijzelf zouden ze 'de computer noemen'. Onderstel dat na een tijd de programmatoren zo goed worden in het schrijven van de programma's en ik zo goed in het manipuleren van symbolen dat mijn antwoorden ononderscheidbaar worden van die van 'native Chinese speakers' (bv. van Deng Xiao Ping). Ik zal de Turing-test voor het verstaan van Chinees uiteraard met succes passeren. Maar niettemin versta ik nog geen woord Chinees en dat geldt evenzeer voor elke digitale computer gezien elke computer hetzelfde doet als ik nl. een formeel programma manipuleren dat geen betekenis toekent aan de gemanipuleerde symbolen. De draagwijdte van het argument is dus: geen formeel programma is uit zichzelf in staat tot 'verstaan' gezien het altijd in principe mogelijk is voor iemand doorheen de stappen van het programma te gaan zonder het relevante 'verstaan' te hebben. En wat geldt voor het verstaan van Chinees, geldt volgens Searle mutatis mutandis voor andere mentale verschijnselen. Daarenboven weerlegt dit argument en passant ook het simplistisch behaviorisme van de Turing-test door er nl. op te wijzen dat een systeem, nl. ikzelf, de Turing-test met succes kan passeren zonder de specifieke overeenkomstige mentale toestand te hebben.

De vele en zeer uiteenlopende reacties van AI-verdedigers op Searle's ingenieus gedachtenexperiment wijzen erop dat daarmee ongetwijfeld een zere plek is geraakt.

Slechts één soort van tegenkritiek vermeld ik even expliciet: een aantal antwoorden (o.m. van Hofstadter en Dennett) onderschrijft het zogenaamde 'Systems reply': de stelling, nl. dat de persoon in de kamer inderdaad geen Chinees verstaat maar wel het totale systeem waarvan hij deel uitmaakt (inbegrepen het instructieboek, de Chinese symbolen, enz.). Daarop antwoordt Searle weer dat het systeem niet in staat is betekenis toe te kennen aan de ongeïnterpreteerde Chinese symbolen (evenmin als de persoon het zelf kan). Het systeem, aldus Searle, heeft evenals de persoon 'een syntaxis maar geen semantiek'. En hij imagineert een subtiele variant van het elastische Chinese-kamer-experiment: Beeld u in dat de persoon het ganse systeem *internaliseert*; onderstel dat hij een supergeheugen heeft en een super-intelligentie zodat hij het instructieboek van buiten kan leren en alle computatiestappen in zijn hoofd doen. Vermits de persoon nog steeds geen Chinees verstaat en vermits er nu niets in het systeem is dat niet ook in de persoon is, is er geen manier waarop het systeem Chinees zou kunnen verstaan.

Ik zal mijn bespreking van het Chinese-kamerexperiment hier beëindigen zonder nog verder in te gaan op allerlei belangwekkende vragen en problemen die door Searle's intrigerende problematisering zelf worden opgeroepen.

IV. Een nieuw licht op oude vragen

Ik meen dat de filosofische discussie rond AI een aantal uiterst belangrijke vragen, die zelf niet geheel nieuw zijn, in een nieuw licht stelt. Ik vermeld er zeer kort enkele van: Wat is (de natuur van) een symbool en (van) een symbolische representatie? Wat is kennis? Hoe komt iets nieuws in de kennis tot stand? Wat is creativiteit? Wat is betekenis? Hoe kan een formeel systeem vasthaken in de wereld? Hoe moet de eenheid van een complex systeem gedacht worden? Wat is emergentie? Wat is reductie? Kan men emergentie van nieuwheid en reductie samen denken en hoe? Zonder op zulkdanige vragen een min of meer voldoeninggevend antwoord te hebben, kan men ook geen filosofische vat krijgen op vragen als: wat is een machine? Wat is een organisme? Wat is een 'mind'? Wat is de mens?

Besluit

De toekomst zal uitwijzen of we met de soort problematiek die ik heb in het licht gesteld, inderdaad staan, zoals sommigen het stellen, aan het begin van een volkomen nieuwe visie op de mens. Een visie die in een bepaalde zin nog fundamenteler (en 'bedreigender' zeggen anderen) zou zijn dan de evolutionistische visie, die enkel de unieke geprivilegieerde plaats van de mens in de biologische natuurorde op losse schroeven zetten. In een extreme AI-ideologie wordt de mens in zekere zin ontdaan van wat hem blijkbaar het meest wezenlijk leek te karakteriseren: zijn creatieve geestelijke vermogens. De vraag is of de ontwikkelingen in het domein van AI en in verwante domeinen en de produktie van intellectus ex machina inderdaad een begrippelijke revolutie zullen teweegbrengen in ons verstaan van de mens als animal rationale; ofwel of we opnieuw

een variant beleven van de nieuwe kleren van de keizer. En daarmee zijn we tenslotte weer bij een sprookje aangeland.

Noten

1. Er bestaat reeds een oeverloze literatuur rond AI. Naast de in de literatuurlijst vermelde boeken, kan verwezen worden naar J. David Bolter, *De mens van Turing*, Kapellen, De Nederlandsche Boekhandel ; en Douglas R. Hofstadter, *Gödel, Escher, Bach: een eeuwige gouden band*, Amsterdam, Contact, 1985. In mijn uiteenzetting heb ik gezien de noodzakelijke beperking recente belangrijke ontwikkelingen (aangeduid met termen als "Connectionisme", "Parallel Distributed Processing", "Neurale Netwerken") buiten beschouwing gelaten. Een aantal zaken werden uitvoeriger uitgewerkt in een bijdrage tot het boek *Over de mens. Vijf filosofische confereties*, Universitaire Pers, Leuven, 1987 (artikel: *Mens en machine*, p. 99-125). Heel wat interessante historische informatie putte ik uit het werk van P. McCorduck, *machines who think*.

2. Zie in dit verband Newell & Simon (1976).

3. Zie bv. Searle (1980).

4. In het licht van recente ontwikkelingen die dikwijls samengevat worden onder de benaming "connectionisme" lijkt althans wat dit punt betreft een aardverschuiving bezig te zijn.

5. Zie Turing (1950).

Literatuurlijst

- HAUGELAND J. (ed.) (1981) *Mind Design*, MIT Press.
- McCORDUCK P. (1979) *Machines who think. A personal inquiry into history and prospects of AI*, San Francisco, W.H. Freeman.
- NEWELL A. & SIMON H.A. (1976) Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the Association for Computing Machinery*, 1976 (19), 113-126. Ook in Haugeland (1981), 35-55.
- SEARLE, J. (1980/82) Minds, Brains and Programs, *The Behavioral and Brain Sciences*, 1980 (3), 417-457 ; en 1982 (5), 338-348.
- TURING, A.M. (1950) *Computing Machinery and Intelligence*, 1950 (59), 433-460. Herdrukt in Anderson, A.R. (ed.) *Minds and Machines*, 1964, 1-30.
- WEIZENBAUM, J. (1984) *Computer Power and Human Reason*, San Francisco, W.H. Freeman, 1976 ; Nederlandse vertaling: *Computerkracht en Mensenmacht*, Amsterdam, Contact.

Artificiële intelligentie menselijk bekeken

Géry van Outryve d'Ydewalle, Laboratorium voor Experimentele Psychologie, K.U.-Leuven

De bedoeling van deze bijdrage is van even te schetsen wat de bijdrage kan zijn van de psychologie als menswetenschap in de ontwikkeling van expert systemen. Wij vertrekken vanuit twee concrete situaties, waarvan een een bestaande werksituatie is en de andere een laboratorium experiment vertegenwoordigt. Die twee situaties vormen een aanloop om de complexiteit van de menselijke kennis, en de moeilijkheden om die te vatten beter te omschrijven. In een tweede deel maken wij een onderscheid tussen situaties met een quasi-exhaustieve kennis van de elementen, en situaties waarbij een dergelijke vorm van kennis niet mogelijk is. In die tweede situatie wordt de problematiek van kennisextractie voor de ontwikkeling van expert systemen cruciaal. De moeilijkheden bij de kennisextractie worden dan besproken in functie van een cognitieve benadering over de werking van de mentale processen: hierbij wordt dan duidelijk waarom kennisextractie een zo'n moeilijke onderneming is. Ten slotte schetsen wij de problematiek van de kennisrepresentatie.

Een zestal maanden geleden zochten wij een concrete situatie waarbij een expert systeem zou kunnen ontwikkeld worden, met de bedoeling ons beter vertrouwd te maken met het hele proces van ontwikkeling van een expert systeem. In het kader van een licentiaatsverhandeling koos P. Van Tricht de situatie van een transport-bedrijf waarbij dagelijks zich het probleem van "dispatching" voordoet. "Dispatching" is eigenlijk synoniem voor "werkverdeling". In het bedrijf dat wij bestudeerden, komt het neer op het volgende: er zijn een aantal werkopdrachten binnengekomen en er zijn dagelijks 40 vrachtwagens ter beschikking. Hoe wordt dit nu samengezet in een zo optimaal mogelijk schema per dag? Optimaal wil hier zeggen: zo snel mogelijk en zonder veel onkosten de opdrachten afwerken. In principe kan de dispatching slechts door een persoon uitgevoerd worden omdat het werkschema echt een kluwen is, waarbij alles met alles in verband moet gebracht worden. Het interesse van de situatie die wij bestudeerden was dat er slechts een "dispatcher" was en dat hij overtuigd was een nieuwe "dispatcher" in maximum tweetal weken te kunnen opleiden. Dus, een ideale situatie om de "dispatcher" te vervangen door een expert systeem. Wij dachten hiermee in een tweetal maanden klaar te komen.

Bij de functionele analyse valt de input-informatie in twee delen uiteen: enerzijds de opdrachten, en anderzijds de beschikbare wagens. Bij de werkopdrachten zit er altijd een deel "vast werk". Dit zijn de vrachten die wekelijks of zelfs dagelijks voorkomen en waarmee dus constant moet rekening gehouden worden. Daarnaast is het mogelijk om tot 9 uur 's morgens contact op te nemen met de "dispatcher" om verdere werkopdrachten door te geven of om wagens te reserveren omdat men er zeker werk voor zal hebben. Om deze taken uit te voeren zijn er nu 40 vrachtwagens beschikbaar. Hiervan zijn er vier die niets anders doen dan in de haven van Antwerpen van kaai tot kaai rijden. Een ander deel van deze 40 rijdt steeds binnen de Benelux. Het overgrote deel

zit echter in het buitenland, met Frankrijk en vooral Duitsland. Daarnaast zijn er nog een aantal chauffeurs voor wie het niets uitmaakt of ze in binnen- of buitenland rijden.

Zoals zo voorgesteld ziet alles nog vrij eenvoudig. Nog een hele reeks factoren beinvloeden echter de werkverdeling. Het is mogelijk dat er te veel werkopdrachten binnenkomen, dat het werk niet allemaal kan uitgevoerd worden door slechts 40 wagens. Dan is het nodig om andere transportbedrijven te contacteren en hen dan in zogenaamde "onderaanneming" te laten rijden. Anderzijds iet het ook mogelijk dat er te weinig werkopdrachten zijn binnengekomen, en dan moet de "dispatcher" zoveel mogelijk contact opnemen met container-bedrijven in de haven om zo meer werkopdrachten te vinden. Verder moet er een afwisseling zijn in de buiten- en binnenlandse reizen. Bij de buitenlandse reizen zijn er duidelijk langere en kortere. Het is nu best om iemand die thuis komt van een lange trip, niet dadelijk een nieuwe lange trip te geven (en omgekeerd is het dan ook zo dat meestal op een korte reis, een lange reis wenselijk wordt). Ook wordt door de wet voorgeschreven dat een vrachtwagen-chauffeur na vier uren rijden, steeds een uur moet rusten (om dit te controleren wordt er in de wagen een "schijf" gestoken, om nagelezen te worden na de reis). Voor buitenlands vrachtvervoer zijn er bepaalde vergunningen nodig, en het is vaak moeilijk om deze vergunningen te bekomen. Bijgevolg heeft niet elke buitenland-chauffeur een vergunning voor elk land. Daarmee moet de "dispatcher" dus ook rekening houden.

Van 6 uur 's morgens tot 9 uur 's avonds worden in de haven van Antwerpen containers opgezet en afgenomen. De snelheid waarmee men laadt en lost verschilt al van bedrijf tot bedrijf, zodat het voor de "dispatcher" nodig is om zijn klanten goed te kennen om dan te kunnen berekenen hoeveel tijd er zal verloren gaan met laden en lossen. Het kan nu gebeuren dat iemand die in het buitenland zit wegens omstandigheden niet voor 9 uur in Antwerpen kan zijn zodat hij zijn container voor de volgende reis niet meer krijgt. Er is nu afgesproken dat in zulke gevallen de chauffeur voor 6 uur 's avonds contact moet opnemen met de "dispatcher". Hij kan dan nog een andere chauffeur naar Antwerpen sturen om de container op te halen en naar het eigen bedrijf te brengen; de chauffeur die uit het buitenland komt, kan dan daar zijn container oppikken. Een container kan vervoerd worden door hem op een chassis te plaatsen en deze chassis dan aan een trekker te hangen. Er zijn nu verschillende maten in containers en dus ook in chassis. Wanneer de overblijvende chassis in het bedrijf niet voldoen, moet de "dispatcher" zich wenden tot bedrijven die chassis verhuren.

