

MORLILLE KAN IKKE FLYVE Natursproglig semantik og datamater

Henrik Prebensen

1. To semantiske grundopfattelser

Semantik er læren om betydning. Semantikkens mål er en teori for relationen mellem sproglige *udtryk* og disse udtryks *indhold*, det udtrykkene *står for*, eller hvad man nu vil kalde det andet relat i betydningsrelationen: »*A betyder B*«.

I grove træk kan vi tale om to forskellige teorier om semantik, som vi kan kalde *repræsentationssprogs-semantik* (RSS) og *denotationssemantik* (DS).

RSS opfatter semantik som studiet af relationen mellem to udtryk. Denne opfattelse kan have sin gode grund i semantisk praksis som kendes fra ordbøger, definitioner osv. Det, *A* betyder, fremtræder altid ikklædt en sproglig form. Der kan argumenteres herfor ved at sige, at man slet ikke kan have noget begreb om betydning uden at kende et sprog, og at det fundamentale spørgsmål »Hvad betyder *A*« ikke kan stilles uden valg af det sprog, hvori svaret skal gives. Spørgsmålet skaber et sproghierarki, hvor *udtrykket A* er *objektsprog* i forhold til det *metasproglige betyder* og det *metasproglige B*, der skal bruges til at vise betydningen af *A* eller at oversætte *A*.

Betydningsbegrebet kan derfor udmøntes i et universelt semantisk *repræsentationssprog*. Dette sprog vil bestå af en mængde semantiske »atomer«, minimale betydningsenheder eller *semer*, der kombineres til komplekse meninger efter bestemte regler. Et universelt semantisk repræsentationssprog vil indeholde en semantisk teori, og det vil være muligt at benytte den til at forklare fx logisk inferens, matematisk kalkyle osv. Den vil også kunne udnytte af andre videnskaber, fx psykologi og måske hjerneforskning. Semerne kan måske bringes til at korrespondere med begreber, der beror på sansning og erfaring, lagres i hukommelsen, kombineres via associationer osv.

DS-tilhængere vil sige, at betydning må være en relation mellem et *udtryk* og noget *ikke-sprogligt*. Betydningsbegrebet må forklare, hvorledes *sprog* og *virkelighed* hænger sammen. Det, vi taler om, er ikke ord eller begreber, men *entiteter*, »værende ting«, hvis eksistens ikke er rent sproglig eller rent mental. Semantikkens opgave er at gøre rede for, hvorledes man kommer fra udtrykket *A* til det, *A* **denoterer**. Den elementære denotationsrelation

kan opfattes som en navngivning eller etikettering af entiteter: »Fido« - **Fido**-relationen har man malende kaldt den: »Fido« **står for** hunden **Fido**. Semantikken skal på basis af denne relation bygge en teori, der operationelt gør rede for, hvorledes man finder denotater til alle udtryk i et sprog, uanset hvor komplekse de er i deres opbygning.

Lingvister vælger ofte et RSS udgangspunkt, logikere et DS. De to synspunkter anses i reglen for at udgøre et *enten-eller*.

2. Repræsentationsprogs-semantik i lingvistikken

2.1. Transformationsgrammatik

Moderne lingvistisk teori har sit udspring i et forskningsparadigme som uløseligt er knyttet til den amerikanske lingvist Noam Chomskys navn¹. Chomskys oprindelige teori betegnes ofte *Transformationsgrammatik* efter en særlig type regler, kaldet *transformationsregler*, som ifølge Chomskys teori spiller en særlig rolle i naturlige sprog.

For Chomsky er det lingvistikken opgave at studere de almene karakteristika ved den menneskelige sprogeevne. Denne evne skildres som en automat, der kan opbygge og håndtere en kompleks algoritme. Lingvistikken kan derfor anvende matematiske modeller, der først og fremmest hentes fra teorien om *formelle sprog*.

Chomskys teori er en matematisk sprogteori med udgangspunkt i velkendte lingvistiske, empiriske data. Den matematiske tilgang har åbnet vejen for nye forskningsområder, både internt i lingvistikken og i relation til andre videnskaber: især matematik, logik, informationsteori og kognitionsforskning.

2.2. Teorien om formelle sprog

Et *formelt sprog* er en velafgrænset, endelig eller uendelig mængde af udtryk, der er opbygget i overensstemmelse med reglerne i en grammatik. Fx er mængden af formler i en matematisk teori eller mængden af udtryk i et programmeringssprog som *Pascal* eller *Prolog* formelle sprog.

I teorien om formelle sprog er en *grammatik* en matematisk struktur, G :

$$(1) G = \langle V_t, V_n, P, [\Sigma] \rangle$$

hvor V_t er det terminale vokabular,
 V_n er det non-terminale vokabular
 P en mængde af produktionsregler og
 Σ , aksiomet, et af elementerne i V_n .

V_t er mængden af de symboler, hvoraf udtrykkene i det sprog, L , som grammatikken G definerer, opbygges. Er sproget Pascal, er V_t altså mængden af

reserverede ord, symboler og *identifiers* som opregnes i Pascal brugermanualen. Er sproget dansk, er V_t mængden af danske ord, dvs. en empirisk given mængde, som ikke findes præcist afgrænset noget sted.

V_n er mængden af symboler for grammatiske kategorier, som indgår i den syntaktiske beskrivelse af sprogets udtryk. Er sproget Pascal, er V_n symboler som $\langle program \rangle$, $\langle block \rangle$, $\langle identifier \rangle$, $\langle record \rangle$. De indgår i syntaksbeskrivelsen og bruges af compileren til at afgøre, om et givet program (et udtryk af kategorien $\langle program \rangle$) er opbygget i overensstemmelse med syntaksreglerne. Er sproget dansk, udgøres V_n af de kategorier, som grammatikere har opfundet til empiriske studier af naturlige sprog, fx *præposition*, *verbum*, *verballed*, *nomen*, *nominalsyntagme*.

P, mængden af produktioner, er regler af formen

- (2) a. $\varphi \rightarrow \psi$ eller
b. $\varphi ::= \psi$ (EBNF-notation²)

hvor og er strenge af terminale og non-terminale symboler, og “ \rightarrow ” (eller “ $::=$ ”), der læses “genskrives som”. Reglen tillader at erstatte udtrykket til venstre med udtrykket til højre, overalt hvor det forekommer.

Drejer det sig om Pascal, er den første produktion i syntaksen således:

- (3) $\langle program \rangle ::= \langle program\text{-}heading \rangle \langle block \rangle .$

hvilket betyder, at et program skal bestå af én del af kategorien $\langle program\text{-}heading \rangle$, én del af kategorien $\langle block \rangle$ og afsluttes med terminalsymbolet “.” (punktum).

Tilsvarende kunne en regel for sætningsbygningen i dansk være:

- (4) $S \rightarrow NP VP$

hvorved tilkendegives, at en streng af kategorien S, fx

- (5) Peter læser en bog

har en struktur med et NP *Peter* og et VP *læser en bog*. Kategorinavnene S, NP, VP er valgt sådan, at de så vidt muligt svarer til velkendte grammatiske kategorier fra lingvistikens traditionelle begrebsapparat: *sentence*, *noun-phrase*, *verb-phrase* (dansk: *sætning*, *nominal- eller navnegruppe*, *verbalgruppe*).

