

‘SESAM SESAM’

eller: Om kognition, sprog, bevidsthed og deres
formodede lykkelige forening i det 21. århundrede*)

Jacob L. Mey

Rasmus Rask Institut for Lingvistik, Odense Universitet

Findes der grænser for KI (Kunstig Intelligens, AI)? Er den ‘ny’ KI bedre end den ‘gamle’? Og hvori ligger forskellen?

Denne artikel undersøger nogle af de påstande, der er fremsat af forkæmpere for den ny, ‘konnektionistiske’ eller ‘parallelt distribueret processerende’ (PDP) model. Jeg rejser spørgsmålet, hvorvidt mennesket på en meningsfuld måde kan sammenlignes med en informationsbearbejdende computer – en antagelse, der går igen i meget, der kalder sig kognitionsforskning.

Det er især to træk ved mennesker, der gør det vanskeligt at modellere dem på en computer: nemlig, intentionalitet og socialitet. Det individuelle menneske baserer sin identitet på det første af de to træk, der sociale menneske på det andet.

Endnu et problem udgøres af den uovervejede brug af begrebet ‘funktionsækvivalens’, som bruges i konnektionist-inspireret modeltænkning. Det er imidlertid ikke kun output’et, der er gyldigt som kriterium for succes: den måde man opnår output’et på, er mindst lige så vigtig.

Endvidere har konnektionistisk tænkning bevidst forsømt at tage de højere niveauer i vores mentale repræsentation af virkeligheden alvorlig, idet den har koncentreret sig om de lavere niveauer (‘det sub-symboliske’).

Et sidste problem har at gøre med muligheden for fredelig sameksistens og gensidig respekterende af territorierne mellem ‘klassisk’ KI (også kaldet GOFAI, ‘Good Old-Fashioned Artificial Intelligence’) og det ny paradigme.

Jeg afslutter med 10 teser til en opsummerende diskussion.

0. Indledning: en sand historie

Kan en computer åbne en dør? Ja, hvorfor ikke? Vores moderne teknik med voice prints og andet fint isenkram og programmel er utvivlsomt i stand til at konstruere systemer, der åbner døre når vi siger det til dem. Vi behøver ikke være nogle Ali Baba’er for at ytre vor moderne tids ækvivalent af et ‘Sesam, luk dig op’ – det er blot et spørgsmål om udvikling, dvs. tid og penge og lønsomhed.

*) Foredrag på symposiet ‘Sprog og Psykologi’ ved Københavns Universitet, 1. februar 1990.

Suzette Haden Elgin har i sin bog *The Judas Rose* (1987) beskrevet en fremtid, hvori sådanne ting er dagligdags. Her er et citat:

»Krol is grinning the satisfied grin of a triumphant man with a triumphant wife at his side as he tells the door to open and it slides silently into its case to let the new couple in.« (p. 264)

I og for sig er det ligegyldigt, hvilken formulering hr. Krol i ovenstående episode vælger at bruge. Han kunne sige: 'Let the door be open', 'Open the door', 'Open door', eller blot 'Open' eller endogså 'O' (ligesom i computer spillet *Adventure*); alt afhænger af, hvilke instrukser der er bygget ind i computeren og den tilhørende servomekanisme. Krol kunne tilmed sige: 'I declare this door to be open', omtrent som når man åbner en bro, eller indvier en ny bygning, park, eller deslige.

1. Er sagen virkelig så enkel?

Ved nærmere eftersyn viser der sig at være nogle vanskeligheder forbundet med vores fremtidsdrøm. Hvis vi spørger os selv, om computeren virkelig har 'forstået' os, så ville nogen måske svare: »Der er da klart: se blot på resultatet. Ordren er pareret, ytringen forstået«. Men for det første: hvad er det for en ytring, vi er kommet med her? Er det at sige 'O' for eksempel, at betragte som en specifik menneskelig, sproglig ytring, eller er det bare et eller andet signal, som det at blæse i en fløjte eller ringe på en dørklokke? Og dette gælder også i de fleste tilfælde hvor jeg bruger en mere elaboreret sprogform; for at tage Leech's eksempel: selv om »declaring a door open is not, normally, a means of opening a door« (1983:180), så kan det udmærket være det, som i vores tilfælde.

Men for det andet, og her nærmer vi os sagens kerne, behøver vi i det hele taget at tale om sprog i vores lille anekdote fra fremtiden? Hvad om vi fortsætter tankeeksperimentet, og forestiller os en døråbningsmekanisme, der opererer, ikke på lydbølger, men på de bølger der udsendes fra en hjerne i aktivitet? Det kan vel tænkes, at der findes, eller vil findes, muligheder for at 'dekodere' sådanne svingninger, omtrent som vi nu er i stand til at dekodere elektroniske svingninger fx. i billed- og anden informationsoverføring. Så kunne man *tænke sig en dør åbnet, og vupti! Så var den åben.*¹⁾

Det at springe sproget over, så at sige, betyder faktisk, at vi giver afkald på hvad Martinet har kaldt den 'dobbelte artikulation', hvorved tankeindholdet bliver symboliseret ved hjælp af et lille inventar af tegnelementer i henholdsvis den anden og den første artikulation. Set fra servomekanismens side sker der det, at den styrende hjerne lægger sig i bestemte folder, der bit for bit stemmer overens med de konfigurationer, der skal til for at mekanismen træder i kraft og åbner den berømte dør. Der er mao. totalt

sammenfald mellem hjernens 'soft- (eller wet-)ware' og den tilsvarende 'ware' i henholdsvis computerens og åbnemekanismens 'hjerne'. Denne tilstand af 'ekvisvingning' ville desuden svare til det tankeeksperiment, som også ofte rumsterer i science fiction (og tilmed alvorlig science): den direkte og totale sammenkobling af to menneskehjerner, og den totale og umiddelbare forståelse, en sådan sammenkobling kunne tænkes at afstedkomme mellem fx. et elskende par.

