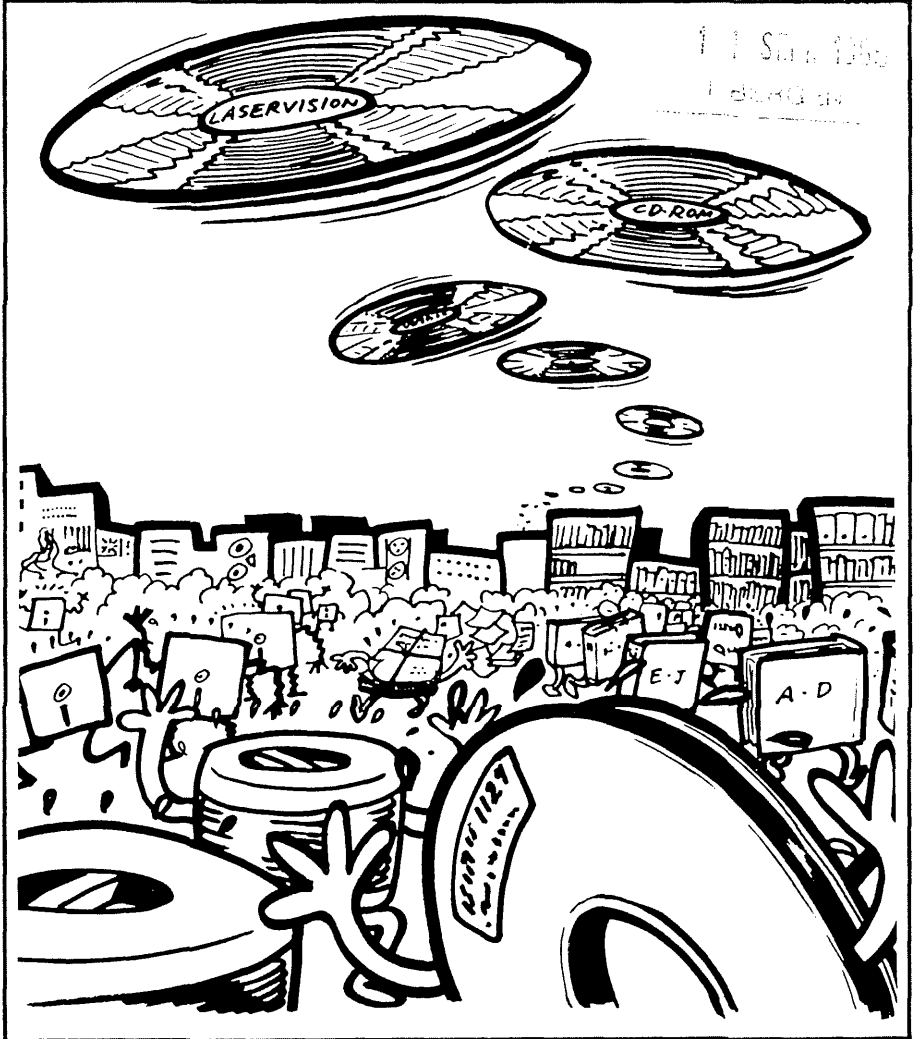


humanistiske data



NAVFs EDB-senter for humanistisk forskning
The Norwegian Computing Centre for the Humanities

2-85

SENTERETS RAPPORTSERIE

Rapporter utgitt f.o.m. 1980

- RAPPORT nr. 13. *Datatjenester for og datasamarbeid mellom kunst- og kulturhistoriske museer*. Februar 1980. 2. opptrykk november 1981. ISBN 82-7283-010-8 Pris kr. 50.
- RAPPORT nr. 14. *NOVA*STATUS HÅNDBOK*
Del 1: Søking. Brukerveiledning. 3. opptrykk 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. opptrykk 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 Lauvskaar: 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.

Forts. 3. omslagsside.

humanistiske data 2-85

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. 9.8.85).

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 (ansv. red.), 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 9. august.

Humanistiske Data is published by The Norwegian Computing Centre for the Humanities. Editorial group: Jostein H. Hauge, Kristin Natvig, 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:

Espen Brodin (sluttet), Jostein H. Hauge, Knut Hofland, 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:

«The Invasion of the Gigabyte Snatchers». (Tegning: Øystein Reigem).