Σ er navnet på den kategori, alle acceptable udtryk skal kunne henføres til. Den kaldes ofte *axiomet*, idet deriveringen af et udtryk i det sprog, grammatikken specificerer, kan opfattes som en slags formelt bevis, startende med grundsymbolet/axiomet som præmis.

producere (eller acceptere) en uendelig mængde af udtryk, der kan henføres til aksiomet, og udstyre hvert udtryk med en niveaudelt syntaktisk struktur, der fortæller, hvilke underkategorier det analyseres i, og hvilke relationer der er imellem underkategorierne på hvert niveau.

Implementeret på datamat kaldes en sådan algoritme en *parser*. En parser bruges i datamatikken til at kontrollere, om et program er korrekt ud fra den formelle grammatik for et programmeringssprog. Parsere bruges af lingvisterne til at afprøve hypoteser om syntaktisk struktur.

2.3. Chomskys sprogmodel

Ifølge Chomsky skal lingvistikken forklare, hvorledes mennesker internaliserer en mental grammatik til naturlige sprog, og hvorledes denne grammatik er opbygget.

Chomsky mener, grammatikken består af forskellige komponenter. **Basis-komponenten** er en formel grammatik. Den består af et **leksikon**, svarende til V_1 , der indeholder sprogets ordforråd, og en **kategorial komponent**, der rummer den øvrige grammatik.

Basiskomponenten producerer til hver sætning en syntaktisk repræsentation, *dybdestrukturen*. Den er i princippet universel, ens for samme sætning i alle sprog.

På disse strukturer arbejder andre komponenter, nemlig en **transformationskomponent**, hvis transformationsregler fx skal sørge for, at ordenes placering i de færdige *overfladesætninger* er rigtig for det enkelte sprog, jf. "Der Professor tauchte in einen Redestrom hinein und kam am anderen Ufer mit einem Hilfsverbum im Munde wieder hervor".

En **fonetisk fortolker** frembringer på basis af den transformerede struktur en fonetisk repræsentation af sætningen, som taleorganerne bruger til at ud-sige sætningen. Lignende forhold gælder ved perception af sætninger og for skriftsprog.

En **semantisk fortolker** producerer tilsvarende en semantisk repræsentation af sætningen. Den danner basis for forståelsen. Den kaldes sætningens *logiske form*. Man må formentlig forestille sig den som et sprog, der ligner prædikatslogikken.

Chomskys teori er en syntaktisk teori. Den syntaktiske komponent, der ofte kaldes autonom, genererer strukturer uden at udnytte anden information end den, der foreligger i de syntaktiske regler og kategorier. Fx. spiller hverken semantisk eller pragmatisk viden, dvs. viden om den kontekst, hvori sætningen ytres, nogen rolle.

Chomskys sprogmodel simulerer derfor sprogevnens som en evne til at op-bygge og udnytte en algoritme, der kan producere grammatisk korrekte udtryk og afgøre korrektheden af et udtryk i forhold til en internaliseret grammatik. Desuden forsyner den hver korrekt sætning med en struktur, der kan danne basis for en semantisk repræsentation.

2.4. Semantikens rolle

Selv om Chomskys teori bærer præg af formel præcision, enkelhed og elegance, selv om den på en tilfredsstillende måde tillader at integrere store mængder af empirisk, lingvistisk viden, især syntaktisk og morfo-syntaktisk, og selv om den viser vej til interessante sammenhænge mellem teoretisk og anvendt lingvistik, så opfylder den ikke de forventninger, som man må stille til en så ambitiøs lingvistisk formalisme: at levere en operationel lingvistik.

Det skyldes, at transformationsreglerne må slette, kopiere og permutere strukturer på vejen mellem dybdestruktur og overfladestruktur. Derved er det umuligt at gå tilbage: givet en overfladesætning at finde dens syntaktiske dybdestruktur. Men for Chomsky er semantikken en projektion af denne struktur, via den logiske form.

Hos Chomsky er syntaksen adskilt fra semantikken. De to følger lineært efter hinanden: semantikken arbejder på output fra den autonome syntaks. Mange forskningsresultater viser imidlertid, at syntaksen må have ikke bare semantisk, men også pragmatisk information til sin disposition.

Det er fx et problem for et syntaktisk analyseprogram, en parser, at mange sproglige udtryk ikke er entydige. Betragt en simpel og let forståelig sætning som:

(6) Jeg så en mand på en bakke med en kikkert.

En mand på en bakke med en kikkert kan syntaktisk analyseres på flere forskellige måder, fx som tre selvstændige substantivsyntagmer (navnegrupper) bundet sammen ved hjælp af *se* (= *se nogen et sted ved hjælp af noget*), men også som ét nominalsyntagme med en underordning (*mand, der står på en bakke, hvor der er en kikkert*). Imellem disse yderpunkter er der mange kombinationer. En rent syntaktisk parser vil tilordne sætningen seks-syv strukturbeskrivelser, som alle er lige gyldige.

Dette er imidlertid ikke tilfredsstillende, hverken i datamatisk eller kognitiv sammenhæng. For det første vil det give en *kombinatorisk eksplosion*. Kontrollen skal holde rede på seks-syv forskellige muligheder, der måske skal ganges op med et lignende antal flere gange. Processen bliver plads- og tidskrævende. Der kan blive overløbsproblemer. For det andet er det ikke sikkert, at det senere ud fra konteksten kan afgøres, hvilken af de syv muligheder, der skal foretrækkes. I så fald kan der ikke genereres én enkelt semantisk repræsentation af teksten, men kun seks-syv.

Dermed har teorien slået fejl som forsøg på at skabe en algoritme for semantisk repræsentation. Den har imidlertid også slået fejl som kognitiv model. Den svarer ikke til, hvad man ved empiriske undersøgelser kan vise om sprogbrug. Ved flertydige sætninger er en sprogbruger kun bevidst om én betydning. Dette ses fx af forsøg med såkaldte *garden path* sætninger, dvs. sætninger, hvor man ledes på syntaktisk vildspor, og først i sætningens slutning opdager det:

(7) the boat floated down the stream sank

Undersøgelser viser, at når sprogbrugeren ved *sank* opdager, at *floated* ikke var sætningens verbal, så oplever vedkommende forvirring. Der er ikke parallelt dannet den alternative syntaktiske struktur, hvori *floated* er et participium knyttet til *boat*. Man må begynde helt forfra på en analyse ad et andet spor.