Ook moet de "dispatcher" lege reizen vermijden om de inkomsten te maximiseren: bijvoorbeeld, een chauffeur rijdt naar Duitsland om daar een container te lossen, maar er is geen andere container die hij van Duitsland moet terugbrengen naar Belgie of naar een ander tussenstation. Ook het tanken impliceert een hele strategie om de onkosten zo laag mogelijk te houden. De vrachtwagens kunnen op het bedrijf zelf mazout tanken. Aangezien deze in grote hoeveelheden aangekocht wordt, kan de prijs vrij voordelig gehouden worden. Het is dus best om zoveel mogelijk thuis te tanken en zo weinig mogelijk buitenhuis om op die manier de kosten te drukken. Nu is het zo dat men de Franse grens maar over mag met ongeveer 250 liter mazout in zijn tank, de Duitse grens met ongeveer 200 liter. Op elke extra-liter wordt tol betaald. De "dispatcher" moet dus de voordelen van een lange reis afwegen tegen de nadelen.

De “dispatcher” werd door ons uitvoerig ondervraagd. Die taak van werkverdeling voert hij ’s morgens uit, en hij vindt ze gemakkelijk. Bovenstaande informatie is wat wij verzameld hadden vanuit onze gesprekken met hem. Hieruit waren wij overtuigd snel een “flow-diagram” te kunnen opstellen, als basis van het expert systeem. Ook de “dispatcher” was de mening toegedaan dat met die informatie hij gemakkelijk “vervangen” kon worden. Wij hoopten met de ontwikkeling van het expert systeem in een tweetal weken klaar te komen. Het was daarop de bedoeling de werkverdeling van het expert systeem te vergelijken met die van de “dispatcher”. De ontwikkeling van het expert systeem werd echter een hachelijke onderneming, en dit niet omwille van de implementatie maar wel omwille van het ontbreken van nog heel wat informatie waarover de “dispatcher” beschikt en hiermee, veelal onbewust, dan ook rekening houdt in zijn werkverdeling. Telkens dat wij “vast” zaten, vroegen wij hem opnieuw uit. Hij was ons die informatie “vergeten” te zeggen. Hij beschikte dus over heel wat impliciete kennis, en die kennis verwoordde hij niet spontaan omdat hij hiermee zo snel, automatisch en op geroutineerde wijze rekening hield in zijn werkverdeling: die kennis vergde geen aandacht van hem.

Het is pas recentelijk dat die vorm van impliciete kennis ook in de experimentele psychologie bestudeerd werd. Hierbij onderzocht men vooral de discrepantie tussen enerzijds de gebruikte kennis in een taak en anderzijds de mogelijkheid om die kennis te verwoorden. Die discrepantie is niet alleen een probleem dat voor de theorie-vorming in de psychologie belangrijk is, maar heeft ook (en misschien vooral) een practische relevantie voor de ontwikkeling van expert systemen. De geprivilegeerde ingangspoort tot de kennis van menselijke experten blijft in de bedrijven dan toch hoofdzakelijk een “opvragen” of “ondervragen” van die experten. Indien zij dan bij die opvraging hun “impliciete” kennis niet kunnen weergeven, dan beschikt men over een informatiebestand die niet voldoende volledig is om het expert systeem te ontwikkelen.

Berry en Broadbent (1984) simuleerden die situatie van impliciete kennis waarbij de proefpersonen de productie van suiker in een suikerbedrijf moesten controleren. In het kader van het I.W.O.N.L. spilprogramma “Expert Systemen” hebben wij die proeven hier in Leuven herhaald. De opdracht was een bepaald niveau van productie te bereiken en dit productie-niveau dan te handhaven. Het productie-niveau bij een bepaalde proefbeurt (Px) werd bepaald door twee parameters, het aantal werknemers (W) en het productie-niveau van de vorige proefbeurt (Py), waarbij $Px = 20W - Py$. Om de formule minder transparant te maken werd ook een toevalsfactor ingebouwd. Uiteraard werden de proefpersonen niet op de hoogte gesteld van de onderliggende formule. Zij moesten het aantal werknemers bij elke proefbeurt instellen. Ofschoon de proefpersonen hun prestatie (t.o.v. het bereiken productie-niveau) systematisch over een lange reeks proefbeurten verbeterden, kwam die verbetering niet tot uiting in een expliciete kennis van de formule. Verschillende types van ondervragingen daarop mislukten om die kennis verwoord te krijgen. Wij hadden last om een verband te vinden tussen de geleverde prestatie op de taak en de antwoorden van de proefpersonen op een aantal gerichte vragen over hoe dat zij te werk gingen.

In een typische bedrijfssituatie moet een expert zijn kennis “doorspelen” aan een novice die de taak moet leren. Indien de expert er niet in slaagt zijn kennis te verwoorden

aan een proefleider zoals in bovenbesproken geval, wellicht is hij wel in staat een novice snel op te leiden. Daarom hebben wij bovenstaande situatie als uitgangspunt genomen om de verwoordingen van een getrainde proefpersoon t.o.v. een novice te bestuderen. Indien een getrainde proefpersoon zijn prestatie om het optimale productie-niveau kan verbeteren, dan blijkt hij dus niet in staat te zijn om de onderliggende kennis bloot te leggen in een ondervraging. Wellicht is hij dan misschien wel in staat om de cruciale aspecten van die kennis door te geven aan een novice. Getrainde proefpersonen die een bepaald niveau van expertise bereikten waren inderdaad wel in staat een novice op te leiden. De prestatie van die novices waren aanzienlijk beter dan de prestatie van de novices die geen "opleiding" kregen van een expert, maar ook beter dan de prestatie van de novices die een opleiding kregen van getrainde proefpersonen waarvan de prestatie nog geen expertise uitdrukte. Wanneer wij de verbale communicatie tussen de succesvolle "experten" en de novices nader analyseerden, konden wij de cruciale aspecten van die informatie met moeite terugvinden.

Belangrijk hier is dus de fase van kennisextractie bij experten. Om de betekenis van die kennisextractie-fase goed te vatten is het wellicht nuttig een onderscheid te maken tussen situaties met een exhaustieve kennis van het fysisch systeem en situaties waar die kennis niet aanwezig is. Wanneer een exhaustieve kennis van een fysisch systeem beschikbaar is (wij spreken verder van quasi-exhaustieve kennis, omdat een volledige kennis uiteraard nooit mogelijk is), dan zal de natuurlijke expert een aantal logisch-opeenvolgende of parallelle operaties uitvoeren om te kunnen inwerken op het systeem. Deze operaties zijn aangepast aan de ergonomische mogelijkheden van het mentaal-cognitief systeem van de natuurlijke expert. Deze operaties kunnen in algoritmen of heuristieken uitgedrukt worden. Heuristieken zijn methodes die informatie omtrent de natuur en structuur van een probleem gebruiken om een probleem eventueel naar een oplossing te brengen zonder dat er garanties zijn dat er altijd een oplossing gevonden wordt. Algoritmen daarentegen kunnen wij definieren als gehelen van regels die, wanneer ze in een bepaalde volgorde gebruikt worden, steeds een oplossing brengen. De formalisaties in algoritmen en heuristieken kunnen dan eventueel leiden tot het ontwikkelen van een expert systeem.

De omweg langs de werkwijze van de natuurlijke expert zal meestal niet nodig zijn wanneer een quasi-exhaustieve kennis van het fysisch systeem voorhanden is. Vanuit die kennis is het waarschijnlijk gemakkelijker een formalisatie na te streven, los van de natuurlijke expert. Uit die formalisatie kunnen aansluitend afleidingen worden gemaakt voor het implementeren van een efficiënt expert systeem. Zo zal bijvoorbeeld het automatisch detecteren van anomalieen in chromosoomkaarten niet noodzakelijk kennisextractie bij een geneticus die een chromosoomkaart inspecteert impliceren. Een omvattende kennis van de genetica maakt de omweg langs de werkwijze van een geneticus overbodig.

In feite definieren wij hierbij quasi-exhaustieve kennis als de situatie waarbij een algoritmische formalisatie van de interactie tussen een natuurlijk expert en zijn (fysisch) systeem althans in principe mogelijk is.

Toch is bij de quasi-exhaustieve kennis van het fysisch systeem de omweg langs

de werkwijze van de natuurlijke expert soms nog aangewezen. In een systeem kan het aantal combinatiemogelijkheden van elementen van operaties of kennisvoorstellingen zo groot zijn dat een volledige formalisatie niet aangewezen is. De natuurlijke expert (met zijn beperkt werkgeheugen) heeft in zijn permanent geheugen een aantal heuristieken beschikbaar om niet exhaustief alles te moeten aftasten. Wij denken hier aan de simulatieprogramma's van een grootmeester in het schaakspel: hij beschikt over een aantal 'shortcuts' waardoor niet alle mogelijke verdere consequenties van een bepaalde zet mentaal bekeken moeten worden. Hierbij rijst de vraag op welk moment de mens algoritmen of heuristieken gebruikt. Verder is het niet zeker dat de situaties die bij de natuurlijke experten eerder door heuristieken in plaats van algoritmen opgelost worden, overeenstemmen met de situaties waarbij expert systemen beter met heuristieken dan met algoritmen geimplementeerd worden. Indien heuristieken aangewezen zijn, dan stellen zich onder andere de volgende vragen: zijn de natuurlijke heuristieken een goede uitgangsbasis voor het ontwikkelen van heuristieken voor expert systemen, en kunnen heuristieken bij natuurlijke experten gemakkelijk geformaliseerd worden?

Zelfs indien algoritmen kunnen gebruikt worden, zullen expert systemen soms sneller en efficiënter, maar ook soms minder snel, minder efficiënt en minder flexibel reageren dan de natuurlijke expert. Hoe kan de natuurlijke expert in bepaalde gevallen karakteristieken vertonen van grotere efficiëntie, flexibiliteit en snelheid? Een grondige studie van de architectuur van het cognitief-mentaal functioneren van de mens kan een aantal aanwijzingen opleveren voor de wijze waarop de technische implementatie in expert systemen soms anders kan geconcipieerd worden. Het is en blijft uiteraard een open- empirische vraag (een vraag waarop wij nu geen antwoord kunnen geven) in welke mate de bijdrage van de gedragswetenschappen op termijn de ontwikkeling van betere expert systemen kan bevorderen.

Wanneer nog geen quasi-exhaustieve kennis van het fysisch systeem voorhanden is, heeft de natuurlijke expert, door zijn langdurige omgang, een kennisbestand opgebouwd dat hem toelaat optimaal met het fysisch systeem om te gaan. Kennisextractie bij de natuurlijke expert wordt hierbij uiteraard belangrijk.

Dit laatste kan op verschillende wijzen gebeuren: men kan steunen op observaties van de expert, men kan hem ondervragen (verbale rapporteringen als data), of men probeert langs indirecte wijze te weten te komen op welk soort kennisbestand hij steunt om te opereren. Wanneer men de natuurlijke expert observeert, dan stelt men handelingen vast die op het fysisch systeem inwerken om bepaalde gevolgen tot stand te brengen (handeling > fysisch systeem > consequenties). Hieruit kan men een aantal regels afleiden die de sequenties 'handeling > fysisch systeem > consequenties' weergeven. Het geheel van de regels zou een intrinsieke structuur moeten kunnen vertonen, omdat de natuurlijke expert in zijn handelingen op zijn kennis steunt om optimaal te reageren. Een inzicht in die intrinsieke structuur zou dan een formalisatie kunnen toelaten. Herhaaldelijk wordt echter vastgesteld dat zulk vastleggen van een aantal regels vanuit observeerbare gedragingen enkel aanleiding geeft tot het verzamelen van een amalgaam van disparate bevindingen die geen bredere conceptualisatie van het kennisbestand van de natuurlijke expert toelaat. Toch moet er ergens een inherente (intrinsieke) structuur zijn in diens mentaal functioneren.

