

# humanistiske data



# SENTERETS RAPPORTSERIE

## Rapporter utgitt f.o.m. 1980

- RAPPORT nr. 13. *Datatjenester for og datasamarbeid mellom kunst- og kulturhistoriske museer*. Februar 1980. 2. opplag november 1981. ISBN 82-7283-010-8 Pris kr. 50.
- RAPPORT nr. 14. *NOVA\*STATUS HÅNDBOK*  
Del 1: Søking. Brukerveiledning. 3. opplag februar 1983. ISBN 82-7283-011-6 Pris kr. 20.  
Del 2: Fil-beskrivelser. Systemdokumentasjon. Utsolgt.  
Del 3: Generering og oppdatering av databaser. Utsolgt.
- RAPPORT nr. 15. *Ivar Fønnes: Tekstsøking på tegnnivå*. Januar 1980. ISBN 82-7283-012-4 Utsolgt.
- RAPPORT nr. 16. *Årsmelding 1979*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-013-2 Gratis.
- RAPPORT nr. 17. *Svein Lie: Automatisk syntaktisk analyse*. Del 1. Grammatikken. Desember 1980. ISBN 82-7283-014-0 Pris kr. 30.
- RAPPORT nr. 18. *Datateknologi og humanistisk forskning*. Bidrag til en NAVF-utredning. Desember 1980. ISBN 82-7283-015-9 Pris kr. 30.
- RAPPORT nr. 19. *Statistiske metoder på arkeologisk materiale*. Rapport fra et seminar på Bryggens museum, Bergen 24.-26. november 1980. Mars 1981. ISBN 82-7283-017-5 Pris kr. 35.
- RAPPORT nr. 20. *EDB-prosjekter i humanistiske fag 1980*. Juni 1981. 2. opplag oktober 1981. ISBN 82-7283-018-3 Pris kr. 45.
- RAPPORT nr. 21. *Rune Johansen: Bruk av EDB i teatervitenskapelig forskning*. Mai 1981. ISBN 82-7283-019-1 Pris kr. 35.
- RAPPORT nr. 22. *Årsmelding 1980*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-020-5 Gratis.
- RAPPORT nr. 23. *Stig Welinder: A program package for archaeological use*. 1981. ISBN 82-7283-021-3 Pris kr. 45.
- RAPPORT nr. 24. *Rapport fra seminar om bruk av edb innen teater og teatervitenskap*. Januar 1982. ISBN 82-7283-026-4 Pris kr. 50.
- RAPPORT nr. 25. *Ole Lauvskar: Diskriminantanalyse i SPSS*. Desember 1982. ISBN 82-7283-028-0 Pris kr. 55.
- RAPPORT nr. 26. *Stig Welinder: Paleodemography*. Oslo 1982. ISBN 82-7283-030-2 Pris kr. 55.
- RAPPORT nr. 27. *Årsmelding 1981*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-029-9 Gratis.
- RAPPORT nr. 28. *Årsmelding 1982*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7284-31-0. Utgått.



# humanistiske data 3-85

---

NAVFs EDB-senter for  
humanistisk forskning

---

The Norwegian Computing  
Centre for the Humanities

---

**NAVF** NORGES  
ALLMENNVIITENSKAPELIGE  
FORSKNINGSRÅD

NAVFs EDB-senter for humanistisk forskning ble opprettet av Norges allmennvitenskapelige forskningsråd i 1972. Senteret har som oppgave å arbeide på nasjonal basis for utbredelse av edb i forskningsarbeidet i de humanistiske fagene. Det er opprettet en samarbeidsavtale med Universitetet i Bergen som bl.a. gir Senteret adgang til edb-tjenester ved Universitetet.

Av sentrale oppgaver kan nevnes utvikling av programutrustning for humanistiske forskningsoppgaver, konsulenthjelp og informasjonstjenester.

Senteret utgir tidsskriftet *Humanistiske Data* (3 nr. pr. år) og en rapportserie (36 er utkommet pr. 12.12.85).

Senteret er sekretariat for International Computer Archive of Modern English (ICAME), og utgir bladet ICAME NEWS.

Senteret driver egne opplæringsprogram for vitenskapelig personale og medarbeidere i den kontor-tekniske gruppen innenfor de humanistiske fag. Det blir også holdt forskjellige kurs og seminar om edb og humanistisk forskning. Tidspunkt og emner blir kunngjort i *Humanistiske Data* og på institusjonene.

Interesserte kan kostnadsfritt bestille årsmelding og *Humanistiske Data* (kr. 60,- for institusjoner).

---

*Humanistiske Data* blir utgitt av NAVFs EDB-senter for humanistisk forskning. Redaksjonsgruppe: Jostein H. Hauge, Kristin Natvig (red.), Espen Ore.

Senterets adresse: Harald Hårfagesgt. 31, Boks 53, 5014 Bergen-Universitetet. Tlf. (05) 212954/55/56

Artikler, rapporter, meldinger mottas. Redaksjonen avsluttet 12. desember.

---

*Humanistiske Data* is published by The Norwegian Computing Centre for the Humanities. Editorial group: Jostein H. Hauge, Kristin Natvig (ed.), Espen Ore.

The journal can be ordered from the address mentioned above. Contributions are welcome. On request the Centre can supply the addresses of contributors to the journal.

---

Medarbeidere fra Senteret i dette nummer:

Erik Fjornes, Jostein H. Hauge, Kristin Natvig, Espen Ore, Per Vestbøstad, Øystein Reigem.

---

*Fotosats i kommunikasjon med Univac 1100/82.*

*Sats: Universitetet i Bergen/NAVFs EDB-senter for humanistisk forskning.*

*Grafisk design og montasje: Kristin Natvig.*

*Trykk: John Grieg A/S*

*Forsidebildet:*

*«Adams fristelse.» Freske fra Maderuelo-eremitasjen i Prado, Madrid.*

*Forst når Adam og Eva har spist frukten fra Kunnskapens tre, er de blitt Mennesker.*

# Innhold

## Artikler

Humanistiske fag og kunnskapsbaserte systemer. <i>Erik Fjornes</i> .....	s. 4
Informasjonsteknologi i Norge – kunnskapsbaserte systemer. Intervju med Jens Erik Fenstad. <i>Erik Fjornes</i> .....	s. 11
How Computers are Changing the Character of Research in the Humanities. <i>Michael Benskin</i> .....	s. 16
Historie og edb – nok en gang! <i>Kåre Andersen</i> .....	s. 26
Kunstig tale på norsk. <i>Ivar Utne</i> .....	s. 38
Databehandling av termtilfanget frå Terminol-prosjektet. <i>Per-Bjørn Pedersen</i> .....	s. 50
Datamater/programmell/og programmer som del af interaksjonen i undervisningen. <i>Erik Meistrup</i> .....	s. 57

## Rapporter

Hva skjer innen humanistisk og samfunnsvitenskapelig forskning? <i>Jostein H. Hauge</i> .....	s. 63
Informasjonsteknologisk satsning ved Universitetet i Oslo. <i>Asbjørn Brændeland</i> .....	s. 65
Seminar om pedagogisk programvare. <i>Gunnar Thorvaldsen</i> .....	s. 67
NORD IoD. <i>Jostein H. Hauge</i> .....	s. 69
International Conference on Data Bases in the Humanities and Social Sciences. <i>Jostein H. Hauge</i> .....	s. 77
Seventh International Conference on Computers and the Humanities (ICCH). <i>Jostein H. Hauge</i> .....	s. 80
The Humanities Research Center Brigham Young University .....	s. 85
Edb og lokalhistorie. <i>Johannes Kvestad</i> .....	s. 88
Humaniora-ukene ved Universitetet i Bergen. <i>Kristin Natvig</i> .....	s. 91
Bibeltekstar og databehandling. <i>Per Vestbøstad</i> .....	s. 94
Nytt fra RHF/NAVF .....	s. 97

<b>Meldinger</b> .....	s. 99
------------------------	-------

<b>Summary</b> .....	s. 105
----------------------	--------



# Humanistiske fag og kunnskapsbaserte systemer

Erik Fjornes

Med kunnskapsbaserte systemer har humanistiske fag både fått et nytt *innsatsområde* og nye *verktøy*. På bakgrunn av dette faktum har NAVFs EDB-senter for humanistisk forskning igangsatt et utredningsarbeid samt tatt initiativet til et nordisk forskersymposium om emnet. Med den foreliggende artikkelen ønsker jeg å gi et riss av på den ene siden hvilke muligheter humanistiske fag har til å *delta i utviklingen* av kunnskapsbaserte systemer og på den andre siden hvilke muligheter humanistiske fag har til å *anvende* kunnskapsbaserte systemer.

For å få dette til må jeg for det første gi en beskrivelse av hva kunnskapsbaserte systemer er, og de problemer som er forbundet med utviklingen av slike systemer. For det andre må jeg si noen ord om humanioras egenart. Dette må innebære en beskrivelse av hvilke typer aktiviteter som står i sentrum innenfor de humanistiske fag. Til slutt vil jeg undersøke hvilke naturlige sammenknytningspunkter det kan være mellom humanistiske fag og kunnskapsbaserte systemer.

## Hva er kunnskapsbaserte systemer?

Kunnskapsbaserte systemer er et ungt, men raskt voksende fagfelt med stort anvendelsespotensiale. Et kunnskapsbasert system er et datasystem som utfører oppgaver som til vanlig krever beherskelsen av menneskelige kunnskaper, f.eks. å kommunisere ved hjelp av naturlige språk, foreta medisinske diagnoser, eller på bakgrunn av sanseinntrykk utføre målrettede handlinger i et fysisk rom. (Fordi det å beherske menneskelig kunnskap er betinget av å ha intelligens, blir fagfeltet ofte kalt anvendt kunstig intelligens.)

## Historikk

Fagfeltet har en tverrfaglig opprinnelse. Det ble skapt for ca. 20 år siden ved at man kombinerte teknikker fra flere områder, bl.a. symbolsk programmering (programmering hvor instruksjonene først og fremst er tilpasset mennesker og ikke maskiner), og kognitiv psykologi. Etter ti års arbeide fremkom det tre primære underområder på feltet: simulering av kompetanse i naturlige språk, ekspertsystemer (systemer som etterligner ekspertise innenfor et fagområde) og robotisering (inkluderer syn, tale og bevegelse.)

## Simulering av kompetanse i naturlige språk

Systemer som simulerer kompetanse i naturlige språk som f.eks. språkforståelse, inneholder informasjon om et gitt språks ordforråd,

grammatikk, semantikk og eventuelt fonetikk og prosedyrer for analyse av tekst i overensstemmelse med språklig kunnskap. Slike systemer har bl.a. følgende anvendelsesområder:

- Kommunikasjon mellom menneske og maskin i spørsmål-svar systemer og ved oppbygging og bruk av ekspertsystemer
- Helt eller delvis automatisert oversettelse
- Systemer for avansert, evt. flerspråklig tekstbehandling
- Informasjonssøking i naturlig-språklige tekstmasser
- Språkpedagogiske hjelpemidler
- Språkkontrollerende hjelpemidler for skribenter

### **Ekspertsystemer**

Et ekspertsystem er et kunnskapsintensivt program som løser problemer som normalt krever menneskelig ekspertise. Det utøver også mange av ekspertens sekundære funksjoner slik som å stille relevante spørsmål og forklare resonnerer. Slike systemer kan:

- Gi mulighet for å resonnerer heuristisk ved å bruke hva eksperter anser som effektive tommelfingerregler
- Gi anledning til interaksjon med mennesker, noe som inkluderer bruk av naturlige språk
- Manipulere med og resonnerer ut fra symbolske beskrivelser
- Fungere med data som inneholder feil ved å bruke usikre domsregler
- Undersøke en mengde konkurrerende hypoteser samtidig
- Forklare hvorfor et spørsmål blir stilt
- Forsvare sine konklusjoner

Bruksområdene til ekspertsystemene er utallige. De viktigste er som følger:

- Være hjerne i en robot ved å slutte situasjonsbeskrivelser fra sanseintrykk
- Tjene som hjelpemiddel i beslutningsprosesser ved å slutte sannsynlige konsekvenser fra gitte situasjoner samt foreslå handlingsplaner
- Gi basis for fremstillingen av læremidler ved å fremlegge lærestoff til studenter samt diagnostisere, finne feil ved og rette på studentoppførsel
- Konfigurere objekter på bakgrunn av bestemte betingelser
- Kontrollere ved å fortolke, forutsi, rette på og føre tilsyn med systemoppførsel
- Gi grunnlag for utvikling av konsultasjonssystemer, f.eks. ved at en overlege på et sykehus legger inn sine spesialkunnskaper slik at assistentleger o.l. kan få tilgang til kunnskapen selv om overlegen ikke er til stede
- Tjene som forskningsinstrument ved at store datamengder blir gjennomgått ut fra en faglig synsvinkel

## Robotisering

En robot vil som oftest ha et ekspertsystem som hjerne og et naturlig språkssystem som kommunikasjonsmiddel. Men selve robotiseringen innebærer sanseapparat, fysiske bevegelsesmuligheter samt utstyr for syntetisk tale.

## Problemsituasjonen for utvikling av et kunnskapsbasert system

En rekke operasjoner må utføres for å utvikle et kunnskapsbasert system. De viktigste er som følger:

- beskrive kunnskapen som ønskes simulert
- analysere den potensielle brukers behov og kunnskapsnivå
- representere kunnskapen i en design for et kunnskapsbasert system
- velge utstyr for omgivelsene til det kunnskapsbaserte systemet
- benytte, tilpasse eller utvikle programvareteknikker
- programmere designen for det kunnskapsbaserte systemet
- teste og evaluere det ferdige system

Kunnskapsbeskrivelsen har en fundamental plass i denne sammenheng.

Formålet med en kunnskapsbeskrivelse er å frembringe best mulig innsikt i egenskapene ved kunnskapen. Men dette må gjøres på en slik måte at den ferdige kunnskapsbeskrivelsen kan omsettes til en systemdesign som er programmerbar.

Både design og programmeringsteknikker blir utviklet og benyttet på bakgrunn av de egenskaper som kunnskapsbeskrivelsen har tatt vare på. Derfor vil egenskaper som kunnskapsbeskrivelsen ikke innfanger heller ikke kunne bli simulert av det ferdige kunnskapsbaserte systemet. Således vil det være mengden av egenskaper som kunnskapsbeskrivelsen evner å innfange som innebærer noen av de viktigste mulighetene og begrensningene for å simulere den aktuelle kunnskapen.

På denne bakgrunn innebærer utviklingen av *mer effektive kunnskapsbeskrivende teknikker* et essensielt innsatsområde for å frembringe fremgang i fagfeltet kunnskapsbaserte systemer. Andre innsatsområder i feltet er:

- utvikle bedre programvareteknikker
- utvikle bedre designprosedyrer
- fremskaffe bedre kunstig intelligensbegreper
- fremskaffe større, raskere og billigere prosessorer og minner
- etablere arkitektur for selve datamaskinen som er bedre egnet til å være en del av et kunnskapsbasert system

Disse innsatsområdene har pekt seg ut på grunn av de barrierer dagens kunnskapsbaserte teknologi stadig møter. De vesentligste uløste





All lærdom begynner med kunnskaper i grammatikk, representert her som en fin dame. I venstre hånd holder hun nøkkelen til kunnskapens tempel. I tårnets nederste etasje finner vi Aristoteles med logikken, Cicero med retorikken og Boethius med matematikken.

Gregorius Reisch: Margarita philosophica, 1583.

problemene har sammenheng med:

- Et behov for fleksibel og generell naturlig språkforståelse. Dette behovet møtes spesielt der hvor brukere trenger å vise initiativ for å styre aktiviteten til et kunnskapsbasert system
- Et behov for å inkorporere kunnskap som er vanskelig å representere: Redskapene for simulering av kunnskap innenfor fagfeltet kunnskapsbaserte systemer har foreløpig vært avhengig av at kunnskapen kan innfanges i teoretiske formuleringer som f.eks. Hvis a og b så c. Men kunnskap som er etablert gjennom lengre tids erfaring innenfor et felt og således innebærer en fortrolighet med det aktuelle saksområdet, har vist seg vanskelig å representere i slike formuleringer. Dette gjelder f.eks. kunnskap for å identifisere en student som et godt lederemne
- Et behov for å kombinere og enhetliggjøre kunnskap fra en mengde eksperter når ingen tidligere standardisering har vært gjort
- Et behov for å benytte en stor mengde kunnskap raskt. Dette kan fremkomme i kommando- og kontrollsituasjoner hvor tidsmomentet spiller en kritisk rolle, f.eks. i en krisesituasjon på et oljeanlegg

### **Humanistiske fag og kunnskapsbaserte systemer**

Humanistiske fag har store muligheter til både å *anvende og delta i utviklingen av* kunnskapsbaserte systemer.

#### *Anvendelsesmuligheter*

Det er spesielt tre bruksområder hvor humanistiske fag kan ha stort utbytte av å benytte kunnskapsbaserte systemer:

- I lagringen av etablerte humanistiske kunnskaper
- I formidlingen av humanistiske kunnskaper
- Som forskningsinstrument

Noen eksempler på anvendelsesmulighetene innenfor disse områdene:

- Dersom en humanistisk forsker trenger å søke i store tekstmasser for å finne relevante informasjoner for en aktuell problemstilling, vil et kunnskapsbasert system kunne gjøre en del av arbeidet med å *fortolke* tekstmassene
- Som lærer. Et datamaskinassistert læringssystem som har et ekspertsystem som kjerne, vil både kunne presentere kunnskap, samt fungere som en erfaren veileder som diagnostiserer studentens forståelse for deretter å foreta de rette skritt for å høyne studentens kunnskapsmengde
- Testing av hypoteser. Et kunnskapsbasert system kan programmeres med en hypotese for deretter å undersøke i store datamengder om det er overensstemmelse mellom teori og virkelighet

### *Humanioras muligheter for å delta i utviklingsarbeidet*

Humanistiske fag har spesielle forutsetninger for å arbeide innenfor en del av innsatsområdene til fagfeltet kunnskapsbaserte systemer.

Humaniora er dypest sett et studium av menneskelige kulturfenomen-er. Dette har medført at humaniora har etablert forskjellige fremgangs-måter for å løse en del spesielle problemer som er forbundet med slike fenomener. De viktigste problemene er som følger:

- Etablere en enhetlig fortolkning av tilsynelatende ikke enhetlige kulturfenomen-er, f.eks. et århundres kunst og litteratur
- Danne og benytte begreper som i vesentlig grad er fundert i en fortrolighet med et saksområde og således ofte er vanskelig å innfange i teoretiske formuleringer. F.eks. er de fleste benyttede begreper innenfor kunsthistorien umulige å definere ved hjelp av nødvendige og tilstrekkelige kjennetegn. På tross av dette benytter en erfaren kunsthistoriker begrepene med stor sikkerhet
- Studere kulturfenomen-er hvor konteksten har betydning for fortolkningens innhold. F.eks. må en diskusjon om hvilken fortolkning av et ord som er den korrekte, hvor to eller flere er mulige, avgjøres med henvisning til den kontekst hvor ordet ble brukt

Dersom den spesielle kompetansen som humaniora har bygget opp gjennom arbeidet med disse problemene, skal komme utviklingen av kunnskapsbaserte systemer til nytte, må humaniora tilføres kunnskap fra i det minste to områder:

- Kunnskap om egne begrepsdannelser. Dette må tilføres fra filosofiske miljøer. Slik kunnskap er allerede i stor grad til stede i filosofiske miljøer, men på grunn av at humanistiske forskere skal benytte sine forskjellige begrepsdannelses-typer på nye områder og på en slik måte at produktet er programmerbart, er det nødvendig at filosofer frembringer innsikt om hva dette innebærer
- Kunnskap om kunnskapsbaserte systemer. Humanioras muligheter for å bruke og delta i utviklingen av slike redskaper vil øke i takt med kunnskapen om kunnskapsbaserte systemer. Det er således helt nødvendig at det settes inn ressurser for å formidle kunnskap om kunnskapsbaserte systemer til de humanistiske fag.

På lengre sikt vil en slik kompetanseoppbygging innenfor de humanistiske fag kunne føre til ny forståelse av problemene som er forbundet med utvikling av kunnskapsbaserte systemer. Dette gjelder spesielt på følgende områder:

- utvikle mer effektive kunnskapsbeskrivende teknikker. For det første kan oppbyggingen av kompetansen i å etablere enhetlige fortolkninger av tilsynelatende ikke enhetlige kulturfenomen-er komme til nytte i arbeidet med å kombinere og enhetliggjøre kunnskap fra en mengde eksperter når ingen tidligere standardise-ring har vært gjort. For det andre kan treningen med å danne og



benytte begreper som i vesentlig grad er fundert i en fortrolighet med et saksområde, komme til nytte under arbeidet med å inkorporere kunnskap som er vanskelig å representere i et kunnskapsbasert system. For det tredje kan kompetansen i å studere kulturfenomener hvor konteksten har betydning for fortolkningens innhold, bidra til frembringelsen av en programmerbar kunnskapsbeskrivelse som oppfanger flere egenskaper ved kunnskapen

- utvikle bedre designprosedyrer

Dersom humaniora evner å frembringe bedre kunnskapsbeskrivende metoder, vil humaniora også være svært velegnet til å frembringe bedre designprosedyrer

- frembringe bedre kunstig intelligensbegreper.

Ved at humaniora frembringer nye synsvinkler på kunnskap samt nye oppfatninger av designproblematikk, vil dette nødvendigvis også føre til ny forståelse av hva simulering av menneskelig kunnskap innebærer. Dette vil innebære muligheter for også å frembringe bedre begreper for kunstig intelligens.

# Informasjonsteknologi i Norge – kunnskapsbaserte systemer

Intervju med professor Jens Erik Fenstad

*Erik Fjornes*

## Bakgrunn

Regjeringens forskningsmelding utpeker informasjonsteknologi som et av de prioriterte områder hvor det er av nasjonal betydning å satse. Med informasjonsteknologi forstås teknikker som baserer seg på bruk av elektronikk og datamaskiner i forbindelse med innsamling, bearbeiding, lagring, distribusjon og presentasjon av informasjon. Informasjon kan her være alt fra målverdier og signaler til tall, tekst, tale, musikk, bilder og symboler. Fagområdet kunnskapsbaserte systemer er et av områdene innenfor informasjonsteknologien som er spesielt prioritert. Et kunnskapsbasert system er et datasystem som utfører oppgaver som til vanlig krever beherskelsen av menneskelige kunnskaper, f.eks. å kommunisere ved hjelp av naturlige språk, foreta medisinske diagnoser, eller på bakgrunn av sanseintrykk utføre målrettede handlinger i et fysisk rom.

Humanistiske Data har tatt kontakt med professor *Jens Erik Fenstad* ved Matematisk institutt i Oslo for å intervju ham om hans syn på informasjonsteknologi generelt og kunnskapsbaserte systemer spesielt.

Det var naturlig å intervju Fenstad av flere grunner: For det første har han en bred erfaring innenfor organiseringen og planleggingen av fagfeltet informasjonsteknologi i Norge bl.a. gjennom sitt verv som ordfører i Rådet for naturvitenskapelig forskning (RNF), Norges delegat i The International Council of Scientific Unions, medlem av den norske komiteen av The International Union of History and Philosophy of Science foruten en rekke andre verv. Dessuten har han en bred forskerpraksis bak seg, både i inn- og utland. Han har f.eks. hatt forskningsopphold i Oxford og Stanford. For tiden arbeider prof. Fenstad bl.a. med et prosjekt knyttet til simulering av naturlig språk.

Vi spurte først Fenstad om hans oppfatning av dagens situasjon innenfor fagfeltet informasjonsteknologi.

– Når det gjelder fagfeltet informasjonsteknologi, er det naturlig å skille mellom den planleggingen som foregår og viktige aktiviteter i feltet. Innenfor planleggingen er det viktig å være klar over det initiativet som industridepartementet tok ved å be Norges Teknisk-Naturvitenskapelige Forskningsråd (NTNF) om å lage et forslag til en

nasjonal plan for informasjonsteknologi (IT-planen). I dette arbeidet kom RNF tidlig inn og hadde et godt samarbeide med utredningsgruppen. Med RNF kom også resten av NAVF-systemet inn i planarbeidet. IT-planen har vært bearbeidet av et embetsmannsutvalg med representanter fra flere departementer. Denne bearbeidelsen er nå i ferd med å bli politisk klarert. Hva den endelige planen vil inneholde vet vi ikke, men vi regner med at den vil være klar i løpet av inneværende budsjettperiode.

Forslaget fra NTNF inneholder mange komponenter og store tall. Planen angår spesielt fire områder av feltet informasjonsteknologi: For det første kandidatutdanningen. For det andre den generelle opprustningen av utdanning og forskningsmiljø. For det tredje forskerutdanning og støtte til grunnforskningen. Og for det fjerde anvendt forskning og da med det siktemål å utvikle industri. Planen er generelt rettet mot å legge forholdene til rette for at Norge skal kunne delta i realiseringen av ideer som i dag er teknisk implementerbare og den inneholder derfor klare konklusjoner om at det må utdannes flere kandidater, opprettes større institusjoner samt fremskaffes mer utstyr.

Vi håper også at planen kan bli et velegnet instrument til å styrke grunnforskning og forskerutdanning. For det er nettopp disse aktivitetene som det er viktig å støtte i dag.

Det er en historisk forklaring på hvorfor det er et så stort behov for å ruste opp grunnforskningen og forskerutdanningen i Norge: Norge var i sin tid i samme situasjon som andre land ved at feltet eksploderte. Plutselig ble det en enorm etterspørsel etter informasjonsteknisk kompetanse slik at det søkte en stor mengde kandidater til feltet. Dette medførte et stort behov for lærerkrefter. Nyutdannede kandidater ble satt til å undervise uten på forhånd å ha fått en bred forskererfaring. Dermed fikk man ikke bygget opp et bredt og internasjonalt integrert forskermiljø i Norge, selv om det har vært gjort internasjonalt betydningsfulle prestasjoner f.eks. gjennom utviklingen av programmeringsspråket Simula. Men hvis Norge i dagens situasjon skal kunne hevde seg internasjonalt innenfor området, er det behov for større miljøer og bedre forskerutdannelse.

Kontakten med internasjonale forskermiljøer er en nødvendig betingelse for å kunne hevde seg i forskningsfronten. Tenk bare på hvilken betydning det har at en vitenskapelig ansatt i Norge har tatt sin Ph.D. ved et anerkjent amerikansk universitet? Han får kanskje som sine studiekamerater noen av de som blir toppene innenfor amerikansk vitenskap, noe som gir adgang til det internasjonale faglaug med sitt nettverk av forbindelser. Når det kommersielle forskermiljøet i Norge heller ikke er spesielt godt utbygget, blir aktivitetene for å utbygge grunnforskningen og forskerutdanningen svært viktige.

*Hva er de mest fundamentale barrierene innenfor jaggfeltet informasjonsteknologi generelt og kunnskapsbaserte systemer spesielt?*

Her kan vi ta den viktigste barrieren innenfor kunnskapsbaserte





*Professor Jens Erik Fenstad*

systemer som et typisk eksempel. For å få innblikk i denne er det nødvendig å være klar over skillet mellom anvendelse av kunnskapsteknologi og grunnforskning som retter seg mot utvikling av nye typer kunnskapsteknologi.

Anvendelsesorientert forskning forsøker å anvende kunnskapsteknologi på et mindre og vel avgrenset område f.eks. ved å få en maskin til å simulere utøvelsen av medisinsk diagnose. Og dette er et viktig og nyttig arbeide. Men disse anvendelser har sine klare begrensninger. Og ofte får man høre oppfatninger angående grensene for hva som er mulig å simulere av menneskelige kunnskaper som bygger på erfaringer fra konkrete anvendelsesprosjekter hvor dagens kunnskapsteknologi er kommet til kort. Men det er lite rimelig å tro at disse teoriene og modellene skulle være sluttsteinen på menneskelige kunnskaper slik at det skulle være umulig å utvikle nye og bedre. Derfor er nettopp utviklingen av nye teorier og modeller det viktigste innsatsområdet innenfor kunnskapsbaserte systemer.

For eksempel hefter det en rekke svakheter ved selve kunnskapsbegrepet som blir benyttet innenfor feltet. Her blir kunnskap alltid på et eller annet nivå forstått som det å følge regler. Men det er ikke sikkert at mennesket alltid følger regler.

En kritisk grunnforskning som tar dette og liknende problemer på alvor, kan lede frem til nye modeller for hva f.eks. kunnskap er. Ved at en deretter forsøker å implementere disse modellene, kan det utvikles kunnskapsteknologi som evner å simulere større deler av den menneskelige kunnskap.

Men det er likevel å gå for langt å tro at en slik grunnforskningsaktivitet vil kunne føre til teknologi som evner effektivt eller algoritmisk å simulere all menneskelig kunnskap. Selv om vi vil kunne skaffe oss bedre modeller, så tror jeg det går en grense et eller annet sted for hva som er mulig å implementere. En datamaskin er jo i bunn og grunn en regnemaskin slik at uansett hva grunnforskningen måtte medføre, så vil

ikke algoritmebegrepets sentrale rolle i beskrivelsen av hva en datamaskin er, bli forandret. Det som vil skje, er at noen av de utviklede modellene er algoritmisk programmerbare, mens andre ikke vil være det.

Men hvor grensen går, kan vi ikke vite i dag. Derfor kan ikke grunnforskningen bare være rettet mot det som er teknologisk implementerbart i dag. Isteden må grunnforskningen diskutere problemene på et fritt og fordomsfritt grunnlag. Og denne aktiviteten er ikke forbeholdt folk som er informasjonsteknologer. Den krever derimot deltakelse av en rekke aktører, som humanister, psykologer, lingvister, naturvitere osv.

*De humanistiske forskere får etter dette også en naturlig rolle i dette arbeidet. Men en slik rolle spiller ikke mange humanistiske forskere i dag. Hva må humanistiske forskere generelt gjøre for å komme sterkere med?*

For det første må de lære seg mer. Det gjelder spesielt bedre skoling i bruk av verktøyet datamaskinen og av modellbygging tilknyttet denne. Datamaskinen er et verdinøytralt redskap på linje med en skrivemaskin. Og det å lære å bruke en datamaskin burde alle humanister gjøre.

For det andre må de få en bredere internasjonal orientering slik at de kan beherske det beste innenfor internasjonal forskning.

For det tredje må lærerne i de humanistiske fag engasjere seg sterkere i en aktiv forskeropplæring og skape miljøer der mester, svenn og lærling aktivt arbeider sammen mot et felles mål.

*Men krever ikke gjennomføringen av alt dette mye penger?*

Jo kanskje for å gjennomføre en almen opprustning av infrastrukturen, men det er mange eksisterende ressurser innenfor de humanistiske fag i form av penger, lærerkrefter og forskerressurser som kan styres inn mot disse områdene.

*De arbeider jo innenfor den delen av fagområdet informasjonsteknologi som går under navnet «Naturlige språkssystemer». Kan De gi et bilde av dette fagfeltet?*

Faget er i startfasen, samtidig som pretensjonene er store. Man har som det optimale mål å utvikle systemer som kan fungere som fullverdige kommunikasjonspartnere. Og det er oppnådd mange interessante ting i form av en del fragmenter av et fullstendig, naturlig språkssystem. Men likevel er dagens systemer altfor enkle.

Vi har f.eks. systemer som fungerer nokså bra når det gjelder syntaks og semantikk, men da på et nøye avgrenset område hvor de utnytter det språkspesifikke innenfor disse områdene. Når slike systemer forsøkes utvidet til å bli generelle slik at systemet skal kunne kommunisere innenfor ikke så veldefinerte områder, så viser de eksisterende redskapene klare begrensninger. Og her fremkommer samme fundamentale problemstilling som i kunnskapsbaserte systemer generelt: Det er modellene for hva kommunikasjon er, samt for hvordan vi reagerer på informasjon, som medfører vanskelighetene.

*De er en sentral kraft i et prosjekt for utvikling av et nytt fundament for simulering av naturlig språk som for øvrig er dokumentert av Dem, Per Kristian Halvorsen, Tore Langholm og Johan van Benthem i arbeidet «Equations Schemata and Situations: A framework for linguistic semantics». Hvor vil De plassere dette prosjektet i det generelle bildet De gav av fagfeltet naturlige språkssystemer?*

Først og fremst prøver vi å arbeide på en slik måte at det vi gjør ikke er avhengig av det språkspesifikke ved en spesiell situasjonstype. Vi forsøker derimot å bygge fra første prinsipper, eller sagt på en annen måte, forsøker å fremskaffe basisegenskapene som et naturlig språkssystem må ha for at det skal kunne fungere generelt. Og mye av det vi har utviklet er implementert. Selvsagt er dette en begynnelse og det gjenstår å se hvor langt vi kan komme langs den veg vi har valgt.