Futile dagdrømme? Jo, til en vis grad. Men de er også nyttige, eller kan gøres nyttige, hvis vi tager dem som yderste konsekvenser af en tænkemåde, der synes at karakterisere meget af den forskning, der for tiden går under navnet 'Cognitive Science'. I det følgende skal jeg diskutere ét sådant tilfælde.

2. 'Mors lille Ole'

»Mors lille Ole i skoven gik ...«. Vi kender alle sangen. Men hvorfor gik Ole i skoven? Hvad ville han der? Hvordan gik han? Alle disse spørgsmål kan besvares på en række niveauer, hvoraf den ene forekommer at være rent fysisk ('Hvordan gik han?' 'Tja, han satte vel det ene ben foran det andet'), et andet er mere 'mentalt' præget niveau ('Han gik med hovedet fuldt af de blåbær han glædede sig til at plukke'), osv. Og lige meget hvor dybt og indtrængende vi analyserer lille Oles blåbærtur, vil der altid være ting, som enten vi har glemt, eller kan tilføjles af andre, eller modsiges, ikke mindst af Ole selv eller af hans psykiater. Vi fristes til at sige, med en anden Ole (Fogh Kirkeby), at den eneste sande beskrivelse af en sådan tur, er en »ikke-beskrivelse« (1989:79).²⁾

Problemet med den gængse måde at beskrive ting på er, at man er nødt til at kondensere erfaring i ord, i symboler ('be-skrive' – kuriøst nok en metafor hentet fra den skriftlige meddelelse, ikke den mundtlige!). Det kunne være nærliggende at mene, at blot man var ude over denne beskrivelses tvang, det digitaliserende mønster, der tilsyneladende er inhærent i alle kommunikationssystemer, så ville man kunne kommunikere intuitivt, i billeder, og skue i fuldkommenhed det, som hidtil kun har været et »gådefuldt spejlbillede« (I Kor. 13:12)³⁾.

Sagen er imidlertid langt fra så enkel. For selv om 'et billede siger mere end tusinde ord', så er det stadig væk et billede som er 'indfattet' i en større tekst, som også indeholder *ord*; og vi véd strengt taget heller ikke, *hvordan* billedet 'siger', dvs. hvordan billedets 'tale' bliver formidlet og kommunikeret.

Søren Brunak og Benny Lautrup, hos hvem jeg har hentet citatet (om billedet og de 1.000 ord), bruger det i en lidt anden sammenhæng. De vil vise, at den sædvanlige verbale kommunikation er en utrolig sendrægtig affære, informationsteknologisk set, hvis man sammenstiller den med fx. bil-

ledkommunikation. Med deres egne ord: »[d]en mængde information der overføres via billedet er altså 1000 gange større end den, der kommer gennem lyden« (1988:136); eller, som Brunak og Lautrup siger sammesteds, »Intuitiv tankevirksomhed er symbolfri og langt rigere på information end symboliserede ræsonnementer« (ibid.:139).

Men: for det *første* er der problemer med sammenligningsgrundlaget: information udtrykt i bits; det holder nemlig ikke. Som velkendt, er informationsvidenskabens informationsbegreb på ingen måde identisk med det, vi i daglig tale kalder ‘information’. Dette kan man udtrykke populært, som Ole Kirkeby gjorde i citatet ovenfor (»en ikke-beskrivelse«); eller man kan med en anden Ole (Togeby)⁴ sige som følger, lidt mere i overensstemmelse med gængs videnskabelig sprogbrug:

»Der er imidlertid den katastrofale fejl i deres [Brunak og Lautrups, JM] ræsonnement, at de sætter ‘antal bits’ lig med ‘mængde af information (som mennesker bruger i deres tankevirksomhed)’ ... [mens] det nærmest [er] omvendt. Det vi mennesker bruger som information er negentropi [negativ entropi, JM], struktur, system, forudsigelighed, regler som reducerer valgfriheden i signalet (bits’ene).«

Kort sagt, de store Ole’r er enige om én ting: den første, lille Ole er alt andet end en computer, der programmeres i et sprog, som kan oversætte(s) til og fra ‘bits’. (Dette vender jeg tilbage til i et senere afsnit).

Men der er også noget *andet*, som ikke passer. Sæt, vi har spændt en computer til lille Oles hovede, mens han drager glad og velfornøjet afsted på sin blåbærtur. Og sæt videre, at vi har en tilsvarende computer koblet til vores egen hjerne. Og lad os nu forudsætte, at alt, der foregår inde i lille Oles hovede, nøjagtigt og præcist og udtømmende kommer frem på ‘skærmen’ hos denne computer, og i sidste instans fremstår som repræsentationer inde i mit hoved, hvor der måske sidder en cartesiansk lille mand og stirrer på min ‘skærm’ for at fortælle mig, hvad lille Ole tænker og føler, eller, mere i pagt med tidens ånd, hvor der produceres nervekoblinger og ‘baninger’, der lige præcis er de samme, som findes hos Ole. Ville jeg så endelig have løst mysteriet: Hvorfor går lille Ole i skoven, ‘What makes him tick’?