Een andere mogelijkheid is het ondervragen van de natuurlijke expert. Ofschoon die werkwijze in sommige gevallen aangewezen kan zijn, stellen er zich hier toch ook problemen (zie de twee bovenbesproken voorbeelden). Vele handelingen van de natuurlijke expert zijn zo geautomatiseerd dat hij het moeilijk zal hebben om die te verwoorden. Meer nog, vele gedragsinterventies lenen zich niet tot verwoording. In tegenstelling tot sommige vormen van gedragswetenschappen staat de experimentele psychologie dikwijls nogal kritisch en wantrouwig ten opzichte van verbale rapporteringen. De mentale operaties van de mens ontsnappen voor een groot deel aan de bewuste beleving ervan, en zulke bewuste beleving is juist een noodzakelijke voorwaarde voor het rapporteren erover. Uiteindelijk zijn de bewuste belevingen en hun rapporteren slechts constructies ('fabricages') vanuit enkele kenniselementen die gekenmerkt worden door hun 'saliencies':

> *...available contents of STM (=onmiddellijk geheugen) are themselves the product of constructive processes. A theory of introspection needs to specify how 'heeding' and 'attending' operate and also how such processes determine (construct) specific conscious contents.; Whenever a conscious content is constructed that seeks to recover an intrapsychic series of events that occurred some time in the past, it should be obvious that such constructions will be influenced by current as well as past activations. As a result, introspective reports will very often reflect not what 'actually' happened, but rather what is the most reasonable construction at the moment of introspection.*
>
> (Mandler, 1985, pp.63-64).

Ericsson en Simon (1984) gaan uit van een expliciet model van de persoon als informatieverwerkend systeem en simuleren een aantal probleemoplossende gedragingen. Vanuit die gegevens stellen zij een aantal regels op voor situaties waarin kennisextractie in verbale rapporteringen al dan niet betrouwbaar kan doorgevoerd worden. Ofschoon zij hun benadering niet specifiek toegepast hebben op de betrouwbaarheid en zinvolheid van kennisextractie voor het ontwikkelen van expert systemen, opent hun werkwijze onderzoeksperspectieven inzake Artificiële Intelligentie. Hiermee wordt onder andere de haalbaarheid aangetoond om, vanuit computersimulaties en vanuit een model van een individu als informatieverwerkend systeem, regels op te stellen voor het afbakenen van de situaties waarin kennisextractie via verbale rapportering tot een betrouwbare informatie leidt. Recent hebben Broadbent, FitzGerald en Broadbent (1986) taakanalyses uitgevoerd om de limieten van de expliciete (beleefde) kennis bij handelingen af te bakenen. Daar waar expliciete kennis faalt, moeten andere technieken aangewend worden om aan kennisextractie te doen. Bij visuele inspectietaken probeerden wij zelf de bruikbaarheid van oogbewegingsregistraties aan te tonen (zie vooral d'Ydewalle, Van Rensbergen, en Pollet, 1987): oogbewegingen worden zo vlug uitgevoerd dat een bewuste beleving van de gevolgde strategie door de natuurlijke expert niet kan gerapporteerd worden. Bij niet-visuele taakopdrachten kan men verder beroep doen op het hele arsenaal van technieken uit de experimentele psychologie om aan de verbale

rapporteringen te kunnen ontsnappen.

De problematiek van de kennisextractie en impliciete kennis kan niet losgekoppeld worden van de problematiek van de representatie van die kennis. Wanneer wij ons de vraag stellen hoe wij uit ons permanent geheugen kennis oproepen, dan moeten wij ook weten hoe die kennis in ons permanent geheugen gerepresenteerd wordt. Er zijn heel wat modellen en theorieen ontwikkeld over de manier waarop kennis opgestapeld is in het permanent geheugen. Wat gemeenschappelijk naar voren komt in al deze opvattingen, is dat elke kennis voorgesteld kan worden als een element in een ruimer netwerk van relaties met andere kenniselementen. Naarmate het permanent geheugen beter georganiseerd is, zal de persoon ook minder "vergeten" omdat bij hem alles met alles in verband staat.

Om evidentie te vinden voor het bestaan van een hiërarchisch netwerk van relaties tussen kenelementen in het permanent geheugen werden heel wat studies verricht. Hiervoor maten Collins en Quillian (1969) verificatietijden voor zinnen die klasselidmaatschap uitdrukken zoals "een kanarie is een vogel" en zinnen die eigenschapsrelaties uitdrukken zoals "een kanarie heeft een huid". Dit paradigma heeft veel nader onderzoek gestimuleerd waarvan we hier de hoofdlijnen zullen weergeven. Voor twee woorden waartussen de semantische relatie-in-kwestie van kracht is–waarop dus het antwoord "ja" of "gelijk" van toepassing is–geldt dat de beslissing meer tijd kost wanneer hun semantische afstand toeneemt. Voor onware semantische relaties (met antwoord "neen" of "verschillend") geldt het omgekeerde verband. Collins en Quillian (1969) vonden in overeenstemming hiermee dat bevestiging van "een kanarie is een vogel" minder tijd kost dan "een kanarie is een dier", terwijl ontkenning van "een kanarie is een struisvogel" langer duurt dan van "een kanarie is een vis".

Cruciaal is hierbij het begrip "semantische afstand". Om "semantische afstand" te definieren, moeten we Collins en Quillians (1969) representatie van informatie in het permanent geheugen verduidelijken. Collins en Quillian (1969) stellen het permanent geheugen voor als een conceptueel netwerk. Er worden twee belangrijke veronderstellingen gemaakt over de aard van dat conceptuele netwerk. Ten eerste is het netwerk hiërarchisch gestructureerd. Dit betekent dat alle knopen die onder een bepaalde knoop x staan er–direct of indirect–mee verbonden zijn, tot de klasse x behoren. Bijvoorbeeld: de concepten "kanarie" en "struisvogel" behoren tot de klasse van de vogels. Men zegt dat een hogere knoop in een superset-relatie staat tot een lagere–bijvoorbeeld, dier staat in een superset-relatie tot kanarie–en dat een lagere knoop in een subset-relatie staat tot een hogere–bijvoorbeeld, kanarie staat in een subset-relatie tot dier. Ten tweede veronderstellen Collins en Quillian (1969) dat in het conceptuele netwerk het principe van overerving geldt. Dit betekent dat eigenschappen maar een keer opgeslagen worden in het netwerk en wel zo hoog mogelijk in de hiërarchie. De eigenschap "heeft veren" wordt niet opgeslagen bij het concept "kanarie", maar wel bij het concept dat hoger gelegen is, nl. "vogel". De eigenschappen die bij het concept x opgeslagen zijn gelden dan voor alle concepten waarmee het concept x in een superset-relatie staat. Men zegt dan dat de eigenschappen overgeerfd worden. Bijvoorbeeld: de eigenschap "heeft een huid" wordt niet expliciet opgeslagen bij de knoop die correspondeert met het concept "kanarie", maar "kanarie" erft de eigenschap "heeft een huid" over van de

superset-knoop “dier”. Alle eigenschappen van hogere knopen worden dus overgenomen in de knopen van lagere orde: men spreekt dan van een “virtuele copie” van de eigenschappen naar de knopen van lagere orde.

Om de zin “een kanarie heeft een huid” te verifieren, moet dus bepaald worden dat een kanarie een dier is, en dat een dier een huid heeft. De semantische afstand tussen twee woorden kan nu uitgedrukt worden in termen van de kortste weg tussen de met die woorden corresponderende concepten.

De interpretatie die Collins en Quillian (1969) aan het semantische afstandseffect geven valt uiteen in afzonderlijke onderdelen voor bevestigende en ontkennende antwoorden. Gemeenschappelijk is alleen de beginfase van het beslissingsproces: het conceptuele netwerk wordt–te beginnen bij de knopen die corresponderen met de inhoudswoorden van de te verifieren zin– doorlopen op zoek naar een pad dat de knopen verbindt. De bij toenemende semantische afstand stijgende reactietijd voor bevestigende antwoorden komt op rekening van de duur van deze zoekfase: deze duur is langer naarmate de knopen verder uit elkaar liggen. Voor ontkennende antwoorden wordt verondersteld dat de zoekfase verbindingspaden aan het licht brengt die bij nadere beschouwing niet blijken te stroken met de relatie die de aangeboden zin uitdrukt. Naarmate de semantische afstand van de inhoudswoorden korter is neemt de kans op dergelijke misleidende paden toe en daarmee de tijd om ze als dwaalsporen aan te merken. Bijvoorbeeld, “schoteltjes hebben oren” zal moeilijker te verwerpen zijn dan “schoteltjes hebben ogen”, omdat “schoteltje” via “kopje” een nauwere relatie met “oor” onderhoudt dan met “oog”.

Tegen het model van Collins en Quillian (1969) zijn een aantal ernstige bezwaren opgeworpen. In hun model behoort een bepaald concept wel (1) of niet (0) tot een superset-klasse en heeft een concept een eigenschap wel (1) of niet (0). In werkelijkheid geldt dit alles-of-niets principe niet. Concepten behoren met een bepaalde gradatie (tussen 0 en 1) tot een klasse en eigenschappen gelden met een bepaalde gradatie voor een concept. Een concept heeft een of meer ideale voorbeelden, prototypen genaamd. Deze prototypen behoren tot de klasse met een gradatie grenzend aan 1. Naarmate een instantie meer afwijkt van het prototype van het concept, neemt de mate waarmee de instantie tot dat concept behoort ook af. Bijvoorbeeld, een hond behoort duidelijk tot de klasse van de dieren, maar het is al veel minder duidelijk en eenduidig dat een bacterie tot de klasse van de dieren behoort. Het concept “bacterie” wijkt immers sterk af van het prototypisch dier. Zo ook benadert het concept “paard” het prototypisch zoogdier veel meer dan het concept “walvis”. Klassen blijken dus niet duidelijk afgelijnd te zin, maar hebben integendeel vage (“fuzzy”) grenzen. Zo’n klasse met een continuum van gradaties van lidmaatschap noemt men een “fuzzy set”. Voor eigenschapsrelaties kan een analoge redenering gemaakt worden als hierboven voor klasselidmaatschap. Zoals bepaalde instanties meer typisch zijn voor een klasse dan andere instanties, zo ook zijn bepaalde eigenschappen meer typisch voor een concept dan andere eigenschappen. Nu is gebleken dat zinnen waarin een typisch lid van een klasse of typische eigenschappen van een klasse gegeven worden, zich veel sneller laten verifieren. Bijvoorbeeld, “een hond is een dier” wordt veel sneller bevestigd dan een “bacterie is een dier” en “een paard is een zoogdier” wordt sneller bevestigd dan “een walvis is een zoogdier”. Deze

bevindingen kunnen niet verklaard worden vanuit het model van Collins en Quillian (1969). De semantische afstand tussen walvis en zoogdier, bepaald vanuit het aantal tussenliggende knopen in het hiërarchisch netwerk, is immers even groot als de semantische afstand tussen paard en zoogdier. Het hiërarchisch model voorspelt dus geen verschil in reactietijd, wat door de empirische gegevens tegengesproken wordt.

Deze data zijn echter in overeenstemming met een andere representatie in het permanent geheugen, nl. deze volgens het kenmerkenmodel (Smith, Shoben en Ribs, 1974). In een dergelijk model wordt een concept opgevat als een bundel kenmerken. De semantische afstand tussen twee concepten wordt dan niet bepaald door het aantal tussenliggende knopen, maar door het aantal gemeenschappelijke kenmerken. Deze voorstelling laat toe dat een concept niet volgens een alles-of-niets principe tot een superset-concept behoort, dat er, met andere woorden, variatie is in de mate waarin concepten tot een klasse behoren. Bijvoorbeeld, het aantal kenmerken dat het concept "paard" gemeen heeft met het concept "zoogdier" is groter dan het aantal gemeenschappelijke kenmerken van "walvis" en "zoogdier".

Binnen het kenmerkenmodel kan men nog een verder onderscheid maken tussen twee theorieen. Deze theorieen verschillen in functie van hun opvattingen over de relaties tussen de kenmerken en het concept. Bij de unitaire theorieen hebben de kenmerken verschillende graden van "typiciteit" (soms ook wel "similariteit" genoemd) tegenover het concept: het behoren tot een bepaald concept hangt dan af van de typiciteitswaarde van de kenmerken. Zo kan men vooraf voorspellen in welke mate een item bij een concept behoort, in functie van het aantal en de beoordeelde typiciteit van de kenmerken.

In de andere theorie (weleens binaire theorie genoemd) maakt men een onderscheid tussen "definierende" en "karakteristieke" kenmerken. Sommige kenmerken zijn meer definierende of essentiele aspecten van een concept, terwijl andere meer karakteristieke kenmerken zijn. Definierende kenmerken bepalen het behoren tot een concept, terwijl de karakteristieke kenmerken de beoordeelde typiciteit tot een concept bepalen. Bijvoorbeeld, van een mus kan je zeggen dat ze tweevoeters zijn, vleugels hebben, vooral in bomen wonen en geen huisdieren zijn. De eerste twee kenmerken zijn meer definierend voor het concept "mus" terwijl de laatste twee eerder karakteristiek zijn. Dit laat variatie toe in de mate waarin een eigenschap meer of minder typisch is voor een concept. Het verifieren van zinnen van het type zoals Collins en Quillian (1969) gebruikten, zou dan als volgt verlopen. Eerst worden alle kenmerken–zowel definierende als karakteristieke kenmerken–van de twee concepten vergeleken. Wanneer de concepten veel gemeenschappelijke kenmerken hebben, kan er snel een positief antwoord volgen, en bovendien sneller naarmate er meer gemeenschappelijke kenmerken zijn. Als er heel weinig gemeenschappelijke kenmerken zijn, volgt snel een nee-antwoord, en bovendien sneller naarmate er minder gemeenschappelijke kenmerken zijn. Wanneer er een middelmatige hoeveelheid gemeenschappelijke kenmerken zijn, zou er een tweede vergelijking volgen, maar ditmaal alleen op basis van de definierende kenmerken, niet van de karakteristieke kenmerken. De tweede vergelijking vraagt vanzelfsprekend enige extra tijd.