Vi har tatt utgangspunkt i at det i en kommunikasjons-situasjon mellom mennesker ofte er selve situasjonen eller sammenhengen hvor ytringen blir gitt som sammen med den situasjon ytringen beskriver, avgrenser de relevante informasjonen fra de irrelevante slik at ytringen får mening. Meningen blir således en relasjon mellom ytringssituasjon og beskrevet situasjon. Dette er en grunnleggende idé i situasjonssemantikken som er et startpunkt for vår analyse.

For å klarlegge hvordan ytringen fungerer i en slik relasjon, har vi videreutviklet en av de ledende grammatikkteorier ved å lage noe vi kaller et situasjonsskjema. Dette skjema innebærer et redskap for å fange opp flest mulig informasjon som ligger i en språklig ytring. Dette skjema koder altså den informasjon som ligger i ytringens struktur og er således utgangspunktet for den videre bearbeidelse.

Vi har en aktiv forskergruppe ved instituttet. Teorien er blitt utvidet i flere retninger, bl.a. er vi interessert i å lage et oversettelsesverktøy som skal kunne analysere mer enn én setning av gangen.

# How Computers are Changing the Character of Research in the Humanities

*Michael Benskin*

*A revised version of a talk given as part of the programme for Humanioradagene in the University of Oslo, 20 March 1985.*

## I

It is stating the obvious to say that computing costs money, and that the mere existence of facilities creates its own clientele. A computing service may be managed by people with a genuinely altruistic concern to put their skills at the disposal of others, but it also has its own administrative and career structures, with their own rationale and dynamics of expansion. The scholar in the humanities is becoming alive to the possibilities that computing offers for his own work, but the providers of computing services have an independent interest in selling him their product. A new category of expenditure is established in the humanities budget.

Allocations of time and capacity on a university-owned machine need not commit a department to the transfer of real money, but they still have to be negotiated, and at some stage the university must balance its books. So the machine may be opened, at a commercial rate, to external users; and they, know it or not, become subsidisers of research in arcane subjects. The principle may extend to the financing of a data-preparation unit, though here the scholar can expect to pay real money for services rendered. Data-preparation is unglamorous, and not an attractor of funds, but it is often the greater part of a research project's expenditure: in the humanities, if only small quantities of data are to be investigated, there is probably no point in using the computer anyway.

The programming effort, by contrast, is often fairly slight: standard routines for sorting records, indexing and reverse indexing, producing concordances and collocation lists, and laying out text, account for a very large share of present usage. Complexities here are likely to stem from excessively bulky data, rather than from the routines themselves. Such programs, with the extensive statistical and cartographic packages developed for geography and the social sciences, are in many universities available «off the shelf» from a program library, to which local users also contribute software. Adaptation of existing programs may demand little expertise or time. New programming is likely to call for money, however, unless the scholar intends to acquire the necessary skills for himself. The prospect of such independence is undeniably

attractive, but the investment can be heavy – if, indeed, the aptitude is there at all. When programs are likely to be complex, proven expertise is worth its price: programs that are not completely reliable are a curse, and make mockery of work schedules. Since, in the humanities, programs are seldom other than a means to an end, the scholar's time may be better devoted to what he really knows about; as a second-rate programmer, he is apt to be expensive.

Outright purchase of user-friendly micro-computers, as their real prices fall and their capabilities improve, looks like becoming standard departmental practice; they need not be so very much more expensive than a first-class electric typewriter (a species whose days look to be numbered). The promise of «improved personal efficiency» has already made the word-processor standard office equipment; for the scholar who types with two fingers and a bottle of correcting fluid, and whose prose crawls off an ever-revising pen, the prospect is less of efficiency than of blessed release. Maybe Jane Austen did do it all on two inches of ivory; but she was not *obliged* to write anything, still less to cope with deadlines and footnotes and bibliography and revisions in the light of what some obscure Bulgarian wrote in 1902. The micro-computer is not, of course, just a sophisticated aid to composition of clean text, though the advantages of incremental correction from first draft to final copy or input for computer type-setting should not be ignored. (Notice how the submission of camera-ready copy has become part of the scholar's duty, so shifting production costs from publisher to author). It is now possible for an individual to have on his desk-top a real computer, not a toy, normally self-contained, but used as a terminal for access to a mainframe when need arises. There are powerful attractions: no queuing for a public terminal on a multi-access system, no congestion from other users, negligible risk of external interference or system failure, books and papers all within reach ... and this is no longer to dream of Utopia.

When compared with the budgets of science departments, the sums required are still very small, but they are not trivial. Expenditure on equipment and its maintenance, libraries apart, has until recently been incidental; the perceived cost of the humanities has been dominated by salaries. This is beginning to change. The humanities in any case have become much more research-oriented during the last quarter-century. (Consider how few of the older teachers in British universities have a Ph.D, and how few of the young ones do not.) Increasingly, funding is sought, and from outside the university – to pay for a sabbatical, perhaps, or to employ a research assistant. These expenditures, however, may well lead to increased demands on the university's computing services; and unless the research grant itself covers the cost, then regardless of any prestige attaching to the award, the university has secured a liability. Some grant-awarding foundations, as a matter of



principle, exclude computing costs from their subventions; some universities are forced to insist that research grants do not commit them to providing free computational support.

Fund-raising in the humanities is an increasingly competitive and skilled activity; the energies of a scholar can be absorbed to such an extent that research itself becomes a therapeutic diversion. Some research projects, inescapably long-term and eminently worthwhile, simply could not be undertaken without grants in aid, and far be it from me to decry such support, or to depreciate the efforts that may be needed to secure it. Yet it is hard to avoid thinking that the transformation of «Let's do some research» into «How do we get a grant?» has become largely automatic; a grant is all too often seen as essential for achieving anything at all. Computers play some part in this. It is easy to imagine that if a job is to be done properly (read «exhaustively») then a computer is necessary, using other than the best possible means is pointless, computing costs money, therefore ... Applying for grants, like computing, is an institution, where formal properties can weigh more heavily than content. It breeds its own excellence: a grant proposal may be of higher quality than any research that its author could accomplish. We may even come to believe, with the Streetsinger in *Die Dreigroschenoper*,

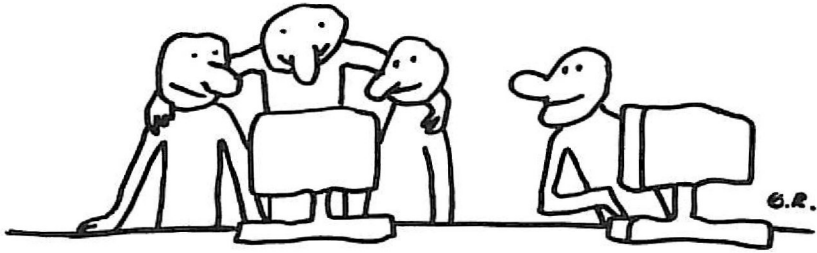
Ist das nötige Geld vorhanden  
Ist das Ende meistens gut.

## II

Of the social changes visited upon the scholar's life by virtue of his commitment to the computer, perhaps the most obvious has been the common need to work at night. During the day, mainframe multi-access systems are normally congested: file-space may be so restricted as to make work on a large corpus impossible, and the system's response to commands may be unacceptably slow. (How infuriating now is the five-second delay, to those who have known the instantaneous!) So, if computing is to be efficient, and in some cases if it is to be done at all, there may be nothing for it but to join the night-shift.

To look for a moment beyond the scholar's horizons, it has been well observed that computing activities have begun to change some of the social attitudes traditional in Britain: the night-shift is a much less reliable indicator of blue-collar working-class employment than it used to be, and it may even become respectable. For the scholar, admittedly, night-work may be nothing new; but the familiar nocturnal habitat has been the study at home, not the public machine-room. In the days of teletypes, it even sounded like a factory.

It is easy to forget just how lonely an occupation research in the humanities can be. Especially is this so for the graduate student: the British tradition has been largely a matter of locking him up in a library for three years, and expecting him to emerge with a thesis – not for him



the intellectual and moral support offered by collaborative research and a place in the laboratory. Involvement with computers changes this in various ways. Learning the elements of computing may account for the only course work he does, and in class he keeps company with the established scholar who has gone back to school. The range of their personal acquaintance is extended, as well as their intellects: they get to know the staff of what may be a very large organisation. The machine-room itself creates its own society. It is a meeting place for people who, though their particular research interests may differ widely, have yet a common interest in certain aspects of computing, and in using similar systems. Pooling of knowledge, and of data, is often informal, but there is organisational support as well: societies, some international, with journals and annual conferences; semi-public text archives, which offer access to contributed data in machine-readable form, and may, like the Oxford University Computing Service, offer special facilities to scholars from all over the country. These organisations, whatever else they do, are an increasingly important social cement.

The scholar's social role may also change, especially in a long-term project. It may become largely managerial, demanding skills not taught on a humanities curriculum or acquired in dealing with academic colleagues. If a data preparation unit is involved, he will have to cope with people whose outlook is very different from his own.

Collaboration with programmers involves some interesting social considerations, though the intellectual adjustment necessary is more likely to be noticed. Programming ability is nothing like so dependent on weight of learning and length of experience as is scholarship in the humanities. Some of the best programmers are very young indeed. The scholar may therefore find himself not only learning from, but deferring to, someone of half his age. A tight research budget may force him to rely on computer science students, part-time or during vacation; if he is lucky, they become colleagues rather than employees, and the dismal barrier between staff and student is transcended. Working with rolled-up sleeves at the same terminal does little in any case to preserve traditional assumptions about status and hierarchy; collaborations that

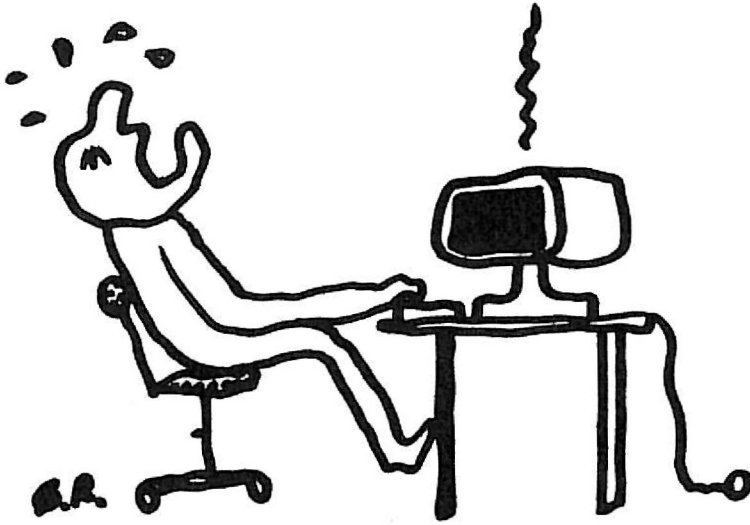
are nowadays commonplace might have been difficult to imagine twenty years ago.

### III

That the computer makes intellectual demands of its users will surprise no-one, though these demands, as will appear, may be more subtle and more pervasive than is generally realised. Moral demands, by contrast, are scarcely recognised, except in so far as they concern access to information that ought to be confidential, and the citizen's right of privacy. Yet it is the moral demands that are most likely to affect the newcomer to computing, especially if his intellectual training is in the humanities as opposed to mathematics or science. The simple-mindedness, and the remorseless logic of the machine, may conspire to persuade the novice that computing is nothing but an expensive way of teaching him that he is stupid. The old assumptions no longer serve. Numbers, for example, are now alpha-numeric strings. A sorting program produces a list like 1, 127, 139, 28, 45, 46, 5, 8, 92, not the numerically-ascending order desired, and so seems to have failed. The program is re-run, the result is the same, and it brings with it a sense of helpless despair. (Remember that the hour is late, and we are tired.) The remedy (we had to ask) turns out to be trivial – recode the data or reset the program – and just because it is trivial, the experience is so much the more demoralising. Were numbers and figures the focus of our attention, the mistake might never have been made; but here they are merely incidental, a little-understood tool, and the obvious is overlooked.

A generation accustomed to computers from its early schooldays may look on such experience with amazement; but today, probably most scholars in the humanities first encountered computers after they had established themselves intellectually: they and their peers had already grown used to the idea that they were beyond the common run of intelligence. For some, such revelation is a blow not lightly taken. From others, it calls forth patience and resilience, and perhaps a new kind of humility.

Other disruptions, unpredictable in their incidence, range from the dreary and demoralising to the amusingly bizarre. Some samples, none of them fiction: The communications network between our remote terminal and the central processing unit breaks down, for the third time in a week. (The engineers have been working on this fault for months, but the bug is elusive). – After a long session, during the final clearing up of interim versions and junk files, the laboriously-obtained final version is inadvertently destroyed. – A workman, digging up a neighbouring street, puts his pneumatic drill through the power cable. Service suspended until further notice. – The building is struck by lightning, and all the terminals announce their demise with a simultaneous pop.



The days of pencil and paper are a comforting vision of reliable progress and no lock-outs. When the system works, it is taken for granted; when it fails, it is reviled.

#### IV

By people who have never done any, research is often imagined as the engagement of great minds by the formidably difficult. In practice, it is very frequently a commitment to long hours of intellectually low-grade labour, which, even though it may verge on the stultifying for the scholar in question, cannot be delegated to anyone less competent. That is nothing new, but in so far as access to computers encourages the setting-up and investigation of large data-banks, such work is likely to account for an increasingly large share of the scholar's time. The incidences and correlations that the investigator seeks to determine are frequently not recognisable by strictly formal criteria. Without an adequate parsing program, the writing of which may well presuppose the knowledge that is in fact being sought, there may be no option but to code the relevant categories by hand. Similarly onerous is the conversion of traditional information-banks – card-indexes and the like – into machine-readable form; recasting into new record categories, rather than mere «flagging» for the key-to-disc typist, is here the likely course. It can take man-years of sheer drudgery, and exacting drudgery too: there is here no indulgence for the ill-formed or the inconsistent. To rely on the machine is to risk a remorseless exposure of incompetence.

So much for commitments to the menial. What of the computer's

impact on real intellectual activity? To judge by present pre-occupations and modes of thinking, it has been profound. Some of the effects are very much in the public eye. The mere possibility of such a machine has raised fundamental questions about the concept of mind, and so, arguably, about the essence of being human. The subject has claimed the attention of science-fiction writers no less than of philosophers; and new thinking in the humanities, research or otherwise, is not always confined within scholarly publications. New disciplines have emerged: on the borders of the humanities as traditionally conceived, stand the cognitive sciences and artificial intelligence. Investment in them is growing fast, as national prosperity is seen to depend increasingly on information technology. I suggested earlier, however, that there are more subtle and perhaps pervasive effects to be observed, and it is rather to these that I wish to draw your attention. New technology may of itself change almost unnoticed the way we think and speak.

Consider, for example, how the European awareness of time has changed, in cities and towns especially. Daily pre-occupation with the passage of time is a legacy from the late mediaeval development of clocks; before that, time passing was bells for church or curfew, and the natural clock of the heavens. Today, the digital clock is breaking up an inherited continuum: we are now made aware of each second as a thing in itself, a discrete unit. We report the time with gratuitous precision: «eleven fifty-four» replaces «nearly five to twelve». (Time past, note, not time to come any more.) Incidentally, our language changes: «ten-fifteen» and «ti-femten» displace «a quarter past ten» and «kvart over ti»; and what will become of expressions like «ti på halv tolv»?

Telling the time from traditional clocks is something that many of today's children will never learn; and they are robbed, for digital readings are explicit, and imagination plays no part. The idea of a continuum indefinitely divisible is no longer the teaching of an everyday object: «Can you see the hands moving?» (Quartz analogues fail: they move in jerks.) Gone, too, an experience of arcs, of angular equalities and proportions. Familiar images may be altogether lost: «Location at ten o'clock»? Clockwise?

In the humanities, modes of measurement are certainly changing, and not before time. As in science, the discovery of significant correlations is frequently an essential part of research, if not its object. By no means all such inquiry lends itself to quantitative treatment, but a good deal of it undoubtedly does. It is accordingly remarkable that statistical techniques have yet to be established as common currency. In part, no doubt, this reflects early specialisation at school, and patterns of recruitment into humanities courses at the universities; lack of aptitude in science or mathematics is often an encouragement rather than a disqualification. Among scholars in the humanities, there is often to be detected a sense that statistical methods perhaps ought to be invoked, but are regrettably beyond reach – a regret diminished, maybe, by a

suspicion that some cherished beliefs would not withstand their application.

Such a position ceases to be tenable once a computational approach is adopted. Initially, the computer may have been attractive just because there is a lot of counting to be done; it was seen, perhaps, as a labour-saver, or as a means of tackling a job too big for kitchen-table methods. Advice on suitable programs and procedures must be sought, and with that advice comes the statistical package. It is simply unthinkable the experts' programming advice should be accepted, but the associated statistical guidance jettisoned in favour of the naive comparisons of percentages accustomed in the humanities. Neither would it occur to the people in charge of program libraries and user support services not to consider the statistical implications of the intended research: their educational background is usually science or mathematics or computing itself. «But why do you want to measure  $x$  in the first place?», when all that was wanted was advice on which program to use, may sound like a threat to the scholar's supposed expertise in his own subject; but it is often the prelude to an education in the elements of logic, as well as statistics. So it is that  $\chi^2$  and Student's  $t$ -test are now and again to be found even in the literary journals. Statistically-based comparisons of style, intended to resolve cases of disputed authorship, were the first foothold, and they pre-date the computer; but such work is greatly facilitated by the machine, and even though its theoretical basis may be questionable, it is becoming more or less routine.

In so far as quantitative data enter into research in the humanities, it is obviously right that their treatment should conform to publicly-accepted standards. Numerically-based claims that would be disbarred if made in biology or the social sciences ought not to be admitted just because the subject happens to be palaeography or literature.

Statistical method is a rather obvious computer-aided intrusion into the humanities. Others are more insidious, and their effects are less easy to assess. So, for example, the idea of a «default value», the value that is to be assumed for a given variable unless some specified condition is fulfilled. («Read 1 year = 365 days, unless year-number exactly divisible by 4 and not ending in 00.» «Read live oysters = unsafe to eat, unless month-name contains letter  $r$ .») The idea of norm and aberration is not of course new, but in a world dominated by routines it is reinforced, and the definitions are sharpened. The conditions in which the default value does not obtain must be specified, and the rules for their recognition made explicit. Possibly diverse phenomena are brought together as a single category, «non-default»; repeated elements in their separate formal descriptions may point to an underlying unity. The epistemological bias shifts, from the atomistic to the unified: the unified explanation itself becomes a default.

A default value implies an algorithm, however simple. The algorithm



is not yet generally established in the intellectual apparatus of the humanities, but in computing it is fundamental. A computer program is itself an embodiment of the idea: a precisely ordered set of logical instructions that transforms, systematically, the information to which it is applied. The scholar who uses a computer, therefore, is in some degree forced to think in terms of transformations. For linguistics, this is nothing new: the idea of grammar as a generative system, comprising a base and transformational rules, was made explicit by Noam Chomsky's *Syntactic Structures* of 1957, and in the course of the last twenty-five years the subject itself has been transformed almost beyond recognition. The computer's world of discourse is limited to data and algorithms, and their products (new data, new algorithms); the grammatical model is likewise bound, for it is supposed to simulate the whole of the speaker's competence. In principle, the machine and the model are independent of one another, or could be; the logic in which they find their common ancestry does not entail the discovery of electricity. Historically, however, they have given each other powerful reinforcement. Moreover, linguistics has itself been an influential discipline – science, some would say – and its models have been imitated by other branches of the humanities.

Closely linked with the idea of the algorithm is the notion of the «well-defined». If a transformation is to be applied automatically, without the need for intervention and particular judgement, then the cases in which it is to apply cannot be left in doubt. Either it applies or it does not. The criteria for decision must therefore be enumerated, and they must be formal and unambiguous. In other words, well-defined. In the more discursive branches of the humanities, such rigour may be novel, and at first confined to explicitly computational work – tighten up the definitions of a few categories, perhaps, so that the machine can do some of the work after all. Yet it may become habitual, and instinctive: What parts of the subject are amenable to this mode of thinking? Are the other parts worth serious attention at all?

There are certain attractions in a discipline that assesses correlations statistically, that admits only decidable propositions, that reduces superficial variousness to the product of algorithms operating on underlying primitives. It invites application to all sorts of inquiry, and the products of its discourse are impressive, because they obey the canons of a rigorous logic, and are in their own terms impregnable. Which is to say that its application can be self-validating. It may also leave out what really matters: a mode of thinking dominated by structure and content is apt to find only those things in what it contemplates – abstractions, bereft of content.

People, the concrete reality – that glint  
of eye and animated gesture which revealed,  
so it seemed, the purposes of myth – seemed



to have been voided from the culture and replaced with a series of binary oppositions coupled to hypothetical history.

Thus the anthropologist Kenelm Burridge, writing of Australian reactions to Lévi-Strauss on myth and totemism (*Encountering Aborigines*, p. 196). The analysis of form and structure can be made routine, with or without computers, and routines of any sort encourage mindless application; even if content is not set aside deliberately, it can be easily forgotten. Yet in decrying a seemingly mechanical formalism, it can no less easily be forgotten that a more precise knowledge of form and structure conduces to a more accurate knowledge of content – this last an insight from the Greeks, as Burridge reminds us.

He has also observed how the binary oppositions of structuralist thought mesh with the logic and bi-polar switches of the computer, and displace the habitual triads of the Classical and Christian traditions. The duality is yet more pervasive. In living matter itself, the fundamental structures are revealed as diads: the DNA molecule is not triune, but a self-replicating double helix. And self-replication is an algorithm.

In response to the computer, research techniques change; so too do perceptions of the content of the humanities. The computer is itself becoming part of that content; but should anyone now object that it is incidental, an analytical engine and nothing more, suffice it to direct him to Douglas Hofstadter's eternal golden braid, Gödel, Escher, Bach. For I have already trespassed too far upon your time.

#### References

- Burridge, Kenelm. *Encountering Aborigines. A case study: anthropology and the Australian aboriginal*. New York, etc.: Pergamon Press, 1973.
- Chomsky, Noam. *Syntactic Structures*. Janua Linguarum IV. 's Gravenhage: Mouton, 1957.
- Hofstadter, Douglas R. *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books, 1979. (A metaphorical fugue on minds and machines in the spirit of Lewis Carroll.)

*Michael Benskin is Professor of Older English Language at the University of Oslo. He is co-author of «A computer-aided system for placing text-blocks on crowded maps», due to be published in ALLC Journal.*

# Historie og edb – nok en gang!

*Kåre Andersen*

## Tilbakeblikk

«Jacob Eriksen var født på Vettalseie i Ullensaker 15. mai 1821, som sønn av husmann Erik Israelson. Om faren vet vi at han ble kastet ut fra plassen Kjensdalen ved underrettsdom i 1840. Han flyttet da til plassen Jahnatuen under Ås, hvor han levde til sin død. Jacob bor hos faren som innerst så seint som i 1851. (...) I 1851 er Jacob Eriksen registrert i lensmannsrapporten som medlem av Borgen arbeiderforening i Ullensaker. (...) Han karakteriseres i lensmannsrapporten som «inderste, givt og tømmermand, men fattig».

Jacob hadde giftet seg første gang i 1845 med husmannsdatter Johanne Gulbrandsdatter, som imidlertid døde før 1850. I 1851 gifter han seg på nytt med en annen husmannsdatter, Marthe Nilsdatter fra Brennieie i Ullensaker, som i tida før 1865 føder ham minst 6 barn. (...)

I 1868 forteller fattigprotokollen at Jacob er reist fra sin familie til Amerika. Hans kone og 5 mindreårige barn er «blottet for levnedsmidler» gis fra nå av 2 Spd. månedlig i understøttelse, med ekstrastøtte til klær, sko og under sykdom.

I 1871 reiser imidlertid også Marthe Nilsdatter til Amerika sammen med to av sine sønner. Billettene er betalt i Amerika.» (S. Langholm; Historisk Tidsskr. 74 s. 254)

I sitatet over finner vi et utdrag av Jacob Eriksens liv og gjerninger. En slik rekonstruksjon av et livsløp er som regel svært arbeidskrevende. Opplysninger må f.eks. samles fra et stort antall forskjellige kilder:

«Opplysningen om farens utkastelse finnes i tingboka, om forliksstevninger og tvangsauksjoner i forlik- og auksjonsprotokoller, om husmannskontrakt i pantaregisteret, om fattigunderstøttelse i fattigprotokollene, osv.» (S. Langholm; HT 74. s. 256)

For lettere å kunne finne fram til samme person i flere ulike kilder, er det nødvendig med en utstrakt systematisering av tilgjengelig materiale. For eksempel vil det være behov for et fornavnsregister til hver kilde, men også andre oppslag må kunne gjøres via registre på: Fødested og fødselsår, patronymikon, slektsnavn/bosted, sivilstand/ektefelle osv.

Nevnte sitat er hentet fra prosjektet «Norsk samfunnsutvikling ca 1860-1900», ledet av prof. Sivert Langholm ved Historisk institutt, Universitetet i Oslo. Ett av målene var dengang (tidlig på 70-tallet) å utprøve bruken av edb i registrering, bearbeiding og analyse av historisk kildemateriale.

Som et resultat av prosjektet vokste det fram et lite miljø med interesse for «edb og historie». I tillegg til at historiske kilder ble tilrettelagt for maskinell behandling (f.eks. *Folketellingen 1875*, for Kristiania, med ca. 77.000 personer), ble det også utviklet en del programvare, bl.a. HISO-pakken (Ivær Fønnes: *Programpakken HISO. Kort presentasjon*. HF-data 1983, 2. utg.).

Dette var altså i Oslo. Omtrent samtidig ledet Jan Oldervoll ved Historisk institutt i Bergen et stort prosjekt for registrering av *Folketellingen 1801*, for hele landet (nærmere 1 mill. personer).

Senere har RHD (Registreringsentral for historiske data) i Tromsø kommet til, og er i dag den *institusjonen* som først og fremst driver masseregistrering av historiske kilder på data.

Situasjonen er altså den at stadig flere kilder blir tilrettelagt for datamaskinell behandling, men hva med utnyttelsen av materialet? Har utviklingen av analyseredskap stått i rimelig forhold til de ressurser som er satt inn på registreringssiden? Er det med andre ord fortsatt slik at vanlig bruk av materialet begrenser seg til manuell gjennomgang av maskinsorterte lister? (Evt. med muligheter for enkel statistisk analyse). I så fall har det ikke skjedd noe *kvalitativt* nytt på dette området siden pioner-prosjektene for snart 15 år siden. Jeg snakker da hele tiden om situasjonen for en *vanlig* bruker, f.eks. en lokalhistoriker eller hovedfagsstudent.

### Lenking av historiske data

Det vi mest ønsker oss er selvfølgelig større automatikk i f.eks. rekonstruksjonen av Jacob Eriksens livsløp. Med andre ord en mest mulig automatisk sammenstilling av opplysninger om personer og familier fra flere ulike kilder – «record linkage»!

Dette er et stort forskningsfelt internasjonalt, men hvor det også i Norge er levert viktige bidrag: Jan Oldervoll (Bergen) og Gunnar Thorvaldsen (Tromsø), begge med en historiskfaglig forankring – Lars Nygård og Mari Vestre med sine hovedoppgaver ved Institutt for informatikk, Universitetet i Oslo.

Feltet ligger – etter mitt syn – opplagt i «grenseland» mellom datafag/informatikk på den ene side, og bl.a. historie, navnegransking og lingvistikk på den humanistiske side. Begreper som «eks-pert-/informasjonssystem» og «kunstig intelligens» burde heller ikke falle unaturlige i denne sammenheng.

Det er planlagt et fellesnordisk prosjekt innenfor dette feltet, og man burde da «satse stort og tenke dristig», både hva *mål* og *ressurser* angår.

Sagt med andre ord: For å komme opp på et nytt *nivå* i utnyttelsen av maskinlesbart historisk kildemateriale, er det helt nødvendig med et *egget* prosjekt – ingen enkeltperson kan etter min mening drive en slik forskning «på si».

Humanistiske fag har i tillegg behov for å «teoretisere» sitt forhold til

informasjonsteknologien. Jeg tenker da ikke bare på *anvendt* edb, men også på den nye teknologien som *objekt* for humanistisk forskning.

Lenking av historiske data inneholder etter mitt syn slike perspektiver.

### **Hva gjør vi i mellomtiden**

Nå er det ikke bare «lenkingsprogrammet» som mangler for at den jevne bruker skal kunne nyttiggjøre seg allerede innskrevet materiale. Det fins meg bekjent f.eks. ingen rutiner for hvordan en bygdebokforfatter på Stord skal kunne «få» en diskett med folketellingsdata fra sitt distrikt, ei heller hvordan en lærer skal få tilgang til 1875-tellingen for Vålerenga – på diskett. (Muligens med unntak av spesialbestillinger fra RHD.)

Jeg antar at det vil være en del formelle hindringer for en slik frigivelse av materialet, men først og fremst er det vel et spørsmål om økonomi og hvem som skal administrere et slikt opplegg. Personlig tror jeg at en oppringbar «database» ville løst de fleste praktiske og administrative problem. Jeg tenker da ikke på en database med avanserte søkemuligheter, men rett og slett et enkelt filsystem med f.eks. sogn/prestegjeld som minste søkbare enhet. (Det største problemet vil antakelig bli avregningssystemet.)

### **Registrering av historiske data utenfor de tradisjonelle miljøene**

Etter hvert som mikromaskiner blir mer utbredt, f.eks. blant lokalhistorikere, i museer/arkiv og også i skolen, må vi regne med at stadig flere ønsker tilgang til registrert historisk kildemateriale. Det er vel også klart at mange vil gå i gang med registrering og bearbeiding av kilder på egen hånd. (Ved HF-data har vi tydelig merket en slik interesse.)

Køen ved Registreringssentralen i Tromsø vil antakelig bare vokse på denne siden av århundreskiftet, og man bør derfor støtte alle initiativ som fører til at nytt materiale blir registrert; men selvfølgelig uten å redusere på de historiskfaglige kravene til resultatet.

### **Hvordan komme i gang?**

Har man bestemt seg for å bruke edb i sitt arbeid med historiske data, så blir neste problem: fins programvaren? Svaret er stort sett «ja», men man skal være klar over at de fleste større f.eks. databasesystem krever en betydelig innsats før man behersker systemet. I tillegg er det vel også slik at kildenes beskaffenhet vil komme i konflikt med programmets krav til formalisering: standardisering av navn, store og små bokstaver, tekst, heltall og desimaltall – hvordan ordnes dette i de forskjellige fasene – f.eks. ved registrering, søking og sortering?

Undertegnede har i en del år arbeidet med bruk av edb i utnyttelsen av historiske kilder – spesielt skiftematerialet. Ett av resultatene av dette

arbeidet er programmet *RUBREG*.

## Programsystemet RUBREG

Alle som har drevet med historiske undersøkelser vet hvor mye arbeid en manuell systematisering av kildematerialet krever. På den annen side er selve det manuelle arbeidet med kilden viktig i forskningsprosessen. Denne «nærheten» til originalmaterialet går lett tapt når datamaskinen får sette *sine* rammer for behandlingen.

Programsystemet RUBREG er utviklet spesielt med tanke på data-maskinell behandling av historisk kildemateriale. Det er tatt utgangspunkt i en historikers «behov», det betyr at f.eks. «arbeidsstil» og «tenkemåte» i liten grad må tilpasses programmet. Hensikten er at brukeren på en enkel måte skal registrere data, foreta søking og utplukking, sortere og lage utskrifter i ønsket format.

Fordi RUBREG overlater til brukeren å sette premissene for data-behandlingen, vil f.eks. en historiker langt på vei analysere sine kilder uten å miste «fingertupp-følingen» med materialet.

RUBREG er *svært* enkelt i bruk: det krever ingen lesing av tykke manualer (hvor data-terminologien vanligvis dominerer) og behovet for opplæring er minimalt: Etter en liten times instruksjon behersker man hovedfunksjonene!

RUBREG kan brukes på de fleste mikromaskiner (skrevet i TURBO-Pascal; 64 kb RAM og en diskett-stasjon), og er i dag i bruk i «historiske» miljøer. I tillegg pågår det forsøk med bruk av programmet i samfunnsfagsundervisningen i skolen (f.eks. ved Ris og Vålerenga skole i Oslo).