En lingvistisk teori må kunne gøre rede for samspillet mellem syntaks og semantik i disse tilfælde. En god hypotese synes at være, at den syntaktiske analyse hele tiden styres af overordnede forventninger om betydningen. Denne hypotese bekræftes af andre forhold, fx ved de såkaldte *indexicaler*, ord hvis denotation er afhængig af ytringssituationen: pronominer og visse adverbier, bestemte og ubestemte nominalsyntagmer, relativsætninger m. fl. Reglerne for deres brug kan ikke formuleres kun syntaktisk. I analysen af text- eller diskursstrukturer er forholdet mellem semantisk forståelse (af meningsindholdet) og syntaktisk forståelse (af tekstens opbygning) dialektisk ligesom ved *garden-path* sætninger: man forstår den syntaktiske opbygning, fordi man forstår meningen, og man gætter meningen ud fra gæt om den fulde syntaktiske opbygning.

Lingvistik kan ikke blot være en teori om syntaktiske egenskaber ved naturlige sprog opfattet som mængder af enkeltstående sætninger. Uden at forkaste Chomskys grundlæggende idé om, at naturlige sprog kan studeres med brug af modeller og metoder fra matematik og logik, skal perspektivet udvides til også at omfatte semantiske modeller og metoder.

3. Denotationssemantik i lingvistikken

3.1. Formelle systemer: Syntaks og semantik

Teorien for formelle sprog er et resultat af en udvikling siden midten af 1800-tallet inden for formel logik og matematisk grundlagsforskning, hvis formål var at udvikle en syntaktisk eller algoritmisk bevisteori og axiomatisere hele matematikken inden for rammerne af ét *formelt system*.

Et formelt eller axiomatisk system defineres som en struktur FS:

(8) FS = $\langle V, L, A, D \rangle$

hvor V er et vokabular,
L, mængden af velformede formler, et formelt sprog over V,
A, aksiomerne, en delmængde af L,
D, deduktionsregler, en mængde af relationer i L.

Hvis man som aksiomer vælger formler, der kan fortolkes som repræsenterende sande udsagn om et eller andet domæne, og hvis deduktionsreglerne

er hensigtsmæssigt udformede, så kan man med et sådant system derivere de teoremer, der repræsenterer sande udsagn om domænet.

Grammatikken for L kaldes også FS's *formregler* og deduktionsreglerne FS's *transformationsregler*. Tilsammen udgør disse FS's *syntaks*. Man siger derfor, at et bevis for et teorem ved hjælp af FS's regler, sker med *syntaktiske midler* alene. FS er en algoritme til at finde teoremer.

Det domæne, som FS's formler repræsenterer, kaldes en *model* eller en *semantisk interpretation* for FS. Modellen siges også at udgøre en *semantik* for FS.

Et velkendt eksempel på et sådant teori-model forhold er Descartes' analytiske geometri. Geometriens grundelementer, punkter, repræsenteres som talpar, geometriske figurer udtrykkes som ligninger, og geometrisk deduktion sker med algebraens metoder. Geometrien interpreteres algebraisk. Algebraisk kalkyle eller syntaks erstatter geometrisk anskuelig tænkning. Ved løsninger af ligningerne behøver man end ikke vide, at de har en geometrisk interpretation. Alt bevisarbejdet sker i et sprog.

I hjerneforskningen beskrives tilfælde, der viser noget om model-teoriens kognitive status. En forsøgsperson får forevist en geometrisk figur sammensat af kvadrater. Forsøgspersonen skal - uden at klippe figuren ud - afgøre, om den kan foldes til en terning, og om to pile, der er tegnet i to af figurens felter i så fald vil pege mod hinanden vinkelret på en fælles kant. En canadisk matematiker med en alvorlig skade i *højre hjernehalvdel* fik forelagt problemet. Det lykkedes ham at løse det, selv om hans rumopfattelse og dermed hans evne til at *forestille sig* figuren anskueligt var borte.

Når man frågede honom hur han gick till väga svarade han: »Det finns ju alltid analytisk geometri«. Med andra ord: han löste problemet med hjälp av den språkliga översättning av rumslige förhållande som den analytiske geometrin erbjuder. På så sätt kunde han använda vänstra hjärnhalvan och dess språkliga tänkande. Det går alltså att lösa de här rymdproblemen språkligt. Men vi vanliga icke-matematiker gör det inte.³

En sådan modelteknik benyttede den tyske matematiker Hilbert til at bevise den euklidiske geometris relative modsigelsesfrihed. Da der til hvert eneste udsagn i geometrien teori svarer et udsagn i den algebra, som geometrien oversættes til, er den euklidiske geometri konsistent eller modsigelsesfri, hvis algebraen er det. Oversættelsen hverken tilføjer eller sletter information.

Hilbert udkastede på dette grundlag tanken om en *metamatematik* eller almen *bevisteori*. Hans idé var at udvikle et sprog for hele matematikken, sådan at man så bort fra begrebernes anskuelighed og indhold og alene opfattede dem som symboler, der manipuleres efter syntaksregler. Gyldigheden af ethvert bevis ville kunne kontrolleres rent mekanisk, omtrent som man kontrollerer, om reglerne er fulgt i et skakspil.

Med metamatematikken kunne man tilvejebringe et absolut modsigelsesfrihedsbevis for matematikken som formelt system⁴. På en kongres i 1900 formulerede Hilbert det som ét af de vigtigste matematiske problemer i det 20. århundrede at finde et sådant bevis. For at dette skulle kunne lade sig gøre, måtte man finde en model, der var tilstrækkelig omfattende til at repræsentere matematikkens syntaks, ligesom algebraen repræsenterede geometrien. Det kunne fx være aritmetikken.

I 1930 lykkedes det faktisk matematikeren Gödel at aritmetisere matematikkens syntaks ved et meget komplekst system af afbildninger af formler ind i tal og udsagn om formler ind i udsagn om tal. På denne måde kunne Gödel vise, at et matematisk system, der er omfattende nok til at udtrykke aritmetikken, ikke er fuldstændigt, dvs. at der i det findes sande udsagn, som ikke kan bevises med formelt syntaktiske midler alene. Desuden beviste han, at hvis man prøver at gøre sådanne systemer fuldstændige, ophører de med at være modsigelsesfri.

Dermed faldt Hilberts formalistiske program til jorden. Det var definitivt vist, at begreberne *bevislig* og *sand* ikke er ækvivalente, og at matematikken ikke kunne nøjes med at anvende rent syntaktiske metoder. Der kræves *noget*, som - på et andet plan - svarer til den geometriske anskuelse, men som aldrig kan oversættes til syntaks.

3.2. Modelteori

Hilberts program og Gödels resultat havde imidlertid en række positive virkninger på forskningen.

På den ene side udvikledes en rig teori om formelle sprog, algoritmer og matematiske automater (Church, Turing, Post, Kleene, von Neumann, Chomsky). Denne forskning er det væsentligste teoretiske grundlag for algoritmeteori og datamaten.

På den anden side udvikledes *modelteorien* (Tarski, Gödel, Robinson, Cohen), en logisk-semantisk teori, der har spillet en afgørende rolle for en række epokegørende matematiske opdagelser, fx Cohens teorem om kontinuum-hypotesen, valgaxiometts uafhængighed, og Robinsons non-standard analyse⁵. Modelteori har siden 70'erne haft en stilling som et af matematikkens mest lovende felter⁶.