Met de kritieken hebben Collins en Loftus (1975) rekening gehouden in een nieuwe voorstelling van kennis in het geheugen, waarbij het hiërarchisch karakter van het vroeger model (tegenover het kenmerken-model) gehandhaafd wordt: het semantisch netwerkmodel. Kennis wordt gerepresenteerd als een verzameling van symbolen of concepten en de specifieke logische relaties daartussen. De symbolen of concepten worden ook knopen of knooppunten ("nodes") genoemd, omdat in elk concept een veelheid van associatieve relaties samenkomt. Het is belangrijk op te merken dat deze betekenisknopen niet noodzakelijk overeenstemmen met woorden over verbale benoemingen, al hebben woorden wel een onderliggende conceptuele structuur. Voor sommigen zullen bijvoorbeeld de woorden "gelei" en "jam" instanties zijn van dezelfde conceptuele betekenisknopen, voor anderen kunnen beide verwijzen naar twee verschillende betekenisknopen, waarvan de associatieve structuur gedeeltelijk overlappend is. Verder wordt aangenomen dat de verzameling knopen en hun logische relaties georganiseerd zijn in een hiërarchisch netwerk, waarbinnen zich een proces van zich verspreidende activatie kan ontwikkelen. Wanneer een bepaalde knoop gestimuleerd wordt, hetzij door een gebeurtenis, hetzij door de voorstelling van een gebeurtenis, dan verspreidt de activatie zich van de gestimuleerde knoop naar de geassocieerde knopen op een automatische en onbewuste wijze. De mate van activatie waarin een bepaalde knoop zich bevindt, heeft invloed op de kans dat men zich van deze conceptuele inhoud bewust wordt.

Allerlei onderzoeksresultaten worden voorspeld en geinterpreteerd met het nieuwe model van Collins en Loftus (1975). Het priming-effect bijvoorbeeld houdt in dat de verwerking van een item de verwerkingssnelheid van een verwant item beinvloedt. Zo vonden Meyer en Schvaneveldt (1971) dat de tijd nodig om te beslissen of een letterketen een woord is of niet (b.v., "boter"), korter wordt, wanneer het voorafgegaan wordt door een semantisch verwant woord (b.v., "brood") dan wanneer het voorafgegaan wordt door een niet verwant woord. De interpretatie is dat het woord "brood" zijn conceptuele geheugenknoop activeert, die activatie verspreidt naar gerelateerde knopen zodanig dat de activatie voor het concept "boter" sneller haar drempel bereikt. Anderson (1983) heeft aan het proces van verspreiding van activatie een centrale rol toebedeeld in zijn cognitieve modellen. Vrij nauwkeurige mathematische formuleringen en computersimulaties zijn ontwikkeld om dit proces af te beelden.

Semantische netwerken zijn voornamelijk gebruikt om de conceptuele structuur van woorden, categorieen en zinnen af te beelden. Sommige auteurs hebben echter getracht representatiesystemen te ontwikkelen voor grotere gehelen van informatie, die een hoger orde-karakter hebben (Schank & Abelson, 1977; Minsky, 1975). Deze systemen kunnen schemabenaderingen genoemd worden en ze sluiten vrij goed aan bij de inzichten van Bartlett (1932; zie vorige paragraaf). De termen schema en netwerk worden soms door elkaar gebruikt, maar in het algemeen verwijst schema, binnen de informatieverwerkingstheorieen, meer naar stereotype elementen van een betekenisstructuur, heeft het een hoger abstractiegehalte en wordt het meestal toegepast om grotere informatiegehelen te representeren. Men zou kunnen zeggen dat een schema een geheel van sterker samenhangende en meer stereotiepe delen is binnen een netwerkstructuur: een soort prototypische representatie.

Het rijke netwerk van verbanden, associaties of connecties in onze permanent geheu-

gen is een van de verstrekkende verschillen tussen mens en computer. Het vermogen van het menselijke brein om met zijn miljoenen neuronen gelijktijdig informatie te zoeken is volstrekt uniek. Spreiding van acti vatie maakt gebruikt van de associaties tussen ideeen of voorstellingen zodat als een gesprek op restaurants komt, alle kennis over dat onderwerp bijna meteen beschikbaar komt. Anders dan bij computers lijkt voor mensen dus te gelden dat hoe meer wij ergens over iets weten, des te sneller we het allemaal ook kunnen oproepen.

Referentes

- ANDERSON, J. R. (1983) *The architecture of cognition.* Harvard University Press, Harvard
- BARTLETT, F. C. (1932) *Remembering.* Cambridge University Press, Cambridge
- BERRY, D. C. en BROADBENT, D. E. (1984) On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- BROADBENT, D. E., FITZGERALD, P. en BROADBENT, M. H. P. (1986) Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, 77, 33-50.
- COLLINS, A. M. en LOFTUS, E. F. (1975) A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- COLLINS, A. M. en QUILLIAN, M. R. (1969) Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- DELHAYE, P. en D'YDEWALLE, G. (1986) Artificiële intelligentie en kennisextractie: Hun implicaties voor natuurlijke intelligentie. In P. J. Janssen, G. Vander Steene en H. Vertommen (Eds.), *Intelligentie: Tussen meten en weten in theorie en praktijk.* Universitaire Pers Leuven, Leuven.
- D'YDEWALLE, G., VAN RENSBERGEN, J. en POLLET, J. (1987) Reading a message when the same message is available auditorily in another language: The case of subtitling. In J. K. O'Regan en A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition.* (pp. 313-321). Elsevier Science Publishers B.V. (North-Holland), Amsterdam
- ERICSSON, K. A. en SIMON, H. A. (1984) *Protocol analysis: Verbal reports as data.* MIT press, Cambridge, Massachusetts
- MANDLER, G. (1985) *Cognitive psychology: An essay in cognitive science.* Erlbaum, Hillsdale, New Jersey.

- MEYER, D. E. en SCHVANEVELDT, R. W. (1971) Facilitation of recognizing pairs of words: Evidence for retrieval operations. *Journal of Experimental Psychology*, 90, 394-398.
- MINSKY, M. (1975) A framework for representing knowledge. In P. Winston (ed.), *The psychology of computer vision*. McGraw-Hill, New York.
- SCHANK, R. C. en ABELSON, R. (1977) *Scripts, plans, goals and understanding*. Erlbaum, Hillsdale, N.J.
- SMITH, E. E., SHOBEN, E. J. en RIPS, L. J. (1974) Structure and process in semantic memory: A feature model for semantic decision. *Psychological Review*, 81, 214-241.

VOETNOOT

Het deel over het onderscheid tussen exhaustieve en niet-exhaustieve kennis is aangepast overgenomen uit Delhaye en d'Ydewalle (1986). De studie is gedeeltelijk mogelijk geworden door onze deelname aan het I.W.O.N.L. Spilprogramma "Expert Systemen". De replicatie-proeven van Berry en Broadbent (1984) zijn door M. Haers en P. Gerits uitgevoerd.

Artificiële neurale netwerken: netwerken met verstand?

M. Van Hulle: Laboratorium voor Neuro- en Psychofysiologie en Departement Elektrotechniek, K.U. Leuven

1. Inleiding

De hersenen vormen een enorm complex netwerk van miljarden neuronen die d.m.v. biljarden synapsen met elkaar in verbinding staan. Toch slaagt de mens erin met dit netwerk te denken, te voelen, te onthouden en creatief te zijn. Sinds jaar en dag proberen onderzoekers a.d.h.v. modellen de geheimen van dit orgaan te ontsluieren. Twee types modellen kunnen worden onderscheiden.

De eerste worden door neurobiologen opgesteld. Neurowetenschappers interesseren zich voor de structuur en functie van de echte hersenen. Modellen worden dan gebouwd om experimentele data uit het domein van het zien, het horen en zelfs het gedrag te kunnen verklaren. Bovendien suggereren deze modellen nieuwe inzichten en experimenten die dan op hun beurt toelaten de validiteit van deze modellen te toetsen.

Daarnaast hebben onderzoekers uit de theoretische biologie, de statistische fysica, de cognitieve psychologie en de ingenieurswetenschappen elkaar gevonden bij het ontwerpen van neuronale netwerken en in de nieuwe concepten en applicaties die hieruit voortvloeien. Aangezien deze onderzoekers zich beperken tot netwerken die een sterke vereenvoudiging zijn van biologische netwerken, worden hun netwerken Artificiële Neuronale Netwerken (ANN) genoemd. ANNs worden dan gedefinieerd als (grote) sterk geïnterconnecteerde netwerken van relatief eenvoudige verwerkingseenheden, wier toestand of "activiteit" het resultaat is van interactie met andere verwerkingseenheden. Het is dus op grond van hun functionele gelijkenis dat deze verwerkingseenheden en hun interconnecties, "neuronen" respectievelijk "synapsen" worden genoemd.

De ontwikkeling van ANNs begon met het werk van McCulloch en Pitts (1943), Hebb (1949), Rosenblatt (1959), Widrow en Smith (1963) en kent nu een hernieuwde belangstelling door het werk van Hopfield (1982,1984), Feldman en Ballard (1982), Amari (1983), Kohonen (1984), Rumelhart et al. (1986), Hinton et al. (1986), Sejnowski en Rosenberg (1986), Grossberg (1986) en vele anderen. De sterk toegenomen interesse voor ANNs is niet alleen toe te schrijven aan de recente successen op conceptueel en fysisch-theoretisch vlak doch vooral aan het groeiend aantal demonstraties die menselijke performantie in perceptie en cognitie simuleren. De ANN-benadering van bv. visuele perceptie, spraak, natuurlijke taalverwerking en cognitieve taken zoals associatie, generalisatie en leren is reeds klassiek geworden in deze domeinen (Lippmann, 1987; Grossberg, 1986).

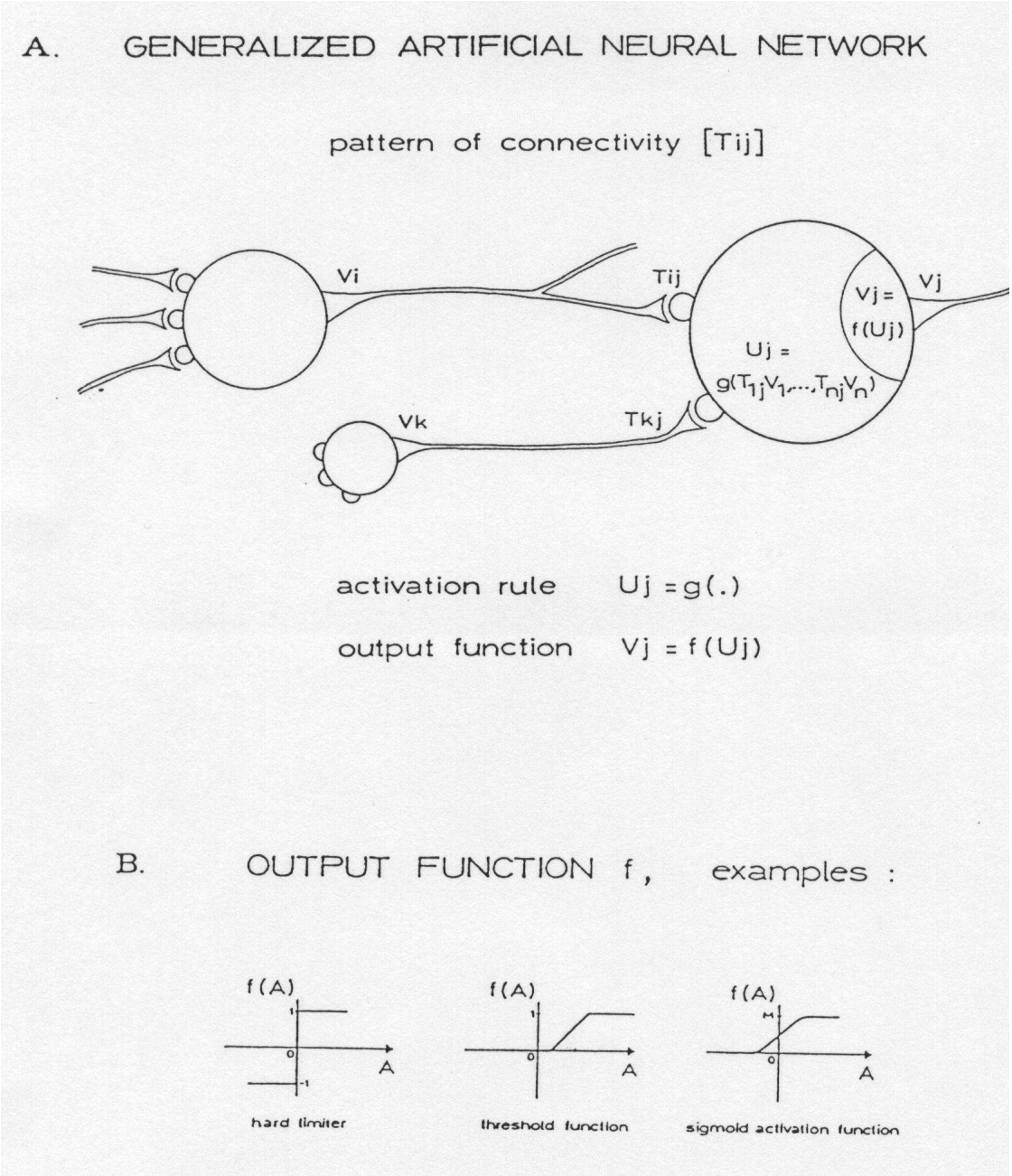
Berekeningen in ANNs gebeuren *parallel* en *numeriek* i.t.t. de formeel sequentiële, symbolische benadering van de klassieke Artificiële Intelligentie (AI). Sinds het midden van de jaren vijftig heeft het ANN onderzoek zich gedistantieerd van de "symbool-regel-representatie"-benadering van de AI en de cognitieve wetenschappen, en er is nog steeds een grote controverse gaande over de rol die ANNs zouden kunnen/moeten spelen in het modelleren van menselijk intelligent gedrag (Prince en Pinker, 1988). Enkele onderzoekers zijn zelfs van mening dat beide benaderingen diametraal tegenover elkaar staan (voor een discussie zie: Smolensky, 1986), doch het is mijn overtuiging dat geen van beide zaligmakend is en dat beide moeten gecombineerd worden om ANNs te kunnen bouwen die echt menselijke intelligentie simuleren.