På de etterfølgende sidene fins en enkel introduksjon til RUBREG, hvor det er lagt vekt på å presentere programdesignet framfor detaljer i ulike ordreoppsett. (Maskinen som er brukt i eksemplene er TIKI-100, bruk av f.eks. funksjonstaster refererer altså til TIKI-tastaturet.)

## Hovedmeny

Etter at RUBREG er lastet inn kommer hovedmenyen opp (se over), og man blir samtidig spurt om **DATABASE-navn** >. På RUBREG-disketten fins en liten testbase med navn **HIDRA**. Denne skal brukes i det følgende, og vi skriver derfor **HIDRA** som svar på spørsmålet om basenavn.

I vårt tilfelle blir nå denne linja skrevet ut:

```
RUBREG(HIDRA) >> Totalt antall poster: 58
```

Teksten til venstre er programmets promter (skrivemerket står like til høyre for de to hakene). Alle hovedfunksjonene iverksettes fra dette nivå. Trykk på f.eks. «?», vil gi hovedmeny.

RUBREG opererer alltid på et *utvalg* av poster. Utgangspunktet er selvfølgelig hele datasettet, men søker man f.eks. på «alle menn yngre



Det er nå viktig å merke seg at *alle* parameterne til de ulike ordrene velges fra skjemanivå.

### Registrering av data (M)

```
*****
* Ordre=M      Registrering/ending av post: 9      Utvid:PÅ      *
*
* Dødsdato: 560601      Sosial status: Søefarende Dræng_____ *
*
* Fornavn: Jens Michael_____ Etternavn: Jensen_____ *
*
* Kjønn: M      Alder: 22_____ Bosted: Fidsel_____ *
*
* Dødsårsak: Døde i Riga af Cholera_____ *
*
* F1:AutoKopi:AV      F2:Kopi:PÅ      F4:FORTSETT      F6:FERDIG      *
*****
```

Det er lagt stor vekt på å gjøre RUBREG til et kraftig verktøy ved registrering av data. Markøren flyttes mellom felt ved hjelp av piltastene, eller ved bruk av RETUR-tasten. Innenfor et felt kan man redigere tekst som i et vanlig tekstbehandlingssystem:

- \* innskriving av tekst
- \* <SLETT>: sletter tegnet over markøren
- \* <UTVID>: hvis aktiv vil man få melding om UTVID:PÅ. Ny tekst vil i så fall settes inn i den eksisterende. Ved UTVID:AV, overskrives gammel tekst.
- \* Trykk på «=» eller «blank» når markøren står i første posisjon i feltet, vil resultere i kopiering av foregående posts felt. Ved «blank» vil markøren i tillegg gå til neste felt. (Disse egenskapene kan «slås av» ved hjelp av «F2»).
- \* Trykk på «>» når markøren står i feltets første posisjon, kopierer innholdet fra forrige felt (samme post).

I tillegg til ovenstående kan brukeren også «merke» («F1» = AutoKopi) et vilkårlig antall felt for automatisk kopiering. Dette er særlig nyttig når man støter på sekvenser av poster hvor ett eller flere felt er like.

Registreringen avsluttes ved «F4» eller «F6». I første tilfelle, «F4», gjør RUBREG klar for ny registrering, mens «F6» gir hovedmeny.

### Søking (L)

Alle de viktigste kommandoene i RUBREG opererer fra det aktuelle skjemaet. I dette tilfelle har ordren «L» resultert i at et tomt skjema er kommet fram på skjermen. Brukeren har her valgt å plukke ut alle



```

*****
* Ordre=L F6:UTFØR Definerings av søkekriterier *
* *
* Dødsdato: 55____ Sosial status: _____ *
* *
* Fornavn: _____ Etternavn: _____ *
* *
* Kjønn: Alder: _____ Bosted: _____ *
* *
* Dødsårsak: _____ *
* *
* F11:+SUB F12:VERSALER+ F13:FRONT F14:SNITT F16:TEKST *
*****

```

personene som døde i 1855. Siden dødsdato er på formen «ååmmdd», er det tilstrekkelig å angi at dette feltet skal innledes med «55».

Nederst på skjermen vil det stå en STATUS-linje for hvert felt. Det vil si at linja endrer innhold alt etter i hvilket felt markøren plasseres. Status-linja over viser «default»-verdiene, men disse kan endres ved å trykke på angitt funksjonstast.

Her er alternativene:

F11: +SUB = angitt tegnsekvens må finnes i feltet  
 -SUB = angitt tegnsekvens må ikke finnes  
 «=» = likhet  
 «<>» = forskjellig fra  
 «<=» = mindre eller lik  
 «<» = mindre enn  
 «>=» = større eller lik  
 «>» = større enn

F12: VERSALER+ betyr: Gjør om til store bokstaver ved sammenlikning  
 VERSALER- betyr: Behandle teksten som den er

F13: FRONT = angitt tegnsekvens må innlede feltet  
 HELE = hvor som helst i feltet  
 FINAL = tegnsekvensen må avslutte feltet

F14: SNITT eller UNION (svarer til «OG»/«ELLER»)  
 F16: TEKST, HELTALL eller DESIMAL

(Kommentar til funksjonstastene: TIKI-100 har 6 funksjonstaster, som kan brukes både SHIFTED og UNSHIFTED. I RUBREG er SHIFT + F1 = F11, SHIFT + F5 = F15 osv.)

Igjen gjelder valgene hvert enkelt felt!!

Søking starter når bruker trykker «F6», og kan avbrytes med trykk på vilkårlig tast. Første funn skrives alltid ut på skjermen, og man må

svare «J» for å fortsette.

Legg merke til at etter at operasjonen er fullført, så er de utplukkede postene utgangspunktet for den videre behandlingen. Dette er det normale, ønsker man imidlertid å slå sammen resultatet av ulike søk, svarer man bekreftende på spørsmål om «summering av utvalg».

Ordren «V» fungerer nesten som «L», men her kan brukeren hente ut aktuelle poster fra flere filer, gjerne lagret på flere disketter. RUBREG spør om «Innfil» helt til brukeren svarer «blankt». Resultatet er en tekstfil med de utvalgte postene, denne kan igjen leses direkte inn i RUBREG med kommandoen «Import» (I).

## Utskrift

```
*****
* F6:UTFØR                Formatering av utskrift                *
*                                                                    *
* Dødsdato:  _____ Sosial status:  _____C_____ *
*                                                                    *
* Fornavn:  _____A_____ Etternavn:  _____B_____ *
*                                                                    *
* Kjønn:    Alder:  ____D__  Bosted:  _____ *
*                                                                    *
* Dødsårsøk:  _____ *
*                                                                    *
* F12:VERSALER- F15:V_ JUST F16:DESIMAL +/-:L=0 '/':S=θ *
*****
```

Etter at vi i forrige eksempel fant fram til de personene som døde i 1855 (i alt 37), skal vi nå skrive disse ut. (Først har vi imidlertid plukket ut alle mellom 25 og 55 år.)

Utskriftsformatet er nærmere spesifisert i skjemaet over.

Vi har valgt å skrive ut feltene i denne rekkefølge: «fornavn», «etternavn», «sosial status» og «alder». Bokstavene angir både *rekkefølge* og *feltbredde*.

Også i denne modus er det mulig å sette parametere. Statuslinja over gjelder alders-feltet, men igjen er det flere valg for hver opplysning:

F12:	VERSALER-	VERSALER+	: Omgjøre til store bokstaver?
F15:	V_	JUST	= Venstrejustert utskrift av feltet
	H_	JUST	= Høyrejustert
	PAK	KET	= Alle etterstilte «blanke» blir fjernet (se under)
F16:	TEKST		= «default»
	HELT	TALL	= RUBREG forsøker å behandle innholdet som et heltall
	DES	IMAL	= desimaltall

+ /-:L = 0

«L» står for linjeskift. Her kan brukeren legge inn et vilkårlig antall linjeskift etter (hvert) felt! («+» øker med 1 for hvert trykk - «-» trekker fra 1 (L>0)). Dette kan være aktuelt når utskriften ikke får plass på en linje, eller man setter spesielle krav til layout'en. F.eks. utskrift av adresselapper.

«/»:S = @

I dette datasettet er «/» brukt som skille mellom felt. Trykker man på dette tegnet vil skrivemerket flyttes til høyre for «=», og man kan nå skrive inn inntil 10 tegn som skille mellom kolonner i utskrift - normalt får man kun en blank (angitt med «@»).

Ved f.eks. å skrive ut alle feltene i pakket form, samtidig som man legger inn noen ekstra «/», kan man utvide antall felt i datasettet. I så fall må skjema omdefineres, utskriften over legges ut på fil og «Importer» igjen. På samme måte kan også feltrekkefølgen endres.

Ved hjelp av alle parameterne til utskriftsrutinen, produseres tekstfiler for de fleste databasesystem.

«Utskrift», «Eksport» og «Import» er ordre som gjør at RUBREG kan produsere og lese filer av et «hvilket som helst» tekstformat: 1801-tellingen bruker f.eks. «%» som skille mellom felt, RHD anvender «<», mens vi i Oslo har brukt «/».

Disketter med data i slike format leses uten videre inn i RUBREG.

I utskriftsbestillingen over, har vi angitt «DESIMAL» på aldersfeltet. RUBREG vil derfor utarbeide statistikk hvor poster med ulovlig tallsyntaks blir forkastet.

Her følger resultatet:

Didrik	Didriksen	Gaardmand	38
Bernt Elias	Tønnesen	Gaardbruger	26
Bernt Johan	Berntsen	Gdbruger	30
Hans Syvert	Olsen	Gaardbruger	30
Lars Andreas	Larsen	Gaardmand	53
Tonette Marie	Tønnesdatter	Gaardmandskone	36
Anne Sophie	Nielsdatter	Konen	51
Didrik	Andersen	Ungkarl	41
Hans Christia	Hanssen	Ungkarl	50
Jens Jakob	Singdalsen	Skibsfører	39
Sivert Cornel	Jensen	Gaardbruger	35.5
Inger Bolette	Torkelsdt	Gaardbrugers ko	43

Statistikk:

=====

Alder:            Behandlet: 12    Forkastet: 0

    Maksimum(pnr. 40):            53.00  
    Minimum(pnr. 37):            26.00

    Sum:            472.50    Antall:        12    Snitt:        39.38  
    Varians:        69.38        Standardavvik:    8.33

Ovenstående kan legges ut på fil, på skjerm eller tas ut som papirutskrift.

### Sortering (O)

```
*****
* Order=O F6:UTFØR Oppsetting av sorteringsnøkkel *
* * *
* Dødsdato: _____ Sosial status: _____ *
* * *
* Fornavn: _____B_____ Etternavn: _____A_____ *
* * *
* Kjønn: _ Alder: ____C__ Bosted: _____ *
* * *
* Dødsårsak: _____ *
* * *
* F12:VERSALER- F13:FRONT F15:SYNKENDE F16:DESIMAL *
*****
```

Her har vi valgt å sortere primært på etternavn, deretter på fornavn og til slutt på «SYNKENDE» alder. Merk igjen det enkle oppsettet på en avansert sortering!

Sorteringsrutinen er svært kraftig, f.eks. er det også mulig å sortere baklengs: For å få gruppert folk etter type gård, kan vi sortere bostedsfeltet baklengs, det vil si at f.eks. alle ..stad-gårdene, ..land-gårdene grupperes sammen i utskrift.

Foretar vi en slik sortering, og deretter skriver ut «BOSTED»

høyrejustert, samt fornavn og farsnavn, kan resultatet bli slik:

Bosted	Fornavn	Etternavn
Bogstad	Lars	Olsen
Hogstad	Bernt Tobias	Tollaksen
Festervold	Rasmus Tobias	Nielsen
Festervold	Nicoline Maria	Marx, f. Bohenvqvist
Hougeland	Astrid	Einarsdt.
Hougeland	Tobias	Hansen
Langeland	Jacob	Andersen
Polland	Margrete	Nielsdt.
Polland	Gunild	Torkelsdt.
Quellandstrand	Tobias	Olsen
Langelandstrand	Hans Martin	Hansen
Berrefjord	Rachel Helene	Villumsdt.
Berrefjord	Mads	Christensen

Her er parameterne til sorteringsrutinen:

- F12: VERSALER+ eller VERSALER-
- F13: FRONT, SNUDD eller FINAL
- F15: STIGENDE eller SYNKENDE
- F16: TEKST, HELTALL eller DESIMAL

### Ordrene «Hent»(H) og «Finn»(F)

Et eksempel vil illustrere bruken:

Anta at vi i en folketelling vil undersøke «familieforholdene» til alle barn under 5 år. Ved hjelp av søkerutinen gjør vi et slikt utvalg, dvs. at RUBREG «holder fast» ved et subsett. Ved hjelp av «Hent» kan vi nå bla oss fram og tilbake blant «barn under 5 år». Samtidig er vi jo interessert i å studere barnets familie, altså kunne bla i det opprinnelige datasettet og i kronologisk orden. «Finn» leter fram en person med et bestemt postnummer fra dette settet og gir deg samtidig anledning til å bla «utenfor» utvalget – uten å «ødelegge» et eventuelt subsett.

På denne måten kan vi gå fra barn til barn ved hjelp av «Hent», og så studere barnets familie i detalj ved hjelp av «Finn».

(Ved bruk av «Finn» og «Hent» er det normalt ikke mulig å foreta endringer i dataene. Denne sperren kan imidlertid oppheves ved trykk på et par funksjonstaster).

### Oppsummering og videre planer

RUBREG er altså primært utviklet med henblikk på bruk innenfor historieforskningen. Det har imidlertid vist seg at selve designet kombinert med kompakt lagring av data, åpner for en mye bredere anvendelse.

Spesielt har det vært artig å se elever i en ungdomsskoleklasse

(Vålerenga skole i Oslo) ha utbytte av å studere eldre folketellingsmateriale ved hjelp av RUBREG.

I det hele tatt så virker det som om programmet vil kunne få en bred anvendelse langt ned i grunnskolen, og da selvfølgelig ikke bare i behandling av historisk materiale. I denne sammenheng er det også viktig å understreke at de fleste initiativ kommer fra lærere med humanistisk bakgrunn.

På bagrunn av den forholdsvis store interessen for RUBREG fra skolehold, er det planlagt å lage klassesett. Disse skal, ved siden av en tradisjonell manual, også bestå av hefter hvor programmet er brukt med utgangspunkt i en *faglig* problemstilling.

For profesjonelle historikere og andre som vil registrere og behandle historisk kildemateriale på egen hånd, er det planlagt en bok som konkret drøfter sider ved overføring til maskinlesbar form. De mest vanlige kildeseriene vil bli gjennomgått.

Det er flere ting ved RUBREG som bør rettes og forbedres. I dagens versjon (ennå bare prototype) er det f.eks. ikke lagt vekt på bruk av indeksfiler. Først og fremst p.g.a. egenskaper ved kildematerialet (navnstandardisering o.l.). Men siden bruken av programmet etter hvert også ser ut til å kunne omfatte mere standardiserte data (brukt i bl.a. museer), så vil oppslag via indekssekvensielle filer bli implementert. Det vil da være mulig å bygge indekser på et vilkårlig antall felt, og på deler og kombinasjoner av et vilkårlig antall felt.

*Kåre Andersen er vitenskapelig assistent ved HF-data, Universitetet i Oslo. Han arbeider med datamaskinell behandling av historisk kildemateriale og programutvikling for humaniora, og underviser i semesteremnet «Edb for humanister».*

# Kunstig tale på norsk

Ivar Utne

En snakkende datamaskin gjengir i tale det som den har fått inn som dataleselig tekst. Den gjør dette uten å kunne formulere språklig riktige setninger med et fornuftig innhold. Utvikling av maskintale dreier seg om å lage språkregler og maskinvare som gjør det mulig å omsette tekst til tale med en akseptabel kvalitet og med en naturlig talehastighet.

Så langt jeg kjenner til, snakker datamaskiner (også amerikanske) med umiskjennelig maskinstemme. Talen er forståelig, men neppe så velklingende at den kan erstatte bruk av mennesket som oppleser i stor grad. Maskintale kan først og fremst benyttes der det vil være svært vanskelig eller lite praktisk å bruke et menneske som oppleser. Mest aktuelt er dette derfor i forbindelse med opplesing av måleinstrumenter og som hjelpeutstyr for funksjonshemmede, særlig blinde. For folk flest vil det derfor ha liten praktisk betydning.

Et prosjekt for kunstig tale på norsk er nå kommet i gang, først og fremst som en følge av at en ved Nordisk institutt, PDS (Prosjekt for datamaskinell språkbehandling) ved Universitetet i Bergen har arbeidd med forskjellige edb-styrte kommunikasjonshjelpemidler for funksjonshemmede.

## Prosjektets gang til nå

Tidligere har en ved PDS prøvd ut talegenkjenningsystem (Votrax) og hatt kontakt med miljø som arbeider med kunstig tale (bl.a. *Gunnar Fant* og talesyntetisatoren OVE III ved KTH (Kungliga Tekniska Högskolan) i Stockholm). Vi ble høsten 1983 kjent med at systemet ved KTH var så godt utvikla at det var i salg i Sverige, under navnet SA-101.

Omkring årsskiftet 1983/84 tok vi kontakt med forskerne *Bjørn Granström* og *Rolf Carlson* ved KTH, og de stilte seg svært positive til et samarbeid for å videreutvikle sitt svenske system for norsk. *Kolbjørn Slethei*, Institutt for fonetikk og lingvistikk ved Universitetet i Bergen, deltar nå i samarbeidet og tar seg av utvikling av de fonetiske reglene.

Det grunnlagsmaterialet som er mest aktuelt for å komme i gang med en utvikling for norsk, er beskrivelse av norsk foneminventar i form av styreverdier for høytaler (FONETISKE REGLER), regler for omdanning fra tekst til lydskrift (TEKST-TIL-TALE-REGLER) og sifferregler (dvs. omsetting fra siffer til de tilsvarende bokstaverte ord). Systemet baserer seg i størst mulig grad på språkregler, slik at verken fonemer, ord eller setninger ligger ferdiglagra. Likevel finnes det en mindre unntaksordliste.

I denne artikkelen vil jeg presentere hovedprinsippene bak disse reglene, og legge størst vekt på tekst-til-tale-reglene som jeg har hatt



mest med å gjøre. Deretter vil jeg presentere prinsippene for programmene og maskinvarene som brukes. Til slutt blir det en liten omtale av det handikaputstyret som dette baner vei for.

## Sifferregler

Sifferreglene er forholdsvis enkle. De innebærer at «2» skal omsettes til ordet «TO», og «22» til «TJUETO». Firesifra tall som kan være både årstall og rekketall, må innafor et visst intervall omsettes til årstall. Det er derfor naturlig at 1985 omsettes til «NITTENÅTTIFEM». Sifferreglene må i størst mulig grad lages generelle og dermed reduseres i antall. Talla fra 20 til 99 omsettes til tieren + eneren. «21» blir til «TJUEEN». Talla fra 101 til 999 omsettes til antall hundre + «HUNDRE» + «OG» + tieren + eneren, i alle fall for tiere fra 2 og oppover. Tilsvarende regler lages for tall så langt de har navn.

## Fonetiske regler

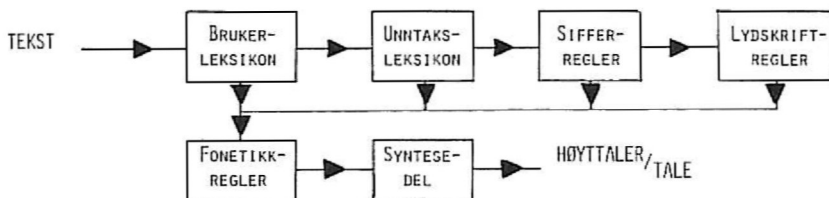
Da en tilstrekkelig detaljert bekskrivelse av norsk foneminventar av denne typen ikke eksisterer, har det flere ganger under denne utviklingen blitt gjort lydopptak for analyse. En slik analyse er avhengig av stemmer med klar artikulasjon. Materialet må gjøre det mulig å plukke ut de lyd-elementene som i sterkest grad bidrar til uttalen av en lyd.

Reglene for lyd kvalitet bygges opp som et knippe av parametre (styreverdier) for ulike enkeltlyder. Vokallydene (vokoidene) består av et sett av lyder som har regelmessige svingninger, mens konsonantlydene (kontoidene) i tillegg består av lyder med uregelmessige svingninger.

En vokal består i dette syntesesystemet av tre formantfrekvenser. Kombinasjonen av de tre formantene gir den aktuelle vokallyden. Konsonantlydene (kontoidene) består av formanter samt en eller flere uregelmessige lyder fra støykilder. Det kan være kilder som gir lukkelyd, friksjonslyd eller nasalisering. Dessuten består enhver lyd av en grunn tone som styrer tonelag uavhengig av fonem.

Alle disse enkeltlydene endres i løpet av den tiden fonemet uttales. Slik endring er dels generell for fonemet, og dels avhengig av den konteksten fonemet står i. Med kontekst mener jeg her dels aktuelle fonemer før og etter, og dels posisjon i morfem (del av ord), ord, ordgruppe og periode (setning). Slik endring over tid er lett hørbar i vokallyder som går over fra åpning til noe mer lukke over tid, f.eks. diftonger i norsk. For konsonanter er dette særlig aktuelt for lukkelyder.

Fonemkonteksten som er omtalt ovafor, kan avgjøre om en konsonant skal uttales stemt eller ustemt, f.eks. /B/ uttales ustemt foran en ustemt konsonant som /S/, slik det vil være naturlig i ordet «ABSOLUTT». Posisjonskontekst avgjør om lyden skal være forholdsvis lang eller kort i forhold til et ideal fonem. Vokaler i bøyingsendinger vil



*Inn i systemet sendes en maskinlagret TEKST. Den kan skrives inn på forhånd med tekstbehandlingssystem på samme maskin som er koplet til taleutstyret.*

*BRUKERLEKSIKONET er den enkelte brukers leksikon med lydskrift for ord som har spesiell uttale, f.eks. navn. Dette kan brukeren redigere selv.*

*UNNTAKSLEKSIKONET er systemets innebygde unntaksordliste. Den inneholder lydskrift for ord der lydskriftreglene ikke produserer rett lydskrift.*

*I SIFFERREGLENE omsettes tall til lydskrift.*

*LYDSKRIFTREGLENE omsetter de resterende ordene fra tekst til lydskrift.*

*FONETIKKREGLENE omsetter lydskrift – som er produkt av leksika, sifferreglene eller lydskriftreglene – til tallverdier for SYNTESEDELEN som sender signaler til en HØYTTALER.*

oftest være korte. Posisjonskonteksten avgjør dessuten trykk og tonelag. Grunntonen bestemmes bl.a. av type setning slik det formelt kommer til uttrykk gjennom tegnssetting (spørsmål, utrop og beretning). Dette kommer i programmet til uttrykk gjennom lydstyrke og grunntone.

### **Tekst-til-tale-regler**

Systemet tar imot dataleselig tekst i såkalt ASCII-kode. Det vil i praksis si det som skrives fra et tastatur eller ei fil som er produsert ved hjelp av et vanlig brukt tekstbehandlingssystem (f.eks. WordStar). Teksten omsettes til lydskrift før den håndteres videre av de fonetiske reglene. Sjøl om lydskrifta i form er ulik annen lydskrift som er i vanlig bruk, dekker den i innhold de mest aktuelle fonema. Lydskrifta sitt utseende bestemmes av de tegna som lett lar seg handtere datamaskinelt. Det vil si bokstaver og tall for å betegne de ulike fonema. Dessuten brukes bl.a. kolon, enkelt og dobbelt anførselstegn. Hovedregelen er at vokallyder, diftonger og konsonantlyder skrives slik vi gjengir dem i skrift.

Trykk og tonelag blir markert med trykktegn foran vokalene i de aktuelle stavelsene. Enkelt anførselstegn markerer trykk og tonelag 1, dobbelt anførselstegn markerer trykk og tonelag 2. «LANDET» i bestemt form av «LAND» skrives i lydskrift /L'ANE0T/ (E0 blir forklart nedafør), og som preteritum (fortid) av «LANDE» som /L"ANE0T/. /K'ASTE0/ («KASTET», substantiv i bestemt form) gir

tonelag 1 og /K"AST'EO/ (infinitiven «KASTE») gir tonelag 2.

Lange vokaler får et kolon etter vokaltegnet. Vokalers lengde bestemmes ut fra konsonantmengden etter vokalen. Lang (dvs. flere konsonanter) konsonant gir kort vokal. Enkelt konsonant eller ordslutt gir i hovedregelen lang vokal. Lengde angis i vokalen, og både enkel og dobbel konsonant skrives i lydskrift med en enkelt konsonantlyd. På denne måten omformes vokal + to konsonanter slik at «AKK» blir til /'AK/, og vokal + en konsonant (foran ny vokal eller morfemgrense) omformes slik at «AK» blir til /'A:K/ (altså lang vokal).

Reduserte vokaler markeres med null bak. For norsk er det særlig aktuelt for schwa, som vil si den slappe E-en som ofte forekommer i bøyingsendinger, f.eks. /K'ASTEOR/ («KASTER»). Dessuten fins egne lange og korte Æ-er og Ø-er som brukes foran R-lyd. Lange Æ-er og Ø-er blir til Æ3 og Ø3, mens korte blir til Æ4 og Ø4.

På konsonantsida fins retrofleksjer, dvs. egne sett R-lyder etterfulgt av bl.a. D og T. RD blir til 2D, RT til 2T osv. «ERT» blir til /Æ42T/, dvs. kort Æ-lyd foran retroflektert R.

Trykk i enkeltord legges oftest på første stavelse. Unntak gjelder for ord hvor utenlandsk opphav som krever annet trykk, lett kan gjenkjennes ut fra skriftbildet. Slik gjenkjenning gjøres ved hjelp av bestemte bokstavsekvenser f.eks. i forstaving, avledning og bøyingsending. Mer om dette senere.

Tonelag kan bestemmes ved hjelp av formelle trekk, dvs. ords utseende. Infinitivsformer med en enkelt stammevokal, som i «KASTE» har tonelag 2. Det samme gjelder substantiv med samme oppbygning – som «HYTTE». En slik framgangsmåte kan brukes til å lage hovedregler, i dette tilfellet endinga -E i ord med en stammevokal. Preteritum (fortid) av verb og bestemt form entall av substantiv ender oftest på -ET. Det er i hovedsak slik at disse substantiva har tonelag 1, mens verba har tonelag 2. Det fins svært få kjennetegn for å holde disse fra hverandre. Program for syntaktisk analyse (parsing) ville kunne være et nyttig redskap til dette. Slik det er nå gir reglene dem det ene av tonelagene. Og en del av de vanligste orda i språket som da får gale tonelag, legges inn i unntaksordlista. I tilfeller med identiske skrivemåter (homografer) med ulike tonelag, som «KASTET», må en velge ett av tonelagene, og dermed få gale tonelag sjøl om en bruker unntaksordliste.

Ord som ender på -ER kan være ubestemt form flertall, presens eller nomen agentis i ubestemt form entall og betegnelser for nasjonalitet. Ubestemt form flertall av substantiv får i regelen tonelag 2, som i «BILER», «GUTTER», «JENTER», «KATTER» og «EPLER». Men unntak fins f.eks. i «BØKER» og «ÅKRER» (som vel kan være både tonelag 1 og 2).

I presens får vi oftest tonelag 2, som i «BAKER», «LØPER», «SYKLER» og «BLOMSTRER». Men tonelag 1 bl.a. i «SKRIVER», «LESER» og «HJELPER».

I nomen agentis får vi ofte tonelag 2, som i «HJELPER» (som skiller seg fra presensforma), «BAKER».

Betegnelser for nasjonalitet får tonelag 1, som i «RUSSER», og «SVEITSER».

I de omtalte tilfella av ord som ender på -ER hadde alle en stammevokal + vokalen i endinga. Dessuten er det innlånte ord. Innlånte ord har et mer komplisert regelverk for både trykk og tonelag. «ARBEIDER» er nomen agentis og presens og har trykk på siste stammevokal og tonelag 1. Slik er det også med «MASKINER» og «DELEGATER». I en del tilfeller kan regler for innlånte ord med visse forstavelser eller avledningskjennetegn plukkes ut. Ord som oppfører seg som de sistnevnte begynner bl.a. på BE- og PRO-, og ender bl.a. på -AT og -ESSE. Slike regler vil føre til en del rett og en del gal markering. Samme forstavelse eller avledning fører til ulike resultater, slik det er med FOR- som forstavelse, jf. «FORBUD», «FORBUDT», «FOREBYGGE» og «FOREVIGE».

Tonelag i setninger bestemmes ut fra skilletegn, og det gjøres under de fonetiske reglene.

Homografer kan skilles fra hverandre med syntaksanalyse. Det fins foreløpig ikke program for syntaktisk analyse for norsk som fyller de krav til hastighet som talesyntese krever. Syntaksanalyse ville først og fremst kunne utnyttes til beregning av setningsintonasjon (se for øvrig omtalen av unntaksordlista der en del såkalte funksjonsord er merka; for syntaksanalyse vises det til Humanistiske Data nr. 3-84).

En del av reglene baserer seg på at det er mulig å gjenkjenne morfem og morfemgrenser. Ved hjelp av lister over aktuelle forstavinger, avledninger og bøyingsendinger er det mulig å identifisere en del morfem. Dette vil lett føre til noen feil, som i ord som «FORMENES» der det for programmet ikke vil være helt klart om FOR er prefiks. Det vil være lettere å avvise det dersom det var færre tegn etter, som i «FORM» og «FORMEN». Det kan legges inn i reglene at det må være f.eks. en vokal etter, og det som kommer etter ikke må være lik noen av de mulige bøyingsendingene. Da vil FOR i «FORMANN» bli tatt som forstaving. «FORMENES» kan også bli tolka rett om en tar inn regler for genitiv.

I tillegg til morfemlister kan delingspunkt i sammensatte ord i noen grad identifiseres ved hjelp av lister over bokstavkombinasjoner som ofte forekommer i sammenfuginger, og som sjelden forekommer ellers. Ordet «FINNMARKSVIDDA» bør deles der det er tre eller flere konsonanter etter hverandre. Det går ikke å uttale NM etter hverandre, og det er derfor ingen sannsynlig start på en ord-del (morfem). Derfor må en dele mellom dobbel N og M. RKSVM må deles. RK og SV fins sammen i mange ord. Det gjelder også KS om det kommer etter vokal. Ut fra argumentene ovafor kunne et naturlig skille være mellom RK og SV. Det går f.eks. i et ord som «PARKSVINGEN». En må gjøre et valg. Det er mest sannsynlig at en S i nærheten av et ord delingspunkt er

en sammenfugings-S, og det faller naturlig med dele og pause (jf. ovafor) etter S-en.

### Unntaksordlista

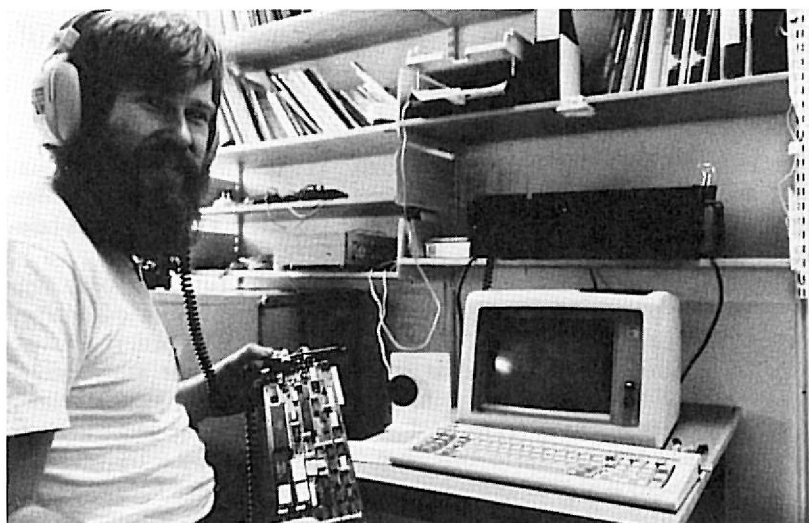
Unntaksordlista har to funksjoner. Dels skal den inneholde informasjon om ord som bidrar til setningsintonasjonen, og dels skal den inneholde høyfrekvente ord som får gal lydskrift etter tekst-til-tale-reglene.

Ord som bidrar til setningsintonasjon er såkalte funksjonsord, som vil si konjunksjoner, preposisjoner, pronomener, artikler og hjelpeverb. Disse blir merka spesielt for at programmet skal gi dem trykksvak posisjon i setningen. Dette bidrar sterkt til at talen blir minst mulig monoton og får en mest mulig naturlig intonasjon på setningsplan.