Ved en modelteoretisk interpretation $\langle \mathbf{L}, \mathbf{M}, \mathbf{I} \rangle$ af et formelt system, $\text{FS} = \langle \mathbf{V}, \mathbf{L}, \mathbf{A}, \mathbf{D} \rangle$, får de velformede formler i FS , \mathbf{L} , tilskrevet såkaldt *ekstensional* eller *detonativ* mening i en mængdeteoretisk struktur, $\langle \mathbf{M}, \mathbf{I} \rangle$, hvor \mathbf{M} er en mængde af genstande og \mathbf{I} en afbildning, *Interpretationsfunktionen*: vokabularet enkelte symboler tages som navne på genstande i \mathbf{M} eller på relationer i \mathbf{M} ; værdien af symbolkomplekser beregnes på basis af de værdier, der er tillagt deres dele, og af de syntaktiske regler, der er anvendt til at danne dem. En formels ekstensionale mening er en *sandhedsværdi*.

Modelteorien og dens sandhedsbegreb bygger på Tarski. Han forudsætter eksistensen af en evne til at afgøre, om basale formler, hvori vi kender betyd-

ningen af de enkelte symboler og den syntaktiske opbygning, korresponderer med en model.

Denne evne til at etablere korrespondenser mellem objektsprog og virkelighed og udtrykke dem i metasproglig form konstituerer den basale sproglige forståelse. Uden den, ingen semantik. Givet den, er det muligt at beregne sandhedsværdien af sammensatte udsagn ved at benytte enkle sandhedsfunktionelle regler for logiske operatører, fx negation, disjunktion, implikation eller kvantorer.

Tarski og andre logikere og matematikere opererer med formelle modeller, i reglen sådan at der er tale om interpretation af en semantisk teoris formler ved en anden teoris entiteter, fx interpretation af geometrien i algebraen, i specielle tilfælde som Gödel-afbildninger endda interpretation af en matematisk teori, aritmetikken, ved teorien selv. Modellen er her et videnskabeligt artefact, som i en eller anden forstand antages at være en repræsentation af noget andet, fx en konkret eller ideal virkelighed, som matematiske teorier kan appliceres på.

3.3. Montaguegrammatik

Logikeren Richard Montague var den første, der anvendte modelteoretisk semantik i en lingvistisk teori. Han definerede sig i opposition til Chomskys syntaksorienterede semantik. Montague indleder en af sine kendteste artikler således:

I reject the contention that an important theoretical difference exists between formal and natural language . . . I regard the construction of a theory of truth . . . as the basic goal of serious syntax and semantics; and the developments emanating from the Massachusetts Institute of Technology⁷ offer little promise towards that end.⁸

Montagues opfattelse gør altså et sprog som engelsk til et formelt matematisk eller snarere metamatematisk studieobjekt på linie med logiske kalkyler. Montague afviser det rene studie af syntaks som vej til en fuldstændig sprogteori. Konstruktionen af syntaks og semantik må gå hånd i hånd efter det *Frege'ske kompositionsprincip*: semantikken tilskriver alle sproglige udtryk ikke-sproglige værdier i en model således:

- entiteter i den givne model tilskrives som værdier til de basale udtryk, fx ordene i et leksikon;
- den værdi, der tilskrives et sammensat udtryk, er en funktion af de værdier, der tilskrives dets dele, og af den syntaktiske operation, der bruges til at danne det;
- sandhedsværdien af en sætning kan bestemmes på basis af den værdi, der tilskrives den.

Et raffineret træk ved Montagues semantiske teori er hans løsning på det gamle problem om ekstensional og intensional betydning. Dette problem opstår bl.a. i forbindelse med temporale og modale operatorer, samt med de såkaldte *attitude*-udtryk, dvs. udtryk der angiver en holdning til sandhedsværdier, såsom *vide*, *tro*, *tvivle*, *ønske*, *forestille sig*. Frege opdagede, at det her undertiden er umuligt at substituere udtryk, der har samme denotation for hinanden. Fx. kan *Morgenstjernen* og *Aftenstjernen*, der begge denoterer planeten *Venus*, ikke substitueres i sætningen:

- (9) Romerne vidste ikke, at Morgenstjernen er den samme som Aftenstjernen

Det er heller ikke muligt at indsætte et denotatum, fx. *Ronald Reagan*, i stedet for udtrykket *den amerikanske præsident* i sætningen

- (10) Den amerikanske præsident dør ofte i embedet

Frege indførte derfor en distinktion mellem *Sinn*, intensional betydning, og *Bedeutung*, ekstensional betydning eller denotation. Han anvendte den samme distinktion til at forklare, hvorfor informationsindholdet i

- (11) a. $2+2 = 4$
b. $4 = 4$

er forskelligt, selvom $2+2$ og 4 denoterer det samme tal, og matematikken tillader substitution af ensbetydende størrelser. Forskellen ligger i, at udtrykket $2+2$ henviser til en anden *måde* at finde tallet på, end 4 gør. Den intensionale betydning rummer altså en anvisning på, hvorledes extensionen findes.

Montague ønsker at løse dette problem uden at operere med Freges to betydninger. I de fleste tilfælde kan vi ganske vist tilsyneladende behandle denotation som en ganske uproblematisk benævnelsesrelation («*Fido*«-**Fido**-relation) mellem sprog og verden: *Venus* er etikette for en bestemt planet, *planet* etikette for en bestemt mængde virkelige genstande, osv. Betydningen er her en funktion fra et udtryk i sproget til en ekstension i modellen (virkeligheden). Men disse enkle tilfælde er kun en slags *forkortede udgaver* af en i teorien mere kompleks relation, som erkendes under særlige forhold: betydning er i grunden en funktion fra et udtryk til en intension, der igen er en funktion fra mængden af mulige verdener til ekstensioner. Udtrykket *Morgenstjernen* denoterer altså ikke en bestemt planet, men en funktion, **intensionen**, der i enhver mulig verden finder en passende denotation til udtrykket. I de mulige verdener, hvor man ikke kender planeten *Venus*, har *Morgenstjernen* en anden denotation end *Aftenstjernen*. Betydning er en forskrift for, hvordan man finder **ekstensioner** i forskellige modeller eller mulige verdener.

Teorien om mulige verdener giver en elegant forklaring på forskellen mellem empirisk (kontingent) og logisk (nødvendig) sandhed: et empirisk sandt udsagn er sandt i den aktuelle verden. Et logisk sandt udsagn er et udsagn, der er sandt i alle mulige verdener.

Et andet elegant træk ved Montagues teori er hans begreb *oversættelse*.

Syntaksen er en algebra, hvor udtryk får tilskrevet entydige syntaktiske kategorier. Semantikken er tilsvarende en algebra, hvor der tilskrives semantiske værdier til syntaktiske udtryk, afhængigt af deres kategori. For ethvert sprog, logisk formelt eller naturligt, kan der gives en syntaks og en semantik. Semantikken er en modelteoretisk semantik, og når den er givet, siges sproget at være (semantisk) interpreteret.