Svaret er selvfølgelig Nej. Og det er det af den simple grund, at dét, Ole føler og ønsker og mener, især om hvorfor han gør som han gør, det er noget han kun selv véd (eller ikke en gang selv véd, om man skal tro dybdepsykologerne)⁵. Vi løber her ind i en række problemer, hvoraf det første er *intentionalitet*, det andet *socialitet*. Med hensyn til det første vil jeg henvise til Searle’s udførlige behandling i sin kendte polemik mod ‘strong AI’ fra begyndelsen af firserne (Searle 1983), og ellers lade det ligge, selv om det selvfølgelig ikke er uden forbindelse til det andet problem, socialiteten.

En væsentlig og grundlæggende fejltagelse ved alle modeller af ‘hjernen’

i moderne kognitive sammenhænge er at de baserer sig på det enkelte, isolerede menneske og hans eller hendes enkelte, isolerede tankevirksomhed, som denne menes at gå for sig i hjernen. I virkeligheden er det, den menneskelige hjerne repræsenterer, på et givet tidspunkt i dens ejers historie, ikke blot dens aktuelle tilstand; det er snarere resultatet af et livs erfaring, hvori der indgår, som uundværlig og væsentlig komponent, menneskets *sociale* situation og samfundsmæssige forhold. Man er, hvad man er, ikke gennem munden eller hjernen, men i kraft af ens ontologiske status som medlem af menneskeslægten, hvor 'her og nu' kun har betydning som 'alles her og nu', ikke den enkeltes. Dybest set er man sig selv kun som en anden, i kraft af en anden; derfor kan min selv-repræsentation kun være en ikke-selv-repræsentation; men det samme gælder alle 'andres' væren-hos-sig-selv. Et sådant knudepunkt i et stadigt ekspanderende net af repræsentationer vil desuden naturnødvendigt ændre sig i det øjeblik, en ny relation træder ind i billedet, fx. ved at der indføres en computersimulation. Resultatet bliver en uendelig regression, og konklusionen må være, at »der er kun én slags computere, der er sig selv, og det er .. mennesker«, som Ole Fogh Kirkeby fyndigt udtrykker det (1989:79). I det følgende vil jeg undersøge to specielle aspekter ved problematikken om menneskets bearbejdelse af 'information'.

3. »Sind die Bienen Reflexmashinen?«

Ovenstående berømte (og vist nok retoriske) spørgsmål, der i denne form først stilledes af den gamle filosof og naturvidenskabsmand Thure von Uexküll en gang i tyverne, kan varieres til at passe på vores problemstilling:

'Er menneskets hjerne
en informationscisterne'?⁶⁾

Jeg vil stille spørgsmålet, og svare på det, ud fra to mulige (og praktiserede) synspunkter, som jeg kort skal karakterisere som 'ES' (ekspertsystemer) og 'FÆ' (funktionsækvivalens).

3.1. ES

Den gængse forestilling om menneskehjernen som informationsprocesse-rende enhed minder mest om et ekspertsystem: et sted, hvor man lægger viden ('information') og kan få fat i den igen, når der er brug for det, eller hvor man opmagasinerer viden således, at den er tilgængelig ikke bare for ejeren af hjernen, men for alle, der har adgang til, evt. har betalt for at komme ind på 'systemet'.

Således beset, reduceres den menneskelige hjernevirksomhed til bearbej-

delse af viden (jvf. 'knowledge engineering', som dens computer-pendant ofte kaldes). Selvfølgelig er der ingen tvivl om, at det er yderst nyttigt at have adgang til en mængde viden, og at det er samfundsmæssigt gavnligt at kunne stille denne viden til rådighed for flere (mod eller uden beregning), men det ville være en mildest talt dårlig ide at gå ud fra, at den menneskelige hjerne og dens måde at arbejde på havde noget til fælles med et sådant system – det være sig da, at hjernen måtte *tilpasses* systemet, for at sidstnævnte overhovedet kunne fungere! Men faktisk er det dette, der finder sted, når 'vidensingeniører' forsøger at 'malke' den menneskelige ekspert for dennes viden; erfaringen viser dog, at det oftest mislykkes, netop fordi et menneskeshovede nu en gang ikke er indrettet som en informationsmaskine. Og det vil omvendt sige, at ekspertsystemers nytte og mulige frugtbarhed er meget mere begrænset end de første entusiastiske rapporter ville have os til at tro; selv om man nu er begyndt at tale om flere 'generationer' af ES, idet man har fjernet sig fra de oprindelige, meget primitive parat-vidensmodeller, og har klaret at indbygge en vis form for ræsonnerende logik (vha. såk. 'inference rules') i modellerne, så er der stadigvæk tale om en yderst primitiv model for den menneskelige tankevirksomhed.

Når det gælder om at tage beslutninger på grundlag af et ES, eller endog at overlade beslutningen til et sådant system, vil de allerfleste mennesker nok betænke sig. Jeg kender af egen erfaring menneskelige eksperter, fx. læger, der godt nok har et ES i deres praksis, men som aldrig ville overlade en diagnose til systemet; og den velkendte anekdote om Hubert Dreyfus, der ikke en gang ville lade sit ES bestemme, hvilken bil han selv skal anskaffe sig, er næsten for velkendt til at bringes her endnu en gang (se Dreyfus & Dreyfus 1986; genfortalt i Mey 1989).

Hvis vi tænker tilbage til det, der blev sagt i forrige afsnit, så er det egentlig også en underlig tanke, at et menneske skulle være i stand til at formulere og udstationere sin viden i system-form, al den stund det slet ikke er sikkert, at jeg véd alt, hvad jeg mener at vide, og i hvert fald ikke véd det på en 'computeriseret' måde. Mht. det første kan vi tænke på den malaise, der ville ramme hele psykolog- og psykiaterstanden, hvis folk virkelig var klar over, hvad der foregik i deres indre (bevidst- og underbevidsthed); når det gælder det andet, kan man blot minde sig selv om, hvor mange gange man har stået med fortvivlelsen i sindet, fordi det ikke lykkedes for én at udtrykke det, man dog vidste så godt – for slet ikke at tale om ens følelser!