Der er ikke realistisk å regne med at en kan lage regler som alltid gir rett lydskrift. En del korte ord som «DEM», «FOR» og «MEN» (konjunksjonen) skal uttales med kort vokal, mens andre som «DAG», «GÅR» og «SER» skal uttales med lang. Dessuten får svært mange ord gal markering av trykk og tonelag, slik det gikk fram ovafor. I fremmedord er det dels gale trykkmarkeringer og dels gale fonem som produseres av tekst-til-tale-reglene. Et eksempel er ord med fransk uttale, som «DEPARTEMENT» med uttalen /DEPARTEOM'ANG/. Vi kunne tenke oss en regel som gjorde «..MENT..» i visse posisjoner om til /..M'ANG../ lydskrift. En slik regel vil også ta med seg ord som skal ha norsk uttale, som «SEMENT». Når en vurderer regler for fremmedord, tar vi hensyn til hva det fører med seg for norske ord. Om det f.eks. er slik at det bidrar til rett uttale for mange fremmedord og regelen samtidig går ut over få ord med norsk uttale, så tas regelen i bruk. Ordet med norsk uttale settes da inn i unntaksordlista.

Unntakslista i den norske og svenske versjonen inneholder ca. 2.500 ord. Den nåværende norske versjonen klarer ved hjelp av unntakslista å få rett uttale på de 5-6000 mest vanlige orda. Foreløpig får om lag halvparten av disse vanligste orda avvik fra en ideell lydskrift, oftest trykk- og tonelagsavvik.

Med det grunnlaget som er nevnt ovafor skal maskinen i prinsippet ta et hvilket som helst ord og gi det en rimelig uttale. Det er målet at unntaksordlista skal sørge for at de 10.000 mest vanlige orda skal få rett lydskrift. De 10.000 mest vanlige orda i språket utgjør rundt regnet 95 prosent av orda i en allmennspråklig tekst. Dette vil naturligvis påvirke helhetsbildet betydelig. De alvorligste feila vil, så langt jeg har erfart, oppstå i fremmedord (der norske uttaleregler ikke passer naturlig inn). Særlig alvorlig er det i lange ord der det får betydelige følger også for trykk og tonelag. Vansker med å identifisere morfemgrenser fører til unaturlige pauser og mangel på slike. Lange norske ord er også et problem av samme grunner.



*Ivar Utne med utstyret for omforming av tekst til lyd. Foto: AV-avdelingen, Universitetet i Bergen.*

### **Arbeidet med et korpus**

Utviklinga av regler både for fonetikk og lydskrift krever at en har et utvalg språkeksempler.

For fonetikkreglene trengs et utvalg av ord som gir flest mulig av språkets lydkombinasjoner i flest mulig av de posisjoner i ord og setninger som er aktuelle.

For å få fram uttalen av lyder i visse kontekster, er det behov for å holde andre faktorer konstante. For å få dette til, har en i noen grad konstruert tøyseord (ord som ikke fins i språket). Slikt materiale er lest inn på lydband, og det er grunnlag for fonetiske analyser høsten 1985.

For lydskriftreglene trengs ideelt sett også et utvalg av ord som representerer alle bokstavkombinasjoner i språket. Da et slikt regelverk kan bli svært omfattende, og mange av reglene vil bli brukt svært sjelden, har en prioritert å lage regler som brukes hyppig. Det får en til med å arbeide med en liste over de mest brukte orda i språket. Vi bruker lista med de 10.042 mest frekvente orda i norsk bokmål som fins i «Norsk frekvensordbok» av *Kolbjørn Heggstad*. Et slikt arbeid forutsetter at en har gjort seg opp visse oppfatninger om hvilke regler som er mest hyppige. Særlig gjelder det fordeling av lange og korte vokaler, uttalen av E foran R, uttalen av O (som O eller Å) og U (som U eller O), uttalen av G (som G, J eller bortfall) og bortfall av D og H. Når en har grunnlag for en del slike regler, kjøres ordlista gjennom et program med

slike regler og produserer et forslag til lydskrift for disse orda.

Disse blir deretter gjennomgått ett for ett på den måten at maskinen uttaler ord for ord med grunnlag i den produserte lydskrifta. Om det er feil, kan disse rettes opp underveis og legges ut på en utfil. Etter en slik gjennomgang blir innfil og utfil sammenlikna. Avvika blir brukt som grunnlagsmateriale for nye regler. På denne måten kan programmet stadig forbedres og utdata fra lydskrift-programmet kan sammenlignes med den kontrollerte fila. En slik kontrollert fil, som for øvrig har gjennomgått flere omganger med manuelle rettinger, tjener som et referansegrunnlag når reglene skal finpusses.

### Programmene som styrer syntesen

Utstyret betjenes av et eget høynivå-programmeringsspråk utarbeidd ved KTH. Det vil si at lingvistene eller fonetikerne forholdsvis lett kan skrive inn regler i en notasjon som ligner mye på oppsett i generativ fonologi. Regler som kan være avhengige av hverandre, kommer i en rekkefølge som gjør at de blir aktivisert riktig. Det vil si at rekkefølgen reglene settes opp i, påvirker resultatet.

Reglene, skrevet i høynivåspråket, går gjennom kompilatoren som oversetter til maskinkode, dvs. program klart for talesyntese.

Reglene for alle nivåer skrives inn etter et fast oppsett. I utgangspunktet er det slik at om X skal endres til Y når det står etter A og før B, så skrives dette slik:

X — Y % A\_B

Til venstre for prosenttegnet «%» står den endringa som skal til. Tegnet til venstre for pila «—» skal endres til tegnet til høyre for pila. Endringa skal skje i den konteksta som står til høyre for skråstreken. Det aktuelle tegnet symboliseres med \_-tegn. Kontekst til venstre står til venstre for \_-tegnet, og kontekst til høyre står til høyre for \_-tegnet. Både tegna til venstre for skråstreken og kontekst til høyre for skråstreken kan stå som ett eller flere bokstavtegn i sammenheng (f.eks. prefikset «FOR»), eller det kan være distinktive faktorer fra generativ fonologi (som i Noam Chomsky & Morris Halle: *The Sound Pattern of English*, New York 1968). Bruk av distinktive faktorer gjør det mulig å inkludere grupper av fonem.

Et konkret eksempel ville være en regel som gjorde bokstaven «I» om til lydskrift-tegna /I:/, dvs. lang /I/. Vi kan forutsette at det kan stå hva som helst foran, men at det enten må være ordgrense eller enkelt konsonant etter.

Siden det er valgfrihet med tanke på høyrekontekst, setter vi opp to regler, slik:

I — I: % — <ORDGRENSE>

I — I: % — <KONSONANT,1,1> <VOKAL,0,\*>



Den første betyr at «I» foran ordgrense blir lang. Når det ikke er nevnt noe om venstrekontekst, står det ingen ting der. Den andre regelen sier at i tilfeller der høyrekonteksten består av minst en og høyst en konsonant og deretter av minst null (dvs. fravær) og ubegrensa mengde (markert med stjerne) vokaler, så gjelder regelen.

## Maskinvarene

Maskinvarene som trengs for brukeren fins til nå i tre versjoner:

1. SA-101 som er en frittstående boks.
2. Kort for IBM PC. Dette monteres inn i maskinen.
3. Ny frittstående boks som settes i produksjon i 1985.

SA-101 er utvikla som en sjølstendig datamaskin med innebygd høyttaler. For å kunne brukes trenger den tilførsel av strøm (2 ampere, 5 volt) og tilkopling til en dataterminal. En dataterminal vil i dette tilfellet si tastatur og skjerm med en såkalt RS232C kabelinngang. Det vil i praksis si at flere av de mindre hjemmedatamaskinene ikke kan brukes. RS232C er for øvrig svært vanlig på større mikromaskiner (f.eks. i prisklassen over 20.000 kroner) og stormaskiner.

Maskinen består av:

- en CPU som er sentralenhet som styrer maskinen og utfører de nødvendige beregninger som andre program har behov for hjelp til,
- et styreprogram som tar seg av tekst og kommandoer, og formidler den nødvendige informasjon videre til språkprogram og fra språkprogram og videre til synteseprogrammet som igjen styrer den tekniske lydproduksjonen,
- et arbeidsminne der programmene får plassere den informasjonen de til en hver tid arbeider med,
- et språkprogram som arbeider etter slike regler som er skissert ovafor; nært knytta til dette programmet ligger ei unntaksliste som i den norske versjonen er på omkring 2.500 ord,
- plass for ei ordliste som kan fylles individuelt i den enkelte maskinen, med plass til 3-400 ord,
- synteseprogram (signalprosessor av typen NEC7720 utvikla ved KTH) som omsetter de taleparametrene (styreverdier) som språkprogrammet leverer med grunnlag i sine regler; dette sendes videre som digitale signal pluss og minus (eller 0 og 1) til en D/A-omformer (se neste punkt),
- en digital/analog-omformer (D/A-omformer) omsetter digitalt koda lyd til analoge lydsignaler, som vil si signaler som en høyttaler kan omsette til hørbar lyd,
- en høyttaler på 1 Watt.

Denne modellen vil med det første bli bytta ut med en annen der transformator er innebygd. Denne blir lansert i 1985.

I 1984 fikk man i stand et kort for IBM PC og IBM PC-kompatible

maskiner. Et slikt kort settes inn i maskinen, og en unngår derfor en boks som tar plass i tillegg til maskinen. Det må koples til en løs høyttaler, sjøl om det også er teknisk mulig å bruke den innebygde høyttaleren i maskinen. Det dreier seg om et såkalt dobbelt Europa-kort. Kortet lastes med ordliste og regler fra diskett. Kortet er det samme for alle språk, mens diskettene inneholder dels programvare som er felles for alle språk og ordliste og regler som er språkspesifikke.

Prisen for et PC-kort er ca. 12.000 kroner, mens en frittstående boks koster ca. 20.000 kroner. Dette kommer i tillegg til en PC som normalt koster fra 20.000 kroner og oppover.

### **Bruk av utstyret**

SA-101 kan koples til terminal slik at det en skriver på et tastatur kommer på skjerm og som tale. Talen kan tas ut som hele ytringer. I det tilfellet kommer hele ytringa fram til et større skilletegn, som vil si punktum, utropstegn eller spørsmålstegn. Det går også an å arbeide med bokstavmodus, ordmodus eller kombinert ord og ytringsmodus. Bokstavmodus fører til at hver bokstav uttales når den er skrevet inn, ordmodus gir ett og ett ord, og kombimodus gir hvert ord under innskriving og hele ytringa til slutt. Dessuten kan en med kommandoer styre fart, lydstyrke og stemmekvalitet.

Men bruk av terminal alene gir få muligheter. Mer brukbart blir det når talen brukes i kommunikasjonen mellom terminal og en annen datamaskin. Det ordnes teknisk med at en kopler inn SA-101 på kabelen mellom maskinen og terminalen. Da blir det mulig å få taleboksen til å lese opp det som maskinen sender tilbake til terminalen. Særlig nyttig er det at en på denne måten kan få lest opp ei fil sammenhengende.

Boksen kan også koples slik at den kan brukes til teksteditorer, helst linjeeditorer, men også fullskjermeditoren WordStar. Den leser opp det som skrives, og det er mulig å få lest opp de linjer en ønsker. Utstyret er bl.a. kopla til mikromaskinene ABC80, ABC800, Kompis (som tilsvare Scandis i Norge), IBM PC og maskiner som er kompatible med IBM PC. I praksis viser det seg at den også kan koples opp mot andre mikromaskiner. Oftest må det skrives et kommunikasjonsprogram som gjør det mulig å gi kommandoer fra vertsmaskinen til taleenheten.

PC-kortet settes inn i en PC og det kan aktiviseres som en egen «device», som i praksis vil si at man skriver til den som man skriver til ei fil. Dette kan gjøres fra et program eller ved hjelp av visse DOS-kommandoer (kommandoer i operativsystemet). Det fins et eget program som gjør at alt som kommer på skjermen går ut som tale samtidig.

Tilkopla WordStar er det mulig å få lest utvalgte avsnitt av skjerm-bildet eller dokumentet bokstav for bokstav, ord for ord eller som flytende tale. Brukeren kan dessuten få lest opp linjenummer og markørposisjon, og evt. store bokstaver og tegn. En slik utvikling er

gjort i samarbeid mellom MicroPro (som leverer WordStar) og INFO-VOX.

### **Organisering av utvikling og markedsføring**

Arbeidet i Norge har til nå vært utført av meg i samarbeid med KTH og Infovox. I januar 1985 søkte KTH-miljøet sammen med Nordisk inst. PDS den Nordiska handikapnemnden om økonomisk støtte for videreutvikling av kunstig tale på norsk. Det er blitt bevilget 150.000 kroner. Midlene brukes dels til utvikling av fonetiske regler av forskerne ved KTH, *Kjell Gustafson* (en norsk språkviter og fonetiker i Oxford) og Kolbjørn Slethei og tekst-til-tale-regler ved Nordisk institutt PDS. Alt videre samarbeid er tenkt mellom de nevnte parter.

Infovox har valgt Standard Telefon og Kabelfabrikk (STK) som sin norske kontakt og viderefører. STK arbeider for øvrig med andre produkter for handikapmarkedet, og kunstig tale vil på den måten komme inn i en naturlig sammenheng.

Den eksisterende versjonen av norsk snakker slik at en trent lytter vil klare seg godt. Det er realistisk med norsk lansering i 1986.

### **Hjelpeutstyr for funksjonshemmede**

Utstyr for tale-ut kan være til hjelp som leseutstyr for blinde og taleerstatning for funksjonshemmede med talevansker. For blinde kan det nyttes til opplesing av dataleselige tekster. Dette innebærer at bøker, aviser og brev som fins dataleselig lett kan leses opp av maskinen. Det er ganske vanlig at bøker og aviser trykkes med grunnlag i dataleselige versjoner. Med mindre modifikasjoner vil disse kunne spres f.eks. på diskett. I Sverige har man f.eks. gjort forsøk med overføring av dataleselige tekster fra aviser til harddisker på brukernes mikromaskiner. Slik overføring foregår f.eks. om nettene enten over modem (telesamband) eller over FM-båndet (radiosamband). Disse forholdsvis store tekstmengdene kan brukerne søke i, og så få lest opp de delene de vil høre. En annen mulighet er å la taleutstyr lese teksten inn på lydassetter, som deretter distribueres. På den måten spares personer til innlesing, og opplesinga kan kjøres kontinuerlig uten stopp. Mengden lest tekst kan bli stor på bekostning av kvaliteten.

Gamle tekster som ikke fins dataleselig, kan gjøres tilgjengelige ved hjelp av optiske lesere, dvs. leseutstyr som leser trykt tekst og skriver den ut i dataleselig form. Slikt leseutstyr er dyrt, og det samme er kjøp av slike tjenester som ofte ikke er mye rimeligere enn nyinnskriving.

Det fins programvare som gjør det mulig å kople tale til alt som blir vist på skjermen, deriblant også tekstbehandlingsutstyr (f.eks. WordStar, se ovenfor) og annen programvare (f.eks. DBase II og III).

For folk med talevansker har det eksistert utstyr som viser på en skjerm det som den funksjonshemmede skriver på tastaturet. Det fins også utstyr med 500-1000 felt på en plate, der hvert felt symboliserer ett

ord eller del av et ord. Utstyret kalles dels SPLINK (der orda står skrevet med bokstaver) og BLISS (med tegn for hvert av orda). Det aktuelle ordet eller et lenger utsagn som bygges opp, sendes til skjerm eller, i dette tilfellet, til taleutstyret. For talehemma og sterkt bevegelseshemma fins det også programvare der en ved hjelp av menyer søker seg fram til ord. Styrima gjennom menyene kan i disse tilfellene gjøres med en, to eller et lite antall knapper (bl.a. såkalte joy-sticks). Disse knappene kan være knyttet til steder på kroppen som den bevegelseshemma har god styring over.

For ytterligere omtale av edb-hjelpemidler for funksjonshemma vises det bl.a. til artikkel av undertegnede i Humanistiske Data nr. 1-84. Samme artikkel er bygd noe ut i Bjarne Norevik (red.): Datamaskinen og språket. Universitetsforlaget 1985.

For ytterligere omtale av kunstig tale vises til artikkelen «Datamaskinen - talemaskinen?» av Kolbjørn Slethei i Bjarne Norevik (red.): Datamaskinen og språket. Universitetsforlaget 1985.

*Ivar Utne er vik. amanuensis ved Prosjekt for datamaskinell språkbehandling, Nordisk institutt, Universitetet i Bergen. Han har tidligere arbeidet med ord- og tekstforståelse, døves mestring av skriftspråk og personnavn, og er nå beskjeftiget med kunstig tale og datamaskinelt ordboksarbeid. Han har skrevet diverse artikler om disse emnene.*

# **Databehandling av termtilfanget frå Terminol-prosjektet**

*Per-Bjørn Pedersen*

Terminol-prosjektet er eit samarbeidsprosjekt mellom Universitetet i Bergen og Statoil om terminologien i handbøkene til Gullfaks A-plattformen. Det tok til 1.1.1984 og er førebels forlenga ut året 1985. Det har ei total økonomisk ramme på vel 5 mill. kroner. Gullfaksfeltet skal administrerast på norsk. Det har i sin tur ført til at alle driftshandbøkene lyt liggja føre på norsk. Desse handbøkene vart fyrst skrivne på engelsk. Ein viktig grunn for det var at det ikkje fanst nokon utarbeidd norsk terminologi på mange av fagområda. Sjølve omsetjingsarbeidet vart utført av Norwegian Petroleum Consultants og MORCO. Universitetet i Bergen har arbeidd med den norske fagterminologien i samarbeid med handboksforfatarane og fagfolk innanfor dei ulike fagfelt i Statoil og Morco.

Prosedyren i arbeidet var først ekserpering og merking av moglege termar i heile handboksmaterialet. Det ekserpererte materialet vart så registrert datamaskinelt, fyrst kvar handbok for seg. Så førte ein på norske ekvivalentar. Etter kvart som tilfanget voks, tok ein til å samla datasetta med terminologi frå dei einskilde handbøkene i ei større fil, som etter kvart kom til å innehalda heile termtilfanget i handbøkene til Gullfaks A. I denne prosessen vart materialet ein god del komprimert ved at dobbel ekserpering frå fleire handbøker vart eliminert. Synonyme nemningar vart førte saman under som same termpost osv. Sidan handbøkene dekkjer ulike system på plattformen, vart det likevel ikkje meir av reine dobbelføringar enn at det var praktisk å arbeida etter måten lenge med termlistene for kvar einskild handbok. Men før ein kunne dra full nytte av det terminologiske arbeidet i omsetjings- og kontrollprosessen, måtte heile tilfanget finnast sortert etter engelsk og norsk hovudoppslag i samla datasett. Dette målet vart nådd utpå våren 1985, og omsetjingsarbeidet hadde då alt gått føre seg i fleire månader. I den tida hadde omsetjaren måtta hjelpa seg med førebels lister frå prosjektet etter kvart som arbeidet gjekk framover. Det førte då og til at det vart noko etterarbeid med å retta inkonsekvensar og feil etter at omsetjarane hadde gjort seg ferdige med sin del av arbeidet.

Omkring årsskiftet 1984/85 starta utviklingsarbeidet med ein MDDBS III database for terminologien i Gullfaks A-handbøkene. Databasen låg føre i ein tidleg prototype den 8.2.85 og i ein ferdig kjøreverisjon i mai 1985.

Av datautstyr disponerer Terminolprosjektet tre terminalar mot universitetets UNIVAC-anlegg. Og det er og kjøpt inn 7 IBM PC datamaskiner og ein PC av merket Apricot. Tre av prosjektet sine

PC-maskiner har utviding med fast platelager. Ei av dei er ein standard IBM XT. Elles disponerer prosjektet ei IBM utvidingseining med 20 megabytes platelager på to stasjonar og ei Bernoulli Box utvidingseining for 10 megabytes patroner på to stasjonar, altså 20 megabytes samstundes i tillegg til diskettstasjonar. Prosjektet har dessutan ein del skrivarutstyr til ulike formål.

Det er kjøpt inn kablar og programvare for snøgg overføring gjennom parallellutgang frå PC til PC, både mellom IBM-maskiner og mellom Apricot og IBM. Av standard programvare til PC disponerer prosjektet m.a. tekstbehandlingssystem, kompilatorar i ei rekkje programmeringsspråk og databasesystemet MDBS III.

Prosedyren for databehandling av termtilfanget frå Terminolprosjektet fylgde i hovudtrekka innarbeidde rutiner frå tidlegare arbeid, som er utført ved Nordisk institutt, Norsk termbank. Det er termbankens format som er nytta, og det same gjeld program for sortering, rydding, plukking og sletting av termpostar.

Innkjøpet av PC-utstyr gjorde det mogleg å overføra all fyrste registrering av termdata frå UNIVAC-anlegget til mikromaskiner. Ein vart då uavhengig av responstid og maskinstabilitet i denne fasen av arbeidet. Til registrering nytta ein termbankens særskilte registreringsformat, som sparer tastslag. Dei registrerte termdata vart så overført til UNIVAC-anlegget over modem og liner. Data vart så omforma til Termbankens eigentlege grunnformat, og den vidare oppdatering og vidareforedling vart så utført på UNIVAC-anlegget med tekstredigeringsprogram.

Termformatet til Norsk termbank har mange moglege felt som kan fyllast ut for kvar term. I Terminolprosjektet er det berre bruka nokre av dei. Det er for det fyrste språkkodane for engelsk og norsk. Så er det koding for hovudoppslag, synonym og term som er nytta, men ikkje tilrådd. Til dette kjem så koding for at ei feltutfylling er ei korting, og ikkje full term. Elles er det teke med definisjonar som finst i materialet og viktige, gjerne forklarande kontekstar. Det vert dessutan markert om termen finst på teikning eller illustrasjon, saman med referansen til teikning/figur. Det vert i alle høve notert bok og sidetilvising for den fyrste staden termen er registrert, og gjerne fleire førekomstar når tydinga vert klårare gjennom fleire kontekstar. Det er sett av eit eige felt til dato for «normering» av termen. Det vil seia dato då ein sakkunnig komite i Statoil eller Morco fastla norsk-engelsk korrespondanse og samling av ulike synonym under eitt oppslag. Ein del termar er og normerte indirekte. Det vil seia at norsk-engelsk samsvar og samling av synonym er fastlagde på grunnlag av tidlegare komitevedtak for nærskylde termar. Slik normering går fram av innført dato.

I arbeidet med den samla fila med termdata til «Gullfaks A» oppstod det nokså snart vanskar med at fila vart for stor for rasjonell behandling med eit tekst-redigeringsprogram. Dette var knytta til dei vanlege problema med å gå «bakover» i fila og tida som går med til å

lesa framom mange datapostar for å nå postar som står langt frå kvarandre. Redigering av den store fila førde til at ho lett kom ut av sortert rekkjefølgje, og måtte gjennom ny sortering for det igjen kunne lagast lister av materialet som var eigna til å slå opp i. Dette vart lett større kjøringar som ikkje utan vidare kunna gå på dagtid. Det representerte heller ikkje særleg rasjonell ressursbruk, sidan etter måten store datasett måtte gjennom sorteringsprogramma endå om berre få postar var ute av rekkjefølgja. Det viste seg dessutan vanskeleg å halda oversikten over kva for endringar som vart utførte på det samla termtilfanget ved kvar oppdatering. To dataregistratorar og nokre terminologar arbeidde parallelt med redigering, både av handboksfilar og hovudfil.

Dette var bakgrunnen for at rutineane for denne meir tradisjonelle databehandlinga i sekvensielle filer vart lagde om seinhaustes 1984. Det var tre mål som ein ville oppnå:

1. Betre trygging av data mot tilfeldig og dårleg koordinert endring.
2. Betinging av responstid ved redigering av sekvensielle datasett.
3. Avgrensing av sortering til mindre datasett.

Dette vart løyst på følgjande måte: Fila med termdata vart aldri redigert direkte. Den vart vidare alltid halden vedlike i sortert rekkjefølgje etter engelsk hovudoppslag. Termpostar som skulle endrast, vart fyrst tekne ut av hovudfila, og lagde opp i eigne datasett. Der vart dei så endra, samanslegne, strøkne osv. Den redigerte fila vart så sortert og fletta inn att i hovudfila. Både redigering og sortering kunne då liggja på etter måten små datasett, der oppgåvene var konsentrerte i ein kort tekst.

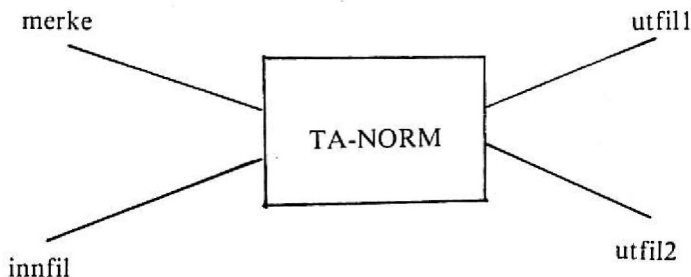
Alle fletteoperasjonar vart nøye bokførte med dato og opplysning om kva som var ført inn i hovudfila av kven. Sikringskopi på band vart tekne med jamne mellomrom, slik at det var råd å gå kontrollert tilbake til tidlegare utviklingssteg dersom ein feil hadde fått utvikla seg over tid.

I samanheng med innføringa av desse rutineane laut det utviklast nokre nye program. Men dei fleste programma til Norsk termbank vart framleis bruka. Eitt viktig nytt program utførte oppdeling av hovudfila i to andre datasett etter spesifikasjonar i ein såkalla «merkefil». Programmet las altså frå to filar og skreiv til to nye. Det heiter «TA-NORM» og er skriva i COBOL.

## **TA-NORM**

Merkefila er i prinsippet «termpostar» med engelsk hovudoppslag og skiljeline. Slike «merkefiler» kan skrivast inn frå terminal med hjelp av ein makro i registreringsprogrammet. Men i dei fleste tilfelle vart dei genererte frå fil med ein makro eller med plukkeprogram, som søkte etter strenger i norsk/engelsk oppslag, synonym og korting, eller eigenskapar ved termposten, så som tilvising til handbok eller illustrasjon. Merkefila var så nøkkelen til å dela hovudfila slik som merkefila





viste. Termar i hovudfila med same oppslag som i merkefila kom i «utfil2», resten kom i «utfil1». Som merkefil kunne ein og nytta fullstendige datasett i termformat. Det gjorde det mogleg å få ut i «fil2» alle termar i hovudfila som hadde same hovudoppslag som termar i t.d. ei handboksfil som skulle leggjast inn. Dette var ein nyttig teknikk i jakta på dobbeltpostar når ein redigerte saman ei handbokstil med hovudfila.

Resultata frå arbeidet i prosjektet vart formidla til oppdragsgjevar, Statoil, og omsetjarfirma, Norwegian Petroleum Consultants og Morco, gjennom lister med termar til høyring og godkjenning i komitemøte, og utskrifter av materialet til bruk for handboksforfattarar og omsetjarar. Ettersom materialet vaks, var det viktig å ha tilgang til kraftige skrivarar tilknytta Universitetets hovudanlegg.

Ettersom arbeidet med terminologien i Gullfakshandbøkene var samla, påført ekvivalentar og normert, trongst det ein effektiv reiskap til å slå opp i tilfanget både etter norsk og engelsk term. Det var viktig å finna fram til hovudoppslag, synonym, frårådd term. Det var ikkje nok med alfabetisering på start av termline, men det var og ynskjeleg å slå opp på termar som inneheldt einskilde ord og sekvensar av ord. For å nå fram til konsekvens i skrivemåte av fuger, endingar og lekkar kunne det vera ynskjeleg med oppslag på ordstart, ordslutt og bokstavstreng i det heile. Dette vart i prosjektet vurdert som ei høveleg oppgåve for eit databasesystem.

Terminolprosjektet har kjøpt databasesystemet MDBS III til bruk på IBM PC-utstyr. Dette er eit system av typen utvida nettverk (extended network) som er ei vidareutvikling av CODASYL nettverk. Ein kan i den logiske strukturen til MDBS III representera alle forhold mellom posttypar som let seg representera i CODASYL nettverk. Men i tillegg kan t.d. mange til mange forhold (N:M) representast direkte i DDL (DATA DESCRIPTION LANGUAGE) til MDBS III. MDBS III tillet elles og rekursive sett (relasjonar). Det vil seia at ein og same posttype kan vera både eigar og medlem av eit sett (relasjon). MDBS III tillet lett multipel eigar- og medlem-sett og er ein kraftig reiskap for å

framstilla både enkle og meir innfløkte logiske strukturar på ein kortfatta og oversiktleg måte.

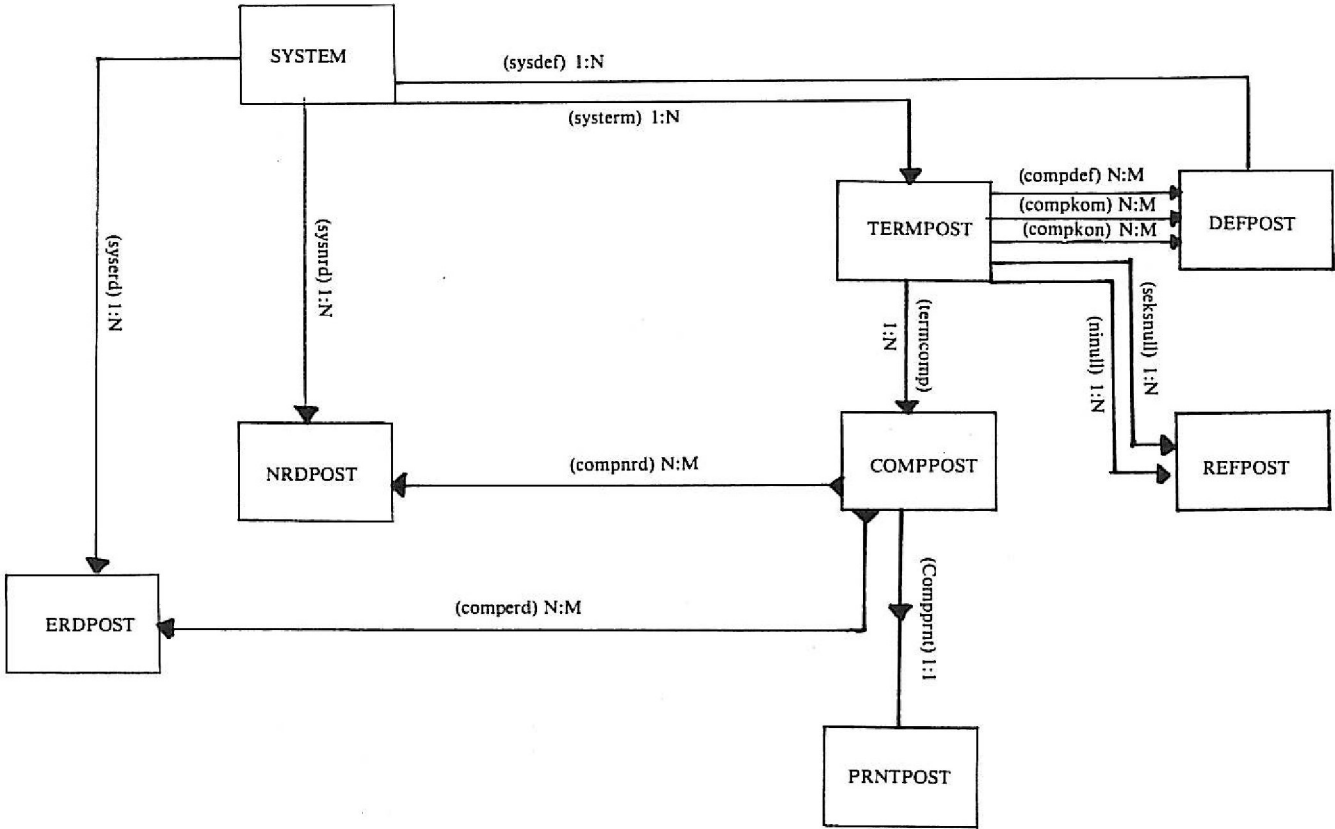
Det arbeidet som førte fram til ein kjøreversjon av ein terminologisk database i MDBS III for Terminols data, tok til ved juletider 1984. Det kom frå fyrst av i stand eit samarbeid med Datarådgivning, *Mike Gillow*, Hjellesstad og Terminolprosjektet. Han utvikla saman med prosjektleiareren ein prototype av ein termdatabase som vart synt fram for oppdragsgjevar den 8.2.85. Seinare har prosjektet sjølv og *Ivar Utne* ved Nordisk institutt, Norsk termbank stått for fullstendig nyskriving, både av laste- og leseprogram, og har dessutan utført modifisering og utviding av den logiske strukturen i databasen. Dei programma som er skrivne mot basen, gjer det og mogleg å oppdatera innhaldet i basen frå skjerm. Men utviklinga av programma kom her i tida etter at den viktigaste redigeringsfasen var fullført. Redigeringsfunksjonen i baseprogramma vart difor ikkje teken fullt i bruk for Terminoltilfanget. Det er likevel nokså klårt at ein kunne ha effektivisert arbeidet og spart ein god del tid, dersom ein hadde hatt høve til å bruka databaseteknikkar heilt frå den fyrste ekserperingsfasen. I to nye prosjekt som Nordisk institutt går i gang med for Statoil, nyttar ein MDBS III frå fyrste stund. Det gjeld fornorsking av MESC (Material and Equipment Standards and Codes) og gjennomarbeiding av ein norsk petroleumstesaurus.