Der findes imidlertid også en *oversættelsesalgebra*, der gør det muligt ved hjælp af rent syntaktiske operationer at overføre udtrykkene i et sprog (kildesproget) til udtrykkene i et andet sprog (målsproget). Hvis nu målsproget er semantisk interpreteret ved hjælp af en model, så *induceres* denne interpretation ind i kildesproget, når dette oversættes. Det er nu muligt at give ethvert naturligt sprog, fx engelsk, en semantisk interpretation via ét interpreteret logisk standardsprog, blot dette er rigt nok på udtryksmidler.

Det *intermediære sprog* er ikke at opfatte som et af de repræsentationsprog, man møder i RSS. Det er et selvstændigt sprog, der benyttes som en slags relæ mellem kildesproget og modellen. Montague konstruerede selv et særligt *intensionallogisk sprog* til en semantisk interpretation af engelsk. Men relæsproget er ikke nødvendigt for en interpretation af engelsk. Det er bare praktisk, fordi det med sin kompakte syntaks er velegnet som metasprog.

Det er her en vigtig pointe, at det intermediære sprog ikke er et »ideal-sprog«. Man kan kreere forskellige sprog til forskellige formål og modeller.

Herefter kan vi definere en modelteoretisk (Montague-)grammatik som et system, MG:

(12) $MG = \langle NL, FL, T, M, I \rangle$

hvor NL er et naturligt sprog med en veldefineret syntaktisk struktur,

FL et formelt logisk sprog med en veldefineret syntaktisk struktur, fx. prædikatslogikkens,

T en *transducer*, dvs. en oversættelsesautomat, der afbilder NL ind i FL,

M en model,

I en interpretation, dvs. en kompleks afbildning af FL ind i M.

M er hos Montague ikke kun en formel model, et artefact, men en virkelig eller mulig verden, som sproget FL interpreteres i.

I består af *tilskrivningsregler*, som fortæller, hvilke elementer fra M, der er værdier for de *basale udtryk* i FL, og *evalueringsregler*, der tillader at beregne

værdierne af *komplekse udtryk* i FL. Disse værdier kan være elementer i M, delmængder af n-tupler af elementer i M, eller værdierne *sand, falsk*, hvis udtrykket er et udsagn.

FL og T er modelteoretiske hjælpekonstruktioner.

Montagues semantik er blevet anerkendt, men ikke ukritisk assimileret. Montagues formelle syntaks har ikke slået an, men erstattes i reglen med andre syntaktiske teorier. Det er en fordel ved den modelteoretiske semantik, at den er åben for sådanne ændringer.

Der er også blevet arbejdet med at forene modelteorien med en mere tilfredsstillende pragmatisk teori. Som matematikere flest beskæftiger Montague sig kun med deklarativ sprogbrug og sandhed. Hvis man også betragter fx ordrer og spørgsmål, er den sandhedsteoretiske synsvinkel for snæver. Ordre og spørgsmål har ingen sandhedsværdi, men de har semantisk mening alligevel.

Montague tager som Chomsky heller ikke fat på tekstlingvistiske problemer, men betragter stort set et sprog som en samling af enkeltstående sætninger. Med disse bygger man den eneste form for tekst, matematikere er interesseret i, nemlig en følgerigtig bevisrække af sande udsagn.

Man har kritiseret begrebet *mulig verden*, fordi det rejser forskellige mængdeteoretiske og forståelsesmæssige problemer. De såkaldte *situationssemantikere* har derfor erstattet det med begrebet mulig situation, samtidig med at de i begrebet *situation* søger at indbefatte hele den pragmatiske kontekst.

Montaguegrammatikken kritiseres endelig for ikke i tilstrækkelig grad at integrere syntaksanalyse og semantisk interpretation. Ganske vist tilvejebringer Freges kompositionsprincip en tæt sammenhæng mellem syntaktisk struktur og semantisk værdi, men der er dog ikke tale om en dialektisk kobling mellem den semantiske interpretation og syntaksanalysen som i mentale eller i datamatisk simulerede sprogprocesser.

4. Semantik på datamat

4.1. Naturligt sprog og datamater

I de seneste år har datamatisk lingvistik flyttet interessen fra de traditionelle applikationsområder: tekstmanipulation (fx stavekontrol), lingvistiske hjælpeværktøjer (ordbogsdatabaser, syntaktiske analyseprogrammer), frembringelse af syntetisk tale og maskinoversættelse, til såkaldt *intelligente sprogapplikationer*, fx sprog- og tekstforståelse. Amerikanske datafirmaer og forskningsinstitutioner har skabt centre og laboratorier, hvor sprogforskere, logikere, matematikere og dataloger arbejder sammen om studiet af sprog og information⁹. Lignende aktiviteter ses i Japan, England, Tyskland og Frankrig.

Denne forskning har både praktisk og teoretisk interesse. Den praktiske

interesse ligger i håbet om at kunne benytte naturlige sprog ved kommunikation med datamater og i behandling af informationer. En fordel herved er den åbenhed, der karakteriserer naturlige sprog i forhold til formalismer. En ikonisk grænseflade opleves således ofte som en begrænsning på brugerens ønsker og initiativrigdom. Databaser i fri tekst, hvor man kan søge ud fra en forståelse af tekstens mening, vil være strukturerede formalismer overlegne; brugerens valgmuligheder er ikke bundet på forhånd. Håndtering af heuristiske data i fri sproglig form er også fristende.

Den teoretiske interesse ligner den interesse, man i al forskning kan have af at benytte datamatiske systemer og modeller i studiet af naturen. Integration af teoretiske antagelser i modeller, der kan implementeres, muliggør en ny form for eksperimenter ved hjælp af simulering.

4.2. Datamatisk sprogforståelse

4.2.1. Datamatiske sprogforståelsessystemer og deres implementering

Et datamatisk forståelsessystem kan opbygges modelteoretisk, dvs. som MG i:

$$(13) \text{ MG} = \langle \text{NL}, \text{FL}, \text{T}, \text{M}, \text{I} \rangle.$$

Et logikprogrammeringssprog som PROLOG er særdeles velegnet til udvikling af sådanne systemer.

PROLOGs formalisme ligger meget tæt op ad de prædikatslogiske formalismer, man plejer at anvende i semantisk repræsentation. Den har samtidig den fordel, at den umiddelbart kan interpreteres af PROLOG-systemet og udføres med dets datamanipulationsoperationer. Derfor kan man i PROLOG definere et semantisk repræsentationssprog, FL, der er tilstrækkeligt rigt på udtryksmidler til at specificere udførelsen af ethvert logikprogram.

I PROLOGs formalisme kan man implementere en *model*, M, som en *database*.

PROLOGs formalisme tillader også at skrive en *fortolker*, I, der interpreterer det prædikatslogiske semantiske repræsentationssprog, FL, i forhold til databasen, M.

Endelig findes der en enkel algoritme for omskrivning af en *formel grammatik* til PROLOG. Den foretager en automatisk syntaksanalyse af sætninger i naturligt sprog. Man kan sætte ekstra betingelser på grammatikkens produktioner, så der sammen med den syntaktiske analyse produceres en oversættelse til det semantiske repræsentationssprog, FL. Man kan derfor også bruge PROLOG til *transducere*, T.