3.2. FÆ

Det er næsten blevet til et dogme i kognitionsforskningens kredse, specielt i U.S.A., at der findes et generelt princip for al kognition: nemlig, at computeren endelig har vist os, at det er ligegyldigt, hvilken *form* vores mentale oplevelser eller tanker har, *konkret*, så længe vi kan definere, eller programmere, en *abstrakt* ækvivalens, hhv. en funktion, som 'oversætter', formid-

ler, fra den ene form til den anden. Således er det fx. ligegyldigt i hvilket programmeringssprog man formulerer sit problem, eller hvordan problemet kører på en bestemt computer: der findes altid rutiner til programkonvertering, og hvis de ikke findes, kan de skabes. Computeren er den universelle Turing-maskine, som i princippet kan simulere alle tænkelige, ekspliciterede tankeprocesser og beregninger. (Læg forøvrigt mærke til den skjulte betingelse, der ligger i ordet 'ekspliciterede': den har sneget sig ind her med vilje! Herom senere).

Som et aktuelt udtryk for hvad man kunne kalde den rene 'funktionalisme' (fordi eftertrykket er på et programs *funktion*, dens form, ikke på dets konkrete realisering), jfr. flg. citat:

»The term 'functional', as used here, communicates the idea that, in order to have a mind, a system did not have to have a specific *physical* makeup; it just had to be so organized that it was capable of realizing a specific sort of *abstract* computational structure. This computational structure might be implemented on different kinds of physical device: electronic circuits or human brains, or on other kinds of more exotic hardwares yet to be discovered. This sort of approach – which might be called 'pure' *functionalism*, or 'pure' *computationalism* – claimed that the mind could be discussed in terms which were entirely *implementation-independent*«

(Sharples et al. 1989:301; udh. i orig.)

Læg mærke til, at der i dette citat bruges to begreber: 'mind' og 'brain', der i den gængse filosofiske forståelse af menneskets mentale evner plejer at blive holdt adskilt, men i citatet bruges uden reel forskel i indholdet. Det er derfor ikke mærkeligt, når forfatterne nærmest automatisk konkluderer, at denne 'pure functionalism' måske vil gøre kål på den evindelige strid om hvorvidt der er noget 'mere' i hovedet end en hjerne, bestående af titals milliarder af celler (neuroner og synapser), og om 'ånden' må siges at være et (simpelt?) materielt fænomen, en epistruktur på en materiel basis, eller om den er noget, som er kvalitativt forskelligt fra materien, måske oven i købet udødeligt og guddommeligt. Alle disse ældgamle stridsspørgsmål vil jeg lade ligge; blot vil jeg her bemærke tre ting:

For det *første* må bevisbyrden for, at den menneskelige hjerneaktivitet 'restlos' kan afbildes ved hjælp af et computerprogram hvile på dem, der påstår dette. Det samme gælder for hypotesen om mennesket som 'information processing device'. Desværre sker der oftest det, at man starter ud med at hypotesere, at menneskets åndelige funktioner kan sammenlignes med en computer, der processerer information, for så et par sider længere henne i teksten at gå ud fra, som en selvfølge, at mennesket *er* en sådan device. (For en mere udførlig kritik af denne fremgangsmåde jfr. Mey & Talbot 1988).

For det *andet* kan vi godt enes om, at der er visse processer og (også menneskelige) aktiviteter, der frit kan oversættes til et andet medium end det, de var 'født i'; fx. kan man spille skak enten ved at sidde overfor sin modstander ved samme bræt, eller man kan spille radio- eller korrespondance-skak: det er stadigvæk skak, man spiller, og det bliver i den sidste ende hip som hap hvorvidt man flytter brikkerne på et bræt, eller slår deres position ind på (fx.) et elektronisk medie. I lingvistikken er denne tankegang jo netop foreviget (gjort virkelig udødelig, kunne man sige) ved Saussure's berømte eksempel med skakbrikkerne, noget der skulle vise den funktionelle ækvivalens mellem sproglige størrelser, uanset deres materielle udseende eller beskaffenhed.

Men for det *tredje* må vi huske på, at selve den saussure'ske, nærmest kanoniserede, metafor i sidste instans bryder sammen, når vi gør alvor af at tage den bogstaveligt, på ordet. Som en hollandsk Hjelmlev-forsker allerede måtte konstatere midt i halvtredserne, kan man ikke spille skak med brikker, der er lavet af vand: »Water will not make chessmen« (Siertsema 1954) (medmindre temperaturen er under 0° C – men så har vi at gøre med is, og ikke vand). Der er mao. naturgivne, ikke-vilkårlige elementer, selv i den mest funktionelle opfattelse af sprogets ækvivalens på tegnplanet. Og endnu værre bliver det, når vi bevæger os højere op i det sproglige hierarki: ingen fornuftige sprogbrugere vil vel påstå, at det er det samme, hvilken sproglig klædebon ens budskab er iført. Pragmatikken har lært os værdien af, og sandheden i det fortærpede slogan, som oprindeligt blev skabt af Marshall McLuhan: »The medium is the message«.