Videoplateteknologien står foran sitt endelige gjennombrudd. Se Øystein Reigems artikkel «Optisk plate-teknologi» og rapporter om BBCs Domesday-prosjekt og fra The 1985 International Videodisc, Optical Disk and CD-ROM Conference and Exhibition. Se også Knut Hoflands rapport fra konferansen ANTEM III.

Innhold

Artikler

Optisk plateteknologi. <i>Øystein Reigem</i>	s. 4
Er datalingvistik del av anvendt språkvitenskap? <i>Lars Sigfred Evensen</i>	s. 23
Prosjekt i automatisk indeksering – lingvistiske metoder. <i>Tove Fjeldvig, Anne Golden og Øystein Reigem</i>	s. 32
Edb-baserte arkivkataloger: Samkatalog for privatarkiv. <i>Elisabeth Koren</i>	s. 50
Edb og arkeologi. <i>Britt Kroepelien</i>	s. 62
DUS-forfatter-systemet: ComCats. <i>Erik Meistrup</i>	s. 70

Rapporter

BBCs Domesday-prosjekt. <i>Øystein Reigem</i>	s. 81
Samisk database. <i>Anders Løov</i>	s. 90
Nytt om Registreringssentral for historiske data, Tromsø. <i>Jostein H. Hauge</i>	s. 96
«High Technology and Computers in Education». <i>Lise Opdahl</i>	s. 99
ANTEM III, Paris 12.-15. mars 1985. <i>Knut Hofland</i>	s. 103
Humanioradagene 1985 ved Universitetet i Oslo. <i>Tellef Kvifte</i>	s. 108
Computers and Art History Workshop. <i>Espen Ore</i>	s. 111
Utviklingsseminaret på Ustaoset 17.-19. april 1985. <i>Espen Brodin</i> .	s. 115
Seminar om pedagogisk informatikk. <i>Per Vestbøstad</i>	s. 117
Computers in English Language Research. <i>Stig Johansson</i>	s. 120
The 1985 International Videodisc, Optical Disk and CD-ROM Conference and Exhibition. <i>Øystein Reigem</i>	s. 123
Nytt fra RHF/NAV F	s. 130

Meldinger	s. 134
-----------------	--------

Summary	s. 143
---------------	--------

Optisk plate-teknologi

Øystein Reigem

Denne artikkelen er ment som en oversikt over optisk plate-teknologi i dag. Den gir på flere områder oppdaterte opplysninger i forhold til Elin Solstrands artikkel i Humanistiske Data 3-83, men erstatter ikke den på alle måter. I denne artikkelen er det lagt vekt på å få med fakta og praktiske opplysninger, spesielt for de typer plater/utstyr som er best etablert og lettest tilgjengelig, såsom Laservision videoplate og CD-ROM.

For de lesere som ikke er vant med dataverdenens mål på lagringskapasitet, er 1 Kbyte (kilobyte) = 1000 tegn, 1 Mbyte (megabyte) = 1.000.000 tegn og 1 Gbyte (gigabyte) = 1.000.000.000 tegn. En skrivemaskinskrevet side inneholder omtrent 2000 tegn, altså 2 Kbyte. 1 Mbyte tilsvarer dermed ca. 500 skrivemaskinsider og 1 Gbyte ca. 500.000 sider. For dem som vil sammenligne optiske mediers lagringskapasitet med konvensjonelle datalagringsmedier, rommer vanlige disketter typisk 360 eller 720 Kbyte, og maksimum 1,2 Mbyte. Fastdisker for mikromaskiner er ofte på 10 eller 20 Mbyte, men også 40 Mbyte disk selges. Diskstasjoner for stormaskiner har gjerne en kapasitet på noen få hundre Mbyte opptil vel 1 Gbyte, men disse inneholder hele stabler av enkeltplater.

Klassifisering av plater: Analog, digital; read-only, write-once, overskrivbar

Optiske plater kan deles i *analoge* plater, som lagrer et *videosignal*, og *digitale*, som lagrer digitale data. De analoge platene er altså de «egentlige» videoplatene (engelsk: videodisc), mens den vanlig brukte betegnelsen på de andre er digital optisk plate (engelsk: digital optical disk eller optical digital disk, forkortet DOD eller ODD). Denne termbruken vil bli fulgt her.