In wat volgt zal ik onder 2. eerst het werkingsprincipe en het leergedrag van ANNs op abstracte wijze formuleren en vervolgens twee soorten ANNs onderscheiden en hun toepassingen met een voorbeeld illustreren. In tegenstelling tot wat de illustraties doen vermoeden, kennen ANNs een veel bredere waaier toepassingen: getuige hiervan de niet exhaustieve lijst van reeds bestaande toepassingsdomeinen, behandeld onder 3. In 4. zal ik twee motieven of doelstellingen voor ANN-onderzoek onderscheiden om dan te besluiten met enkele bedenkingen omtrent ANNs en verstand, het modelleren van menselijk gedrag en menselijke performantie en de bijdrage die klassieke AI en ANNs door samenwerking hieraan kunnen leveren.

2. ANNs: Werking, soorten en leergedrag

Voor complexe systemen is het niet altijd mogelijk een mathematisch model te formuleren dat een realistische voorstelling van het systeem vormt en bovendien analytisch kan opgelost worden. Indien deze analytische benadering niet mogelijk is, dient het complexe systeem te worden ontbonden in subsystemen, nl. in componenten wier mathematisch model wel een analytische oplossing toelaat. Het gehele systeem wordt dan opgebouwd met behulp van de verbindingen, de connecties tussen deze componenten: er ontstaat een netwerk van relaties tussen de karakteristieken van de individuele componenten. In het geval van een ANN zijn componenten "neuronen" en worden de relaties tussen deze "neuronen" door "synapsen" gelegd (Fig. 1A). Met elke synaps stemt een gewicht, nl. een "connectiesterkte" overeen en met elk neuron een numerieke waarde nl. de "activiteit". Het uitgangssignaal van een neuron is dan een niet-lineaire funktie van de activiteit van het neuron (Fig. 1B) en de activiteit zelf is funktie van de interacties, gewogen volgens de connectiesterkte, met de uitgangssignalen van de neuronen waarvan het neuron invoer ontvangt.

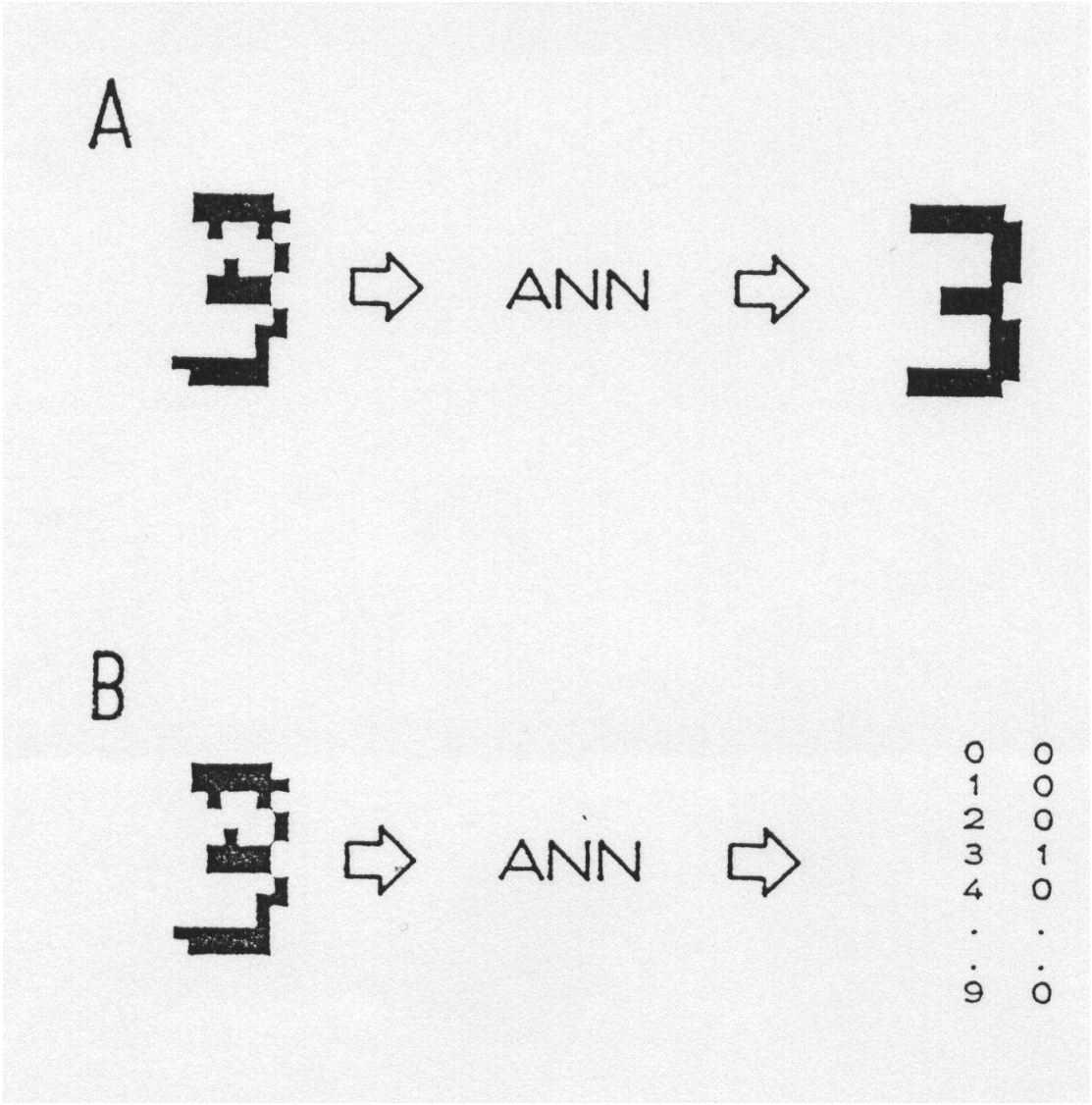
Beschouw nu een arbitrair netwerk als een "black box" bestaande uit N_u neuronen en N_c synapsen (een continue of veldbenadering is eveneens mogelijk, zie: Van Hulle en Orban, 1988b). Selecteer twee niet overlappende deelverzamelingen van n ingangs- en m uitgangsneuronen. De activiteiten van deze neuronen worden dan ingangs- respectievelijk uitgangspatroon genoemd. De fundamentele werking van elke ANN is *associatie* en wordt gedefinieerd als de transformatie van een verzameling $\mathbf{I}$ van p ingangspatronen in een verzameling $\mathbf{O}$ van p uitgangspatronen. Indien nu de beide verzamelingen



A Artificieel Neuronaal Netwerk (ANN) bestaande uit n neuronen, met toestand U_j, en een connectiviteitspatroon $[T_{ij}]$ nl. de matrix van de connectiesterktes. U_j is het resultaat van gewogen interacties met de $n-1$ andere neuronen volgens de activatieregel g. De uitgangsfunktie f is een niet-lineaire funktie van de argumenten.

B Voorbeelden van niet lineaire uitgangsfunkties f met argument A.

Figuur 1:



Ingangspatronen (links) worden aan een ANN gelegd en het ANN genereert dan de aangegeven uitgangspatronen (rechts). Alle in- en uitgangspatronen zijn binaire patronen: een wit blokje stelt de activiteit "1" voor.

A Voorbeeld van een auto-associator.

B Voorbeeld van een hetero-associator, gebruikt voor herkenning. Het uitgangspatroon is als bij een decoder opgevat: het uitgangsneuron dat met 3 overeenstemt is aktief, de andere uitgangsneuronen zijn niet aktief.

Figuur 2:

I en **O** identiek zijn, dan spreekt men van *auto-associatie* ; in het andere geval spreekt men van *hetero-associatie*. Opdat het ANN het gewenste associatief gedrag zou vertonen, dient het ANN te worden getraind. Verschillende leerstrategieën voor ANNs zijn mogelijk (Lippman, 1987) doch "leren" of "trainen" bestaat er meestal in de connectiesterkte te wijzigen zodat de discrepantie tussen het *actueel* uitgangspatroon en een *gewenst* uitgangspatroon (als antwoord op hetzelfde ingangspatroon) vermindert. Aldus wordt informatie omtrent de gewenste associatie *niet* als toestand in de neuronen zelf opgeslagen maar in de synapsen *tussen* deze neuronen. I.t.t. de klassieke computerparadigma's, is informatie niet gelokaliseerd doch gedistribueerd over het hele systeem. Via de connecties wordt het associatief gedrag door meerdere neuronen ondersteund en vertoont het ANN hierdoor ook een aanzienlijke robuustheid en fouttolerantie: een synaps meer of minder zal de associatieve funktie niet noemenswaardig schaden.

Het voornaamste toepassingsdomein van een ANN als auto-associator is patroonvoltooiing (Fig. 2A). Veronderstel dat een ingangspatroon $I_i \in \mathbf{I}$ met ruis wordt vervuild of slechts gedeeltelijk aan het getrainde ANN wordt aangeboden. De bedoeling is nu dat het ANN het onvervuilde patroon *reconstrueert* of het gedeeltelijk aangeboden patroon *voltooit* en ter beschikking stelt als uitgangspatroon. Intussen blijft het vervuild of onvolledig ingangspatroon aan het netwerk aangelegd. Het toepassingsdomein als hetero-associator is breder en kan verdeeld worden in twee categorieën: herkenning en klassifikatie. In herkenning wordt aan het getrainde ANN een onvolledig ofwel met ruis vervuild ingangspatroon $I_i \in \mathbf{I}$ aangelegd (Fig. 2B). Het netwerk genereert dan het met I_i *geassocieerde* uitgangspatroon O_i dus zelfs indien I_i niet op correcte wijze werd aangeboden. In het geval van klassifikatie bevat de trainingsset de vertegenwoordigers van de categorieën of klassen waartoe de ingangspatronen kunnen behoren: het uitgangspatroon laat dan toe de afzonderlijke klassen te onderscheiden (bv. als decoder). Veronderstel dat aan het getrainde netwerk een ingangspatroon $I_q \notin \mathbf{I}$ wordt aangeboden. De bedoeling is nu dat het netwerk de klasse waartoe I_q behoort, voorspelt volgens het klassifikatiemodel bevat in de trainingsset. Het succes van de voorspelling hangt af van de mogelijkheid tot generaliseren van getrainde ANNs: i.p.v. alle mogelijke patronen op te slaan, wordt het ANN alleen met een deelverzameling hiervan getraind en wordt aldus het ANN gebruikt om te generaliseren over *alle* patronen. Dankzij deze generalisatie eigenschap is een ANN ook in staat op robuuste wijze regel-gebaseerd gedrag te vertonen zonder dat het ANN deze regels op expliciete wijze bevat: regelgebaseerd gedrag is gewoon het resultaat van associatie tussen in- en uitgangspatronen en wordt bijgevolg gespecifieerd door de trainingsset. Daar een ANN geen expliciete representatie van regels dient te bevatten om regel-gebaseerd gedrag te vertonen, zijn ANNs niet alleen geschikt voor toepassingen met goed gedefinieerde regels zoals de "past tense" regel in het Engels (Rumelhart en McClelland, 1986) maar ook voor toepassingen waarvoor de regels onvolledig, onbetrouwbaar of zelfs onbekend zijn. Bijgevolg is het mogelijk met ANNs systemen te ontwerpen die geen "*a priori*" specificatie van het toepassingsdomein vereisen: een eigenschap waarin klassieke kennisgebaseerde systemen (vaak) tekort schieten.