4. »Redningsmanden er opstanden ...«

Som allerede antydet i sidste afsnit, er der mange, som tror, at den nyere udvikling i computervidenskaberne vil hjælpe os ud af, eller endog gøre det helt af med, århundrede gamle *crux*'er som spørgsmålet om 'mind' og 'brain', og hvordan de forholder sig til hinanden. Den løsning, der foreslås bl.a. af fortalere for den ny retning kaldt 'connectionism', går ud på, at spørgsmålet endegyldigt likvideres ved en *reductio*, ikke *ad absurdum*, men snarere *ad inane*: spørgsmålet vises ikke at have nogen som helst betydning, fordi dets materielle grundlag er bortfaldet. Hvis man siger 'brain' på den rigtige måde, behøver man slet ingen 'mind', for nu at fremstille det lidt populært.

Men hvori består så denne nymodens løsning på vort gamle problem? Det er en fejltagelse, siger man, at tro at den menneskelige hjerne nødvendigvis er organiseret efter det 'symbolske' princip: dvs. at tro, at der for verden omkring os (genstande, egenskaber, forhold (abstrakte såvel som konkrete) osv.) nødvendigvis findes en eller anden 'genspejling', en symbolisering, i hjernen. Specielt er det unødvendigt, og tilmed forkasteligt, som

nogle har gjort, at tale om et repræsentationssprog, en *lingua mentalis* (Wierzbicka) eller *mentalese* (Fodor), som skulle bestå af sådanne symboler, forbundet ved visse regler (som man også kender det fra en vis lingvistisk retning, hvor symbolerne for de sproglige størrelser forbindes ved abstrakte (genskrivnings- og transformations-) regler efter forskellige systemer, afhængigt af hvilken variant af den chomsky'ske skole, man beskæftiger sig med).

Alt dette, siger konnektionisterne, er unødvendigt og komplicerer hele foretagendet. Vi har slet ikke disse symboler behov: hvem har vel nogensinde *set* et symbol, og hvordan kan man vide at hjernen bruger dem i sin forarbejdelse af informationsinput'et? Hvad værre er: symbolerne er en unødvendig hæmsko for vores opfattelse af de mentale processer. Først når vi skiller os af med dem, og lægger mærke til, hvordan hjernen selv fungerer, kan vi gøre os nogen forhåbninger om at forstå tænkning og sprogbrug og alle de andre mentale processer. Symboltænkningen er en besværlig omvej: lad os i stedet skyde genvej, og lære af hjernens egen måde at processere sit input på.

Alt dette kan lyde en smule fantasifuldt og ikke særlig seriøst: for hvad ved vi egentlig om hjernen og dens måde at gøre tingene på? Men det er netop her, hunden ligger begravet: vi véd faktisk en del om hjernen og dens processer, bl.a. om de sidstnævnte hastighed. Som jeg tidligere antydede, er hjernen, *hvis* den skal sammenlignes med en computer, i hvert fald ikke én af de hurtigste; og hvis man måler hastigheden i de neurale baner og sammenligner den med det, der behøves for at processere symboler én efter én i små 'skridt', ja så er der faktisk et misforhold i størrelsesordenen af 1:1.000! (Feldman & Ballard 1982; for en redegørelse på dansk, se Andersen 1988). Dvs. for at kunne klare selv de simpleste mentale opgaver måtte hjernen løbe ca. 1.000 gange hurtigere end den faktisk gør for at opnå den tid, der i virkeligheden bruges. Noget stemmer ikke her.

Hvis nu hjernen var en computer (læg mærke til *hvis*'et!), hvad skulle der så til for at den kunne klare opgaven på 'normeret' tid? To ting: opgaven skulle fordeles ud på enkelte 'underleverandører', og disse skulle arbejde samtidig, i parallel, hellere end i serie.⁷⁾ Denne sidste teknik kaldes ofte 'PDP', 'parallel distributed processing'; jeg vil her ikke komme nærmere ind på den. Mht. den første taler man ofte om det 'subsymbolske' niveau, dvs. et niveau der ligger under det symbolske og adskiller sig fra dette ved at være fuldt ud ekspliciteret og formaliseret, og således beskrivbart og processerbart ud til de mindste detaljer.

5. Livet under tegngrænsen

Idéen med at dele tegn op i mindstedele, der ikke selv er tegn, er selvfølgelig ikke ny: jeg har allerede nævnt Martinet's 'double articulation'; man kunne

også henviser til Hjelmlev's berømte 'glossemer', mindste-enhederne i sprogsystemet. Kendt er også Jakobson og Halle's opsplitning af sproglydene i 'distinctive features', der kan kombineres i temmelig frie mønstre til bundter, kaldt fonemer.

For alle disse mindstestørrelser gælder imidlertid, at de (selv om de ikke indgår i nogen form for tegnrelation *direkte*) alligevel har et vist *fundamentum in re*, fx. ved at tilknyttes en vis indholds- eller udtrykssubstans, eller ved at 'symbolisere' (om jeg så må sige) bestemte lyd kvaliteter – selv om det ikke altid var let at få øje på (eller skulle jeg sige 'have øre for'), hvad fx. et træk som 'forward-flanged' stod for, for slet ikke at tale om de forskellige skikkelser et træk som 'grave' kunne optræde i.