Siden analoge plater baserer seg på et videosignal, kan de lagre stillbilder eller levende video av (meget god) TV-kvalitet, men med de begrensninger i oppløsning et videosignal har. Lyd kan også lagres, enten som stereo eller som to separate spor, f.eks. tospråklige kommentarer. Digitale plater kan i prinsippet lagre hva som helst på digital form – inntastet tekst, programmer, digitaliserte bilder, digitalisert lyd, osv. For å gjøre situasjonen mer komplisert, finnes det videoplateutstyr som kan lagre digitale data som video. Slikt hybridutstyr vil altså ha både analoge og digitale muligheter. Ved å digitalisere lyd og legge den inn på denne måten, kan en også presse inn adskillig mer lyd på en videoplate enn det det vanlige lydsporet kan romme.

En fordel med analoge plater er at kapasiteten er langt høyere for vanlige bilder og levende film enn for digitale plater. En kritisk faktor ved valg av teknologi er da om videokvalitet er tilfredsstillende for den aktuelle anvendelse, dvs. om oppløsningen er god nok.

Med digitalt utstyr vil en i prinsippet kunne lagre bilder med vilkårlig god oppløsning. I praksis vil digital plate kunne brukes dersom bildematerialet er fattig på fargenyanser/gråtoner, eller hvis mulighetene er store for komprimering, f.eks. dokument sider eller strektegninger, som inneholder store hvite flater.

Digitalisert informasjon har en rekke fordeler framfor analog: Vha. av *feilkorrigerende koder* kan informasjonen gjøres meget robust overfor støy og skader på platen. – Informasjonen kan *kryptifiseres* for å gjøre den utilgjengelig for uvedkommende. – Informasjonen kan lettvis *viderebehandles* på forskjellige måter. Digitaliserte bilder kan f.eks. panoreres, zoomes, klippes og mikses på en grafisk skjerm, kontraster kan skjerpes, farger endres, osv.

En inndeling som går på tvers av skillet mellom analog og digital representasjon, er grupperingen i *read-only-media* («kun lesbare» for den som ønsker en norsk term), *write-once-media* (egentlig *write-once-read-only*, på norsk «skriv-én-gang») og *overskrivbare media* (*erasable* på engelsk). Read-only-teknologien er i dag i bruk i en rekke anvendelser. Det finnes også flere operative write-once-systemer, men denne teknologien må kunne sies å være mer i en oppstartingsfase når det gjelder praktisk bruk. Systemer med overskrivbare optiske plater spås å være kommersielt tilgjengelige i 1987 – noen sier 1986.

Read-only-plater

Read-only-plater produseres ved en prosess som er svært lik framstillingen av gramfonplater. Alt materiale som skal inn på en read-only-plate, må ferdiggjøres på forhånd og legges over på videobånd (for videoplate) eller magnetbånd (for digital plate). Dette såkalte *masterbåndet* sendes til en pressefabrikk som produserer en *masterplate* og presser opp det ønskede antall kopier av platen. Hver kopi er forholdsvis rimelig, men selve masteren er relativt kostbar. Imidlertid vil oftest tilretteleggingen av materialet som skal inn på platen være den største utgiften.

For å kunne bruke en read-only-plate kreves en eller annen form for platespiller/diskstasjon koblet mot ens datamaskin. Videoplatespillere kan også brukes alene. Slikt utstyr er forholdsvis rimelig.

Read-only-teknologien er den eldste og best etablerte, og platene er forholdsvis standardiserte. En viktig anvendelse av read-only-plater vil være publisering av materiale som er relativt statisk, eller som ikke krever *for* hyppig oppdatering. Med et visst opplag blir platekopiene såpass rimelige at man kan tenke seg abonnementsordninger på f.eks. oppslagsverk. Read-only-videoplaten har et stort anvendelsesområde

innen datamaskinassistert undervisning, og også innen andre områder hvor det er ønskelig med en fleksibel multimedia-kommunikasjon. F.eks. vil vi om noen år i butikker finne systemer som selger oss varer via berøringsfølsomme skjermer.

Hvor godt read-only-plater vil slå gjennom som medium for elektronisk publisering er ennå usikkert. At potensialet er til stede er utvilsomt, men det hele er et spørsmål om «kritisk masse». For at institusjoner, bedrifter og enkeltpersoner skal være villige til å investere i utstyr, må det nemlig være et visst tilbud av publisert materiale. Samtidig vil det være lite å tjene på publisering via plate før utstyret har en viss utbredelse, altså et klassisk egg-høne-problem. BBCs Domesday-prosjekt kan være den typen tiltak som skal til for å skape gjennombruddet for teknologien. (Se artikkel om Domesday-prosjektet annensteds i dette bladet.)