En fordel ved logikprogrammering i alle systemmodulerne er den, at de enkelte moduler kan kalde hinanden som subrutiner. Således kan transduceren benytte fortolkeren til at undersøge databasens tilstand under omstændigheder, hvor der kræves »viden om verden« for at vælge mellem flere mu-

lige oversættelser. Omvendt kan transduceren lade passager henstå delvis uoversatte i den semantiske repræsentation, hvorefter fortolkeren kan aktivere transduceren, når den skal fortolke meningen.

En yderligere fordel ved at benytte et sprog som PROLOG er, at man løbende meget let kan ændre sit program. Sproget er derfor meget velegnet til eksperimentelle programmer, hvor ændrede hypoteser ustandselig skal indlægges i programmet.

4.2.2. Transduceren

Transduceren, T, er en automat, der afbilder sætninger i naturligt sprog på den logiske formalisme, der udgør den semantiske repræsentation. T har naturligvis den samme egenskab som andre automater: den specificerer en velafgrænset klasse af input.

Det er derfor en tilsnigelse at hævde, at et forståelsessystem forstår *naturligt sprog*. Systemet kan syntaktisk analysere og generere semantiske repræsentationer til en uendelig mængde af velformede formler tilhørende et formelt sprog, NL. En dygtig lingvist kan udnytte sin viden og erfaring til med stadigt stigende perfektion at bringe NL til at ligne naturens egne sprog, fx engelsk eller dansk. Men NL er selv i bedste fald stadig en formelt afgrænset mængde af input, ikke et naturligt sprog. Lingvistikken, der bygges på, er som andre formelle videnskaber og teorier kun i stand til at give et idealiseret billede af et udvalgt aspekt af verden.

Input til T behandles først af en scanner, der foretager leksikonopslag, og derved bl.a. kontrollerer, om de anvendte ord overhovedet genkendes af systemet. Genkendelse forudsætter, at der sker en morfologisk analyse af det enkelte ord.

Leksikon kan i praksis anskues som delt i to klasser. Den ene klasse består af *fuldord*, dvs. de ord, der har denotativ mening, fx. adjektiver, substantiver og verber. De kan i leksikon enten være defineret ved hjælp af andre leksikalske ord eller deres betydning kan være *grundet*, hvorved forstås, at deres denotative betydning i forhold til modellen er angivet, fx i form af konstanter i FL, der er navne (*»Fido«*-Fido-betegnelser) for entiteter i modellen M. Den anden klasse består af *syntaksord*, dvs. ord som artikler, præpositioner, pronominer, visse adverbier. Disse ord har ikke nogen fast betydning i forhold til M, men repræsenteres i leksikon ved navne på processer, der under oversættelsen tillader at kalkulere deres bidrag til betydningen.

Efter scanningen foretager T en syntaktisk analyse af input på basis af den indbyggede grammatiks regler. Hver regel specificerer, hvilke kategorier input er opbygget af, og hvilke relationer der er imellem kategorierne. T kan nu flytte rundt på kategorierne, således at deres rækkefølge passer med de regler, der gælder for målsproget, FL. Endvidere erstatter T inputtets ord med de repræsentationer, som de til syntaksanalysen knyttede semantiske regler tilskriver hver af de betydningsbærende kategorier.

T skal oversætte input til en *grundet repræsentation* i FL. At repræsentationen er *grundet*, betyder, at alle udtryk, der denoterer entiteter i modellen M, skal være oversat med konstanter.

4.2.3. Fortolkeren

Repræsentationssproget FL minder om prædikatslogikken. Grundelementer er da *termer*, der står for entiteter, *prædikater*, der står for egenskaber og relationer, og *funktorer*, der står for afbildninger. Endvidere findes der *logiske operatoren*.

Modellen, M, kan implementeres som en PROLOG-database. Entiteterne i M er da posterne i databasen med identifikationer, egenskaber og indbyrdes relationer. Hele databasen kan opfattes som et univers eller en verden, hvis tilstand ændres ved, at der slettes eller tilføjes poster i basen eller ændres ved disse. Til basen kan være knyttet et programmodul, der viser situationen, ligesom man i et skakprogram viser brikernes stilling på brættet løbende.

Fortolkeren, I, har til opgave at tilskrive værdier i M til konstanter i FL-oversættelserne af sætningerne i NL, samt at evaluere de således interpreterede sætninger i forhold til modellen. I princippet sker dette ved, at en termkonstant, fx »Fido« instantieres med entiteten **Fido** i M, hvis den findes. Tilsvarende skal en prædikatskonstant fx »run« instantieres med en mængde af entiteter i M, nemlig sandhedsmængden for **run**. Og endelig skal det evalueres, om *Fido* er element i *run* eller ikke, dvs. om inputsætningen *Fido runs* er sand eller ikke i M. I PROLOG-sammenhæng kan dette foregå ved, at den producerede oversættelse, *run (Fido)*, gives som argument til et meta-prædikat i PROLOG, *call*, der omdanner det til *goal* og eksekverer det, dvs. undersøger om det er sandt i databasen.

Imidlertid er PROLOGs indbyggede meta-prædikat *call* ikke godt nok til et avanceret sprogforståelsessystem. PROLOG er som logikprogrammeringssprog behæftet med samme mangler som modelteoretisk logik, nemlig kun at operere sandhedssemantisk. Normalt er naturligt sprog rettet mod andre pragmatiske funktioner, først og fremmest de handlingsorienterede: at afgive ordrer, at stille spørgsmål, at berette om ukendte sagforhold. Disse sproghandlinger kan repræsenteres af imperative performativer: **EXECUTE (command)**, **VERIFY (question,answer)**, **KNOW (fact)**. Disse og muligvis andre performativer udløser forskellige aktioner i forhold til databasen:

(14) bind Fido! – > bind (Fido) – > EXECUTE (bind(Fido)),

der fx kan medføre, at informationen *fri (Fido)* slettes og erstattes med informationen *bundet (Fido, hundehus)*, hvilket simulerer verdens ændring som følge af kommandoens udførelse.

(15) hvor er Fido? – > position (Fido,X) – >
VERIFY (position(Fido,X),X)

der fx kan søge efter informationen *position* (*Fido*, ?) og returnere svaret *bundet til hundehuset* på basis af den tidligere indskrevne information *bundet* (*Fido*, *hundehus*).

(16) *Fido har fundet et kødben*. – > fundet (Fido, kødben) – >
KNOW (fundet(Fido, kødben))

der fx medfører opdatering af databasen. Denne proces er kompliceret. Måske skal M udvides med en ny entitet *Fidos kødben*. Kødbenet skal have en identifikation og registreres under forskellige relationer i databasen, fx at det er en *genstand*, der kan *ejes af en hund*, osv.

En interpretation af en sætning i en model M er altså ikke bare en undersøgelse af en sandhedsværdi i forhold til en database, men en langt mere dynamisk manipulation med informationer i denne database.

4.3. Reference-problemer

Modelteoretisk semantik indeholder mange interessante hypoteser og rejser mange spændende problemer. Vi vil her se nærmere på et enkelt sæt af problemer i forbindelse med *reference*.