Strukturen i Terminoldatabasen er etter måten enkel. Han dekkjer ikkje alle dei måtane Termbankens formåt kan fyllast på, men er skreddarsydd for den bruken Terminolprosjektet gjer av termpostformatet. Det inneber ei forenkling i høve til det generelle formatet.

Til interesserte vert det her vist til vedlagde struktur-skisse over databasen. Kort kan ein seia at setta «system», «syserd», «sysnrd» inneheld kjedar med alle termomgrep og alle engelske og norske ord i termlinene. «sysdef» inneheld alle ord i definisjonar og kontekstar. «termcomp» inneheld kjedane med liner i kvar termpost, og «comperd/compnrd» inneheld kjedar med engelske og norske ord i termlinene. Tilsvarende gjeld for resten av nettverket. Tekst vert lagra med så lite redundans som råd. Alle ulike ord vert lagra ein gong, saman med opplysningar om kor ofte og i kva rekkjefylgje dei kan plukkast ut for å forma termliner, oppslag, definisjonar, tilvisingar.

Databasen inneheld til saman kring 10.000 hovudoppslag. I lasta form er databasen på 5.5 megabytes. Dei sekvensielle data er på vel 60.000 liner, og linene er gjennomsnittleg 25 teikn lange med koding.

I starten tok programmeringa mot MDBS III til med MS PASCAL. Etter ei tid tok ein til å bruka C. Grunnen var frå fyrst av at C vert levert med eit bibliotek med gode strengfunksjonar, og er særskilt MDBS-venleg på den måten at kalla til DML (DATA MANIPULATION LANGUAGE) lett let seg tilpasse kontrollstrukturen i C. Etter slutten



av mars 1985 skjer all programmering mot MDBS III ved Terminol-prosjektet i C.

Terminoldatabasen har vist seg særst nyttig, ikkje minst i samanheng med språkleg (terminologisk) gjennomarbeiding og retting av dei ferdig omsette handbøkene. Det har blitt meldt interesse for prosjektets materiale og terminologiske database, både frå andre delar av Universitetet og frå industrien. Universitetet arbeider med å få gjort termtilfang-et frå Terminol offentleg tilgjengeleg gjennom ein publikasjon snarast råd.



*Forsteamanuensis Per-Bjorn Pedersen er leder for Terminol-prosjektet ved Nordisk institutt, Universitetet i Bergen. Han arbeider bl.a. med utvikling av databaseverktøy for prosjektet. Sammen med Kåre Hegland har han utarbeidet prosjektrapporten «Petroleumsterminologi», Rogalandsforskning 1982.*

# Datamater/programmel/og programmer som del af interaktionen i undervisningen

*Erik Meistrup*

Ethvert teknisk orienteret undervisningssystem af en vis kompleksitet indeholder begrænsninger i underviserens frihed til frit at bestemme det faglige indholds udformning overfor den konkrete opgave og overfor de bestemte mennesker, der skal modtage undervisningen.

Man kan altså roligt vedgå, at enhver brug af datamater lægger nogle pædagogiske hindringer i vejen for læreren, men der stilles omvendt også en række nye muligheder til rådighed for undervisningssituationen.

Generelt må man sige, at det i høj grad er op til lærerens pædagogiske planlægning, hvordan datamater, programmel, programmer og anden teknik bliver udnyttet bedst muligt i forhold til de sociale betingelser for (ind)læringen. Informationsteknikken integreret i den almindelige undervisning kan altså efter placering og hensigt både have en hæmmende og en frigørende effekt i de forskellige undervisningssituationer.

## Forudsætning/mål/arbejdsbetingelser

Det er af yderste vigtighed, at indførelse og brug af ny teknik både på det overordnede plan (beslutningsplanet) og på det underordnede plan (udførelsesplanet) er underkastet følgende hovedpunkter:

### Forudsætning

1. at teknikken indgår i en overordnet filosofisk, politisk og pædagogisk holdning, der indeholder en videnskabelig teori. At individet og dets udvikling sættes i centrum.
2. at undervisningen ses både som en helhed og som en løbende proces. Planlægningen skal derfor tage udgangspunkt i sammenhængen og ikke i de enkelte dele.

### Målsætningen

Er en konfliktmodel, hvor individet skal være/bør blive selvstændigt og dermed

1. Problemformulerende
2. Løsningssøgende (kunne stille forslag)
3. Handlende

## **Arbejdsbetingelserne**

Er i grove termer:

1. Samarbejde! Socialudveksling med skiftende arbejdsformer men med vægten på gruppearbejde.
2. Procesarbejde! Både projekt- og problemarbejde ses som forløb i en livslang læring, der løbende medfører ændringer, og derfor har hovedvægten på arbejdsformer og metoder frem for fagligt indhold (ud over den basale indlæring).
3. Egne betingelser! Den vigtigste drivkraft i læringsprocessen er nysgerrigheden. Den må derfor stimuleres mest muligt.
4. Leg/eksperimenteren! Indebærer frihed til at (af)prøve og lære gennem forsøg.

Alle undervisningsaktiviteter har to yderpunkter, der kan formuleres således:

### **Aktiviteter, der øger specifikke færdigheder og viden**

Det er en problemløsende adfærd, hvor andre har defineret problemerne.

De vigtigste arbejds måder er: Øve og træne. Konkurrence mellem deltagerne spiller automatisk en væsentlig rolle.

Målsætningen bliver et enkelt og afsluttet forløb, hvor «fejl» skal undgås eller fjernes.

### **Aktiviteter, der øger generelle muligheder for den enkelte til at udvikle sig**

Det medfører en selvstændig problemformulerende og afprøvende adfærd, hvor det er modeller/strategier for løsninger, der er det væsentligste.

Arbejds måderne vil være: lege og eksperimentere. Samarbejdsformer vil være væsentligere end konkurrence.

Målsætningen er et åbent og uafsluttet forløb (selve procesarbejdet), hvor «fejl» også er brugbare resultater.

Undervisningen vil være en form for bølgebevægelse mellem disse ydre punkter, der dermed tilfører nye aspekter til hinanden. De væsentligste resultater af den samlede undervisningsaktivitet vil være:

En bevidsthedsudvidelse (dannelse af ny bevidsthed).

En blokeringsbearbejdelse (overvindelse af frygt/angst for det ukendte/nye).

En begrebs- og sprogudvidelse (dannelse af nye udtryk).

Disse faktorer er indbyrdes afhængige af hinandens forudsætninger.

### **Datamaternes placering**

Disse overordnede (og i denne sammentrængte form lidt luftige)

begreber, er det altså vigtigt hele tiden at vurdere i forhold til, når man indfører datamaterne i undervisningen, så teknikken ikke sluger perspektiverne.

Der sker noget specielt alene ved indførelsen af maskinerne (især datamaterne), idet undervisningsmaterialet – indholdet – vil blive beskrevet og opfattet via en skærm. Selve iagttagelsesfeltet vil blive begrænset til et lille lysende vindue, der kun repræsenterer noget andet. Dette andet er ofte en mere direkte fysisk kontakt med forskellige materialer (papir, skriveredskaber, bøger, instrumenter osv.), selvom disse igen er symboler for andre ting eller handlinger. Datamater fremmer også brugen af en formallogisk tænkning, der indebærer en nedprioritering af sanselige og følelsesmæssige erkendelsesmåder (disse må derfor styrkes/generhverves på andre felter).

Maskinerne er dermed endnu et lag i en fremmedgørelseseffekt – noget der også gælder alle andre former for teknik.

### **Resultater af datateknikken**

Når vores brug af værktøjer ændrer sig, så medfører det også ændringer i vores måde at tænke på. Det fremmer en bestemt tænkning<sup>1</sup>. Derved vil senere både vores sprogbrug<sup>2</sup> og handlingsmønstre også ændre sig. Disse ændringer vil også omfatte de personer, som ikke selv har personlig kontakt med den nye teknik. Tænk bare på, hvordan vi alle har vænnet os til at acceptere, at vi skal bruge et personnummer for at kunne opnå noget både hos de offentlige myndigheder og hos store dele af det private erhvervsliv. Vi er f.eks. blevet registrerbare og kontrollerbare numre hos banker og sparekasser, og det offentlige får automatisk oplysninger tilsendt om vores økonomi i forbindelse med skattefastansættelsen.

Vi er siden fjernsynets indførelse, og senere forstærket med brugen af video, blevet vænnet til at acceptere en af andre udvalgt og manipuleret virkelighedsgengivelse via en skærm. Denne accept gør det lettere for især unge at godkende et nyt teknisk medie som datamaskiner.

Selvom de specielle forhold omkring selve brugen af maskiner ikke skal berøres yderligere i denne sammenhæng, vil jeg godt støtte en tanke om, at skolen burde give mulighed for at lære at betjene og dermed gennemskue teknikken. Det kan ske i form af medieundervisning på medieværksteder (tv, video, bånd, dias osv.) og i informatikundervisningen (datamaternes rolle, også i forhold til de andre audiovisuelle medier).

### **Programmets placering i interaktionen**

Når man anskuer selve maskinlets rolle, så er dét programlet, der skal til for at få maskinerne til at yde noget, af yderste vigtighed. Historisk synes maskinerne og deres muligheder og begrænsninger at have været den vigtigste styring for udvikling af programlet. Det



gælder i de fire generationer af maskiner og tilsvarende fire perioder og niveauer af programmet. I øjeblikket arbejder man så primært i USA og Japan på femte generation: «kunstig intelligens», der kræver en parallel udvikling af de to sider for at kunne blive et virksomt produkt.

Et af de vigtigste forhold, hvor de teknisk avancerede samfund finder størst udnyttelse af datamaten og dets programmet, er muligheden for at kunne håndtere stadigt stigende informationsmængder<sup>3</sup>.

Disse stigende informationsmængder er både udtryk for en stigende samfundsmæssig kompleksitet og i sig selv en igangsætter af ønsker om at kunne vurdere og kontrollere stadigt nye områder.

Når vi får mulighed for adgang til stadigt flere informationer, kan vi opnå større indsigt og dermed sikre bedre beslutningsmåder. Dette gælder også, og måske især, uddannelsesstederne, og disse må derfor sikre sig forskellige former for programmet, der sætter dem i stand til let både at få adgang til oplysningerne og til at kunne lade dem indgå i undervisningen på brugernes (elevernes/studerendes) vilkår.

### **Programmerne udgør interaktionen**

Programmet skal altså sikre, at det er rimeligt enkelt at lave de nødvendige programmer og, at disse kan påvirkes (manipuleres) af brugerne. Denne påvirkning kan f.eks. ske ved at stille programmet direkte til rådighed for brugeren, noget der skulle være muligt – og ønskeligt – med overgangen fra programmeringssprogene til forfatter-systemerne.

Programmet skal altså opfylde det hovedsigte at skabe programmer (programudvikling), der skal kunne bruges af andre (programafvikling).

Hvis vi skal bestemme interaktionsgrundlaget i de to faser, så ser det således ud:

#### *Programudvikling:*

Forfatteren (subjekt)    program/maskine    brugeren (objekt)

#### *Programafvikling:*

Brugeren (subjekt)    program/maskine    forfatteren (objekt).

I denne skematiske model ses mellemeddet dvs. «program/maskine» som et sæt af pseudohandlinger, der udgør dialogperspektivet eller interaktionen. Denne model kan vi udvide til at gælde indlæringsforståelse dvs. undervisning som helhed (se figuren).

Afslutningsvis skal det påpeges, at i fremtiden bliver det vigtigt, at undervisningsverdenen på alle måder er med til at fremme så brede kreative arbejdsmåder som muligt for også derigennem at være med til at fremme en *teknologisk fantasi på et humanistisk grundlag*, der fremover kan give andet og bedre materiel/programmet, hvor brugeranvendeligheden er fremmet mest muligt.

PROGRAMUDVIKLING OG LÆRINGSFORSTÅELSE

FORHOLDET I UNDERVISNINGEN MELLEM:		FILOSOFISK/PÆDAGOGISK		BRUG AF DFU/DUS	MAGTRELATIONER
Lærer/System	Elev	GRUNDHOLDNING	MÅL	Undervisningsform	
Subjekt ->	Objekt	<u>MATERIALISME</u> Individerne er ved fødslen identiske. Det er omgivelserne, som former, og skal forme, dem.	Værdifulde Individer	Liniære og forgrenet undervisning. (Programmeret undervisning)  Klasse- og individuel arbejde (gruppearbejde).	Sofistikeret udformning kan udviske/tilsløre og fremmedgøre.  Autoritet = magt/ret.
Objekt <-	Subjekt	<u>IDEALISME</u> Individerne er ved fødslen forsynet med en grundmenneskelighed. Den skal de selv forme og udvikle.	Selvstændige individer.	Lukkede info-søgnings- og simuleringsprogrammer  Tema- og emnearbejde (gruppe/individuel/klasse).	Eleven vælger forløbet - læringen ligger i valgene.  Bestemmelsen ligger i programforfatterens strukturering (parameteropbygning).
Subjekt <->	Subjekt	<u>DIALEKTISK</u> Individerne har ved fødslen en vist grundberedskab. Dette skal formes gennem deres interaktion med omverdenen.	Selvstændige værdibevidste individer - en bedre omverden.	Åbne/dynamiske database- og simuleringsprogrammer (ekspertsystemer/kunstig intelligens).  Projektarbejde og problemløsning.	Erkendelse->Ret->Autoritet  Åbne -> mulighed for fortløbende ændringer via ny viden og dialog.  (Udbygget demokrati).

Note: Skemaet er en udvidet version af oplæg af Peter Bøgh Andersen. Humanistisk Datalog, Århus Universitet

## Noter

### 1. Tænkning

Menneske – maskine – interaktion medfører det, man kalder «procedural tænkning». Denne tænkning fordrer præcis viden om det sæt af instruktioner, der skal til for at styre maskinen/programmet. Siden 1700-tallet har denne tænkemåde vundet stigende indpas, da den passer sammen med den industrielle produktionsmåde. Brugen af kunstig intelligens vil ikke ændre på dette, snarere forstærke tendensen.

Man kan kort definere procedural tænkning som en problemløsning foretaget via et sæt af dellosninger.

Der er også andre tænke måder som, selvom de bliver «undertrykt» af den procedurale alligevel stadig er gældende og vigtige. Det gælder f.eks. den paradimatiske eller prototypiske, hvor problemløsningen sker ud fra erfaringer og eksempler. Det betyder, at et nutidigt problem ansues ud fra fortidens erfaringer, hvorefter beslutninger/ønsker om fremtiden træffes.

### 2. Sproget

Da både maskinel og programmel hovedsagelig er udenlandsk (alt overvejende amerikansk) ser vi alene deraf en sproglig ændring, der vil fortsætte. Det er f.eks. blevet almindeligt også i ikke datamatiserede sammenhænge at tale om «input» og «output».

Sproget ændres også derved, at man f.eks. i produktionsprocesser overgår til at tale om maskinlets måde at styre/overvåge produktionsprocessen på frem for at tale om selve produktets bearbejdning. På nuværende stadi er det hovedsagelig mennesker, der kender selve forarbejdningsfasen som noget konkret – og derfor kan forstå per fornemmelse – der betjener datamaterne. Hvad vil der ske med vores forståelse af omgivelserne, når det bliver mennesker uden «gammeldags» kendskab til arbejdet, der skal overvåge komplicerede og måske risikable produktionsprocesser?

Endelig ændres sproget, hvad angår omtalen af individerne i f.eks. produktionen. Disse mennesker og deres behov aflæses og opfattes som begreber eller ting. Tidsfaktorer, der indeholder helt forskellige oplevelsesmønstre, kan på skærmen aflæses som «intervaller», «forbrug», «akkorder» eller «friproduktions perioder» (altså pauser evt. med medmenneskelige kontakter). Tilsvarende kan man foretage forskellige former for medicinske dvs. lægelige undersøgelser direkte fra patient til datamat og evt. stille en endelig diagnose.

### 3. Data/information/viden

Data er en formaliseret repræsentation af kendsgerninger eller forestillinger på en sådan måde/form, at de kan videregives/omformes ved en proces.

Information er det særlige betydningsindhold, den enkelte tillægger data, de kan altså ikke omdannes ved edb.

Viden er helheder (gestalter) af forståelse, der indeholder mere end summen af sine enkelte elementer. For at kunne søge og behandle informationer må man derfor være i besiddelse af viden. Der er en viden – «den tavse viden» – som ikke er specificerbar og som f.eks. er tilstede ved mellemmenneskelige kontakter.

Det er viden, forståelse og overblik som sammenhængende dele, der kan stimulere fantasien og muliggøre fantasiernes omsættelse til handling.

*Erik Meistrup er ansatt ved «Danmarks EDB-Center for Forskning og Uddannelse», Århus Universitet, hvor han arbejder med datamaskin-formidlet undervisning. Han har skrevet en række artikler om dette emnet.*

# RAPPORTER

## Hva skjer innen humanistisk og samfunnsvitenskapelig forskning?

*Jostein H. Hauge*

Den humanistiske og samfunnsvitenskapelige forskningen i Norge er spredt på en lang rekke arbeidssteder og består av forskning innenfor mange ulike fagområder. Selv i et lite land som vårt er det vanskelig å skaffe oversikt over den forskning som pågår.

Slik er det både for forskerne, de forskningsplanleggende og forskningsfinansierende institusjoner og dem som vil formidle forskningen gjennom tidsskrifter, aviser, lyd- og bildemedier.

Som meldt i Humanistiske Data 1-84 har NAVF som et 5-års forsøksprogram opprettet NAVFs informasjonstjeneste for forskningsprosjekter i Bergen.

Informasjonstjenesten består av to fagtjenester for humanistisk og samfunnsvitenskapelig forskning og en felles, koordinerende edb-sentral. De to dokumentasjonstjenestene er opprettet av tre av de faglige råd i NAVF og har som hovedformål a) å dokumentere den forskning som NAVF finansierer og b) å kartlegge all humanistisk og samfunnsvitenskapelig forskning i vårt land.

I tillegg skal fagtjenestene (og den felles edb-sentralen) kunne påta seg dokumentasjonsoppgaver som betalte oppdrag. Før tiden slutføres oppdrag knyttet til dokumentasjon av forbruksforskning og idrettsforskning.

I løpet av det første driftsåret har informasjonstjenesten lagt ned mye arbeid i å utarbeide formålstjenlige skjema for prosjekt- og litteraturregistrering og å etablere praktiske samarbeidsordninger med distrikthøgskolene, universitetene m. fl.

Til i dag har fagtjenestene gitt ut oversikter over humanistisk og samfunnsvitenskapelig forskning finansiert av NAVF i 1984. Tilsvarende kataloger for 1985 er for tiden under utgivelse. I høst har den humanistiske fagtjenesten også gitt ut et eget prosjektheft over forskning knyttet til samisk og kvensk språk, historie og kultur.

I samarbeid med de forskningsutførende institusjonene holder fagtjenestene nå på med en heldekkende registrering av forskningen innenfor humanistiske fag og samfunnsvitenskapelige fag i 1985.



Universitetene og distriktshøgskolene vil i samarbeid med fagtjenestene registrere den forskning som pågår ved institusjonene. Dette materialet vil bli supplert gjennom en spørreskjemaundersøkelse blant ca. 600 institusjoner som kan tenkes å drive eller finansiere humanistisk og/eller samfunnsvitenskapelig forskning.

Som resultat av denne kartleggingen vil det foreligge relativt detaljerte opplysninger om flere tusen forskningsprosjekter rundt om i landet. Også den litteratur som prosjektene (foreløpig) har resultert i, blir registrert. Prosjektinformasjonen fra den heldekkende kartleggingen vil, i likhet med opplysningene om NAVF-forskningen, bli lagt til rette i en database til bruk for interesserte. For tiden arbeides det med å utarbeide informasjonsmateriell og brukerrettledninger. De institusjoner som ønsker å utgi separate, trykte kataloger over sin egen forskning vil få materiale fra databasen som grunnlag.

Informasjonstjenesten vil for sin egen del utgi en oversiktskatalog over hele materialet som en veiviser til databasen.

For at de bibliografiske opplysningene som samles inn skal kunne komme best mulig til nytte i bibliotekenes arbeid, er det opprettet kontakt med Norske avdeling ved Universitetsbiblioteket i Oslo med tanke på en regelmessig overføring av opplysninger dit.

Det er et omfattende og nitid arbeid å holde forskningsinformasjonen tilnærmet komplett og ajourført. Men målet med arbeidet er først nådd når informasjonen er kommet ut til de ulike brukergruppene i en meningsfull form. Derfor vil fagtjenestene satse på varierte formidlingsformer, også i form av sammenfattende oversyn over forskningen på ulike fagfelt, og tilby en rekke brukermiljøer prøveabonnement på

databasene i Bergen.

Nærmere opplysninger om informasjonstjenestens arbeid og datakilder kan en få ved henvendelse til:

konsulent Karin Nordenstam, tlf. (05) 21 29 58

Fagtjenesten for informasjon om humanistiske forskningsprosjekter  
Harald Hårfagresgt. 31, 5014 Bergen-Univ.

konsulent Helge Simon Møll, tlf. (05) 21 21 19

Fagtjenesten for informasjon om samfunnsvitenskapelige forskningsprosjekter

Hans Holmboesgt. 22, 5014 Bergen-Univ.

## **Informasjonsteknologisk satsning ved Universitetet i Oslo**

*Asbjørn Brændeland*

Det skjer store ting på IT-fronten ved landets universiteter og høyskoler – i alle fall når det gjelder planlegging og utredning. (Legg merke til at man nå snakker om IT – for InformasjonsTeknologi – og mindre om EDB eller Informatikk.) Et av de viktigste utspillene i 1985 kom fra Kuvås-utvalget – et NTNf-utvalg som på oppdrag fra Industridepartementet utarbeidet forslag til et satsningsprogram for IT i Norge. Forslaget omfatter bl.a. en økt utdanning av kandidater med IT-relevant kompetanse på universitetsnivå og en kraftig opprustning av utstyrsparken for forskning og utdanning. Den økonomiske rammen er på nesten 500 mill. kr pr. år i en 4-årsperiode.

Som et direkte svar på denne utfordringen, har man ved Universitetet i Oslo utarbeidet en egen handlingsplan for IT-utbyggingen. Planen, som omfatter anskaffelse av utstyr og tilføring av stillinger for forskning, undervisning, tjenester og drift, har en ramme på nærmere 700 mill. kr for perioden 1986-90 (ca. 137 mill. kr pr. år). I planen er det lagt stor vekt på en bred utbygging, både når det gjelder å skaffe forskere og studenter fra ulike fag tilgang på utstyr, og når det gjelder kompetanseoppbyggingen og økningen i IT-undervisningen ved de ulike fakultetene.

Et sentralt element i planen er ODIN-prosjektet («ODIN» står for «Oslouniversitetets Distribuerte InformasjonsNett»). Målsetningen for prosjektet er å bygge opp en informasjonsteknologisk infrastruktur ved UiO med mer eller mindre avanserte arbeidsplass-stasjoner, flerbruker-anlegg og spesialressurser i et omfattende nettverk.

Et annet interessant innslag i aktivitetene i Oslo er samarbeidsdrøftelsene med ND (Norsk Data). Direktør Rolf Skaar i ND foreslo i mars

overfor UiO et samarbeid om IT-utbyggingen ved universitetet og om utviklingen av IT-produkter og -konsepter spesielt rettet mot universitetsmiljøer. Forbildet for samarbeidet var ulike samarbeidsprosjekter mellom universiteter og dataindustrien bl.a. i USA. Skaar antydte en tidsramme på 3-5 år og en økonomisk ramme på ca. 100 mill. kr, der ND og UiO skulle finansiere hver sin 1/3 og gå sammen om å skaffe ekstern dekning for den siste 1/3. Drøftelsene mellom ND og UiO er ennå på et forberedende stadium, men bør – om de fører frem – kunne gi konkrete resultater alt fra begynnelsen av 1986.

Sett fra de humanistiske fag, må Kuvås-instillingen alene synes noe skuffende. Den betoner alt overveiende den type hard teknologisk virksomhet som tradisjonelt forbindes med edb, og har liten forståelse for betydningen av humanistisk kompetanse innenfor den framtidige informasjonsteknologi. Mer oppløftende er det derfor å lese de utredninger som kommer fra UiO, der bl.a. samfunnsfagene og de humanistiske fag har fått vesentlig bedre plass, både når det gjelder den tiltenkte ressursfordeling, og når det gjelder å dra nytte av ikke-realvitenskapelig kompetanse innenfor IT.

#### **Referanser**

Stortingsmelding 60, 66 og 83

NTNF: Nasjonalt IT-program – et forslag til satsningsprogram på informasjonsteknologi i Norge (Kuvås-innstillingen), Oslo april -85.

NAVFs utredningsinstitutt: Behovet for datautdanning, notat nr. 1/85 i serien utredninger om forskning og høyere utdanning, Oslo -85

Kollegiets informatikkutvalg, UiO: Instilling nr. 1, Oslo november -85

Kollegiets informatikkutvalg, UiO: Universitetet og informasjonsteknologien, et strateginotat og en dokumentsamling, Oslo juni -85

Det historisk-filosofiske fakultet, UiO: Elektronisk databehandling ved Det historisk-filosofiske fakultet, Oslo desember -83

Det historisk-filosofiske fakultet, UiO: Innstilling om grunnfag i EDB for humanister, Oslo mai -85

Det samfunnsvitenskapelige fakultet, UiO: Samfunnsinformatikk, innstilling om innhold og struktur i nytt studietilbud ved Det samfunnsvitenskapelige fakultet, Oslo november -84

Pressemelding fra UiO og ND: Norsk Data og Universitetet i Oslo planlegger omfattende teknologisk samarbeid, Oslo juli -85

*Asbjørn Brændeland er amanuensis ved HF-Data, Universitetet i Oslo.*



## Seminar om pedagogisk programvare

*Gunnar Thorvaldsen*

Kirke- og undervisningsdepartementet arrangerte 2. til 15. juni 1985 sitt andre seminar for forfattere av pedagogisk programvare. Seminaret ble holdt på Teleskolen i Grimstad, som byr på egnede lokaler i flotte omgivelser. Utgangspunktet er at utvalget av edb-program for norsk skole er lite, og kvaliteten vekslende. Det er – stort sett – enighet om at man ikke skal importere utenlandske skoleprogram, men satse på oppbygging av egne ting som er tuftet på verdigrunnet i norsk skole. Derimot synes det klart at vi har mye å lære fra utlandet mht. metoder for konstruksjon av pedagogisk programvare. Hensikten med Grimstad-seminaret er å få til en overføring av slik metodekompetanse til norske pedagoger.

Kursopplegget kan sammenfattes i begrepet «programdesign verksted». Det vil si at man la stor vekt på deltakernes aktivitet i gruppearbeid. Med utgangspunkt i de metoder som ble presentert i forelesningene, skulle hver gruppe utarbeide systemskisse og tegne skjermbilder for sin egen programvare-ide. Gruppene ble veiledet av lærere som hadde deltatt på fjorårets kurs. Vi hadde en stab av programmerere til disposisjon, og flere av gruppene fikk se prototyper av sine programutkast fungere på mikromaskinen før de reiste hjem. Deltakerne ble oppfordret til å søke økonomisk støtte fra Datasekretariatet i KUD for å videreutvikle sine program.

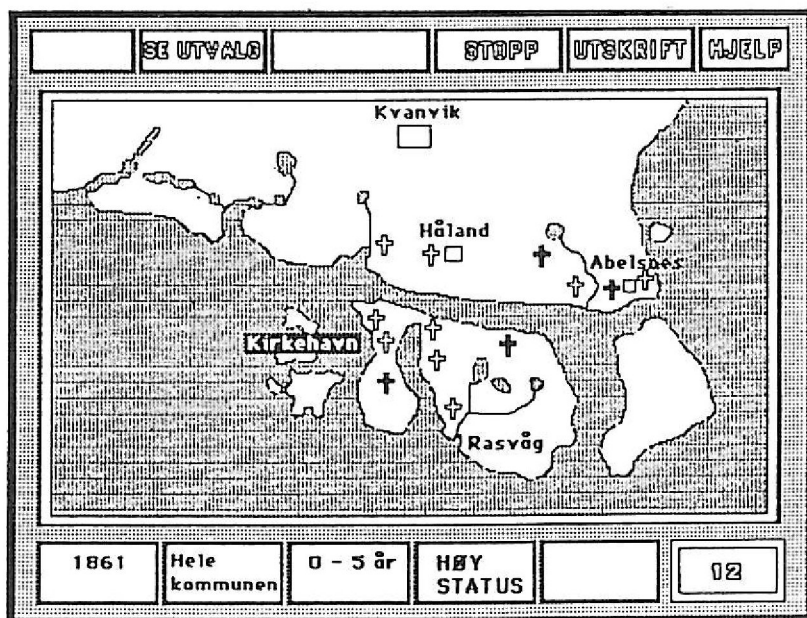
Hovedforelesere var kanadieren *Les Green* og *Dan Daniel* fra USA. Målet for deres metodeundervisning var todelt. For det ene skal programvaren aktivisere elevene. «What can the pupil do?», lød refrenget. Det andre punktet de terpet var design av skjermbildet. Man kan si at de her preket Aristoteles' gamle ide om enhet i tid, sted og rom. Elevene skal møte *ett* hovedbilde. Når innholdet på skjermen endrer seg, skal de ulike bildene være mindre endringer i dette hovedbildet. Etter endt sesjon skal så eleven forbinde det hun har lært med hovedskjermbildet og varianter over det.

Å ha et klart fokus, et punkt hvor man samler trådene, er et pedagogisk prinsipp med lang hevd. De fleste er da vel også enige om at det ikke er klokt å bruke dataterminalen til å presentere side etter side med tekstlig informasjon, det gjør man bedre i bokform. Metodisk sett har vi derfor mye å lære av de prinsipper for programdesign som ble presentert og som er utviklet i forbindelse med alle de spill man kan kjøpe til sin mikromaskin.

Faren ved dette er selvsagt at man tar med seg deler av innholdet i spillene. Visse temaer er enklere å presentere visuelt på en dataskjerm enn andre. Velger man så tema ut fra metodens fortrinn, vil man fort

kunne oppleve at nissen har blitt med på lasset. Altså at man ikke bare har importert metoder for programdesign, men fått med innholdet i de utenlandske undervisningsprogrammene også. I historiefaget opplevde vi at foreleserne var svært åpne for å simulere tenkte historiske forløp, dvs. la elevene «velge» historiens gang. Dette er ytterst betenkelig for en faghistoriker, særlig fordi det blir uklart hvilke forutsetninger som har reell historisk bakgrunn, hvilke programdesigneren har bygget inn, og hvilke som er valgt av eleven. Derfor var det skuffende at foreleserne var mindre åpne for programvare som gir elevene mulighet til å sette seg inn i konkrete historiske temaer. I det hele tatt var nok mange av deltakerne på kurset mer opptatt av innholdet i undervisningsprogrammene.

Flere av gruppene arbeidet med programvare for humanistiske skolefag. Bl.a. tok en gruppe for seg lese/skriveopplæring for svake førsteklassinger og en annen det nordiske samfunnet på Grønland i middelalderen. Til sist må jeg få nevne min egen gruppe med det tradisjonsrike navnet «Data Heimsins». Siktemålet med vår prototype er å gi elevene innsikt i årsakene til at folk døde i det gamle samfunn,



Copyright ©



samtidig som elevene lærer sitt nærmiljø bedre å kjenne. Illustrasjonen viser kartet over Hidra kommune på Sørlandet, hvor vi har plottet inn utvalgte dødsfall fra 1861. Elevene gis mulighet både for å registrere data, og for å «spille av» de demografiske hendelsene.