3. Toepassingsdomeinen

ANNs worden niet alleen gebruikt om cijfers te herkennen of te reconstrueren. Het toepassingsdomein is veel ruimer en wordt enkel beperkt door het vinden van een geschikt netwerktype, het bepalen van een geschikte representatiestruktuur (Van Hulle en Orban 1988a,c) en het samenstellen van een trainingsset. Hiervoor een aanvaardbaar compromis vinden is geen sinecure en soms wordt er veel inventiviteit van de ontwerper verwacht: er bestaat immers geen standaard-ontwerpprocedure voor ANNs doch enkel een reeks vuistregels die vaak in de originele publikaties niet worden vermeld (!). Ik zal nu ter illustratie een niet-exhaustieve lijst van reeds bestaande ANN-toepassingen geven, een verklaring van de gebruikte terminologie geven zou mij echter te ver voeren.

- intelligente robotcontrole
- adaptieve regelaars
- combinatorische optimizatie
  - scheduling, sequencing
  - travelling salesman probleem
  - partitionering van grafen
- patroonherkenning, signaalverwerking
  - spraakherkenning
  - spraaksynthese
  - artificieel zicht
- natuurlijke taalverwerking
- optische neurocomputers
- connectionistische regelgebaseerde systemen
- diagnostiek

4. Motivatie

Globaal gezien kunnen voor het ANN onderzoek twee motivaties worden aangehaald.

Ten eerste worden ANNs geïntroduceerd enkel om de computationele kracht ervan te benutten, zelfs indien de struktuur van deze netwerken kenmerken zou bevatten die niet in overeentemming zijn met de wetenschappelijke evidentie over hersenstrukturen. Deze ANNs vertonen vaak meer overeenkomst met spin-glass modellen (zoals Hopfields model, zie: Hopfield, 1982, 1984): ze vormen een studie-objekt op zich.

Een tweede motivatie is trachten te verstaan hoe een ANN menselijke performantie kan halen in visie, spraak, motorische controle en cognitieve taken zoals associatie, generalisatie, enz... Hiervoor is het nodig inzicht te krijgen in de computationele kracht van ANNs en in het belang van neurobiologische getrouwheid.

Deze laatste opmerking brengt ons dichter bij de hamvraag van deze tekst. Ik zal er nu op ingaan.

5. Bedenkingen en vragen: ANNs en verstand

ANNs kunnen een complex, dynamisch gedrag vertonen zodat verschillende types van patroonvorming in ANNs kunnen voorkomen, afhankelijk van de initiële stimulering en toestand van het netwerk, nl. de activiteiten van de N_u neuronen en de N_c gewichten (Kishimoto en Amari, 1979; Amari, 1983). Dynamica van patroonvorming vormt de basis van corticale informatieverwerking. Anderzijds bieden de ANNs de mogelijkheid, dankzij hun generaliserend vermogen, te ontsnappen aan de rigiditeit van klassieke symboolsystemen, zoals bestudeerd in de klassieke AI, en zouden bijgevolg de *mogelijkheid* bieden om zonder weerga menselijke intelligentie te simuleren. Doch tot nu toe is dit slechts "wishful thinking": het is niet omdat bepaalde ANNs neurobiologische overeenkomsten vertonen dat deze ANNs dan ook in staat zouden zijn menselijke intelligentie te simuleren! Inderdaad, is het noodzakelijk de werking van neuronen en synapsen na te bootsen om menselijke intelligentie te simuleren? Klassieke Ai-ers zullen op deze vraag natuurlijk een negatief antwoord geven, misschien zich hierbij gesterkt voelend door Newell's uitspraak (Newell, 1982):

> "The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system"

Wat er dus op neer komt dat menselijke intelligentie met een symboolmanipulerend algoritme te vatten zou zijn. Newell's uitspraak is echter een ideologische uitspraak i.p.v. een testbare hypothese en ze vindt haar oorsprong in de klassieke ideologie van de cybernetica: "the brain is a computer". Doch de hersenen zijn gelukkig veel meer dan een computer en intelligentie is niet zomaar in een algoritme te vatten. Volgens het categorisch dualisme van Descartes behoren de "geest" en de hersenen tot twee verschillende logische types en zou het bijgevolg onmogelijk zijn de relatie tussen beide empirisch te onderzoeken. Moet artificiële intelligentie als een Cartesiaanse automaat worden beschouwd of als Gods schepping nl. een intelligent "wezen" met eigen strategieën en motieven? Is een biologisch netwerk meer dan de som van de eigenschappen van de individuele componenten nl. niet alleen van individuele "neuronen" en "synapsen" doch ook van membranen, ionenekanalen, neurotransmittervesikels, ionenconcentraties, enz... Zijn bijgevolg alleen de hersenen zelf de beste simulatie van zichzelf? Bovenstaande discussie is echter klassiek en ressorteert onder het "brain-mind" probleem. Het weze in elk geval duidelijk dat het ondoenbaar is een biologisch netwerk op een computer te simuleren. Wat is dan de *complexiteitsbarrière* die de

biologische realiteit onderscheidt van het simuleerbare nl. welk niveau van abstraktie kan men zich veroorloven om een ANN te bouwen dat op kwalitatieve wijze de eigenschappen van biologische netwerken bevat? Ook dit is nog niet duidelijk. Wat is wel realiseerbaar? Enerzijds bezitten ANNs unieke eigenschappen zoals generalisatie, associatie en fout-tolerant gedrag en anderzijds heeft de klassieke AI inzichten geleverd in het soort problemen dat moet worden opgelost om intelligente systemen te creëren. Doch symboolsystemen zijn inherent rigide door hun gebrek aan generalisatie en fouttolerantie nl. eigenschappen die ANNs wel bezitten. Indien ANNs enkel zouden dienen om snellere symboolsystemen te bouwen dan zal men de boot missen. Het is duidelijk: beide benaderingen moeten gecombineerd worden om systemen te kunnen bouwen die de menselijke intelligentie benaderen (voor een discussie, zie: Van Hulle en Orban, 1988d).

Referenties

- AMARI S. (1983) Field theory of self-organizing neural nets. *IEEE transactions on Systems, Man, and Cybernetics,* vol. SMC-13, No. 5, September/October, pp. 741-748.
- FELDMAN J.A. en BALLARD D.H. (1982) Connectionist models and their properties. *Cogn. Sci.,* 6: 205-254
- GROSSBERG S. (1986) *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm,* and *The Adaptive Brain II: Vision, Speech, Language, and Motor Control,* Elsevier/North-Holland, Amsterdam.
- HEBB D.O. (1949) *The organization of behaviour,* Wiley, New York.
- HINTON G.E., McCLELLAND J.L. en RUMELHART D.E. (1986) Distributed representations. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1.: Foundations.* D.E. Rumelhart & J.L. McClelland (eds.), MIT Press, pp. 77-109.
- HOPFIELD J.J. (1984) Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA,* 81: 3088-3092.
- KISHIMOTO K. en AMARI S. (1979) Existence and Stability of Local Excitations in homogeneous Neural Fields. *J. Math. Biol.,* 7: 303-318.
- KOHONEN T. (1984) *Self-organization and associative memory.* Springer-Verlag, Berlin.
- LIPPMANN R.P. (1987) An introduction to computing with neuronal nets. *IEEE ASSP Magazine,* April 1987, 4-22.
- McCULLOCH W.S. en PITTS W. (1943) A logical calculus of the ideas imminent in nervous activity. *Bull. Math. Biophys.,* 5: 115-133.

- NEWELL A. (1982) The knowledge level. *AI,* 18: 87-127.
- PRINCE A. en PINKER S. (1988) Rules and connections in human language. *TINS,* Vol. 11, 5: 195-202.
- ROSENBLATT R. (1959) *Principles of Neurodynamics.* New York, Spartan Books.
- RUMELHART D.E., HINTON G.E. en McCLELLAND J.L. (1986) A genral framework for parallel distributed processing. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations.* D.E. Rumelhart & J.L. McClelland (eds.), MIT Press, pp. 44-76.
- RUMELHART D.E. en McCLELLAND J.L. (1986) Psychological and Biological Models. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2: Foundations.* D.E. Rumelhart & J.L. McClelland (eds.). MIT Press, pp. 216-271.
- SEJNOWSKI T.J. en ROSENBERG C.R. (1986) *NETtalk: a parallel network that learns to read aloud.* Johns Hopkins Univ. Techn. Report HU/EECS-86/01.
- SMOLENSKY P. (1986) Information processing in dynamical systems: Foundations of harmony theory. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations.* D.E. Rumelhart & J.L. McClelland (eds.), MIT Press, pp. 194-281.
- VAN HULLE M.M. en ORBAN G.A. (1988a) Entropy driven artificial neuronal networks and sensorial representation: a proposal. Special Issue on "Neural Computingöf the *Journal of Parallel and Distributed Computations,* in press.
- VAN HULLE M.M. en ORBAN G.A. (1988b) A field approach to higher order ANNs for sensorial representation. *Proc. First European Conference on Neural Networks,* Paris, June 6-9.
- VAN HULLE M.M. en ORBAN G.A. (1988c) Correlative value/variable binding in higher order ANNs. *Proc. IEEE International Conference on Neural Networks,* July 24-27, San Diego, Ca., USA.
- VAN HULLE M.M. en ORBAN G.A. (1988d) Sensory ANNs and Symbolic Decomposition. *Proc. 1988 DECUS Europe Symposium,* September 5-9, Cannes, France.
- WIDROW B. en SMITH F.W. (1963) Pattern recognizing control systems. *Computer and Information Sciences (COINS) Symposium Proceedings,* Spartan Books, Wash. DC.

Linguistiek en artificiële intelligentie: Theorie en praktijk van een haat-liefde-verhouding

G. Adriaens, Siemens N.V. Brussel en Departement Linguistiek, K.U.-Leuven

Inleiding

Als er iets is wat de mens zowel van het dier als van de machine onderscheidt, dan is het wel zijn vermogen om via natuurlijke taal met zijn medemens te communiceren. Linguisten proberen al eeuwenlang het fenomeen natuurlijke taal te doorgronden, waarbij vooral de theoretici onder hen op zoek gaan naar de algemene principes die natuurlijke talen karakteriseren. De meeste theoretici wensen hun werk ook niet los te zien van de mens als taalverwerkend wezen, en stellen dan ook dat de algemene principes die zij ontdekken de kern van de menselijke taalkennis of het menselijk taalvermogen uitmaken. Onderzoekers uit het veel jongere domein van de artificiële intelligentie (verder AI-onderzoekers genoemd) van hun kant houden zich bezig met theorievorming over de mentale vermogens van de mens met als belangrijkste methodologisch instrument de computer. Met een meer praktische formulering kan men ook gewoon zeggen dat ze programma's schrijven die "gedrag" vertonen dat we bij de mens intelligent noemen. Ook AI-onderzoekers zijn van in het begin gefascineerd geweest door het fenomeen van de mens als *homo loquens,* en ze hebben dan ook geprobeerd programma's te ontwikkelen die talig gedrag vertonen (in de eerste plaats taalverstaan, tegenwoordig ook meer en meer taalproduktie).

Daarmee is de achtergrond geschetst van wat in deze uiteenzetting besproken zal worden. Het wordt een typisch voorbeeld van een "weinig-over-veel"-tekst, waarbij een heel ruim gebied verkend wordt. Eerst zal de wetenschapstheoretische kant van de relatie tussen de twee geïntroduceerde wetenschappen kort besproken worden, omdat het om een vrij ongewone relatie gaat. Daarna wordt geleidelijk overgegaan naar de praktijk (wat gebeurt er op het domein waar linguistiek en AI elkaar ontmoeten?). De universitaire onderzoekspraktijk komt daarbij eerst aan bod, en er wordt geëindigd met de industriële praktijk: de markt van produkten die uit dat onderzoekswerk gegroeid zijn (of nog moeten groeien), met een voorbeeld van een succesvol produkt.