Write-once-plater

Write-once-utstyr tillater innlegging av materiale på stedet. Hvis det er bildemateriale i en eller annen form (foto, videoopptak, dokumentsider m.m.) som skal lagres, kreves dermed at utstyret har en input-enhet for bilder – et videokamera eller en scanner.

Siden platene ikke kan brukes om igjen, vil write-once-plater bare delvis kunne erstatte magnetiske lagringsmedier (disk, diskett). De vil finne sin plass i anvendelser der materialet vokser, men ellers i liten grad blir endret, og i anvendelser der platens kapasitet er så høy i forhold til behovet at man kan tillate seg å «sløse» med plass.

Et uttrykk man ofte hører i forbindelse med write-once-plater er DRAW (Direct Read After Write). DRAW er en teknikk som brukes for å unngå støy på platen, og består i umiddelbart å kontroll-lese det som skrives på platen. Hvis det ikke er samsvar mellom det som ble skrevet og det som ble lest, blir informasjonen slettet og skrevet på nytt. Enkelte bruker forresten termen DRAW nesten synonymt med write-once.

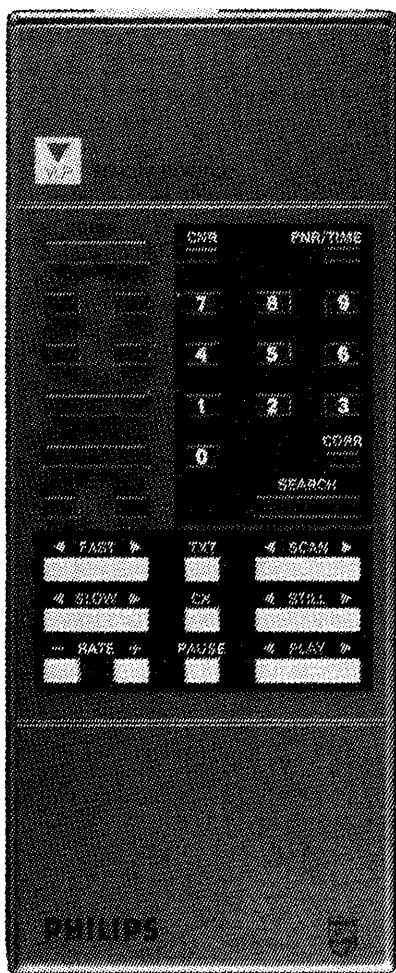
Write-once-utstyr er relativt dyrt. Standardiseringen er ikke kommet langt.

Overskrivbare plater

På read-only- og write-once-plater lagres vanligvis informasjonen som hull i en metallfolie. Folien er igjen dekket av et beskyttende lag med plast. Ved avspilling avleses informasjonen vha. en laserstråle. I write-once-plater brukes også laseren til å brenne hullene i folien. En så brutal prosess kan selvsagt ikke brukes for overskrivbart utstyr. Flere alternative teknikker for en reversibel prosess er under utprøving. Den såkalte magneto-optiske platen ser ut til å ligge best an, men Matsushita (Panasonic) satser på faseforandringsprinsippet.

Read-only videoplater

Videoplaten ble lansert som et medium for spillefilm, altså som en konkurrent til videobånd. Av forståelige grunner (read-only) har ikke denne typen plate fått det store gjennombruddet på hjemmemarkedet. Videoplaten omfattes imidlertid med stor interesse innen den anvendelsen som kalles *interaktiv video*. I interaktiv video brukes materialet på platen i et samspill med en «bruker». Det kan dreie seg om undervisning (alt fra drill og ren presentasjon til simulering og dialogbaserte systemer), gjenfinning (oppslagsverk, manualer, bildedatabaser), spill, salg av varer, osv.



Fjernkontroll til videoplatespiller. Fjernkontrollens funksjoner kan også utføres via mikromaskin.

Nebraska-skalaen

Man snakker om tre nivåer av interaktive videoplatesystemer, klassifisert etter den såkalte Nebraska-skalaen. I nivå 1-systemer styrer man videoplatespilleren fra en fjernkontroll, helt analogt (sic) med styring av en vanlig videobåndspiller. Man henter selv fram de stillbilder og videosekvenser man vil ved å taste inn bildenummer på fjernkontrollen. Videoen vises på en monitor (TV) som står tilkoblet, akkurat som med en videobåndspiller.