Grimstad-seminarene har skapt mange ideer til pedagogisk programvare. Mange søker nå KUD om støtte til videreutvikling av sine prototyper, og det blir spennende å se programmene etter hvert som de blir ferdige. Hvis Datasekretariatet også i framtida skal arrangere tilsvarende seminar, bør man i større grad ta sikte på å tilpasse metodene fra tidligere Grimstadkurs til norsk pensum og mønsterplan. Dessuten kunne seminaret vinne på at man tok for seg noen bestemte skolefag.

*Gunnar Thorvaldsen er daglig leder ved Registreringsentralen for historiske data i Troms.*

## **NORD IoD6**

**Helsingfors, 19.-22.8. 1985**

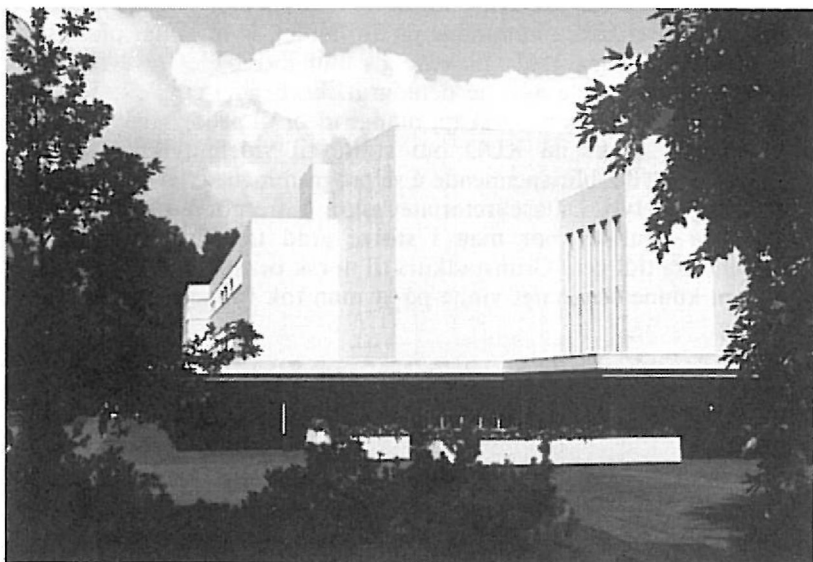
*Jostein H. Hauge*

De nordiske konferansene om informasjon og dokumentasjon (NORD IoD) kan føre sine tradisjoner tilbake til 1953. Etter en serie mindre arrangementer på 50- og 60-tallet begynte man fra 1970 av å arrangere større nordiske konferanser i et praktisk samarbeid med de enkelte lands IoD-foreninger (IoD = Informasjon og Dokumentasjon).

I Norge hadde Norsk Dokumentasjonsgruppe ansvaret for konferansen i Trondheim i 1982. Årets konferanse i Helsingfors samlet over 400 deltakere (ca. halvparten fra Finland og bl.a. 40 fra Norge). Den var arrangert av Samfundet för Informationstjänst i Finland. Nordisk samarbeidsorgan for vitenskapelig informasjon (NORDINFO) støttet arrangementet økonomisk. Arrangementsoppgavene ble usedvanlig dyktig løst av vertsskapsorganisasjonen innenfor rammen av det moderne og funksjonelle Finlandia-huset.

Temaet for årets konferanse var «Information Resources Management», av de faginnvidde oftest kalt IRM.

Omkring dette temaet var det organisert et 60-talls foredrag hvorav de fleste ble fremført i parallelle sesjoner. At informasjon er en ressurs, som også i stigende grad gir basis for dannelsen av kommersielle organisasjoner og firma, viste konferansen med all tydelighet.



*Finlandia-huset.*

På konferansen deltok ca. 30 utstillere fra nordiske og sentral-europeiske land. Her ble bl.a. utstyr og programvare for informasjonsbehandling vist. En rekke kommersielle informasjonsleverandører (information brokers) og vitenskapelige dokumentasjonsorganer demonstrerte sine tjenester og produkter.

Både utstillingen og programmet gav et helt klart inntrykk av at interaktive informasjonsdatabaser nå er på full fart inn overalt i biblioteks- og dokumentasjonssektoren og i et økende antall forskjellige former. Det ble også demonstrert et omfattende nordisk og internasjonalt samarbeid på feltet.

Nedenfor vil det i korte glimt bli gitt en omtale av enkelte av de tema som ble tatt opp på konferansen.

### **Information Resources Management**

I innledningsforedragene første dag ble begrepet Information Resources Management (IRM) behandlet fra en rekke synsvinkler.

*Forest Woody Horton* fra et konsulentfirma i Washington redegjorde i sitt gjesteforedrag for bruken av IRM i USA. Gjennom en serie studier i offentlige og private institusjoner på 60- og 70-tallet i USA ble det klart at de tradisjonelle informasjonsoppgaver holder på å endre karakter på en gjennomgripende måte. Det konvensjonelle informasjonsarbeidet (motta – lagre – spre informasjon) må suppleres med nye tjenester for *informasjons-forvaltning*.

Informasjon er i ferd med å bli anerkjent som en sentral ressurs også i bedriftslivet. Dette viser seg bl.a. ved at en serie større bedrifter tillegger den sentrale informasjonsforvalter (chief information officer) en viktig stabsfunksjon på linje med økonomidirektører osv. På samme måte som vi trenger «computer literacy», må vi også ha «information literacy», dvs. en storstilt satsing på opplæring i bruk av informasjon og ikke minst kyndighet i samordning av de ulike informasjonsressursene som finnes, slik at de kan brukes ved nyskaping av produkter og tjenester.

For å kunne føre an i denne utviklingen må de tradisjonelle biblioteks- og informasjonsmedarbeidere utvide sine kunnskaper med viten fra informatikk, kostnadsanalyse, personalforvaltning m.v.

### **Information Literacy**

Behovet for å dyktiggjøre den oppvoksende generasjon i informasjonsbehandling var også temaet for foredraget for *Jón Erlendsson* fra det islandske forskningsrådet. I mange land satser man nå på store undervisningsprogrammer i edb-teknologi, men fremover må det satses like mye på å dyktiggjøre brede lag av befolkningen i å utnytte de edb-baserte informasjonsproduktene som vil prege det kommende samfunnet. Et viktig sted å begynne er i skoleverket, også på lavere trinn, men dette vil stille krav om undervisningsopplegg som på en pedagogisk gjennomtenkt og ikke-triviell måte tar i bruk informasjonsteknologien. Sagt på en annen måte: Å lære elevene oppslag i interaktive databaser er ikke tilstrekkelig.

### **Databasene må forandres – ikke brukerne**

*C.N. Jansen van Rosendaal* fra Direktorat B i EF (Information Market and Innovation Commission) sa i sitt foredrag at EF har sin hovedinteresse innenfor feltet spesialisert informasjon slik den defineres av fagfolk i industri, handel og forskning. Det finnes i dag over 1000 databaser av denne type i EF-landene. Typisk for situasjonen i dag er at databasene er beregnet på profesjonelle informasjonssøkere og ikke på de periodiske brukere uten spesielle forkunnskaper/trening. Systemene er rett ut sagt så komplekse at den sporadiske bruker ikke klarer å nyttiggjøre seg dem på en siktemessig måte. Derfor må det gjøres kraftanstrengelser for i første rekke å forbedre *informasjonssystemene* – ikke brukerne – for dem er det ikke noe i veien med, sa foredragsholderen. Det er altså systemene som må tilpasses behovene og ikke omvendt.

I dag må kundene vite hva de vil ha og hvor de skal lete etter opplysningene («the get principle»). I tiden som kommer må «the bring principle» settes i fokus, dvs. systemene skal veilede og komme brukerne i møte. Vi må ta som et faktum at brukerne ofte har relativt



*Sentrale personer bak NORD IoD6:*

*F.v. Helena Basilier, ordfører for Samfundet för Informationstjänst, Elin Törnudd, ordfører for arbeidsgruppen for NORD IoD6 og Harriet Lönnqvist, konferansens generalsekretær.*

vage forestillinger om hva de egentlig søker når de starter informasjonsinnhenting, mente Rosendaal.

Faren er også til stede for at systemene er for perfekte:

«Another consideration is whether the information profession is too 'purist' in its approach to its task and consequently in its training. This can manifest itself in the pursuit of completeness in for example the cataloguing and indexing of information. While this may appear laudable as an objective there is a danger of losing sight of the real requirement of the user. In addition, compared to the attitude of the commercial supplier of information handling software, the purist approach may prove too expensive to be acceptable in cost/benefit terms».

I planleggingen av neste generasjons informasjonssystemer må vi bli flinkere til å velge ut den informasjon brukerne trenger og mer aktivt sikte mot en samordning av ulike informasjonstjenester. Dette mente foredragsholderen var viktigere enn å skape systemer for elektronisk dokumentformidling (document delivery systems).

### **IRM i praksis**

I øvrige plenumsforedrag ble bl.a. IRM og den nasjonale informasjonspolitikk tatt opp av *Nils Gram*, Riksbibliotekstjenesten, Norge. IRM

står for idealet om en koordinert utnyttelse av en organisasjons interne og eksterne informasjonsressurser. Det gjenstår å se hvordan IRM blir som praktisk administrasjonsform i fremtiden. En koordinert og tverrfunksjonell utnyttelse av informasjon vil bl.a. forutsette en nær kontakt mellom fagfeltene informasjonsbehandling, telematikk og databehandling. Det blir også viktig å vurdere hvilken rolle de tradisjonelle informasjonsbehandlere kan få ved utformingen av de nye metoder og arbeidsmåter og som policy-skapere på ulike nivåer.

### **Får vi de databaser vi fortjener?**

IRM i går, i dag og i morgen var temaet for *Mats Lindquist*, Paralog, Stockholm. Han stilte det provokatoriske spørsmålet om vi har fått de databasene vi fortjener og om hvem det er som bestemmer hvilke databaser det er behov for.

Foredragsholderen så mange mangler ved dagens elektroniske informasjonskilder. De har bl.a. den ulempe at de bare er tekstbaserte – informasjon er tross alt langt mer enn tekst. De inkluderer også for mye tekst i forhold til brukernes behov.

Den sterke tro på PC-basert informasjonsverktøy ble kraftig imøtegått. I dag er de personlige datamaskiner nærmest for leketøy å regne, men de vil utvikle seg til kraftige arbeidsstasjoner med mulighet for bl.a. bildelagring og interaksjon med databaser ved hjelp av naturlig tale. Men en grunnleggende forutsetning for bruk i praktisk informasjonsarbeid er at de knyttes til effektive kommunikasjonsnett.

### **Skal informasjonsvaren ha en pris?**

Spørsmålet om prissetting og markedsføring av produkter og tjenester ble bl.a. tatt opp av *Karsten Weis*, Teknologisk institut, København. Mens spørsmålet i næringslivet gjelder hvor mye varen skal koste (og svaret her er ofte tilfeldig nok, ifølge foredragsholderen) er den vanlige varianten av spørsmålet innenfor BDI-sektoren: Skal varen i det hele tatt ha en pris? Foredragsholderens syn var klart nok. Informasjonstjenestene skal prissettes og får paradoksalt nok en høyere verdi utifra tankegangen om at det som er gratis, har liten verdi.

Mens folkebibliotekstjenestene i Danmark ennå stort sett er gratis, er forskningsbibliotekene på vei mot prissetting av visse tjenester. De moderne informasjonstjenestene rettet mot næringslivet er derimot avgiftsbelagte – og at det skal være så er hevet over diskusjon. Næringslivet betaler i dag ofte store summer for å få utført regelmessige litteraturovervåkingstjenester av private informasjonsforetak. Weis mente at det var liten fare for at prising ville føre til mindre etterspørsel etter informasjons-tjenestene. Men det forutsetter øket og mer moderne markedsføring av gamle og nye tjenester og utvikling av nye informasjonsprodukter. Gratisprinsippet søkes ofte beholdt på grunn av faglig og administrativ treghet og tilstivning, mente foredragsholderen.



Det er uråd å gi mer enn en antydning av de tema som ble tatt opp i de nærmere 50 foredrag som ble holdt i parallellsesjoner.

### **Nordisk databaserevy**

I en egen programsesjon kalt *Nordisk databaserevy* ble en serie nordiske databaser gjennomgått i rask progresjon. Ifølge en utredning fra SCANNET fins det i dag ca. 200 on-line databaser i de nordiske land. Som et resultat av denne kartleggingen utkommer «Nordisk databaseguide» ved årsskiftet 1985/86 som NORDINFO-publikasjon 9. På konferansen ble en rekke databaser knyttet til økonomiske, tekniske, naturvitenskapelige, medisinske og samfunnsvitenskapelige fagfelt gjennomgått.

De fleste databasene dokumenterer bøker og tidsskrifter. Men på konferansen kom det også frem en økende interesse for å dokumentere pågående forskning innenfor flere av de ovenfornevnte felt.

En stor gruppe foredrag tok opp ulike deltemaer knyttet til databasebruk. Flere la vekt på viktigheten av å planlegge nøye før man satte i gang med å etablere en ny database. Bl.a. må en ha detaljerte kunnskaper både om sluttbrukernes informasjonsbehov og forventninger. Det må også klarlegges hvilke yrkesroller de nye informasjonstjenestene skal sikte mot, og hvordan de kan tilpasses den vanlige brukers arbeidssett.

### **Fremtidens databaser**

De nye databaser gir også i stigende grad kildedata som et supplement til referanser til kildene. På hvilken måte forandrer dette brukerprofilen – skal sluttbrukerne selv i fremtiden være den primære anvendergruppe? I enkelte foredrag ble spørsmålet om å utvikle såkalte strategiske informasjonssystemer drøftet. Dette er en type beslutningsstøttesystemer (decision support systems). Karakteristisk for slike systemer er at de er langt mer dynamiske og skiftende enn de vanlige bibliografiske informasjonssystemer og at de trekker på kunnskapsstoff fra langt flere kildetyper enn de tradisjonelle.

Av særlig interesse for de humanistiske og historiske fag var redegjørelsen for forsøk ved Det kongelige bibliotek, København med å lagre håndskrifter, plakater, kort m.v. i digitalisert form på optisk plate.

### **Databaser i praktisk bruk**

For dem som arbeider praktisk med informasjonsoppgaver i større bedrifter og forskningsinstitusjoner, er det verdifullt å få presentert konkret bruk av databaser i en organisasjons daglige arbeid. Fra Statens tekniska forskningscentral i Tammerfors ble det rapportert om omfattende bruk av interaktive databaser (40 søkesystemer med flere

hundre enkeltbaser). Siden 1980 har det foregått en kraftig desentralisering av søketjenesten i institusjonen, bl.a. slik at informasjonsarbeidet er flyttet helt ut i laboratoriene, hvor også forskerne selv tar del i informasjonssøkingen. Tidligere var dette en spesialisert arbeidsoppgave lagt til informasjonsavdelingen.

Økende krav hos sluttbrukerne til utformingen og bruk av de produkter som databasesøkingen gir, kom også frem i flere foredrag. Fra Kungliga Tekniska Högskolan i Stockholm ble det redegjort for hvordan mange brukere i dag ønsker å opprette egne spesialiserte databaser på sin mikromaskin av det materialet som de har funnet i de større databasene.

For at databasene skal få den bruk de fortjener, må også programvaren videreutvikles i betydelig grad. I dag føles både opp- og nedkøplingsprosedyrene og selve søkingen som altfor besværlig. Således ble nye krav til programverktøy tatt opp i flere foredrag. Fra Volvo, Göteborg ble det bl.a. omtalt programverktøy som også gir brukerne adgang til å tekstbehandle resultatene fra søkingen slik at opplysningene kan få en tjenlig form for sluttbrukeren.

### **Er informasjonssystemene verd prisen?**

I samme grad som informasjonssystemene blir mer anerkjente ressurser i private og offentlige organer, blir det også nødvendig med metoder og teknikker for evaluering av biblioteks- og informasjonssystemer. Sentrale begreper er her «cost effectiveness», dvs. hvor kostnadseffektivt de ønskede mål for virksomheten oppnås, og «cost benefit», dvs. om systemet fyller de tiltenkte funksjoner (f.eks. i en organisasjon). Å definere presist mål og kostnader i informasjonsarbeid kan være vanskelig, ikke minst i det offentlige.

Hvordan en kan få de lavest mulig direkte kostnader ved informasjonssøking i databaser, var også temaet for flere innledninger. I dag kan en ofte finne at de samme databasene tilbys til ulike priser (f.eks. i USA og Europa). Men da er det lett å bli lurt – de lavere priser gir ikke alltid de billigste tjenester. Av faktorer som bør tas i betraktning, ble bl.a. nevnt kvaliteten på tjenestene, åpningstider for databasen, bruksvennligheten til systemene osv.

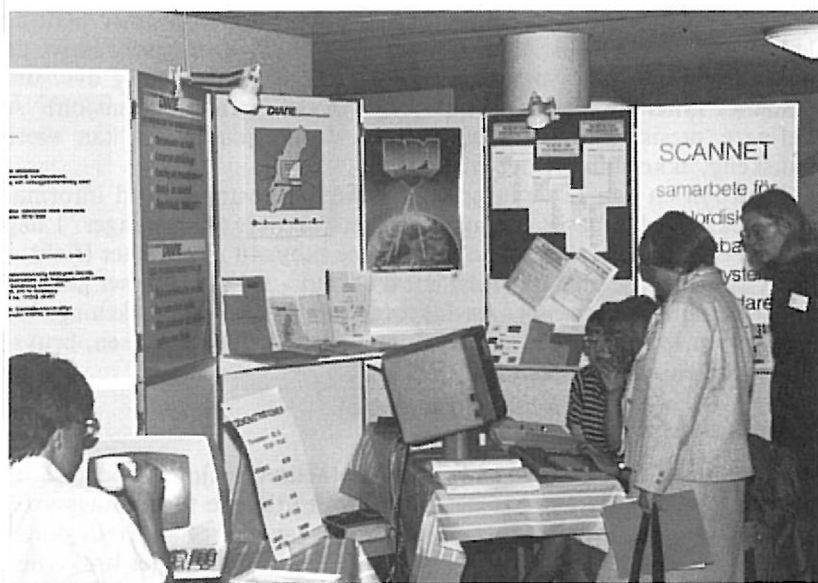
### **Infrastruktur blir viktig**

For å kunne utnytte det økende antall databaser nasjonalt, nordisk og globalt er det avgjørende at det finnes godt utbygde formidlingssystemer og koordinerende organer. Konferansen gav interessante redegjørelser for virksomheten til en serie organisasjoner som hjelper brukerne i utnyttelse av relevante kilder. F.eks. ble det orientert om DOCLINE ved Chalmers tekniska högskolas bibliotek, Göteborg som i dag knytter sammen over 75 skandinaviske bibliotek (de fleste svenske) innenfor

tekniske fag. DOCLINE er et databasert system for formidling av dokumentbestillinger fra nordiske bibliotek til British Library Lending Division. DOCLINE har i dag også noen nordiske leverandører (og forhandlinger føres med TIB Hannover (Technische Informationszentrum und Bibliothek)).

Innenfor MEDLARS-systemet ved Karolinska Institutets bibliotek och informationscentral, Stockholm gis brukerne tilgang til søking i alle de større medisinske databaser og som en ekstra service får man opplysninger om hvordan de relevante tidsskrift kan lokaliseres og bestilles fra ca. 400 nordiske bibliotek.

I fremtiden vil de såkalte DIANE-sentra stå sentralt i arbeidet med å spre informasjon. De første DIANE-sentra for rådgivning, koordinering og brukeropplæring ble etablert innenfor EF i tilknytning til opprettelsen av EURONET. Det danske senteret startet sin virksomhet i 1981 med det formål å øke bruken av on-line tjenester i Danmark. I parentes kan det her bemerkes at det nylig er inngått en avtale mellom Norge og EF-kommisjonen om norsk deltakelse i EURONET og at spørsmålet om opprettelse av et norsk DIANE-senter er under utredning.



*Mange nye informasjonstjenester ble presentert på konferanseutstillingen på NORD 10D6. Her demonstreres DIANE-sentra og SCANNET-tjenestene.*

## En omfattende mønstring

Avslutningsvis må det sies at NORD IoD6 gav en omfattende mønstring av de nordiske aktiviteter innenfor BDI-området. Som ved alle større konferanser kunne trolig en god del av deltakerne også her oppleve en slags frustrasjon: De uinnvidde kan gjerne komme til å føle seg overveldet og forvirret av all den informasjon de får del i, mens de erfarne fagmedarbeidere gjerne føler at de får for lite tid til å sette seg skikkelig inn i nyutviklingen på sine felt. Dette var etter referentens mening særlig følbart under Nordisk databaserevy hvor orienteringen om de enkelte databasene nettopp passerte revy – og i altfor høyt tempo.

Men det som man ikke rakk innenfor programmet, ble det muligheter til å ta igjen utenfor de faglige rammer. Her gav ikke minst de sosiale arrangementer utmerket anledning til å opprette og styrke personlige og faglige bånd.

Den neste IoD-konferansen blir holdt i Danmark i 1989.

## International Conference on Data Bases in the Humanities and Social Sciences

Grinnel College, Iowa, June 22-24, 1985

*Jostein H. Hauge*

Det er ikke uvanlig at man tilkaller eksperter til åpningen av en faglig konferanse for å kaste glans over den innledende sesjon og få en – som oftest positiv – beskrivelse av virksomheten på feltet.

Til åpningen av denne konferansen hadde man fått prof. *Jean Claude Gardin* fra Ecole des Hautes Etudes en Sciences Sociales, Paris. Prof. Gardin har vært aktiv på feltet siden slutten av 50-årene og en del i Norge vil kjenne til hans tidligere arbeider med edb-tilpasset klassifisering av arkeologisk keramikk. Oppgaven som prof. Gardin hadde satt seg denne gangen, var å gi en faglig vurdering av utviklingen innenfor humanistisk databehandling – og den var meget nedslående.

Etter gjennomlesing av konferansesammendragene var han kommet til at det hadde skjedd en relativt liten faglig nyutvikling i de siste 20 år.

De aller fleste foredragene kretset om de samme faglige målsettinger som tidligere, konstaterte han, og bar bud om bruk av stort sett de samme metoder, selv om de teknologiske forutsetninger naturlig nok har endret seg radikalt.

Det har med andre ord ikke skjedd noe gjennombrudd for en ny og kvalitativt annerledes humanistisk databehandling. Metodebruken og programutviklingen bærer ikke preg av at humanistene har brakt sine særlige faglige forutsetninger inn i det nye feltet. Prof. Gardin tenkte her særlig på en kvalitativt annerledes databehandling der kognitiv viten, lingvistisk og semantisk kompetanse inngår sentralt i databehandlingen.

Prof. Gardins observasjoner var kanskje riktige når det gjaldt denne konferansen, men beskrivelsen kan virke urettferdig overfor feltet som sådant. Bl.a. fikk undertegnede anledning til i private samtaler å nyansere bildet av utviklingen ved å henvise til virksomheten f.eks. innenfor datalingvistik.

Databasekonferansen i Iowa hadde samlet inn på 200 deltakere fra 14 land. De fleste deltakerne hadde tilknytning til et humanistisk fagfelt. Det er derfor trolig riktig å si at som samfunnsvitenskapelig edb-konferanse har denne konferansen nå bare marginal interesse.

Emnemessig fremsto konferansen som et konglomerat der ytterpunktene kan illustreres ved henholdsvis et foredrag om studiet av arkeologiske funn i forbindelse med tyding av Indus-skrifter og et foredrag om metoder for studiet av valgstrategier i organisasjoner.

Dette viste seg også i organiseringen av sesjonene, der det ble nærmest umulig å skape tematisk sammenhengende sesjoner av de ca. 60 foredragene.

Begrepet *database* er gjennom årene blitt så depresisert at det er blitt nærmest meningsløst. I dag omfatter konferansen tiltak der forskningsmateriale, kildedata eller referanseopplysninger foreligger i en edb-form – og da faller lite edb-virksomhet utenfor den faglige ramme.

Men når dette er sagt, må det også understrekes at konferansen inneholdt en lang rekke interessante foredrag som også gir bud om nye utviklings- og interesseretninger. Nedenfor vil et utvalg av de humanistisk orienterte foredragene bli nærmere omtalt.

For undertegnede var det verdifullt å få innblikk i de interaktive bibliografiske tjenester som nå etableres også for de humanistiske fag. I flere foredrag ble det redegjort for hvordan humanistene kan gjøre seg nytte av databaser, som Modern Language Association (MLA) Bibliography og MARK og REMARK (retrospektive) databaser. MLA-bibliografien inneholder referanser til artikler fra ca. 3000 tidsskrifter og dessuten en rekke andre typer materiale. De to andre databasene inneholder referanser til ca. 6 millioner verk, katalogisert av Library of Congress. Nylig er også The Arts and Humanities Citation Index (A & HCI) gjort søkbar on-line. (Nærmere opplysninger om dette kan fås av referenten.)

Fra Frankrike ble det orientert om den store FRANCIS databasen (ca. 1 million innførsler) som drives av Télésystèmes i Paris. I subdatabasen FRANCIS-H er de fleste humanistiske fag representert og med sin brede bibliografiske dekning (artikler, bøker, avhandlinger osv.) er



denne databasen en sentral, men antakelig også utilstrekkelig brukt vitenkilde.

I mange land blir det arbeidet flittig med å skape store samlinger av kildedata for humanistisk forskning. Fra Rutgers University ble det meldt om fremdriften i arbeidet med å lage en internasjonal oversikt over edb-tilgjengelig tekstmateriale, særlig innenfor litteraturfagene. På spørsmål ble det sagt at kopi av registeret senere ville kunne distribueres på magnetbånd, og det vil dessuten være tilgjengelig som en on-line database.

Internasjonalt finnes det i dag en lang rekke datasamlinger for språkforskning. På konferansen ble det bl.a. orientert om den omfattende virksomheten ved Institut für deutsche Sprache i Mannheim, som omfatter en lang rekke tekst- og ordkorpora. Ett korpus er direkte innrettet mot å studere den ulike utviklingen av tysk språk i Øst- og Vest-Tyskland. Her ligger sikkert mange spennende oppgaver og venter.

Som tidligere var også arkiv- og gjenstandsfagene godt representert. Her ble det bl.a. orientert om arbeidet i det kanadiske arkiv- og museumsvesen med å skape «a global methodology that allows for the complete machine-readable description of its visual records». Arbeidet drives innenfor rammen av The Canadian Heritage Information Network (CHIN) som sikter mot å skape en omfattende sentral databank for offentlige museumssamlinger i Canada. Særlig har avdelingen for bildebehandling laget et system som nå vekker internasjonal oppmerksomhet. Nærmere opplysninger kan fås av referenten.

Bruk av edb til informasjonssøking i arkivsektoren er aktuelt i mange land, jf. virksomheten i vårt land innenfor automatisert arkivinformatjon (Senteret/Riksarkivet). utfordringene er metodisk stort sett de

samme i f.eks. USA, men at innsatsfeltet er langt større forstår man når man får vite at det finnes ca. 16.000 arkiver av ulike typer i USA. En første oppgave blir derfor å tilrettelegge oversiktsinformasjon over hva som finnes i de ulike samlingene. Å tilrettelegge selve kildematerialet for databehandling vil være en håpløs oppgave.

Japanerne er langt fremme på en rekke områder innenfor humanistisk databehandling, jf. Humanistiske Data nr. 3-84 side 63 til 69. På denne konferansen ble dette vist gjennom et foredrag om «An Archaeological Survey Database and the Development of its Application Software» (A. Oikawa). På en overbevisende måte demonstrerte foredragsholderen hvordan en kan utnytte metoder for datafangst i felten, databaseløsninger, grafiske teknikker og resultater fra satellittovervåking til å skape et integrert informasjonssystem for arbeidet med arkeologiske utgravninger i Japan (pr. 1984 ca. 8.000 ulike utgravningsfelt). Kopi av foredraget er tilgjengelig.

Til slutt kan det nevnes at det også på denne konferansen ble presentert informasjonstjenester som dokumenterer den forskningsvirksomheten som pågår. Her ble det orientert om større systemer i Canada (samfunnsvitenskapelig forskning), Frankrike (heldekkende forskningsdokumentasjon) og Norge (humanistisk og samfunnsvitenskapelig forskning).

## **Seventh International Conference on Computers and the Humanities (ICCH)**

**June 26-28, 1985, Provo, Utah, USA**

*Jostein H. Hauge*

Den syvende ICCH konferansen var lagt til Brigham Young University i Provo, en relativt liten by som i høy grad er preget av virksomheten ved dette kjente mormoner-universitetet. Arrangementet var tatt hånd om av Humanities Research Center under ledelse av direktør *Dr. Randall J. Jones*. I løpet av tre dager fikk ca. 350 deltakere anledning til å velge mellom godt over 100 foredrag som foregikk i 4 parallelle sesjoner.

Det settes undertiden (og med noen rett) spørsmålstejn med verdien av denne type arrangementer hvor ofte svært forskjellige emner presenteres på løpende bånd (dvs. hvert 30. minutt).

Når sammendragene av foredragene bare foreligger ved starten av konferansen, blir det også i høyeste grad et kappløp med tiden for i det



hele tatt å finne ut hvor en skal delta. Foredragstitlene kan være svært så dårlige veiledere i så henseende.

Allikevel forsvarer disse arrangementene sin plass. Ved sin størrelse og emnefordeling gir de mulighet til å blinke ut særlige tema dersom en har spesielle interesser. For den som vil skaffe seg oversikt over hva som holder på å skje innenfor (hovedsakelig amerikansk) humanistisk databehandling, gir det brede programmet utmerket anledning til det og ikke minst viktig: Gjennom personlige kontakter i pauser og fritid og ved studium av sammendragene i ro og mak etter konferansen, får en mulighet til faglig fordyping.

For denne anmelder er det også viktig at den dokumentasjon man har fått med seg, blir spredd så bredt som mulig. Leserne av Humanistiske Data gis herved anledning til å få utfyllende opplysninger om konferansen, kopi av sammendrag m.v.

Så over til selve konferansen.

For undertegnede, som ikke har vært til stede ved de to foregående arrangementene, var det særlig interessant å finne ut hva som er nytt og hva som er videreføring av gamle arbeidsformer innenfor humanistisk databehandling.

### **ICCH har endret profil**

Først en generell observasjon om konferansens innhold. ICCH er i ferd med å forandre karakter fra å være en generell edb-konferanse for de humanistiske fag til å bli en konferanse om databehandling i språk- og



*Brigham Young University.*

litteraturforskning. Foredrag om edb i historie og arkeologi, musikk, teologi, kunsthistorie, teatervitenskap osv. glimret stort sett med sitt fravær. Slik var det ikke før 1980.

Grunnen kan være at en rekke disipliner har fått sine mer spesialiserte konferanseserier slik som vi finner det i historie, museumsforskning, arkeologi o.fl. Datalingvistikk er nå også praktisk talt ute av programmet. Innenfor den gjenværende språklige delen er de filologiske prosjektene i flertall, dvs. virksomhet hvor datamaskinen tas i bruk for å hjelpe forskeren i løsningen av tradisjonelle (men også nye) filologiske arbeidsoppgaver.

I USA er det alt etablert en tradisjon for å foreta litterære og stilistiske analyser ved hjelp av datamaskin, og dette ble også gjenspeilet på denne konferansen.

### **Mikrodatamaskinene kommer**

Det var flere nye innslag på konferansen. For det første har den store spredningen av mikrodatamaskiner i de humanistiske akademiske miljøer endret forutsetningene radikalt for å utnytte datamaskinelle metoder. Tidligere var de store data-anlegg vanskelig tilgjengelige for store grupper innenfor humanistisk forskning i USA til forskjell fra situasjonen i Europa og for den del i Norge.

En rekke foredrag tok således for seg hvordan forskere kan ta i bruk standard programvare eller lage nye programmer for sitt faglige arbeid.



*Edb-konsulent Roald Skarsten og førstelektor Magnar Brekke, Universitetet i Bergen, studerer de siste fremskritt innenfor datastøttet oversettelse.*

I en rekke tilfeller var utgangspunktet de muligheter som mikromaskinene gir for datastøttet undervisning. En lang serie foredrag redegjorde for hvordan en hadde utviklet/fått utarbeidet program for datastøttet språkopplæring i både klassiske og moderne språk. Amerikanske universiteter og college legger også som man vet, stor vekt på skrivetrening (composition).

De datastøttede undervisningssystemenes mulighet for å individualisere, repetere, forgrene, visualisere og animere læringsstoffet gjør at svært mange av studentene er motiverte for denne type læring. Og resultatene av innlæringen er bedre enn hos kontrollgruppene, sies det. Om dette er en tendens som vil holde seg, blir spennende å se. Kanskje er det nyfikenheten som gir bedre motivasjon og dermed bedre innlæring?