Med det såk. 'subsymbolske' niveau og dets indbyggere forholder det sig imidlertid ikke sådan. For eksempel er de fonetiske træk, som forekommer her (de såk. 'Wickelfeatures') udelukkende defineret ved en *position*: en 'Wickelphone'-repræsentation af det engelske ord *melt* ville bestå af sættet (eller 'map'-en) [# me, mel, elt, lt #] (hvorved faktorernes orden strengt taget bliver uvedkommende (Pinker & Prince 1988:109), netop fordi de individuelle 'træk' er fuldstændig bestemt ved deres position).⁸⁾

For nogle (fx. Smolensky 1988) er hypotesen om det subsymbolske niveau hovedkendemærket for hvad ofte kaldes 'ny AI' ('gammel AI', eller 'GOFAI', er så den 'good old-fashioned artificial intelligence', der baserer sig på en symbolsk behandling af stoffet: hertil hører de kendte AI-skoler såsom Schank's (tidligere ved Yale-universitetet)). I modsætning til den 'klassiske' opfattelse hævder connectionisterne, at vi ikke behøver et uafhængigt niveau, hvor symboler repræsenterer 'virkelighed': faktisk behøver man slet ikke at tale om 'repræsentation' på det symbolske niveau – hvis vi ellers vil kalde det et 'niveau' overhovedet! Problemet med, hvordan man kommer frem til de trods alt nødvendige enheder (såsom fonemer, morfemer, ord, osv.) løses let og elegant ved, at de subsymbolske elementer associeres til hinanden ved vægtning af deres forbindelser; denne vægtning fremkommer ved, at man sender impulser gennem systemet i form af korrekt input, dvs. subsymbolske frekvenser der faktisk forekommer, og derfor præger de 'rigtige' forbindelser fremfor de forkerte og mislykkede.

Hvis man nu forestiller sig de subsymbolske enheder placeret i store netværk, der består af indbyrdes fuldt forbundne celler (eller, alternativt, af en menneskehjerne med dens forbundne neuroner og synapser), kan indkodningen og afkodningen af elementer og deres sammensætninger foregå ved at man aktiverer visse forbindelser, mens man lader andre hvile. I næste omgang vil der så optræde en selvforstærkende effekt, hvorved de rigtige bliver endnu rigtigere, og den korrekte forbindelse efterhånden cementeres i netværket, resp. hjernen.

Konnektionisterne gør gældende, at en subsymbolsk 'repræsentation'⁹⁾ er den eneste rigtige og mulige, hvis man skal bygge en rimelig model af den menneskelige hjerneaktivitet (fx. i børns tilegnelse af uregelmæssige verbal-

former); konkret menes der her en konnektionistisk-inspireret computer-model, der faktisk simulerer, med en vis grad af succes, hvad der sker i børn. Men hovedbegrundelsen er af en mere formel karakter: de operationer, som en *symbolsk* repræsentation nødvendigvis må forudsætte, er alt for komplicerede til at lade sig simulere på en maskine. De er, med andre ord, ikke tilstrækkeligt *eksplicite*; deraf hvad Smolensky kalder «the connectionist hypothesis» (som han pudsigt nok formulerer negativt):

»The intuitive [dvs. human] processor is a *subconceptual* connectionist dynamical system that doesn't admit a complete, formal, and precise *conceptual*-level description«

(1988:7; min fremh.)

6. Den økumeniske løsning

Hvis det er rigtigt, som Smolensky siger, at den symbolske model principielt ikke er i stand til at levere den »complete formal account of cognition«, som videnskaben rettelig har krav på, så er der ingen tvivl om, at den konnektionistiske model er den symbolske overlegen. Men bemærk, at det her drejer sig om en teoretisk (eller mere præcist, en *meta-teoretisk*) påstand, der ikke direkte kan afprøves eksperimentelt. Hvis vi nu for et øjeblik antager, at konnektionisterne har ret, for så vidt deres modeller faktisk kan fremvise gode, måske tilmed imponerende resultater i simuleringen af menneskelige åndelige aktiviteter, kunne man så alligevel ikke antage, at mennesket (dette '*animal symbolicum*', som Cassirer kaldte det) brugte symboler og symbolske repræsentationer i sådanne aktiviteter, og at der kunne være noget ved forsøgene at repræsentere og simulere en sådan symbolik på en computer? Kunne man endvidere ikke tænke sig, at symbolsk og sub-symbolisk repræsentation havde hver sit domæne, dvs. at de hver for sig var gode til nogle ting, men måske kunne suppleres af hinanden? Fx. synes en digital, sekventiel repræsentation at passe meget dårlig sammen med den lynhurtige opfattelse som vi mennesker har gennem øjet: forskningsfeltet 'computer vision' har da også i stigende grad været præget af konnektionistiske metoder og programmer. Andre menneskelige aktiviteter synes derimod i højere grad at være præget af 'step-by-step processing' (selv om det intuitive 'blik' i løsning af matematiske opgaver, eller psykiaterens 'gaze', tilføjer ræsonnements- og slutningsprocesserne en værdifuld 'visuel' komponent).

Den økumeniske løsning på kontroversen mellem konnektionister og symbolikere kunne da enten bestå i en deling af territoriet, eller (som mange har foreslået) i en re-interpretation af konnektionistisk, hhv. symbolsk processering som forskellige *niveauer* i en videnskabelig diskurs. Fra fysikken ved vi, at kvantemekanikkens størrelser og begreber ikke uden videre

kan lade sig oversætte til den klassiske mekanik: begrebet 'naturlov' har en helt anden vægtning når vi taler om himmellegemer end når vi bruger termen i en subatomisk kontekst. På denne måde kunne man (uden at forfalde til strikte hierarkiske betragtningsmåder) opfatte det subsymbolske som en 'mikroteori', som forholdt sig til 'makroskopisk' kognitionsvidenskab som kvantefysikken til den newton'ske.¹⁰⁾

Imidlertid ser konnektionisterne ikke med blide øjne på sådanne forsøg på at stifte fred mellem hytterne. For dem er modellerne væsentligt og uigenkaldeligt forskellige og uforenelige; ikke blot er de inkonsistente, men de er gensidig inkompatible. Der er imidlertid et andet problem med niveauernes indbyrdes forhold, som rejser spørgsmålet om, hvad det egentlig er, man er ude på at (sub)symbolisere.