In theorie

Uit de boven geschetste bezigheden van linguisten en AI-onderzoekers zou men kunnen opmaken dat ze in feite op hetzelfde terrein actief zijn (ze zijn allebei bezig met het menselijke taalvermogen en taalgedrag), en dat er dus interessante interacties tussen de respectieve wetenschappelijke activiteiten zouden kunnen plaatsvinden. Voortgaande op de op het einde van de jaren 70 in vooraanstaande tijdschriften (*Cognition* en

Cognitive Science) gevoerde polemiek tussen taalkundigen en AI-onderzoekers zou men echter denken dat er absoluut geen ruimte is voor dialoog. Waar draait de hele discussie rond? (Theoretische) linguisten (niet meteen de meest bescheiden onderzoekers) begonnen de polemiek met AI-onderzoekers te verwijten dat ze eigenlijk alleen bezig waren met het schrijven van computerprogrammaatjes voor natuurlijke-taalbegrip vanuit een "als-het-maar-werkt"-instelling, dat ze niets serieus over taal te vertellen hadden, kortom, dat er veel praktijk bij kwam kijken, maar nauwelijks ernstige theorievorming. De AI-onderzoekers die met natuurlijke-taalverwerkingsprogramma's bezig waren namen deze kritiek niet, en brachten op hun beurt de kanonnen in stelling. Taalkundigen waren volgens hen altijd al ivoren-toren-wetenschappers geweest, die veel theorie verkochten, maar naar de praktijk toe van het eigenlijke menselijke taalgebruik weinig te vertellen hadden. Daarenboven hadden ze zich altijd meer geïnteresseerd voor de vorm en de structuur van de taal (het syntactische aspect), en nauwelijks voor hetgeen waar taalgebruik volgens de AI-onderzoekers echt om gaat, namelijk betekenis en context (de semantische en pragmatische aspecten). Ik moet hier helaas volstaan met de polemiek te vermelden, en kan niet ingaan op vele interessante kanten ervan (een gevestigde wetenschap tegenover een nieuwkomer, het gevecht om onderzoeksfondsen, enz. - zie Bresseel 1987 voor een gedetailleerde bespreking). Alleen nog even dit: in feite speelt een derde wetenschap ook een belangrijke (zij het passieve) rol bij de polemiek, met name de psychologie. Het theoretische raakvlak van taalkunde en AI (de menselijke talige capaciteiten) is eigenlijk ook het domein van de psychologie, een wetenschap waar zowel linguistiek als AI altijd graag naar gelonkt hebben (door psychologische claims rond hun theorieën), maar die door hen te weinig au sérieux genomen wordt. Maar vóór mijn uiteenzetting een verhaal wordt over een wetenschappelijke driehoeksverhouding, kan ik beter het theoretische gedeelte afsluiten met te verwijzen naar de manier waarop alle genoemde wetenschappen met elkaar verzoend zouden kunnen worden. Er is namelijk een nieuw would-be wetenschapsparadigma in opkomst (*cognitive science*) dat probeert een aantal bestaande paradigma's met elkaar te verenigen in een ambitieuze interdisciplinaire onderneming. Linguistiek, AI, psychologie, filosofie, antropologie en neurowetenschappen zouden hun krachten moeten bundelen om het uiterst complexe domein van de menselijke cognitie te bestuderen. Laten we hopen dat ook linguisten en AI-onderzoekers bereid zullen zijn om in een dergelijk kader samen te werken om onze kennis van de mens als taalgebruikend wezen te vergroten (zie b.v. Gardner 1985 voor een goede inleiding tot wat *cognitive science* inhoudt, of Adriaens 1986 voor een specifieke visie op de integratie van de taalkunde binnen *cognitive science*).

In de praktijk

Terwijl het meestal zo is dat dingen er in theorie wel mooi uitzien, maar dat de praktijk veel te wensen overlaat, is het (gelukkig) voor de verhouding tussen linguistiek en AI net andersom. Hoewel er theoretisch onoverkomelijke meningsverschillen schijnen te bestaan, is er in de praktijk heel wat interessante wederzijdse beïnvloeding merkbaar. Het praktische domein waarop taalkundigen en AI-onderzoekers elkaar ontmoeten is dat van de computerprogramma's voor natuurlijk-taalverwerking (verder NTV). Vóór ik verder inga op dat domein van de NTV, wil ik om de praktische integratie van de

linguistiek en AI te illustreren mog even vermelden dat bijna één kwart van de ongeveer 1200 pagina's tellende *Encyclopedia of Artificial Intelligence* (Shapiro 1987) handelt over topics uit het domein van de natuurlijke-taalverwerking!

Voor linguisten vormen de NTV-programma's een ideale methode om hun theorieën over de structuur van natuurlijke talen mee te testen: de vereiste dat een theorie ook omgezet kan worden in een werkend programma dat taal analyseert volgens de principes van die theorie creëert namelijk de noodzaak van grote expliciethеid, formalisering en volledigheid. Ik zou zelfs kunnen stellen dat het voor taalkundigen vandaag de dag een must geworden is om de *computationele realiseerbaarheid* van hun theorieën te bewijzen aan de hand van een computersimulatie. Het is waarschijnlijk zelfs zo dat die aanpak sterk beïnvloed is door het belang dat precies in domeinen zoals dat van de AI aan computers gehecht wordt @ldots@

AI-onderzoekers van hun kant hebben computerprogramma's ook altijd gebruikt om hun theorieën te testen (dat zijn dan echter veeleer theorieën over de *processen* die actief zijn bij natuurlijke-taalverwerking, en in mindere mate benaderingen van de *structuur* van talen). Zoals boven echter reeds uit de twee omschrijvingen van AI-activiteiten duidelijk geworden zal zijn, is er ook steeds de bekommernis geweest om praktisch werkende systemen te bouwen, vooral met het oog op de verwezenlijking van een vlottere interactie tussen de mens en de onmondige machine.

Enkele van de dingen die de twee onderzoeksgroepen aan elkaar te bieden hebben zijn zeker de volgende. Enerzijds zal een werkend NTV-systeem altijd terug moeten vallen op een belangrijk gedeelte linguistische beschrijving van en theorievorming rond taalfeiten; AI kan m.a.w. niet langs het eeuwenlange werk van grammaticaschrijvende en woordenboekmakende taalkundigen heenkijken. Anderzijds is het ook zo dat AI een belangrijke bijdrage geleverd heeft tot de horizonverruiming van de linguisten. AI-onderzoekers zijn namelijk altijd sterk begaan geweest met de problemen rond kennisrepresentatie (zie ook de bijdrage van prof. Willems), en niet alleen representatie van syntactische kennis. In AI heeft men onmiddellijk gezien dat een computer over veel meer kennis moet beschikken dan alleen die over structuur van een taal: kennis over betekenis, over de context waarin taaluitingen gebruikt worden is onontbeerlijk opdat een computer ook maar enige blijk van taalbegrijpend gedrag zou kunnen geven. Behalve op de noodzaak van meer studie op de domeinen van semantiek en pragmatiek, hebben makers van grote praktische NTV-systemen ook gewezen op het belang van exhaustieve taalbeschrijving. Meer nog dan de (wetenschaps)mens zelf kan de computer meedogenloos aan het licht brengen dat b.v. zinsstructuren of -wendingen nergens in een theorie of beschrijving aan bod gekomen zijn. Kortom, er valt op het terrein van het NTV-onderzoek nog veel weg af te leggen.

De praktijk in het algemeen

Wat zijn nu de concrete onderzoeks- en toepassingsdomeinen binnen het ruime domein van de NTV, waarop zowel taalkundigen als AI-onderzoekers actief zijn?

Als we taal opsplitsen in gesproken en geschreven taal, dan is er vooreerst het subdomein van de spraakverwerking: spraakverstaan (met inbegrip van akoestisch-fonetische spraakherkenning) en spraaksynthese. De aard van het bestudeerde object (het spraaksignaal), met alle variaties en varianten die het karakteriseren maakt dat het domein van de spraakverwerking (en ik beperk me hier even tot spraak*herkenning*) tamelijk apart staat binnen het ruime domein van de NTV, en minder goed gekend is dan de subdomeinen rond geschreven NTV. Vandaar dat ik er even in wat meer detail op inga. Wie op geschreven taal werkt (waarmee dan wel "niet-handgeschreven" taal wordt bedoeld, dus getypte of gedrukte taal), heeft het voordeel van een standaard-tekenset (het alfabet) en spellingsconventies. Gesproken taal is echter een veel grilliger gegeven: er zijn niet alleen dialectische verschillen tussen groepen mensen, maar ook veschillen tussen de mensen onderling, en zelfs verschillen bij éénzelfde persoon naargelang van zijn humeur of gezondheidstoestand! Verder is het ook zo dat klanken verschillend uitgesproken kunnen worden naargelang van de hen omgevende klanken, wat voor herkenning de zaken er niet makkelijker op maakt. Al deze problemen brengen met zich mee dat men voor spraakherkennende systemen opdelingen maakt tussen sprekeronafhankelijke en sprekerafhankelijke systemen, en ook tussen systemen waarin men alleen afzonderlijke woorden herkent en systemen waarin men "normale" (vloeiende) spraak herkent. In functie van de waarden voor die parameters krijgt men dan een bepaalde woordenschatgrootte. Wie sprekeronafhankelijke vloeiende spraak wil herkennen, zal zich moeten beperken tot een paar tientallen woorden. Wie het houdt op herkenning van de spraak van één spreker, en meteen ook alleen afzonderlijke woorden, kan een woordenschat bestrijken van enkele honderdtallen/duizendtallen woorden. Het zal duidelijk zijn dat commerciële toepassingen voorlopig alleen sprekerafhankelijke, afzonderlijke-woordherkenningssystemen zijn. Concreet gaat het om toepassingen waarbij b.v. een persoon een rapporteringstaak moet uitvoeren terwijl hij tegelijk letterlijk de handen vol heeft (zoals bij kwaliteitscontrole van onderdelen op een band, of bij voorraadbeheer in een pakhuis). (Wie verder wil kennismaken met het domein van de spraakverwerking, kan b.v. de bundel van het BIRA [Januari 1988] over dit onderdeel van de NTV raadplegen).

Zoals gezegd neemt spraakverwerking een aparte plaats in binnen de NTV, en staat het naast het veel bredere domein van de studie van "geschreven" taal. Wat dit laatste betreft, moet ik volstaan met kort de verschillende deeldomeinen ervan te vermelden. (In de bibliografie kan u een lijst vinden van inleidende werken hierrond.) Reeds toen de computer nog in de kinderschoenen stond, was er al onderzoek i.v.m. *automatische vertaling,* een domein dat vandaag de dag ook weer in de lift zit (niet alleen in Europa met zijn veelheid van talen, maar ook in de Verenigde Staten en Japan). Verder is er het subdomein van de *document/content scanning* of ook *information retrieval,* waar het b.v. gaat om sleutelwoorden uit teksten te halen, of zelfs om teksten te proberen samen te vatten. NTV-onderzoekers houden zich ook bezig met het ontwerpen van *dialoogsystemen*, die de mens willen toelaten in zijn eigen taal met een computer te interageren. Het gaat hier om interfaces met gegevensbanken of expertsystemen, waar de mens d.m.v. natuurlijke taal informatie uit kan halen of aan kan toevoegen. (Zie het concrete voorbeeld aan het einde van deze uiteenzetting.) Verder zijn er wat men de *schrijfomgevingen* noemt: tekstverwerkingssystemen die een heleboel interessante

hulpmiddelen bevatten om het schrijven van teksten te vergemakkelijken (on-line woordenboeken, spellingscorrectoren, splitsingsprogrammatuur, stijlverbeteraars, enz.) Ten slotte dient ook nog het domein van de *computer-aided/assisted instruction/learning (CAI,CAL)* vermeld, waarin onderwijstoepassingen bestudeerd en ontwikkend worden.

De praktijk in België: universiteit en industrie

Hoewel het interessant zou zijn een (algemeen) beeld te hebben van de NTV-bedrijvigheid op wereldvlak, zijn er te weinig gegevens beschikbaar om een dergelijk beeld te schetsen. Van de Amerikaanse en Engelse NTV-activiteiten uit de commerciële sfeer bestaat wel een studie (Johnson 1985, zie ook verderop), maar andere uitvoerige studies zijn mij niet bekend. De EG schijnt wel plannen te hebben om een dergelijke studie voor Europa uit te voeren m.m.v. de bedrijfswereld, alsook om het onderzoek dat zij mee gefinancierd heeft (in het kader van EUROTRA, ESPRIT, enz.) stilaan "de industriële wereld in te schoppen" (LIFE, Language Industry For Europe). In het kader van deze uiteenzetting wens ik het echter beperkt te houden door een aantal projecten te vermelden die aan Belgische universiteiten en aan in België gevestigde bedrijven lopen op het domein van de NTV. Ik merk hierbij meteen op dat naar een zekere volledigheid gestreefd is, maar dat lacunes niet uitgesloten zijn; ik verwijs trouwens ook naar de bibliografie voor verdere informatie rond vermelde projecten.

Aan de K.U.-Leuven is er vooreerst heel wat bedrijvigheid rond de ontwikkeling van automatische-vertaalsystemen. Aan het departement linguistiek loopt zowel meer "zuiver" wetenschappelijk onderzoek in het kader van het ruimere EUROTRA-project van de EG, als meer produktgericht onderzoek in het kader van het METAL project (samenwerking van Siemens-Brussel met o.m. de K.U.-Leuven). Verder loopt er ook onderzoek rond taalkundig-gefundeerde lexicale databanken (PROTON), parallelle analysesystemen (WEP/PEP), onderwijstoepassingen (COMENIUS) en spraakverwerking (dit laatste niet alleen op het departement linguistiek, maar ook bij ESAT). Aan de R.U.-Gent zijn ook onderzoekers actief rond spraakverwerking (zie de boven vermelde BIRA-bundel voor meer details). Morfologische taaltechnologie wordt bestudeerd aan de V.U.-Brussel (Daelemans 1987). Aan de U.I.-Antwerpen wordt ook rond onderwijstoepassingen gewerkt. Ten slotte is er de Université de Liège, waar aan taalanalyse en -generatie gewerkt wordt in het kader van dialoogsystemen (SABA, HERMES). Al dit onderzoek samen bestrijkt dus een vrij groot gedeelte van de verschillende subdomeinen van de NTV.