### Nye teknologier

Bruk av ny teknologi ble også tatt opp i et par av foredragene. Ved University of Iowa foregår et prosjekt hvor videoplater med lysbilder tas i bruk i opplæringen i kunsthistorie i kombinasjon med i dette tilfellet en Prime minidatamaskin.

Fra vertsskapsinstitusjonen, Brigham Young University, ble det redegjort for et videoplateprosjekt knyttet til språkopplæring der videoplaten inneholder en spillefilm som er inndelt i sekvenser og gjort til basis for et undervisningssystem på mikrodatabasemaskin. Prosjektet



*Demonstrasjon av et interaktivt videoplatebasert undervisningsopplegg om nazismens fremvekst.*



*Et foredrag i humanistisk databehandling kan kreve mye utstyr. Her presenteres et prosjekt hvor videoplateteknikk brukes i språkopplæring.*

kan allerede melde om interessante innlæringsresultater, ikke minst når det gjelder idiomatiske uttrykksett.

Til forskjell fra i alle fall situasjonen i Norge har også optisk lesing (OCR) langt sterkere plass som teknikk for å overføre kildemateriale til datamaskinen. I flere foredrag ble det berettet om en stor og økende bruk av de etter hvert billigere Kurzweil-lesere. De brukes i dag i forbindelse med tilrettelegging av store ordboksarkiv, leksika og litterære tekstsamlinger. Det ble opplyst at leseren ved Brigham Young University ble kjørt gjennomsnittlig 18 timer pr. døgn med slike oppgaver (jf. separat rapport om «The Humanities Research Center» ved Brigham Young University).

Også på denne konferansen ble det holdt en serie foredrag om behovet for bedre programvarer for de humanistiske fag i forbindelse med undervisning og forskning. Behovet for opplæring av forskere i moderne former for tekstfremstilling ble også omtalt. Her var det bl.a. interessant å merke seg de initiativ som National Endowment for the Humanities har tatt for å gjøre humanistiske forskere fortrolig med moderne publiseringsmetoder.

Konferansen var lagt til Brigham Young University Center, et moderne og funksjonelt konferanse-senter hvor det meste fantes unnatt kaffe (av religiøse grunner), og fordi det ikke var godt for oss, som kantinebetjeningen så velmenende belærte oss.

## **The Humanities Research Center Brigham Young University**

The Humanities Research Center provides the technological tools, resources and support to foster high-quality research and scholarship in the College of Humanities at Brigham Young University.

### **IBM 370/138 Mainframe Computer**

A mainframe computer provides the extended resources of data storage and computation capacity necessary for large-scale projects and databases.

*Concordances.* The HRC maintains two general-purpose concordance generating software packages on the IBM 370 computer (WATCON and ARRAS), as well as specialized concordance software for particular projects. Several major projects are in various stages of completion including concordances in English, German, Portuguese, French, Finnish and Chinese. Concordance software is also used as a tool in such areas as language acquisition, authorship and readability studies and proofreading.

*Language Databases.* The HRC has three general text databases for use in studying grammar and usage: the Brown Corpus of American English, the Limas Corpus of German, and the Weidner Corpus of Portuguese. A comparable corpus of Spanish is being planned. We also have the TLG collection of pre-Christian Greek texts and some ancient Latin texts of interest to scholars in Humanities and Classics.

*LDS Databases.* The HRC has a complete English Scriptures database, updated with the new chapter/section headings and other changes in the 1981 editions. We have recently scanned the English Topical Guide and the complete Spanish Scriptures in a joint effort with the LDS Translation Department to produce a Spanish Topical Guide that will soon be ready for publication.

*Dictionaries.* Dictionaries of South American Indian languages and a Serbo-Croatian Verb dictionary are being developed on the IBM 370 computer. In addition, the Longman Dictionary of Contemporary English is available on the system and dictionaries and word lists in Portuguese and Finnish are being scanned. There are plans to acquire the Oxford English Dictionary in computer form as it becomes available and a project to scan the 1755 and the 1773 editions of the Samuel Johnson Dictionary is in the initial stages.

### **Kurzweil Optical Scanner**

A mini-computer with optical character recognition capability manufactured by Kurzweil Computer Products (a Xerox subsidiary) is the «front-end» processor for a large portion of the computing projects in

the College of Humanities. Since many faculty members involved in computing need to work with large data bases, a method faster than manual input was necessary for their creation. With the optical scanner, data bases can now be created six to twelve times faster than before.

*Scanner Capabilities.* The scanner can be taught to recognize the alphabets of virtually all languages except Arabic, Chinese, Japanese and Korean. It can scan both proportionally (typeset) or uniformly (typed) spaced material. Font changes, such as italics and bold, and point size changes are handled easily. Once the scanner has been trained to recognize a particular type style, that training can be saved and used in the future on a document of a similar type style.

*Scanner Projects.* Over the past two years since the acquisition of the scanner, over a million words have been scanned for professors with the college and for other departments and outside agencies:

- Complete works of the German author Goethe.
- LDS Topical Guide and Spanish Scriptures (for LDS Translation Services in the creation of a Spanish Topical Guide.)
- LDS missionary folklore.
- Portuguese frequency word lists.
- German word lists.
- a Finnish-English dictionary to be used in teaching students to conjugate Finnish verbs.
- French-English translation database for translation research.
- Spanish Attrition study data.
- Several PhD dissertations and other manuscripts for further editing and publication.

## **Humanities Learning Resource Center**

The HLRC is designed to provide instructional support to students and faculty members in the College of Humanities. In the Center are facilities that allow for individual as well as group study. Approximately six to eight hundred students use the facility each day.

Since a large proportion of the students in the College are in language programs, both second/foreign and first language classes, a great deal of emphasis has been focused on support for these programs. There are facilities that allow students to listen to language tapes both on an independent basis or with a teacher monitoring their performance. In addition to being able to listen to tapes in the Center, copies of these programs are made available to students to use outside of the Center (depending upon copyright restrictions).

In addition to audio, or listening capabilities, the Center has various types of video equipment. This equipment is used by students from all disciplines within the College. Several video monitors are connected to the campus-wide TV cable network, allowing professors to request a particular program in the HLRC without having to spend extra in-class

time. There are also video recorders available where students are able to control the video type, replaying portions as necessary. Other video capabilities include slide/sound and filmstrip/sound projectors.

A rather new, yet very important, academic program at BYU is the Spanish interpretation and translation major. The HLRC has two simultaneous interpretation stations that simulate as nearly as possible the actual live interpretation experience. We are able to train students to do the kinds of translation and interpretation activities that are required at the L.D.S. General Conference and at the United Nations.

In order to assist in the preparation of language tapes, in the analysis of acoustic phonetics and in other teaching and research related concerns, a state-of-the-art recording lab has recently been installed.

Another important function of the HLRC is to assist in the production of teaching and research materials. The recording lab in the Center is equipped with machines capable of providing this service, including signal mixers, amplifiers, speech analysis equipment, isolation booth, etc.

Since language training is becoming increasingly linked with computer technology, a section of the Humanities Learning Resource Center is devoted to humanities computing. Our micro-computer lab is constantly in use by students doing, among other things, computer-aided translation, foreign language word processing and interactive computer/video lesson development.

### **Microcomputers in the Humanities Research Center**

*Humanities Learning Resource Center Computer Lab.* The HLRC computer lab provides computer support for both faculty and students who use computers in research and assignments. The lab has ten Apple II+ microcomputers, four IBM Personal Computers, one IBM 2378 terminal and two printers. The installation of a Corvus Omninet system will allow three of the IBM computers to share a 20-megabyte hard disk and the Epson dot matrix printer. The center provides training in use of the computers and software. Students use the lab for word processing, computer assisted instruction, statistical analysis, and programming. Lab assistants are on duty to help students with problems. Currently over 300 students are signed up for use of the lab facilities.

*Microcomputer support in the HRC.* The professional faculty of the HRC are available to the college faculty for consultation on microcomputer purchase, training and programming assistance. There are faculty members currently using the following computers: the IBM Personal Computer, the IBM PC Jr., the Apple II+, the Apple //e, and the Apple Macintosh. The HRC personnel help the faculty with the programming and operation of various programs on these computers. Foreign language display and printing are ongoing concerns of the Center.



## Computer-assisted Language Instruction

During the past few years Brigham Young University has become a world leader in the area of computer-assisted language instruction. Extensive courses have been written for French, German, Spanish and English as a Second Language. Courses are currently being written for Dutch, Italian, Japanese and Swedish. In addition, computerized language tests have also been written. Beginning in March, 1985, BYU will begin using the modern Micro TICCIT system.

One of the most exciting areas of computer-assisted language instruction now being developed at BYU is in interactive videodisc. Under the direction of the new organization CALICO (Computer-Assisted Language Instruction Consortium), the Humanities Research Center is assisting in developing interactive lessons. The computer not only presents foreign language text material, but also video and audio material from the videodisc. An authentic context for the language is thus created, and the student has the feeling that he is actually experiencing the foreign culture. At the present time we are working in German, Hebrew, Korean and Spanish.

## Edb og lokalhistorie

### Referat frå eit kurs skipa til av NLI

*Johannes Kvestad*

Den første halvdel av 80-åra vil truleg av framtidige historiegranskarar bli rekna som tida då datateknikken slo gjennom for alvor; tida då alle skulle læra seg å bruka den nye teknikken; tida då ansvarsmedvitne fædrar kjøpte heimedatautstyr til tusenvis av kroner for å sikra seg at avkomet ikkje skulle falla av karrierestigen; tida då alle lærarar skulle læra seg den nye teknikken i løpet av nokre kurs osv. osv.

Mykje tyder på at andre halvdel av 80-åra vil bli prega av ei langt meir nøktern holdning til den nye teknikken. Mange heimedata-anlegg skal visstnok alt stå og støva ned i ein krok; det var ikkje så nyttig likevel å halda styr på innhaldet i fryseboksen pr. data. På den andre sida kan det sjå ut for at dei som verkeleg ser seg nytte i det nye, no ser ut for å vakna. Blant desse er dei mange lokalhistorikarane.

Det er alt over 5 år sidan ein del lokalhistorikarar tok til å sjå nærare på den nye teknikken, og det var i første rekkje tekstbehandlinga som verka tillokkande. Etter kvart er det ikkje så få som sit og arbeider med

ei eller anna form for tekstbehandling, og nokre få har også teke i bruk datateknikken på andre felt. Men til denne tid har denne gruppa vore lita. Det store fleirtalet har hatt liten kontakt med det nye. Og mange har nok følt at dei trong meir informasjon og ikkje minst råd om kva dei skulle gjera.

Det var såleis tydeleg eit populært tiltak Norsk Lokalhistorisk Institutt (NLI) gjorde då dei skipa til eit «Seminar om edb og lokalhistorie» på Skiphelle utanfor Drøbak 19.-21. august i år.

Godt og vel 50 deltakarar var det på seminaret, visstnok langt fleire enn det var tenkt, og dei fordelte seg godt ut over landet frå Finnmark i nord til Agder i sør. Kurset var delt på fem emne: 1) Tekstbehandling, 2) Kva edb-hjelp kan lokalhistorikaren få frå sentrale institusjonar, 3) Kva kan lokalhistorikaren gjera sjølv med mikromaskinen?, 4) Kobling/lenking av data frå ulike kjelder og 5) Program.

Ein viktig del under emnet tekstbehandling var at dei deltakarane som ikkje hadde vore borti «det nye» før, skulle få prøva seg litt i praksis. Frå ein forhandlar (Scanvest) og frå Statens datasentral vart det stilt ein del maskinar til rådvelde, og det vart halde eit «minikurs» for dei uinnvigde. Sidan det var underskrivne som var kursleiar, skal eg avstå frå andre kommentarar enn at sjeldan har vel så mange prestert så mange rare tastetrykk på så få maskinar.

«Frå tekstbehandling til bok» er eit sentralt emne for mange lokalhistorikarar. *John Aage Gjestrum* frå Toten bygdeboknemnd la her fram røynsler frå arbeidet med trykkinga av band 4 av Toten-boka. Dei hadde skrive manus inn på eit tekstbehandlingsanlegg og teksten vart så vidare overført til trykkeriet pr. elektronikk. Gjestrum vart supplert av *Rolf Eriksen* og *Terje Huuse* frå Centraltrykkeriet og deltakarane fekk såleis mykje nyttig informasjon med seg omkring dette sentrale emnet. For oss som har baska på med dette med overføring av tekst, var det befriande å høyra kor mykje det hadde skjedd på dette feltet dei siste åra. Om ikkje så alt for lang tid vil dette vera reine rutinen.

*Ståle Dyrvik* tok for seg sjølv arbeidsmetoden som er knytt til bruk av tekstbehandling - korleis teknikken påverkar oss. Edb-teknikken kan verka imponerande: me kan skyta inn tekst og me kan retta etter kvart og me kan heile tida sitja med ein oppdatert versjon. På den andre sida kan me missa ein del av oversynet me hadde då me sat med manusbunken. Skjermen gjev oss berre ein liten glytt inn i manus. Han kom deretter inn på ulike måtar å arbeida med eit bygdebokverk på og korleis dette vert påverka av edb.

Neste emne var kva hjelp me kan venta oss frå sentrale institusjonar og *Gunnar Thorvaldsen* orienterte om arbeidet til Registreringssentralen for historiske data (RHD) der dei i fleire år har arbeidd med å registrera kykjeboksmateriale og folketeljingar på edb. Arbeidet er i gang, men med den noverande kapasiteten ville det ta 40 år før landet er dekkja. Det skal såleis ikkje undra nokon at det vart spørsmål om

samarbeid mellom sentralen og meir lokale prosjekt. Thorvaldsen peika òg på ein del andre viktige databaser som kan vera til nytte for lokalhistorikarar. Han peika serleg på basen til Norsk Samfunnsvitskapleg Datateneste (NSD) som inneheld store datamengder men som ikkje er serleg innretta på lokalhistorikaren og evt. andre som er interesserte i mange ulike datatypar frå eitt område.

*Ivar Fønnes* frå Riksarkivet orienterte om arbeidet som f.t. er i gang i arkivverket og om planane for publikumstenesta. Etter hans meining kom dei sentrale miljøa til å bidra med *data, program og konsulenttenester*.

Dei to neste emna gav deltakarane døme frå ulike arbeidsfelt der edb var teken i bruk innan lokalhistorisk arbeid, alt frå datamaskinell behandling av skiftetilfang til oppbygging av lokalhistoriske databasar. For dei som har drive litt med gards- og ættesoger var det interessant å sjå at databaseprogrammet DBASE II var teken i bruk slik at kvar person fekk sin post.

Siste emnet var spørsmålet om kva program som skulle nyttast. Gunnar Thorvaldsen informerte om ulike programpakkar som er i handelen. Serleg gjekk han inn på dei såkalla integrerte programpakkanane som skal innehalda både tekstbehandling, database og rekneark. Konklusjonen var vel at det enno er langt att før ein kan kasta spesialprogramma på dør og erstatta dei med eitt program som greier alt. Også det gamle spørsmålet om den som skal nytta edb bør kunna laga sine egne program eller om me kan satsa på standardprogram vart teke opp av Ståle Dyrvik. Konklusjonen var vel at i mange høve vil ein ha nytte av å kunna laga sine egne program, – standardprogramma vil ikkje kunna dekkja alle eventualitetar like godt.

Kva kom ut av seminaret? Truleg var utbyttet varierende alt etter kva føresetnader ein møtte med, men det er grunn til å tru at dei fleste visste meir då dei reiste frå seminaret enn då dei kom. Eit spørsmål som mange var opptekne av var kva maskintype dei burde satsa på. Nokon konklusjon vart ikkje gjeven på seminaret, men mann og mann imellom var vel svaret at ein burde halda seg innanfor standardane (MS-DOS eller CP/M) for å vera sikker på at ein har tilgang på program og at eventuell overføring av tekst skal gå greit. Heilt konkret vart det frå NLI signalisert at dei ville setja ned eit utval som skal ta seg av spørsmålet om edb og lokalhistorie, noko som er i ferd med å skje.

*Johannes Kvestad er styrar ved Voss Folkemuseum.*

# Humaniora-uken ved Universitetet i Bergen

## Humaniora i dataalderen

*Kristin Natvig*

Den første «humaniora-uke» ved Universitetet i Bergen gikk av stabelen i tida 22.-26. september. Hovedtemaet for uken, «tradisjon og fornying», ble belyst ut fra et stort antall felles- og instituttarrangementer under hovedtitlene «Humaniora mot tusenårsskiftet», «Arbeid og arbeidsmarked», «Dannelse» og «Tradisjon». Programmet var svært variert men så tettpakket at det var umulig å få med seg alt av interesse.

Denne rapporten skal konsentreres om sesjonen «Humaniora i dataalderen», som bestod av tre foredrag og et delvis parallelt «edb-verksted» med ni ulike demonstrasjoner.

Først ute blant foredragsholderne var førsteamanuensis *Jan Oldervoll* ved Historisk institutt. Oldervoll skisserte utviklingen hittil i bruk av edb ved Det historisk-filosofiske fakultet og presenterte noen framtidsperspektiver på humanistiske forskeres rolle i forhold til edb-teknologi.

Det er bare en liten del av den historiske forskningen som drar nytte av edb, ifølge Oldervoll. Et generelt trekk ved HF-fakultetet har vært at de forskere som har tilegnet seg kunnskaper i databehandling, fungerer som «orakler» for ansatte ved andre institutter, t.o.m. ved andre universiteter. Økningen i bruk av tekstbehandling de siste årene har imidlertid lettet overgangen til edb-bruk for forskningsformål. Det er dessuten blitt betraktelig lettere å lære edb siden mikromaskinenes inntog og p.g.a. en økende grad av brukervennlighet (i form av f.eks. menyer og «mus»). Tilbudet på kurs i regi av Edb-seksjonen ved HF-fakultetet har også virket ansporende på mange forskere. Alt dette har bidratt til at humanistiske forskere generelt er blitt mer aktive edb-brukere.

Inntil nylig har forholdet mellom de humanistiske fagene og edb-industrien vært ensidig: industrien har levert utstyr og til dels programvare til humanister. Disse har altså kun fungert som teknologikjøpere. Utviklingen er begynt å gå i motsatt retning nå. Dels er humanister blitt en pressgruppe på industrien (de krever f.eks. brukerdokumentasjon på norsk), dels er de selv blitt kunnskapsleverandører. Konkrete eksempler på det siste er utvikling av håndbøker og maskinlagrede hjelpeordlister.

En målsetning for dagens dataindustri er at maskinene skal bli mest mulig «menneskelige», f.eks. at de kan forstå naturlig tale. Humanistisk kunnskap blir derfor stadig mer interessant for denne industrien, avsluttet Oldervoll.

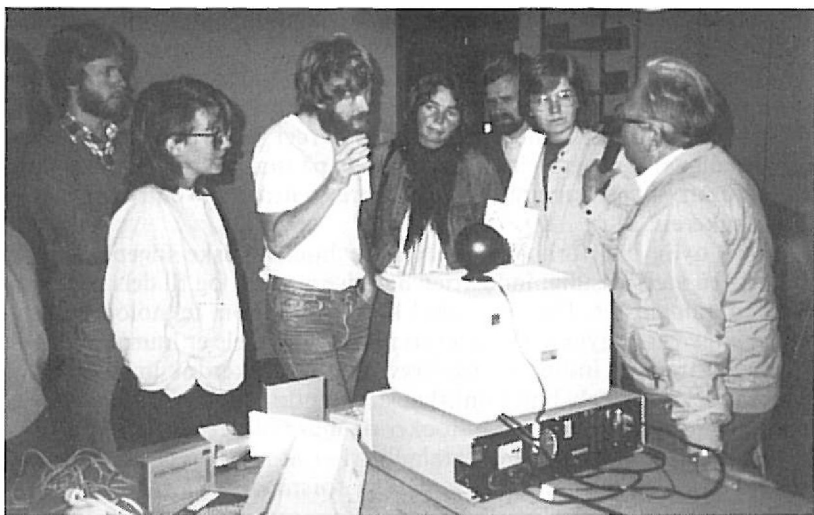
I sitt foredrag «Maskinen og språkopplæring» orienterte førstelektor *Magnar Brekke* fra Engelsk institutt om et prøveprosjekt i grammatikk på mellomfagsnivå. Brekke vil redegjøre nærmere for dette prosjektet i en egen artikkel i neste nummer av *Humanistiske Data*.

Amanuensis *Kolbjørn Slethei* holdt et foredrag om det fonetiske utviklingsarbeidet i tilknytning til et norsk tekst-til-talesynteseprosjekt ved Universitetet i Bergen. Slethei gikk gjennom trinnene i prosessen fra inntastet tekst til output i form av tale. Han orienterte om fortrinnene ved å lagre tale på datamaskin fremfor på lydbånd – på datamaskin er talen ikke-sekvensiell og en kan stille betingelser til den.

Bruksområder for talesyntese er bl.a. en automatisk rikstefontjeneste, «telefonpost» og talende flaskeautomater. Etter Sletheis oppfatning kommer vi til å oppleve en rekke både talende og lyttende datamaskiner i vår tid. (Se for øvrig Ivar Utnes artikkel i dette nummer av HD.)

Både det norske og et utenlandsk system for talesyntese ble demonstrert ved sesjonens «edb-verksted». Det er den svenske tekst-til-tale syntetisatoren OVE III som nå blir tilpasset norsk tale. Foruten for svensk er denne syntetisatoren programmert for engelsk, tysk, spansk, italiensk og mandarin. Den andre syntetisatoren, DECTALK, er utviklet kun for engelsk (ved MIT).

Edb som hjelpemiddel i andre fagområder ble også demonstrert, som informasjonshåndteringssystemet MAPPER anvendt på data om de arkeologiske funnene fra Bryggen i Bergen. Ved hjelp av MAPPER er det blitt utviklet en relasjonsdatabase hvor ulike typer funnopplysninger kobles sammen via en primærnøkkel. Historiefaget var representert ved en demonstrasjon av søking i 1801-folketellingen.



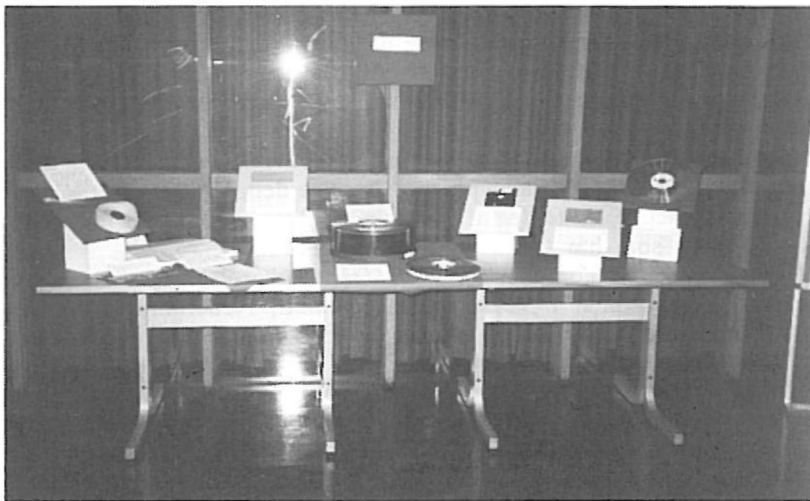
*Mange tilskuere fant veien til demonstrasjonen av talesyntese.*

Generelle hjelpemidler for humanistisk forskning som ble vist, var tekstbehandling og søking etter litteratur ved universitetsbibliotekene i Oslo og Trondheim. Det var dessuten adgang til å prøve et undervisningsprogram i historie på ungdomstrinnet - «Afrika-meister» - utviklet av Jan Oldervoll.

NAVFs EDB-senter for humanistisk forskning stod for flere demonstrasjoner, den ene av edb-metoder for rekonstruksjon av papyri (se Espen Ores artikkel i HD 1-85). I tilknytning til en demonstrasjon av videoplateteknologi for humanistiske forskningsformål ble det orientert om teknologien generelt og om Senterets eget prosjekt (se omtalen i HD 1-85). Ettersom Senterets egen plate ikke er presset ennå, tok demonstrasjonen utgangspunkt i en prøveplate fra University of London som bl.a. inneholder en rekke enkeltbilder av museumsgjenstander. Et utvalg av disse bildene ble vist på skjerm sammen med fiktive katalogdata lagt inn i en mikromaskinlagret database. Kombinasjonen videoplate/mikromaskin ble styrt av et program skrevet i Dataflex for lokal registrering og søking i museenes gjenstandsregistre.

Senteret var også ansvarlig for en demonstrasjon av fritekstsøkesystemet SIFT i en database sammensatt av katalogdata fra fem kulturhistoriske museer. I tillegg viste Senteret en utstilling som ga dels informasjon om Senterets virksomhet generelt, dels en oversikt over edb-ansettelser i de humanistiske fag eksemplifisert ved konkrete prosjekter.

Også Edb-seksjonen ved HF-fakultetet viste en utstilling. Her ble ulike lagringsmedier - fra papirutskrifter til videoplate - og deres kapasitet anskueliggjort med utgangspunkt i 1801-tellingen.



*Edb-seksjonen ved HF-fakultetet stod for en utstilling om ulike datalagringsmedier.*

Både «edb-verkstedet» (utenom demonstrasjonen av talesyntese) og foredragene i sesjonen «Humaniora i dataalderen» var dessverre svært dårlig besøkt, til tross for en del forhåndsomtale i media. Til og med fakultetets egen lærere og studenter glimret ved sitt fravær. Det er antagelig flere årsaker til den manglende oppslutningen – bl.a. at sesjonen var lagt til en uheldig tid på dagen, og at det samlede tilbudet under humaniora-ukene var overveldende. Kanskje mange humanister inntar en rent defensiv holdning m.h.t. forholdet mellom edb og humaniora, slik som Johan Fjord Jensen, professor i litteratur ved Århus universitet. I sitt foredrag «Humaniora ved tusenårsskiftet» ga han uttrykk for at humanister må verne seg mot presset fra vårt høyteknologiske samfunn. Humaniora-ukens edb-sesjon viste imidlertid klart at humanistiske verdier og kunnskap både kan fremmes ved bruk av edb som et redskap i forskningen, og er nødvendig i utforming- en av en mer «menneskelig» teknologi.

## Bibeltekstar og databehandling

### Første internasjonale kollokvium

*Per Vestbøstad*

Seminaret var arrangert av organisasjonen «Association Internationale Bible et Informatique», som spring ut frå miljøet kring det katolske «edb-klosteret» i Maredsous, 80 km sør for Brüssel. Her har ein dei siste 15 år bygd opp ein bibel-databank til hjelp i arbeidet med nye omsetjingar og bibel-ordbøker.

Det katolske universitetet i Louvain-la-Neuve sto som vertsinstitu- sjon, og tok vel i mot dei opp mot hundre deltakarane på alle vis. Louvain-la-Neuve synte seg å vera ein ny by, bygd berre for dette universitetet, som har flytta sørover frå Leuven.

Flest deltakarar kom naturleg nok frå dei katolske landområda i Sentral-Europa, men også frå Vest-Tyskland, Nederland, USA og Israel kom ein del bibel-forskarar og edb-spesialistar. Elles kom ein skilde deltakarar frå Sør-Afrika, Øvre Volta, Zaire, England og Finland.

Dei over 30 presentasjonane tok opp både nytestamentlege og gamaltestamentlege emne, men sistnemnde emne var i fleirtal.

Dekanus ved det teologiske fakultetet i Louvain-la-Neuve, professor *R. Gryson*, sa i helsingstalen at forskarane no har gjeve kyrkja ein påliteleg grunntekst for nytestamentets vedkomande, medan tilhøvet



mellom variantane av gamaltestamentlege tekstar ikkje alltid er like klårt.

Ulike metodar for lagring og framhenting av tekst og grammatikk, så som databasar med og utan ordbok, konkordansar i ulike format og parallell lagring av tekstvariantar var sentrale emne på konferansen.

*Eep Talstra*, Vrije Universiteit Amsterdam, hevda at bibelforskarar ofte ikkje tek det lingvistiske aspektet ved emnet alvorleg nok. Han skisserte ein database for bibeltekstar med eit logisk nivå for kvart språkleg nivå (morfem, leksem, uttrykk, setningar og setningsgrupper). Ein slik struktur gjer det lettare å forma relevante spørsmål og gjev snøggare søking.

Ved Vrije Universiteit Amsterdam brukar ein denne modellen til m.a. å «nykonstruera» den hebraiske grammatikken. Dei bearbeidar språket uavhengig av tradisjonell grammatikk og simulerar så å seia leseprosessen. Tilnæringsmåten liknar på den bibelmisjonærar må nytta når dei skal trengja inn i eit nytt og ubeskrive språk.

Konferansen hadde eit overraskande sterkt innslag av statistikarar og interessen for statistisk bearbeiding av bibeltekstar var stor. Statistikaren *Dean Forbes* frå Palo Alto hadde føgjande råd å gje til dei som vil bearbeida ein tekst statistisk: 1. Slå deg ikkje til ro med første resultatet. Prøv ulike analysemåtar. Leik med data og sjå dei frå ulike synssteder! 2. Søk kvalifisert hjelp for å unngå dei fallgruvane ein lett hamnar i.

צפרדע	00013	צפרדע	צפרדע	06531	1	-----	
		צפרדע	צפרדע	06532	1	-----	
		צפרדעים	צפרדעים		1		
		צפרדעים	צפרדעים		1		
			צפרדעים		1		
			צפרדעים		1		
צפרה	00003	צפרה	צפרה	06533	1	-----	
צפרן	00002	צפרן	צפרן	06534	1	-----	
		צפרני	צפרני		1		
צפה	00001	צפה	צפה	06535	1	-----	
צפה	00001	צפה	צפה	06536	1	-----	
צקלב	00015	ציקלב	ציקלב	06537	1	-----	
			ל-ציקלב		1		
		ציקלב	ו-כ-ציקלב		1		
		צקלב	צקלב		1		

Hebraisk indeks frå Bible Research Associates i Wooster, Ohio.

Finn ein statistikar som er interessert i problemet ditt, ikkje ein som berre vil bearbeida tala dine.

Demonstrasjonane viste interessante løysingar på presentasjonsproblema: Greske og hebraiske teikn på papir og skjerm var ønskemålet for dei fleste. Mikrah Computer Research Systems synte ein database med heile GT på hebraisk, berekna for bruk på personlege datamaskiner. Systemet skreiv hebraisk frå høgre mot venstre på skjermen og kostar \$5.000 inklusive tekstane.

Jack Abercrombie frå Pennsylvania University hadde laga ein grafisk verktøykasse («toolbox» i Pascal) og synte korleis tastaturet kunne programmerast til å gje greske, hebraiske, kyrilliske eller fonetiske teikn på skjermen. Programmet går under Turbo-Pascal på personlege datamaskiner, så eg vil tru at svært mange språkforskarar vil kunna gjera seg nytte av det. (NAVFs EDB-senter vil skaffa seg programma denne hausten.)

Elles vart det demonstrert læreprogram for gresk- og hebraisk-studentar og i bibelske emne for skuleverket og folk elles.

## Oppsummering

I sluttforedraget peikte professor Emmanuel Tov frå det hebraiske universitetet i Jerusalem på dei veldige gevinstane ein kan oppnå og dei nye forskingsoppgåvene som kjem innafør rekkjevidde når ein tek datamaskiner i bruk. Men han åtvåra også mot det store og aukande



Micra GT med hebraiske teikn på skjermen.

gapet som finst mellom forskarar med og utan datamaskin-tilgang. Utviklinga mot kraftigare mikromaskiner vil kunna gje ei jamnare spreiding av datamaskin-kraft, men då må ein også visa større vilje til samarbeid og utveksling av data, program og metodar.

Konferansen viste stor variasjon når det gjeld å ta i bruk datamaskiner i bibelforskinga, og det bør ikkje overraska nokon om dette forskingsfeltet gjev opphav til nye metodar som har allmenn interesse innom språkvitskapane.

## Nytt fra RHF/NAVF

### Eiendomsforhold og oppbevaring av data fra NAVF-prosjekter

NAVFs styre har oppnevnt et utvalg som skal utrede forholdene omkring eiendomsforhold og oppbevaring av data fra NAVF-finansierte prosjekter. Det er den stadig økende bruken av edb som lagringsmedium som har aksentuert behovet for en slik utredning, men utvalget skal vurdere forholdene omkring data fra NAVF-prosjekter uansett hvordan de blir lagret.