7. Mellem to stole?

Overskriften kunne lige så godt have heddet: 'mellem to niveau'er', for det er dem, det drejer sig om her. Problemet ligger, kort sagt, i de *analytiske* operationers forhold til *syntesen*. Hjelmlev plejede at sammenligne disse operationer med en elevators køren frem og tilbage mellem etagerne: det er ligegyldigt om man kører op eller ned, stuen er stuen og 5. etage er 5. etage, uanset fra hvilken højde man nærmer sig dem. Metaforen skal tjene til at vise os, at analyse og syntese er to sider af samme sag, inverse funktioner, så at sige.

Men er de nu også det? I lingvistikken er det nærmest anerkendt som en trosartikel, at når man analyser en sætning i dens mindste bestanddele, dvs. morfemer (evt. fonemer), så er man også i stand til at syntetisere den 'samme' sætning ved at bruge genskrivnings- og transformationsapparatet. Imidlertid kan lingvistikken kun gøre dette, og har lov til det, så længe den (bevidst eller ubevidst) skærer alt det 'ekstralingvistiske' bort: dvs. lingvisten ser på de sproglige fænomener som et immanent system, og overlader anvendelsen til andre. (Bemærk i parentes, at Saussure aldrig gjorde sig umagen at fortælle os, hvordan men egentlig kommer fra 'langue' til 'parole').¹¹⁾

Men i det øjeblik, det drejer sig om størrelser fra virkeligheden, såsom fysiske, biologiske, biokemiske osv., bliver sagen en anden. I det foreliggende tilfælde er problemet hvordan vi med nogenlunde sikkerhed kan afbilde den formale beskrivelse fra det subsymbolske niveau på de begreber, menneskene bruger i deres daglige omgang med verden. Eller: givet, at der er en mulighed for at kortlægge menneskets begrebsverden udtømmende og tilfredsstillende (hvilket forresten ikke er sikkert), hvordan kan vi være visse på at, og i hvilken forstand, denne kortlægning, i al sin formale pragt, svarer til den menneskelige virkelighed, og kan genkendes som sådan af mennesker?¹²⁾

Man kunne også spørge omvendt (idet man lader elevatoren køre ned, i stedet for op): hvordan ved vi, at det subsymbolske niveau lige præcis ligger i stueetagen, og ikke fx. i kælderens? Hvilken garanti har vi for, at en modellering, der baserer sig på idéen om et neuralt netværk, ikke selv er en slags 'symbolisering' af det, der foregår længere nede? Lad mig citere de to allerede nævnte forskere, Dietrich og Fields:

»If there is something to the claim that concept-level descriptions are fuzzy in principle, what prevents us from using the same argument to show that subsymbolic descriptions are only fuzzy approximations of neural descriptions, or that neural descriptions are only fuzzy approximations of biochemical descriptions, and so forth?« (1988:30)

Det forekommer mig at vi her er i gang med at åbne den proverbiale 'can of worms', en pandora-æske af repræsentations- og legitimationsproblemer – vel at mærke, hvis vi forsøger at løse disse problemer på det abstrakte, for ikke at sige apodiktiske, niveau.

8. En slags konklusion

Det er på tide, vi opsummerer, efter at have stiftet et (indrømmet, hastigt og noget overfladisk) bekendtskab med de konnektionistiske påstande og nogle af resultaterne. Følgende teser vil jeg fremlægge som bidrag til en konkluderende diskussion:

1. Konnektionisme, eller PDP, er et interessant forsøg på at tilføre debatten om maskinel simulering af menneskelig adfærd nye impulser.
2. Den konkrete virkeliggørelse af modellen bygger dels på et 'maskinelt' (hardware) argument, dels på mere videnskabsteoretiske og filosofiske tankerækker.
3. Mht. det første ('hardware-argumentet' – eller skulle vi måske snarere tale om 'wetware'?) må man anerkende, at en seriel simulation er utilstrækkelig, og at en eller anden form for parallelitet og distributivitet er påkrævet.
4. Mht. det andet, mere filosofisk prægede ræsonnement hersker der berettiget tvivl, hvorvidt de eksklusive påstande (som bl.a. Smolensky har fremsat) om det formelle og eksplicite holder stik, når man opfatter dem som specifikt gyldige for simulering af menneskelig adfærd.
5. Hovedgrunden for denne tvivl ligger i, at forestillingen om mennesket som 'information processing device' strengt taget er en metafor uden beviskraft; bevisbyrden for alle påstande, der (direkte eller indirekte) påberåber sig denne metafor, ligger stadig hos dem, som fremsætter påstandene, hhv. indfører metaforen.

6. Mht. deres praktiske anvendelighed er konnektionistiske modeller svage, da de ikke tillader fejlkorrektion 'undervejs' (dvs. inde i simulationen), idet modellens processer er principielt ugennemskuelige.
7. I de tilfælde, hvor simulationerne tænkes brugt i en samfundssammenhæng (fra bilregistrering til SDI, eller 'Star Wars') kan det medføre problemer at placere ansvaret for fejl korrekt, når beslutningerne er taget i medfør af en bestemt, ureproducerbar netværkskonstellation.
8. I det hele taget er en model for menneskelig aktivitet, som baserer sig på den enkelte hjernes processer af 'information' dårligt egnet til at fange kompleksiteten af menneskets sociale adfærd. (Denne kritik gælder selvfølgelig ikke udelukkende konnektionismen).
9. Det må være muligt at forene konnektionistisk og symbolsk modellering af menneskelig hjernevirksomhed ved at antage en fredelig koeksistens af det subsymbolske og det symbolske niveau.
10. Denne sidste antagelse er især af betydning, når man skal til at simulere mere udviklede kognitive processer end dem, konnektionismen påberåber sig som empirisk støtte for teorien (hovedsagelig sprogindlæring på fonemisk og morfemisk niveau).