Kun i de færreste tilfælde benytter brugere af naturlige sprog navnet »Fido« til at referere til hunden **Fido**. Man bruger beskrivelser: *min onkels hund*, *det kære dyr*, anaforer: *han*, *den*, eller kombinationer heraf: *hunden*. Det almindeligste er endda, at vi slet ikke i vores begrebsverden har navne på entiteter, men kun kan beskrive dem: *sommerfuglen*, *der i søndags satte sig på moster Dagmars næse*. Undertiden svarer beskrivelser slet ikke til virkelige entiteter: *det første menneske på Mars*.

Dette rejser nogle komplicerede problemer med hensyn til, hvorledes transduceren, T, og fortolkeren, I, skal fungere, hvor megen sproglig viden, hvor megen logik og hvor megen encyklopædisk viden, de skal disponere over. Og hvorledes modellen, M, skal se ud, fx om den skal oprette en slags *personnummer* på hver eneste entitet, der tales om eller kan tales om: *sømkassen*, *de 37 søm i sømkassen*, *hovederne på sømmene i sømkassen*, *kanten på hovederne på sømmene i sømkassen*, osv. Af de 37 søm kan desuden udtages 2^{37} forskellige delmængder, der hver især også kan optræde som en entitet, osv. En del af disse problemer kan løses teoretisk og praktisk, en del kun teoretisk, mange indtil videre *ad hoc* eller slet ikke.

Vi vil her kigge på den sproglige side af referenceproblemerne. Transduceren, T, har som opgave at finde en *grundet* repræsentation af referentielle udtryk, dvs. erstatte

- (17) a. den hund, der har hængeører
b. hunden
c. den/han
d. det kære dyr

med *Fido* eller hvilken anden konstant, der identificerer den pågældende entitet.

Hvis referencen er af type a.: *beskrivelse med en relativsætning*, kan transduceren omforme relativsætningen (*den hund*), *der har hængeører* til et spørgsmål: *Hvilken hund har hængeører?*, oversætte spørgsmålet til FL og give det til I's fortolkerprædikat **VERIFY**. **VERIFY** vil returnere navnet som svar: *Fido*, eller svarene *flere hundelungen hund*. I det første tilfælde har T fundet referenten, i sidste tilfælde må T konstatere, at afsenderen har forudsat en (encyklopædisk) viden hos systemet, som ikke matcher med universet. Den sproglige viden består i dette tilfælde i at kunne finde, hvilke spørgsmål der skal stilles for at *forstå* beskrivelsen. Den encyklopædiske viden er den viden, der skal til at *besvare* spørgsmålene. Hvis disse processer er implementeret i systemet, vil det kunne reagere »intelligent« på beskrivelsen, fx ved at forklare: *Det er ikke klart, hvilken hund du siger til*.

Den krævede *sproglige viden* er langt mere omfattende, end det her kan antydes, bl. a. fordi der kan stilles 3 forskellige typer af spørgsmål, svarende til tre forskellige typer af relativsætninger. Hvilke typer, der skal vælges, kan igen afhænge af brugen af bestemt og ubestemt artikel osv. osv.

Hvis referencen er som i (17)b.: *substantiv med bestemt artikel*, må T undersøge om *hunden* refererer til en unik entitet i modellen M, altså om *Fido* er den eneste hund, systemet har viden om, eller den eneste hund, brugeren har viden om, eller om *hunden* har været omtalt i det foregående på en måde, der gør det tilladeligt at referere *anaforisk* til den. Dvs. at systemet, ud over at holde styr på det univers af entiteter, der udgør modellen, også skal holde styr på et begrænset, anaforisk univers af entiteter, som er relevante, fordi de har været omtalt på en sådan måde, at de opfylder betingelserne for *anaforisk genoptagelse*. Det komplicerer implementeringen af et sådant univers, at der også skal holdes regnskab med mængder, som om de var entiteter i modellen:

- (18) Der var engang tre prinser. Den første hed Skrat, den anden Skrat-skratterat og den tredje Skratskratteratskratskrirumskrat.

Her er der *fire* entiteter, nemlig foruden hver af de tre berømte prinser også selve mængden *de tre prinser*, selv om den ikke optræder som benævnt entitet i modellen, M.

I (17)c. bruges *bestemt pronomen*, *den* eller *han*. Her er informationsmængden lille. Der er derfor stor sandsynlighed for, at entiteten findes i det anaforiske univers. *den* eller *han* kan bruges som ikke-anaforisk identifikator, men det er et meget specielt stilistisk virkemiddel, som næppe umiddelbart har betydning for datamatiske forståelsessystemer.

(17)d. er et eksempel på den retoriske figur, *metonymi*, »brug af et navn for et andet«. Den altid kræver speciel ikke-sproglig viden, som i reglen er fælles for en klike eller mindre sproggruppe. Den har næppe heller stor be-

tydning. Systemet vil blot reagere som på en ikke forstået beskrivelse af typen (17)a.

En særlig type reference-problem optræder ved fantasi-verdener, dvs. når der tales en tænkt situation:

- (19) Hvis vi bare havde en hund. Så skulle den hedde Fido. Hver dag ville jeg gå tur med ham. Jeg ville kaste pinde, som han kunne løbe efter. Vore katte og hunden ville være venner og spise af samme tallerken.

Her oprettes et midlertidigt diskursunivers, hvori der anvendes anaforiske referencer, som om der fandtes en model og et anaforisk univers. Men de forsvinder igen, når der tales om noget andet. Et forståelsessystem må derfor, men også af andre grunde, have et fantasimodul, der kan oprette og nedlægge midlertidige universer. Det kan implementeres ved, at hele systemet kalder sig selv som subrutine. Dermed kan man også få midlertidige universer inden i midlertidige universer i vilkårlig dybde, som kinesiske æsker. Mange af den slags konstruktioner er velkendte fra studier i litterær komposition.

En særligt vanskelig reference opstår, når en sprogbruger henviser tilbage til tidligere diskurser eller tidligere stadier af den samme diskurs for at beskrive: *Den hund, Fido, jeg fantaserede om i går*. Derved opstår problemet om at få et sprogforståelsessystem til også at håndtere et resumé af forudgående diskurs(er) med en passende mængde af detaljer. Dette resumé udgør et *memoriseret diskursunivers*.

4.4. Reference-problemer og repræsentationsprogssemantik

Studiet af referenceproblemer viser, at de to semantiske grundopfattelser, der blev omtalt i begyndelsen, RSS og DS, ikke står i et eksklusionsforhold, men at begge er nødvendige for at realisere et avanceret datamatisk forståelsessystem.

DS er i det *modelteoretiske* forståelsessystem en basal del af hele systemarkitekturen. Men forholdet mellem **tale** og **omtalt virkelighed** (model) må nødvendigvis artikuleres som et afbildningsforhold mellem NL og M, hvor hverken FL, T eller I kan undværes.