Ook de industrie waagt zich op het terrein van de NTV, zij het vaak in samenwerking met de universiteiten. Siemens is reeds vermeld als de industriële partner in het METAL automatische-vertaalproject. Verder is er BIM, die rond een dialoogsysteem werkt voor gegevensbanken (LOQUI), samen met het departement computerwetenschappen van de K.U.-Leuven. UNISYS is onlangs ook gestart met een project rond automatische vertaling i.s.m. de ULB. Ten slotte zijn er ook ACEC en PHILIPS, die met spraakherkenning bezig zijn.

De NTV-markt: vooruitzichten en een voorbeeld

Als we de universitaire onderzoekswereld verlaten (hoewel die nooit ver weg is, getuige de samenwerkingsverbanden met de industrie) en nog verder de pure (industriële) praktijk van het NTV-raakvlak tussen linguistiek en AI beschouwen, dan is het interessant om marktvooruitzichten te bekijken. Veel zijn er niet beschikbaar, maar wat er is toont wel de niet onbelangrijke plaats van NTV binnen de AI-markt in het algemeen, alsook een optimisme i.v.m. de groei voor de komende tien jaar. Een overzicht van de Amerikaanse AI-markt vermeld in "*01 Informatique*" (mei 1987) ziet er als volgt uit (NTV vetjes):

SECTOR	MARKTAANDEEL
gespecialiseerde	
AI-hardware	41%
Visiesystemen	20%
Expertsystemen	19%
Natuurlijke taal	**11%**
Spraakverwerking	**6%**
Programmeertalen	3%

JAAR	TOTALE MARKT
1986	1,046 mld $
1990 (schatting)	4,2 mld $

Als "natuurlijke taal" en spraakverwerking (die hier apart beschouwd zijn) weer samengenomen worden tot NTV in het algemeen, dan is er sprake van een niet onaardig aandeel van 17% in de totale AI-markt (waarvan de omzet op vier jaar tijd trouwens zou verviervoudigen . . .). (Als ik me even mag laten gaan in het natte-vinger cijferwerk, en verwijs naar de 25 % van de *Encyclopedia of AI* die over NTV-onderwerpen gaat, dan komen we dus aan een gemiddelde van interesse voor NTV tussen onderzoekswereld en industrie van ongeveer 20 % . . .). In hetzelfde tijdschrift wordt ook nog vermeld dat het Amerikaanse aandeel in de AI-wereldmarkt geleidelijk zou dalen van de huidige 72 % tot 65 % in 1990, terwijl het Japanse en het Europese aandeel zouden toenemen (respectievelijk van 11 tot 17 % en van 17 tot 19 % — dit laatste dus niet zo spectaculair voor Europa).

Wat de groei van het NTV-domein op zichzelf betreft, kunnen nog even kort de spectaculaire prognoses uit Johnson 1985 voor de Amerikaanse en Engelse NTV-markt vermeld worden. (Het gaat ruwweg samengenomen om de subdomeinen van de dialoogsystemen, machinevertaling, content scanning, schrijfomgevingen, spraakverwerking; zie Johnson 1985, 39-65 voor de details). De Amerikaanse markt zou groeien van een omzet van 15 miljoen $ in 1985 tot 1500 miljoen $ in 1995 (verhonderdvoudiging!), de Engelse markt van een omzet van 1,1 miljoen $ tot een omzet van 230 miljoen $ (. . .). Kortom, als we de cijfers mogen geloven, zit NTV duidelijk in de lift.

Om als laatste onderdeel van deze uiteenzetting nog een idee te geven van de concrete praktijk, zou ik willen afsluiten met een korte beschrijving van een (Amerikaans) produkt dat al op de markt is en het vrij goed schijnt te doen. Vooral interessant eraan vind ik de manier waarop mensen geprobeerd hebben om de afstand te overbruggen tussen wat in een onderzoekswereld ontwikkeld wordt en wat de gewone computergebruiker kan aanspreken. (Hoewel dat een ander theorie-versus-praktijk verhaal is, is het inderdaad zo dat in het algemeen de afstand tussen de universitaire onderzoekswereld en de industriële wereld (helaas) nog altijd vrij groot is ...)

De *Intelligent Assistant* is een natuurlijke-taal-interface bij een databasesysteem (&A van de firma *Symantec*; zie *Byte* van januari 1986 en december 1987). Ze laat toe in het Engels gegevens uit de databank te halen *en* ook gegevens eraan toe te voegen. Typisch is dat het gaat om een produkt dat oorspronkelijk in een onderzoeksomgeving ontwikkeld is (LIFER/LADDER, Hendrix 1977), en dat dan aangepast is om op (IBM) PC te kunnen draaien. (Kleine noot terzijde: Hendrix is nu vice-president van een afdeling van Symantec; merk op dat tussen het onderzoeksresultaat en het aanbieden van een afgeleid produkt op de markt minstens acht jaar verstreken zijn!). Precies die noodzaak van portering naar kleine machines heeft geleid tot een aantal keuzes die normaal niet de eerste bekommernis zijn van onderzoekers. Zo moest alle programmatuur uiteindelijk geschreven worden in talen (C en machinetaal) die gewoonlijk niet tijdens onderzoeksfases gebruikt worden (Lisp of Prolog zijn eerder basiswerktuigen van de onderzoeker). Verder mochten de ambities rond de *coverage* van het systeem (d.w.z. de verscheidenheid en complexiteit van talige structuren die door het systeem correct "begrepen" worden) niet al te hoog zijn: de responstijd na een vraag in natuurlijke taal moet aanvaardbaar blijven (hoe groter de *coverage*, hoe complexer het systeem, en hoe langer een analyse-plus-antwoord duurt). Verder is ook veel aandacht besteed aan de robuustheid van de *Intelligent Assistant*, wat inhoudt dat hij semi- of zelfs niet-grammaticale input aankan (ook weer een belangrijk praktijkgegeven waar onderzoekers vaak overheen kijken: een mens gebruikt zijn taal lang niet altijd zoals het "volgens de boeken" zou moeten; er zijn dus regels nodig voor kromme input). De *Intelligent Assistant* gebruikt hiervoor de interessante techniek van de gelaagde afzwakking van beperkingen: eerst worden de (herschrijf)regels van het systeem toegepast met al hun restricties, waarna bij falende analyse geleidelijk geprobeerd wordt de beperkingen te laten vallen om alsnog een aanvaardbare analyse te vinden voor een inputzin. Een greep uit de zinstypes die de *Intelligent Assistant* aankan als varianten op de vraag naar salarissen van vrouwelijke werknemers (uit *Byte*, december 1987, 252):

```
Female salaries.  Show the female salaries.
Salaries for females.  Salaries of the women.
What do we pay the women?
What are the salaries of the female employees?
What are women paid?
How much pay do the women get?
What salaries do the female employees have?
Can I have the salaries from the records for females?
Please find the earnings of our women employees and present them
```

```
for me.
Get the salaries of the employees who are women.
Get the salary data on all females.
I want to see the salaries of the females.
Let us see the salaries of the female employees.
Please make a report that shows the values from the salary field
from all forms where the value in the sex field is female.
For female employees, make a list of the salaries.
If an employee is female, I want to see her salary.
```

Er valt nog veel meer te bespreken in verband met een produkt als de *Intelligent Assistant*, of soortgelijke produkten, maar deze korte presentatie moet volstaan. Eén tekortkoming van het systeem is nog het vermelden waard, omdat ze typerend is voor een zeker gebrek aan globale visie op NTV-produkten en hoe die met elkaar geïntegreerd moeten worden. Men kan stellen dat een database-interface een "hoger" type is van NTV-toepassing (tegenover een "lager" type zoals een spellingscorrector). Wie echter een hoger type aanbiedt, zou er eigenlijk voor moeten zorgen dat zo'n type lagere types bevat (hiervoor zorgen vergt precies een globale visie op NTV-toepassingen). Het is namelijk zo dat de *Intelligent Assistant* geen spellingscorrector bevat, zodat fout gespelde input niet behandeld kan worden (erger nog, bij opnieuw ingeven van de gecorrigeerde invoer wordt het hele analyseproces opnieuw gestart –eventuele bruikbare deelanalyses uit de eerste poging worden niet bewaard). Als een gebruiker met een "hoger" NTV-produkt werkt, zal hij er van uitgaan dat het tenminste de eenvoudiger problemen ook aankan. Als dat niet zo is, kan de ergernis bij gebruik vrij groot worden...

Besluit

In deze uiteenzetting over de theorie en de praktijk van de samenwerking tussen taalkundigen en AI-onderzoekers heb ik geprobeerd een zo breed mogelijk beeld van de werkzaamheden op het domein van de natuurlijke-taalverwerking te schetsen. Hopelijk zal gebleken zijn dat het om een vrij actueel domein gaat, waar zowel de universitaire als de industriële wereld heel wat interesse voor betonen.

Hoewel te verwachten is dat in de komende tien jaar heel wat vooruitgang geboekt zal kunnen worden in het NTV-domein (mede door de komst van parallelle machines), zou ik graag willen eindigen met een relativerende noot rond de hele AI-bedrijvigheid. Het gaat om wat Gardner (1985) de *computational paradox* noemt: hoe meer we werken met de computer als hulpmiddel (en metafoor) bij de studie van de menselijke cognitie, hoe meer we moeten inzien dat die menselijke cognitie totaal verschillend is van die computer. Meer zelfs: het naïeve geloof dat de computer de mens op termijn zou kunnen evenaren heeft stilaan plaats gemaakt voor respect voor de complexiteit van het menselijke cognitief functioneren. En wat het menselijk taalvermogen in het bijzonder aangaat: het is niet voor niets dat een belangrijk werk rond psycholinguistiek in zijn allerlaatste zin moest besluiten dat "there is no substitute for human intelligence" (Fodor, Bever & Garrett 1974)...

Bibliografie

Tekstverwijzingen

- ADRIAENS, G. (1986) *Process Linguistics: the Theory and Practice of a Cognitive-Scientific Approach to Natural Language Understanding.* Dissertatie, departementen linguistiek en computerwetenschappen Leuven.
- BRESSEEL, M. (1987) *Natural Language Processing within Generative Linguistics and Artificial Intelligence. A Critical Overview and Discussion of a Long-standing Debate.* Verhandeling, departement linguistiek Leuven.
- DAELEMANS, W. (1987) *Studies in language Technology. An Object-Oriented Computer Model of Morphophonological Aspects of Dutch.* Dissertatie, departement linguistiek Leuven.
- FODOR, J. et al. (1974) *The Psychology of Language. An Introduction to Psycholinguistics and Generative Grammar.* McGraw-Hill, New York.
- GARDNER, H. (1985) *The Mind's New Science. A History of the Cognitive Revolution.* Basic Books, New York.
- HENDRIX, G.G. (1977) LIFER: a Natural Language Interface Facility. *Sigart Newsletter* 61 (February 1977), 25-26.
- JOHNSON, T. (1985) *Natural Language Computing: the Commercial Applications.* OVUM Report, London.
- SHAPIRO, S.CP (ed.) (1987) *The Encyclopedia of Artificial Intelligence* (2 volumes). Wiley, New York.
- SRAAKHERKENNING (1988) *Bundel n.a.v. een lessenreeks rond NTV, spraakverwerking en computer vision.* Belgisch Instituut voor Regeltechniek en Automatisering, Werkgroep AI en Expertsystemen.
- TAALONDERWIJS en TECHNOLOGIE (1987) - Overzicht van NTV-onderzoek aan het departement linguistiek van de K.U.-Leuven.

Enkele inleidende werken rond NTV

(stijgende moeilijkheidsgraad)

- McTEAR, M. (1987) *The Articulate computer.* Basil Blackwell, Oxford.
- BRANDT-CORSTIUS, H. (1978) - Computer-taalkunde. *Coutinho.*
- KING, M. (ed) (1983) *Parsing Natural Language.* Academic Press, London.
- GRISHMAN, R. (1986) *Computational Linguistics. An Introduction.* Cambridge University Press.

- TENNANT, H. (1981) *Natural Language Processing: An Introduction to an Emerging technology.* Petrocelli, New York.
- WINOGRAD, T. (1983) *Language as a Cognitive Process.* Adisson-Wesley, Reading, Mass.
- ALLEN, J. (1987) *Natural Language Understanding.* Benjamin/Cummings, Menlo Park CA.