Rådet for samfunnsvitenskapelig forskning (RSF) og Rådet for humanistisk forskning (RHF) har tidligere fattet vedtak som dekker deler av sakskomplekset. RSF har allerede som vilkår for bevilgning at prosjekter som har samlet inn eller tilrettelagt data, ved prosjektets avslutning skal overføre kopi av samtlige datasett med dokumentasjon til Norsk samfunnsvitenskapelig datatjeneste.

Utvalget har fått følgende mandat:

- å utrede spørsmålet om eiendomsretten til data fra NAVF-prosjekter
- å utrede spørsmålet og komme med forslag til regler/rutiner for oppbevaring av slike data
- å utrede spørsmålet og komme med forslag til regler/rutiner som sikrer fornuftig gjenbruk av slike data
- å avklare forholdet til det offentlige arkivverket
- å komme med forslag til avtaler med aktuelle arkivinstitusjoner
- å vurdere behovet for opprustning av aktuelle institusjoner og i den sammenheng også vurdere de økonomiske konsekvensene av en slik opprustning
- å komme med forslag til revisjon av NAVFs bevilgningsvilkår.

Utvalget har denne sammensetningen:

Direktør Bjørn Henrichsen, NSD, formann

Konservator Natascha Heinz, RNF

Førsteamanuensis Eivind Smith, RSF

Professor Leiv Mjeldheim, RHF

Førsteamanuensis Lorentz M. Irgens, RMF

Avdelingsarkivar Ivar Fonnes, Riksarkivet

Leif T. Andressen, RHF, er sekretær for utvalget.

Utvalget skal levere utredningen i løpet av våren 1986 slik at eventuelle tiltak kan iverksettes fra 1987.

*Leif Andressen*  
*Konsulent, RHF*

# MELDINGER

## **Driftsklart databasesystem for musea**

Eit fullt operativt lokalt nettverk med 3 arbeidsstasjonar, ein liten gjenstands- og ein litt større foto-database vart demonstrert på NKKMs landsmøte i Trondheim 12.-13. september.

Databasane inneheldt katalogopplysningar registrert etter NKKM og Sekretariatet for fotoregistrering sine standard skjema. Registreringane kan søkjast fram på skjermen etter 4-5 nøkkelfelt, og via utskriftsfunksjonen kan ein søkja i alle felte.

Knytt til databasen med gjenstandsopplysningar var også eit arkiv med bilete av gjenstandane på videoplatespelar. Deltakarane fekk såleis høve til å sjå korleis ein frå gjenstandsdatan kan styra videoplatespelaren til å syna fram eit bilete av gjenstanden samstundes som ein har opplysningane om gjenstanden på dataskjermen.

Dataflex databasesystem er nytta i programmeringa, samstundes som det er laga ein enklare versjon med identiske skjermbilete i Datastar. Denne versjonen gjev berre høve til å søkja fram registreringar etter eit nøkkelfelt, men krev mindre lagerplass og eignar seg difor betre for bruk på frittstående diskettmaskiner.

Databasesystemet er utvikla av Per Vestbøstad og Jon Birger Østby innanfor NAVFs edb-prosjekt for dei kunst- og kulturhistoriske musea.

## **Edb-kurs for medarbeidarar ved musea**

10 medarbeidarar frå like mange kulturhistoriske museum landet over var samla på Norsk folkemuseum i tida 30.9-4.10. for å gjera seg kjent med dei nyutvikla databasesystema for gjenstands- og fotoregistrering. I øvingane vart det nytta 5 arbeidsplassar i Folkemuseets lokale nettverk (Corvus Omninet med Scanvest Corona mikromaskiner). Her var også rikeleg med lagerplass og skrivarkapasitet til å få gjort realistiske prøver både med Dataflex- og Datastar-systemet. Alle deltakarane kom frå museum som alt har skaffa seg datautstyr til registreringsoppgåver. Kursleiar var forskar-NAV *Jon Birger Østby*.



## THESAURUS LINGVAE GRAECAE

### Nytt fra Thesaurus Linguae Graecae

Thesaurus Linguae Graecae ved University of California, Irvine, gir nå ut et meldingsblad hvert halvår (i juli og desember).

I det siste nummeret meldes det at TLG er i ferd med å produsere en laserplate (CD-ROM) med maskinleselige tekster og annet materiale. Platen skal inneholde 27.5 mill. ord fra TLGs database over gammelgreske tekster, en elektronisk versjon av TLG Canon, alle de greske dokumentariske papyri som er tilgjengelige ved the Duke Data Bank of Documentary Papyri, en rekke tekster på latin, en del hebraisk, koptisk og gresk epigrafisk tekstmateriale og ordindekser til ulike TLG-tekster.

Meldingsbladet fås gratis ved henvendelse til: *Thesaurus Linguae Graecae, University of California, Irvine, Irvine, California 92717, USA.*

### «Manuscripts and Technology»

I 1982 og 1983 ble det i regi av European Science Foundation samlet informasjon om bruk av moderne teknologi i arbeid med manuskripter. Prosjektene man var interessert i å fange opp, spenner fra fysisk/kjemiske materialanalyser over metoder til å gjøre utydelig tekst mere leselig og til elektronisk databehandling av manuskriptenes tekster.

Det kom inn 72 besvarte spørreskjemaer fra 24 land. En femtedel av svarene var fra Norden.

En sluttrapport fra undersøkelsen er under arbeid. Nærmere opplysninger gis i HD etter utgivelsen av rapporten.

## **Directory of Computer Assisted Research in Musicology**

The Center for Computer Assisted Research in the Humanities i USA har nylig utgitt *Directory of Computer Assisted Research in Musicology*. Katalogen er et resultat av en spørreundersøkelse utført ved senteret vinteren 1985.

Innholdet favner om en rapport over pågående forsøk med å registrere, behandle og trykke musikalske data, en liste over pågående og nylig avsluttede edb-prosjekter i musikologi, en kort beskrivelse av forsøk på å etablere standarder for utveksling av musikalsk informasjon, og omtaler av studieprogrammer, nyere litteratur, og respondentenes syn på behov innenfor dette fagfeltet.

Katalogen er gratis og fås ved henvendelse til senteret ved følgende adresse: *525 Middlefield Road, Suite 120, Menlo Park, California 94025, USA*.

## **Aktuelle tidsskrifter**

### **SCOPE**

Ifølge utgiverne er SCOPE (Scholarly Communication: Online Publishing and Communication) et forum for informasjonsutveksling mellom lærere, forskere, administratorer, informasjonsmedarbeidere, programmerere og utstyrsleverandører. SCOPE er innrettet på samfunnsvitenskapene og humaniora og utgis annenhver måned. Det inneholder meldinger om nye edb-metoder utviklet ved universiteter, og om nettverk som letter utvekslingen av informasjon, data og programmer mellom universiteter. Dessuten meldes møter og konferanser relatert til edb-anvendelser i forskningsmiljøer, og utstyr og programvare for undervisning omtales. Det gis også en utvalgt bibliografi i hvert nummer.

Et abonnement på SCOPE koster \$52 pr. år. Adresse: *SCOPE, Paradigm Press, Inc, P.O. Box 1057, Osprey, Florida 33559, USA*.

### **news from RILA**

RILA - en amerikansk tjeneste for sammendrag av og indekser til litteratur om kunsthistorie - utgir på uregelmessig basis meldingsbladet *news from RILA*. Dette bladet inneholder ofte stoff om edb-anvendelser i kunsthistorie. Det fås ved henvendelse til: *RILA, c/o Clark Art Institute, Williamstown, Massachusetts 01267, USA*.



# Konferanser

## History and Computing

I forbindelse med stiftelsen av «Association for History and Computing» skal det avholdes en konferanse om edb og historie ved Westfield College, London, 21.-23. mars 1986. Konferansen skal ha fire hovedtema: 1. Edb-utstyr, programvare og konsulenttjenester for historikere, 2. Pågående og avsluttede prosjekter, 3. Metoder og teknikker: kilde-materiale, datatilrettelegging, kunnskapstilegnelse og -utvikling, kvantitative og kvalitative metoder (statistikk, historisk lenking, familierekonstitusjon osv.), og 4. Opplæring i edb for historikere og datastøttet undervisning. I tillegg til foredrag om disse emnene skal det gis demonstrasjoner og holdes en utstilling.

I tilknytning til konferansen skal foreningens stiftelsesmøte avholdes. Foreningen skal fungere som et forum for historikere som er interessert i/allerede er brukere av edb-metoder. Den skal stå for møter, seminar og kurs, og utgi et tidsskrift med den foreløpige tittelen «History and Computing». Innholdet i dette tidsskriftet er tenkt å bestå av artikler og meldinger om pågående arbeid og vurderinger av utstyr og programvare. Stoffet vil være rettet mot et bredt publikum, både eksperter og amatører i edb.

Forespørsler om konferansen kan stilles til: *Dr Deian Hopkin, Department of History, Hugh Owen Building, The University College of Wales, Aberystwyth, Dyfed SY23 3DY, UK.*

## International Conference on Machine and Machine-Aided Translation

Denne konferansen skal avholdes ved Aston University, Birmingham, 7.-9. april 1986. Arrangørene er Aston University i fellesskap med the Natural Language Translation Group of the British Computer Society. Konferansen er innrettet mot både de som utformer og de som bruker edb-baserte oversettelsessystemer. Det skal holdes både felles- og parallelle sesjoner samt en utstilling over operative systemer.

Påmelding: *Mrs A Kukulska-Hulme, Department of Modern Languages, Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom.*

## **Computers and Writing**

University of Pittsburgh, USA er vert for denne konferansen i tida 2.-4. mai 1986. Temaene for foredrag og programevaredemonstrasjoner vil være: datamaskinstøttet skriveundervisning, edb-basert tekstanalyse, tekstbehandling i klasserommet, databehandling av naturlige språk, og intelligente læresystemer. Flere opplysninger kan fås fra: *Terri Yousko, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.*

## **CALICO - Third Annual Symposium**

CALICOs tredje symposium skal finne sted 12.-16. mai 1986 i Annapolis, Maryland, USA. Symposiet henvender seg til lærere, programadministratorer, programmerere, forskere, representanter for offentlige organer og program- og utstyrsleverandører. Det tas sikte på å skape et forum for en diskusjon om teknologi i ulike disipliner med formålet å tilpasse denne teknologien til en mer effektiv undervisning, læring og behandling av språk.

I tilknytning til konferansen skal det holdes «workshops», demonstrasjoner og en utstilling. Programmet er ikke fastsatt ennå, men flere opplysninger kan fås fra: *CALICO Symposium 1986, Attn: H. Leon Twyman, 3078 JKHB, Brigham Young University, Provo, Utah 84602, USA.*

## **Seminar om flerspråklig leksikografi og terminologi**

Universitetet i Bergen, Norges Handelshøyskole og Norsk Termbank organiserer et seminar i leksikografi og terminologi i Bergen, 26.-27. mai 1986. Seminaret vil først og fremst ta opp praktiske sider av arbeidet med flerspråklige ordbøker og terminologier. Et av emnene det vil bli lagt stor vekt på, er datamaskinstøttet leksikografi og terminologi: en orientering om og presentasjon av databasesystemet MDDBS-III anvendt på terminologiprojekter og relaterte projekter - f.eks. ved utarbeidelsen av Norsk-fransk ordbok.

Flere opplysninger kan fås fra: *Lars Otto Grundt, Romansk institutt, Universitetet i Bergen, HF-hygget, 5014 Bergen-Universitetet.*

## **COLING 86**

XIth International Conference on Computational Linguistics - COLING 86 - skal avholdes ved Universitetet i Bonn i tida 25.-29. august. Konferansen skal omhandle en rekke emner i tilknytning til datalingvistikk. Etersom programmet ikke er klart ennå, blir det gitt flere opplysninger i HD 1-86.

## Nytt i biblioteket

### Edb-emner

- Clark, K.L./Tarnlund, S.A. (eds.): Logic Programming. London, 1982. 366 s.
- Dale, Nell/Lilly, Susan C.: Pascal plus data structures, algorithms, and advanced programming. USA, 1985. 490 s.
- MacLennan, Bruce J.: Principles of programming languages. USA, 1984.
- National interactive video centre/Council for educational technology: An introduction to interactive video. London, 1985. 38 s.

### Edb i undervisningen

- Centre d'Etudes des Systèmes et des Technologies Avancées: ANTEM III. Pedagogical applications of the image. Paris, 1985. 88 s.
- Centre d'Etudes des Systèmes et des Technologies Avancées: ANTEM III. Pedagogical applications of new communication technologies. Paris, 1985. 64 s.
- Fjornes, Erik: Datamaskin-assistert læringssystem til bruk innenfor faget kunsthistorie: Konsept for utvikling. Bergen, 1985. 21 s.
- Society for Applied Learning Technology: Sixth annual conference on interactive videodisc in education and training, August 22-24, 1984. Virginia, 1984. 84 s.

### Historie

- Nygaard, Lars: Fra historiske kilder til persondatabase. Et edb-basert system for registrering, bearbeiding, lagring og bruk av personhistoriske data. Institutt for informatikk, Universitetet i Oslo, 1985. 268 s.
- Vestre, Mari: Foneq et navnstandardiseringssystem for personnavn i historiske kildetekster. Institutt for informatikk, Universitetet i Oslo, 1985. 111 s.

### Språk

- Akkerman, Erik/Masereeuw, Pieter/Meijs, Willem: Designing a computerized lexicon for linguistic purposes. Amsterdam, 1985. 80 s.
- Hein, Anna Sägval: Foredrag vid de nordiska datalingvistikdagarna 1983. Uppsala, 1984. 199 s.
- Zifonun, Iradj/Plastwich, Konrad (red.): Dokumentation Textkorpora des neueren Deutsch. Mannheim, 1982. 141 s.

# SUMMARY

## **Humanistiske fag og kunnskapsbaserte systemer**

### **The humanities and knowledge-based systems**

The Centre has started to investigate the combination of the humanities and knowledge-based systems and is planning on arranging a Nordic symposium on this topic. Erik Fjornes outlines the possibilities for the humanities both to take part in the development of knowledge-based systems and to apply such systems.

First a description is given of knowledge-based systems and the problems connected with their development. Three main sub-fields have emerged: simulation of competence in natural languages, expert systems and robotization. Fjornes also seeks to define the specific character of the humanistic disciplines, including the types of activities that are central to these disciplines.

The humanities can take advantage of knowledge-based systems in especially three fields of application: in the storage of established humanistic knowledge, in the dissemination of such knowledge, and as a research tool.

The humanities have established various procedures for the solution of special problems in the study of human cultural phenomena. If this particular competence is to be utilized in the development of knowledge-based systems, humanists must, as a minimum, acquire knowledge about such systems and investigate how they form their basic concepts. Such competence can result in a new understanding of the problems connected with the development of knowledge-based systems, enabling more efficient techniques for describing knowledge, improved design procedures, and better concepts of artificial intelligence.

## **Informasjonsteknologi i Norge – kunnskapsbaserte systemer**

### **Information technology in Norway – knowledge-based systems**

Erik Fjornes of the Centre has interviewed Professor Jens Erik Fenstad, Dept. of Mathematics, University of Oslo. Prof. Fenstad has played an important part in the organization and planning of information technology in Norway and at present is working on a project concerning the simulation of natural language.

According to Prof. Fenstad, Norwegian information technology must be strengthened by educating more researchers, establishing larger institutions and supplying more equipment. As for knowledge-based systems, more emphasis must be put on developing new theories and models, e.g. for what knowledge is. This requires the joint participation of humanists, psychologists, linguists and scientists.

Prof. Fenstad believes that in order to take part in this type of basic research, humanities scholars must be better trained in the use of computers and model building. They also must be more internationally orientated and help promote the education of researchers.

Prof. Fenstad also gives an overview of the field of natural language systems,

which is in its infancy. Present systems are far too simple and limited. Current models for what communication is, and how we react to information, must be improved.

The participants of the project on simulation of natural language are attempting to isolate the basic characteristics such a system must have in order to function in a general way. Their point of departure is that meaning is a relation between the situation of the expression itself and the described situation.

### **Historie og edb - nok en gang!**

#### **History and computing - again!**

Research assistant Kåre Andersen of the University of Oslo starts this article by emphasizing the need for record linkage in connection with historical research. Access to machine-readable historical sources can also be facilitated by establishing a simple on-line data base.

Andersen himself has developed the program RUBREG, which enables the historian to state the terms of the computing process and is easy to learn. RUBREG can be run on most microcomputers and is in use both for research purposes and in schools.

In his introduction to RUBREG Andersen gives details of the program's functions: Main menu, form design, data entry, searching, print-outs and sorting. Plans for improving RUBREG are also mentioned.

### **Kunstig tale på norsk**

#### **Artificial speech in Norwegian**

A project on artificial speech has been initiated by Ivar Utne of the Section for Computational Linguistics at the Department of Scandinavian Languages and Literature, University of Bergen. The project is a co-operative venture with the Royal Technical College in Stockholm, which has produced a Swedish speech synthesizer. Kolbjørn Slethei of the Department of Phonetics and Linguistics, University of Bergen, is responsible for the development of the system's phonetic rules.

The basis of the Norwegian system consists of phonetic rules and rules for converting both text to phonetic transcription and digits to their alphabetic versions. Neither phonemes, words or sentences are stored, but a small exemption vocabulary can be activated.

In this article Ivar Utne presents the main principles of these rules (particularly the ones for text-to-speech conversion) and the software and hardware used in the system. The exemption vocabulary is also described in detail, as is the work on the corpus of language examples that forms the basis of phonetic analysis and the rules for phonetical transcriptions.

Utne also outlines the technical possibilities for use of the equipment. The Norwegian version will appear on the market in 1986.

Finally, details are given of how speech synthesis equipment can aid both the blind and people suffering from speech impediments.

## **Databehandling av termtilfanget frå Terminol-prosjektet**

### **Data processing of the term vocabulary in the Terminol Project**

The Terminol Project is a co-operative venture between the University of Bergen and the Norwegian government's oil company Statoil. It deals with the terminology used in the manuals of the Gullfaks A oil platform.

The project was started on January 1st, 1985 and so far has been extended until the end of 1985. Since the Gullfaks oil field is to be administrated in Norwegian, operating manuals must be available in this language. The manuals were first written in English, as corresponding Norwegian terminology was lacking. Translation of the manuals was carried out by the firms Norwegian Petroleum Consultants and Morco.

Senior Lecturer Per-Bjørn Pedersen describes in detail the procedures involved in the data processing of the project's term vocabulary. These mainly followed routines established in the course of previous projects conducted by the Norwegian Term Bank at the Department of Scandinavian Languages and Literature, University of Bergen.

The results of the work on the Terminol Project were eventually entered into a data base system, MDDBS III. This system will be used in two new projects commissioned by Statoil: a translation of MESC (Material and Equipment Standards and Codes) and the development of a Norwegian petroleum thesaurus.

## **Datamater/programmel/og programmer som del av interaksjonen i undervisningen**

### **Computers and software as a part of the interaction in education**

The point of departure for this article by Educational Consultant Erik Meistrup of the University of Århus is the fact that information technology integrated in teaching can have both an inhibiting and liberating effect.

According to Meistrup the introduction and use of new technology in education must be subjected to certain pedagogical conditions and goals. This is particularly important because computers in themselves have an alienating effect.

At the same time young people easily accept new technical media such as computers. It is therefore important that pupils learn about the various roles of computers. Access to the increasing amounts of computer-stored information must be facilitated so that teachers can make this information a part of education on pupils' premises.

Meistrup also shows how the combination software/hardware can be seen to constitute the interaction in education. Finally, he claims that it is important that the field of education develops creative modes of work that can contribute to advancing an imaginative technology on a humanistic basis. This in turn can result in qualitatively different and better software with a high degree of user-friendliness.

## **Hva skjer innen humanistisk og samfunnsvitenskapelig forskning?**

### **Current research in the humanities and the social sciences**

As reported in HD 1-84 the Norwegian Research Council for Science and the

Humanities has established an information service for research projects as a 5-year pilot program in Bergen. The service is to document all Norwegian research within the humanities and social sciences and can also document special fields as paid commissions. At present commissions concerning both consumer research and sports research are being concluded.

Up to now surveys of research in the humanities and social sciences financed by the Council in 1984 have been published. Corresponding catalogues for 1985 will appear soon. This autumn an overview of projects within Lappish and Finnish-Norwegian language, history and culture have also been published.

In co-operation with research institutions comprehensive documentation of research within the humanities and social sciences in 1985 is now being carried out. Both this information and data from Council-funded research will be prepared for entry into a database. A survey catalogue of the whole material will be published as a guide to this database. The bibliographical information will be regularly transmitted to the University Library in Oslo.

The information service will disseminate data in a variety of forms, including comprehensive overviews of research in special fields, and will offer trial subscriptions to the database.

### **Informasjonsteknologisk satsning ved Universitetet i Oslo**

#### **Investment in information technology at the University of Oslo**

Lecturer Asbjørn Brændeland of the University of Oslo reports that a committee from the Royal Norwegian Research Council for Scientific and Industrial Research – appointed by the Department of Industry – has suggested an investment program for information technology in Norway. This plan involves increasing both the number of university candidates with competence relevant to this field and the amount of computer equipment available for research and education.

In answer to this challenge the University of Oslo has prepared its own plan of action for developing information technology. This plan, which encompasses the procurement of equipment and supplying new positions for research, education, services and operation, has an economic framework of 700 mill. Nkr for the period 1986-90.

The University of Oslo is currently also discussing a joint venture with the firm Norsk Data concerning both the development of information technology at the university and products specially aimed at the university's needs.

The government-appointed committee's suggestion is disappointing from the point of view of the humanities. It emphasizes «hard» technological activities and shows little understanding of the importance of humanistic competence in the information technology of the future. However, according to the University of Oslo's own reports, not only will the humanities be allotted increased resources – humanistic knowledge will be utilized in the development of this field.

### **Seminar om pedagogisk programvare**

#### **Seminar on pedagogical software**

In June the Ministry of Church and Education arranged its second seminar for



authors of educational software. The basis of these seminars is the fact that the selection of software for Norwegian schools is too limited and of varying quality. Instead of importing foreign software, programs suited to Norwegian educational values should be produced. On the other hand, much is to be learnt from design methods developed abroad. This year's seminar can be designated a «program design workshop», reports Manager Gunnar Thorvaldsen of the Norwegian Historical Data Archives.

The main speakers were Les Green and Dan Daniels from Canada and USA, respectively. According to them, software should activate pupils, and screen design should be based on one main picture. Thorvaldsen comments on the pro's and con's of the design principles presented by these speakers.

Presentations were also given of work carried out by several groups of participants on software for the humanities, particularly history.

### **NORD IoD6**

The 6th Nordic conference on information and documentation was arranged in Helsinki in August for 400 participants. The theme of this year's conference was «Information Resources Management». Director Jostein H. Hauge of the Centre reports that 60 lectures on this theme were given. 30 exhibitors showed hardware and software for information processing, and a number of information brokers and scientific documentary agencies demonstrated their services and products. The conference's activities clearly showed that interactive information databases are becoming more and more common in the library and documentation sector and in an increasing number of forms.

In the plenary session the term Information Resources Management was treated from different points of view. Two speakers dwelled on the importance of «information literacy» – of qualifying both present staff and future generations in utilizing computer-based information products. However, other speakers stressed that information systems should be adapted to the needs of users, not vice versa.

The questions of charging fees for information services and the marketing of products were also dealt with in this session.

In a separate session a series of Nordic databases was presented. Databases were also the main theme of the 50 lectures in parallel sessions, many of which stressed the importance of careful planning before establishment.

In some lectures the development of so-called strategic information systems was discussed. The increasing demands of users on both retrieval systems and the design and use of the products acquired via database retrieval was also pointed out.

The topic of several lectures was how to retrieve information at the lowest possible cost. In addition, accounts were given of the activities of several organizations that help users to gain access to relevant sources.

### **International Conference on Data Bases in the Humanities and the Social Sciences**

This conference took place at Grinnell College, Iowa in June. Director Jostein H. Hauge of the Centre reports that it attracted nearly 200 participants from 14 countries, mainly connected with the humanities. 60 lectures on a very wide variety of topics were given.

The term data base clearly has become a common denominator for computer-stored research material, source data and reference information. In Hauge's opinion the conference offered a large number of interesting papers that demonstrated new developments and interests within computing and the humanities.

Interactive bibliographical data bases are now being established also for the humanities. In several lectures accounts were given of how humanists can employ American and French data bases of this kind.

In many countries large collections of source data for humanistic research are being created. Rutgers University is compiling an international survey of computer-stored text material. A presentation was also given of the text and word corpora available at Institut für deutsche Sprache in Mannheim.

As at previous conferences computing activities at archives and museums were well represented. A presentation was given of work being carried out within the framework of the Canadian Heritage Information Network, which aims at the creation of an extensive central data bank for public museum collections. In the USA a survey of the contents of the country's 16,000 archives is in the process of preparation.

A. Oikawa's lecture «An Archaeological Survey Database and the Development of its Application Software» demonstrated the high level of arts computing in Japan.

Information services on on-going research in Canada, France and Norway were also presented.

### **Seventh International Conference on Computers and the Humanities (ICCH)**

The 7th ICCH conference was arranged at Brigham Young University, Provo, Utah, in June by the Humanities Research Center. About 350 participants could choose between more than 100 lectures in four parallel sessions.

Director Jostein H. Hauge of the Centre notes that ICCH is in the process of changing from a general computing conference for the humanities to a conference on computing in languages and literature. Projects concerning computer tools for philology and literary analysis dominated the conference this time.

Many lectures dealt with how humanistic researchers can apply standard software or devise new programs, often based on the possibilities microcomputers give rise to for computer-assisted teaching, particularly in modern and classic languages.

The use of new technology was also taken up in a few lectures. At the University of Iowa videodiscs with slides are currently being tried out in the teaching of art history. An account was given of an interactive videodisc project in language training at Brigham Young University. This training system consists of a film on videodisc split into sequences and with CAI programs on a microcomputer linked to the videodisc player.

In several lectures a widespread and increasing use of the Kurzweil OCR was documented. It is currently used in connection with the preparation of large dictionary archives, encyclopaedias and collections of literary texts.

A series of lectures also dealt with the need for both better software for the humanities and for training researchers in modern text production (electronic publishing).

## **Edb og lokalhistorie**

### **ADP and local history**

Local historians in Norway started to use computers for word processing over five years ago, but few have tried computing for other purposes. For this reason the Norwegian Institute for Local History arranged a seminar on ADP and local history in August.

More than 50 people took part in this seminar, which was divided into five subjects: word processing, kinds of computing assistance that can be obtained from central institutions, what local historians can do themselves on microcomputers, record linkage, and programs.

The seminar was lead by the Manager of Voss Folk Museum, Johannes Kvestad. He reports that in connection with the first subject lectures were given on the transmission of processed texts to printers and how researchers' work is influenced by the use of word processing.

As for central institutions, a presentation was given of the Norwegian Historical Data Archives and other important data bases of interest to local historians. The National Archives were also presented.

Lectures given on the next two subjects gave the participants examples of computer applications in different fields within local history.

In the final section of the seminar information was given on available program packages, especially integrated ones. However, it was stressed that specially designed programs are still necessary and that designing one's own programs can in many cases be very useful.

## **Humaniora-uken ved Universitetet i Bergen**

### **The «Humanities Week» at the University of Bergen**

The first «Humanities Week» at the University of Bergen took place from September 22-26. Information Officer Kristin Natvig of the Centre reports on the session entitled «Humanities in the Data Age».

Senior Lecturer Jan Oldervoll of the Department of History spoke on the development so far in the use of computing at the Faculty of Arts and the future role of humanities scholars in relation to data technology.

Senior Lecturer Magnar Brekke of the Department of English presented a trial project in the teaching of grammar. A more detailed account of this project will appear in the next issue of HD.

Lecturer Kolbjørn Slethei gave an overview of the phonetic development work involved in a Norwegian text-to-speech synthesis project. Both this system and DECTALK (developed for English at MIT) were demonstrated at the «computing workshop».

Computing tools for archaeology and history were also demonstrated, along with general aids such as word processing and a retrieval system for literature held at Norwegian university libraries.

The Centre held several demonstrations – on computer methods for the reconstruction of papyri, videodisc technology (in particular the Centre's own project), and using the free text information retrieval system SIFT in a data base of museum catalogue data. In addition, the Centre showed an exhibition of its projects and general activities. Another exhibition, prepared by the Computing Section of the Faculty of Arts, showed different kinds of storage media.

## **Bibeltekstar og databehandling**

### **Bible texts and computing**

This is the subject of an international seminar – the first of its kind – arranged by «Association Internationale Bible et Informatique» at the University of Louvain-la-Neuve in Belgium.

Senior Computing Officer Per Vestbøstad of the Centre reports that the 30 presentations given at the seminar dealt with both the New and the Old Testament. Central topics were various methods for the storage and retrieval of text and grammar, such as databases with and without vocabularies, concordances in different formats and parallel storage of text variants. Statistical methods for processing bible texts were also presented. Vestbøstad supplies summaries of several lectures on these topics.

Demonstrations showed interesting solutions to problems of presentation on both screens and paper, including a data base of the Old Testament in Hebrew for personal computers and a toolbox for screen representation of Greek, Hebrew, Cyrillic and phonetic signs. Educational programs were also shown for students of Greek and Hebrew and on biblical topics for school pupils and the general public.

### **Nytt fra RHF/NAVF**

#### **News from the Council for Research in the Humanities**

A committee has been established in order to investigate questions concerning ownership and storage of data from projects financed by the Norwegian Research Council for Science and the Humanities. The increasing use of computers as storage media has made the need for such a report more acute.

#### **Meldinger**

##### **News**

A fully operative local network with work stations and databases of objects and photos has been developed for demonstration purposes by Per Vestbøstad and Jon Birger Østby of the Council's computing project for museums of art and cultural history. The databases contain information on the objects which in turn may be connected with a videodisc archive of pictures of the objects.

These database systems were the subject of a course arranged at the Norwegian Folk Museum in October, attended by 10 members of staff from museums of cultural history.

Thesaurus Linguae Graecae at the University of California, Irvine has started to publish a newsletter twice a year. In the last issue TLG announces that it is in the process of producing a laser disc (CD-ROM) of machine readable texts and other material.

In 1982-3 the European Science Foundation sponsored a survey on the use of modern technology in work on manuscripts. 72 questionnaires from 24 countries have been answered. A final report will be published soon.

The Center for Computer Assisted Research in the Humanities, California, has recently published *Directory of Computer Assisted Research in Musicology*.

Two American publications presented in this issue are: the bimonthly SCOPE (Scholarly Communication: Online Publishing and Communication), which is aimed at the humanities and social sciences, and *news from RILA*, which often contains items on computing in art history.

*Forts. fra 2. omslagsside.*

- RAPPORT nr. 29, 30, 31, 32: *Stig Welinder et al.: STAR I-IV A* program package for archaeological use. Bergen 1983. Samlet pris kr. 180. (Rapportene kan også kjøpes enkeltvis).
- nr. 29 STAR I Introduction and Star manual. ISBN 82-7283-033-7  
Pris kr. 50.
- nr. 30 STAR II Student textbook and STAR examples. ISBN 82-7283-034-5  
Pris kr. 60.
- nr. 31 STAR III Archaeology for statisticians. ISBN 82-7283-035-3  
Pris kr. 60.
- nr. 32 STAR IV STAR algorithms. ISBN 82-7283-036-1  
Pris kr. 30.
- RAPPORT nr. 33. *Årsmelding 1983*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-038-8  
Gratis.
- RAPPORT nr. 34. *Jostein H. Hauge: Tutorial on Machine Translation*. Rapport fra en konferanse i Lugano 2.-6. april 1984. ISBN 82-7283-039-6.  
Pris kr. 60.
- RAPPORT nr. 35. *Ole Lauvskar: Bruk av statistiske metoder i språk- og litteraturforskninga*. Rapport frå ei spørjeundersøking, September 1984. ISBN 82-7283-041-8  
Pris kr. 50.
- RAPPORT nr. 36. *Årsmelding 1984*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-042-6.  
Gratis.

# C

Av innholdet:

Erik Fjornes: Humanistiske fag og kunnskapsbaserte systemer

Erik Fjornes: Informasjonsteknologi i Norge – kunnskapsbaserte systemer. Intervju med Jens Erik Fenstad

Michael Benskin: How Computers are Changing the Character of Research in the Humanities

Kåre Andersen: Historie og edb – nok en gang!

Ivar Utne: Kunstig tale på norsk

Per-Bjørn Pedersen: Databehandling av termtilfanget frå Terminol-prosjektet

Erik Meistrup: Datamater/programmell/og programmer som del af interaksjonen i undervisningen