I nivå 2-systemer har videoplatespilleren en innebygd prosessor, slik at en kan legge inn programmer som styrer gjennomgangen av plate-materialet. Nivå 2-utstyr har som regel begrenset hukommelse, og dermed begrensning i hvor avanserte programmer som kan kjøres.

I nivå 3-systemer styres videoplatespilleren fra en ekstern datamaskin, som regel en vanlig mikromaskin (PC o.l.). Dette krever at spilleren har en inngang som kan ta imot datasignaler fra en datamaskin. Det finnes en rekke spillere som har dette. Det vanlige er en RS-232-inngang. Vanlige mikromaskiner – i hvert fall over hobbynivå – har oftest én eller flere RS-232-utganger, og det byr da på små problemer å koble. Også mange mindre maskiner kan utstyres med RS-232-utgang.

På The Emerging Technologies – Audiographics/Videographics Conference i Houston 13.-15. april 1983, framkom forslaget om å definere et nivå 4-system som et nivå 3-system med følgende tilleggsegenskaper: 1) Mulighet for å dekode digitale data lagret som videobilder på platen. 2) Mulighet for å legge grafikk fra mikromaskinen (i form av enda et videosignal) over videosignalet fra platen. BBC/Philips' utstyr for Domesday-prosjektet kan vel betraktes som nivå 4-utstyr.

Nivå 4 er også blitt brukt til å betegne andre utvidelser i forhold til nivå 3.

Forskjellige videosignal-standarder

Da videoplaten lagrer et videosignal, vil dessverre det lagrede signalet avhenge av hvilken fjernsynsstandard som er brukt. Som kjent består levende video (fjernsyn) av et visst antall enkeltbilder («frames») pr. sekund. Hvert av disse enkeltbildene skapes så igjen av at en elektronstråle (tre for fargefjernsyn) med varierende intensitet sveiper linjevis over skjermen. Strålen får et fluorescerende belegg på skjermen til å lyse opp. I USA brukes standarden NTSC med 30 enkeltbilder eller «frames» pr. sekund og 525 linjer pr. «frame». I Europa brukes PAL (Vest-Europa) og SECAM (Frankrike, Øst-Europa), som er 25 «frames» pr. sekund og 625 linjer pr. «frame». Det er også andre forskjeller mellom de tre standardene.

Dermed vil ikke en NTSC-plate uten videre kunne spilles på en PAL-spiller og omvendt. Derimot tar europeiske spillere gjerne både PAL og SECAM. Når det gjelder *monitører*, finnes utstyr som tar både amerikansk og europeisk standard.

PAL, som er systemet som brukes i Norge, er forresten heller ikke helt ensartet. Et britisk PAL *antennesignal* adskiller seg fra vanlig PAL, noe man skal være oppmerksom på hvis man kjøper videoutstyr i Storbritannia. Derimot er selve *videosignalene* like, så *videoutganger* og *-innganger*, *videobånd* og *videoplater* er som vanlig PAL.

Laservision og VHD

Siden videoplaten ble lansert, har en rekke systemer dukket opp, og i stor grad forsvunnet igjen. I dag er det i praksis to systemer på markedet – VHD og Laservision. Begge bruker 12 tommers plater, dvs. på størrelse med LP-plater. VHD-plater kommer i beskyttende plastkassetter, Laservision-plater i vanlige grammofonplateomslag. Bortsett fra skillet mellom NTSC og PAL/SECAM, er de to systemene godt standardisert, slik at plater fra forskjellige produsenter kan spilles på spillere av forskjellige merker.

VHD-platen er egentlig ikke en *optisk* plate, da det er en slags stift som avleser informasjonen på platen. Nettopp derfor må man regne med en tanke slitasje på platen ved bruk. Generelt er kvaliteten på VHD-platen noe lavere enn på Laservision-platen. En spesiell egenskap ved VHD-systemet er at man ikke kan stoppe på ett enkelt videobilde. Spilleren vil alltid vise to påfølgende «frames». Dette har sammenheng



Philips Laservision videoplatespiller.

med at hver spor-runde på platen inneholder to «frames». Ved lagring av stillbilder må derfor disse dupliseres. En annen konsekvens er at frysing av levende video vil gi flimrende bilder. VHD-platens kapasitet er 45.000 enkeltbilder eller én time levende video pr. side. (Det er naturlig å oppgi kapasiteten pr. side, da platespillerne bare kan spille av én side om gangen.)

Fordelene med VHD-systemet er en lavere pris enn Laservision, samt at det finnes utstyr som kan spille *både* amerikanske og europeiske plater. Trolig finnes det ikke Laservision-utstyr som kan dette.