NOTER

- 1) Også dette motiv har været udnyttet af science fiction-forfattere: Ursula Le Guin beskriver i sin bog *The Lathe of Heaven* en mand, der bogstaveligt talt kan drømme sig til hvad som helst, med alle de konsekvenser det medfører, ikke bare for ham selv, men for alle i hans omgivelser, der pludselig bliver forsat til en fjern fortid eller en aldeles ukendt fremtid.
- 2) Denne Ole var på vej til bageren, forresten, bymenneske som han er.
- 3) »Nu ser vi ved Hjælp af et Spejl på gaadefuld Vis«; Erikke Rosenørn-Lehns oversættelse (Det Nye Testamente, Roskilde 1932).
- 4) I sin kritik af Brunak og Lautrups formuleringer. (Det er i øvrigt hos denne Ole nr. 3, jeg har fundet citatet fra de to herrer).
- 5) Men det må vi hellere spørge endnu en anden Ole om (den fjerde i rækken) (Andkjær Olsen 1988).
- 6) For at få rimet med, har jeg måttet give afkald på forarbejdsaspektet – medmindre man henleder tankerne på en helt bestemt slags cisterne; men nok derom, inden min lille vittighed begynder at lugte.
- 7) Forbavsende nok har denne tanke i årtier været kendt hos økonomerne under betegnelsen 'omvejsproduktionens synkronisering'; som sådan er begrebet specielt udviklet af den såk. 'østrigske' skole. (For referencer, se Meij 1943, 1945).
- 8) Dette minder ikke så lidt om tidlige forsøg inden for maskinel tekstproduktion; disse var baseret på 'Markov'-regler, hvorved sandsynligheden af et elements forekomst udelukkende bestemmes af, hvad der sidst var processeret.
- 9) Til trods for den oprindelige modvilje mod at bruge ordet 'repræsentation' i konnektionistiske sammenhænge, er der nu flere, der accepterer denne sprogbrug. Værdien af konnektionisme-prægede opfattelser af fx. menneskelig indlæring hænger ikke på en bestemt sprogbrug; for mange konnektionister er «connectionism nowadays ... fully representational» (Jordan Pollack, pers. medd.).

- 10) Formuleringen er hentet fra et bidrag til konnektionismediskussionen af Dietrich & Fields, som svar til Smolensky's oplæg, i samme nummer af *Behavioral and Brain Sciences* 11(1), 1988.
- 11) Tanken skyldes Lene Fogsgaard i sin disputats fra 1989 om *Ser og Estar* (København: Basilisk).
- 12) Her ligger forøvrigt et helt andet problem: mangel på intuitiv gyldighed, og i det hele taget rimelig genkendelighed hos de konnektionistiske modeller. Det ville føre for langt at komme ind på dette her; der henvises til Mey (1989).

REFERENCER

- ANDERSEN, H.B.. 1988. At modellere tavs viden. *Psyke og Logos* 9(2):230-252.
- BRUNAK, S. og B. LAUTRUP. 1988. *Neurale netværk: Computere med intuition*. København: Munksgaard.
- DIETRICH, E. and C. FIELDS. 1988. Some assumptions underlying Smolensky's treatment of connectionism. *Behavioral and Brain Sciences* 11(1):29-31.
- DREYFUS, H.L. and S.E. DREYFUS. 1986. *Mind over machine*. Hillsdale, NJ: Ablex.
- ELGIN, S.H. 1987. *The Judas Rose* (Native Tongue II). New York: DAW.
- FELDMAN, J.A. and D.H. BALLARD. 1985. Connectionist models and their properties. *Cognitive Science* 6(3):205-254.
- FOGSGAARD, L. 1989. *Ser/Estar og udsigelsen*. København: Basilisk.
- KIRKEBY, O.F. 1989. *Ægte intelligens: Om bevidsthedens program*. København: Munksgaard.
- LE GUIN, U.K. 1973. *The lathe of heaven*. New York: Fontana.
- MEIJ, J.L. 1943, 1945. *Leerboek der bedrijfseconomie I, II*. Wassenaar: Delwel. (Lærebog i driftsøkonomi; på nederlandsk).
- MEY, J.L. 1989. The proper proper treatment of connectionism. *Springer Lecture Notes on Computer Science* 308:1-15. Heidelberg/Berlin/New York: Springer. (= Proc. 4th Portuguese AI Conference, Lisbon 1989).
- MEY, J.L. and M. TALBOT. 1988. Computation and the soul. *Semiotica* 72(3/4):291-340.
- OLSEN, O.A. 1988. *Ødipuskomplekset*. København: Reitzel.
- PINKER, S. and A. PRINCE. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73-193.
- SEARLE, J.R. 1983. *Minds, brains and computers*. London: BBC (The Reith Lectures).
- SIERTSEMA, B. 1954. *A study of glossematics: Critical survey of its fundamental concepts*. The Hague: Nijhoff.
- SHARPLES, M., D. HOGG, C. HUTCHISON, S. TORRANCE and D. YOUNG. *Computers and thought: A practical introduction to Artificial Intelligence*. Cambridge, Mass./London: Bradford/MIT Press.
- SMOLENSKY, P. 1988. On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11(1):1-23.
- TOGEBY, O. 1989. *PRAXT: Pragmatisk tekstteori*. Københavns Universitet (upubliceret ms.).