Ganske vist så vi, at Montague, der er den egentlige »opfinder« af lingvistisk, modelteoretisk semantik, mente, at FL kun var et metodisk relæ, en slags værktøj, der gør det nemmere at definere (programmere) det komplekse afbildningsforhold mellem NL og modellen M. Imidlertid kan det vises, at en tilfredsstillende løsning af bare elementære referenceproblemer, forudsætter eksistensen af FL, der derved skifter status fra at være et rent metodisk kneb som hos Montague til at være et selvstændigt systemelement. Det er nemlig via FL, at de modelafledte, men nødvendige universer, *det anaforiske univers* og *det memoriserede diskursunivers*, konstitueres. Elementerne i det anaforiske univers er hverken NL-udtryk eller M-entiteter, men

FL-udtryk, der refererer til mængder af entiteter i M. Og det memoriserede diskursunivers består af propositioner, der resumerer NL-diskursen, dvs. ikke beskriver M-tilstande, men beskriver, hvad der er *forstået* ved NL-diskursen i relation til M-tilstandene. Disse propositioner kan kun udtrykkes i FL.

5. Semantik og kognition

Jeg har i denne artikel beskrevet to semantiske grundopfattelser, *RSS* og *DS*.

Jeg har vist, hvorledes moderne teoretisk lingvistik, der siden 50'ernes slutning har arbejdet på et formaliseret logisk-matematisk grundlag, har tolket de to opfattelser som modstående.

Jeg har endelig antydnet, hvorledes et datamatisk forståelsessystem, med udgangspunkt i en speciel udformning af den denotative semantik, nemlig *modelteorien*, ved løsningen af nogle af de komplicerede referenceproblemer tvinges til at kombinere de to opfattelser, dvs. at udnytte både semantisk *repræsentation* og semantisk *interpretation*. Dette er et eksempel på erkendelser, som man ved eksperimentelt at implementere lingvistiske teorier på datamat kan komme frem til. Dermed er ikke sagt, at man ikke kunne komme til de samme erkendelser ad anden vej.

Et interessant aspekt ved denne form for eksperimentelt-teoretisk arbejde med datamat er de overvejelser, man tvinges til at gøre sig, om ligheden mellem de i elektronik realiserede processer og det menneskelige kognitive apparat.

Det er selvfølgelig slående, at man i sit arbejde med et teknisk medium møder strukturer, som man fristes til at benævne som og dermed begrebsligt subsumere under kognitive størrelser og processer: *intelligent, forstå, samtale-partner, memorisere, referere*, osv. Men er fristelsen ikke en fælde?

På en måde ligner situationen den, man var i, da den formelle, algebraiske logik blev opfundet i sidste århundrede. Man *psykologiserede* logikken. Et af de første og berømteste værker inden for området var *George Boole: An Investigation of the Laws of Thought, 1854*, hvis titel illustrerer, hvorledes man opfattede logisk algebra: som en udforskning af tankens love. Dette århundredes forskning har afsløret illusionen heri. Logikken fremstår nu som et spil, som vi kan spejle sider af menneskelig adfærd i, et meget specielt og nyttigt spil, der tillader at kontrollere gyldigheden af visse inferenser. Men logik er ikke tanker, ikke en metode til kreativ slutning eller bevisførelse, så lidt som Matador er finansteknik eller skak militærstrategi og -taktik. Lingvistik, og specielt datamatisk lingvistik, tillader at opstille modeller og kalkyler for sprog og sproglig adfærd, men de *er* ikke sproglig adfærd, heller ikke selv om de automatiseres.

Eller *er* de?

Tjah. Udviklingen i vores teknologi afsætter sine spor i vore sprog. I mit 200 år gamle eksemplar af det Franske Akademi's Ordbog er *flyvning* define-

ret **biologisk**: med anvendelse af baskende vinger, og **ballistisk**: med anvendelse af bue (til pil) eller *krudt* (til raketter, kanonkugler og deslige). Disse definitioner lægger tilsyneladende sprogbrugen så fast som i syllogismen fra *Erasmus Montanus*:

- (20) En sten kan ikke flyve
 Morlille kan ikke flyve
 Ergo er morlille en sten

Men ifølge min moderne ordbog *kan* morlille faktisk flyve, og det uden at skulle optræde som kanondronning på markedspladser. Og *luftkaptajn Petersen kan flyve hende til Mallorca*, selv om Petersen ikke har vinger som en stork. Teknologien har ændret vores ordbog. I den moderne ordbog har *flyve* tre definitioner: *som en fugl, som et missil, som* eller **med en flyvemaskine**.

Morlille *kan* altså flyve. Men datamater kan endnu ikke forstå. Det er *ganske vist*.

NOTER:

1. Dette paradigme blev først fremlagt i værkerne *Syntactic Structures* (1957) og *Aspects of the Theory of Syntax* (1965). Chomsky har siden ændret sin teori. Detaljerne i disse ændringer ses der her bort fra.
2. EBNF står for *Extended Backus-Naur Form*, en særlig formalisme som anvendes i grammatikker for højere programmeringssprog.
3. A. Ellegård: *Språket och hjärnan*, 1982, p. 161. Se også P.J. Davis & R. Hersh: *The Mathematical Experience*, 1980, pp. 314-16.
4. Se B. Maegaard, H. Prebensen, C. Vikner: *Matematik og lingvistik*, 1975, Kap. II.
5. Se P.J. Davis & R. Hersh: *The Mathematical Experience*, 1980, pp. 223 ff.
6. Se J. Barwise & S. Feferman: *Model-Theoretic Logics*, 1985, Kap. I.
7. Hvor Chomsky arbejder.
8. *English as a Formal Language* i *Formal Philosophy*, 1974, p. 180, cf. *ibid.* p. 222.
9. Et af de mest ambitiøse og spændende er nok *Center for the Study of Language and Information, CSLI* i Stanford, etableret af Stanford, SRI og Xerox.

LITTERATUR

- BARWISE, J. & S. FEFERMAN (ed.): *Model-Theoretic Logics*, 1985.
 BARWISE, J. & J. PERRY: *Situations and Attitudes*, 1983.
 CHOMSKY, N.: *Syntactic Structures*, 1957.
 CHOMSKY, N.: *Aspects of the Theory of Syntax*, 1965.
 CHOMSKY, N.: *Lectures on Government and Binding*, 1981(82).
 DAVIS, P.J. & R. HERSH: *The Mathematical Experience*, 1980.
 DOWTY, D.R., R.E. WALL, S. PETERS: *Introduction to Montague Semantics*, 1981.
 ELLEGÅRD, A.: *Språket och hjärnan*, 1982.
 JACKENDOFF, R.S.: *Semantic Interpretation in Generative Grammar*, 1972.
 MAEGAARD, B., H. PREBENSEN, C. VIKNER: *Matematik og lingvistik*, 1975.
 MONTAGUE, R.: *Formal Philosophy*, ed. R. Thomason, 1974.
 PREBENSEN, H.: *BLOCKHEAD - misunderstanding Natural Language in The PC-Prolog Advanced User's Guide*, 1988.
 STERLING, L. & E. SHAPIRO: *The Art of Prolog*, 1986.
Sproget og datamaskinerne, 1988, udgivet af Modersmålselskabet.