VHD-systemet er utviklet av JVC. I Storbritannia leveres utstyr av Thorn-EMI. Artikkelforfatteren kjenner ikke til at noen i Norge har gått til anskaffelse av VHD-utstyr.

Laservision-plater finnes i to formater, CLV og CAV. CLV er beregnet på spillefilmer og CAV på interaktiv video. CAV betyr *constant angular velocity*, dvs. at platen spinner rundt med konstant vinkelhastighet. Hver runde sporet gjør på platen inneholder én «frame», noe som gjør det lett for spilleren å finne ett bestemt bilde. Kapasiteten for én side på en CAV-plate er ca. 36 minutter levende video, noe som tilsvarer ca. 54.000 enkeltbilder.

CAV-systemet sløser med plass på platen. Selv om det ytterste sporet på platen er mye større enn det innerste, må begge inneholde samme mengde informasjon. CLV betyr *constant linear velocity*, og nettopp fordi CLV-platen avleses med konstant *lineær* hastighet, blir det plass til mer informasjon i de ytterste sporene enn de innerste. En CLV-plate kan inneholde opptil én time levende video på hver side. Til gjengjeld kan det ikke stoppes på enkeltbilder. De funksjonene som er tilgjengelig for en CLV-plate er foruten vanlig avspilling, bare «scan» (dvs. rask «spoling») og «finn kapittel» (materialet på en Laservision-plate kan være delt i opptil 99 «kapitler» pr. side).

CLV-plater brukes til spillefilm, musikkvideoer, m.m. og CAV til interaktiv video-anvendelser. De fleste allment tilgjengelige plater er dermed CLV-plater, men det finnes noen CAV-utgivelser for hjemmemarkedet også, med temaer som varierer fra golf og gamle jernbaner til Asterix og Pink Floyd. CAV- og CLV-plater kan spilles om hverandre på alle spillere som artikkelforfatteren kjenner til. CLV kalles også «long play» og CAV «active play».

Laservision-systemet er utviklet av Philips. Andre utstyrsfabrikanter er bl.a. Sony og Pioneer. I Norge har noen få institusjoner og bedrifter gått til anskaffelse av Laservision-utstyr. I Norge selges trolig bare Philips og Pioneer. Sony regner med å selge spillere fra høsten av.

Digitale data og lyd på videoplate

Som nevnt finnes utstyr som kan lagre *digitale* data på en (CAV-)videoplate. Kapasiteten er da i størrelsesorden en halv til én Gbyte pr. side (dvs. omtrent det samme som for en 12-tommers *digital* plate). Kapasiteten for digitalisert *lyd* vil være avhengig av hvilken

lydkvalitet man vil ha. Med 4 Kbyte pr. sekund lyd kan en komme opp i 3 sekunder lyd pr. «frame», eller mer enn 50 timer pr. plateside. (Til sammenligning lagrer den digitale lydplaten Compact Disc bare én time lyd på en halv Gbyte, men da med meget høy kvalitet og i stereo.)

Laservision-utstyr

På utstyrssiden skiller en mellom hjemmespillere og profesjonelle spillere. De profesjonelle er raskere og mer robuste enn hjemmespillerne. De profesjonelle spillerne har en inngang (som regel RS-232) slik at de kan kobles mot og styres fra en datamaskin, eller de kan ha egen prosessor. Etter hvert har også mange hjemmespillere RS-232-inngang. Philips' rene hjemmespiller koster ca. kr. 7000 og hjemmespiller med RS-232-inngang ca. kr. 10.000. Philips' profesjonelle spillere koster fra ca. 16.000 og oppover. Sony opplyser at deres (profesjonelle) spiller vil starte i pris på ca. kr. 18.000. I Storbritannia selges hjemmespillere helt ned til £229 (pris vinter 1984).

En meget interessant spiller er den nye hjemmemodellen fra Pioneer til ca. £500 (pris Storbritannia). Pioneer-spilleren er kun 42 cm bred og har frontmating av platen, slik at den kan stables. Med tilhørende mikromaskin, klar for kobling mot fjernsynet, koster den ca. £800.

Aksesstiden, eller tiden spilleren bruker for å finne fram et bestemt bilde fra platen, varierer med de forskjellige spillere og med den fysiske avstanden på platen til det aktuelle bildet. Gjennomsnittlig aksesstid ligger typisk på 2-5 sekunder. Aksesstid for de fysiske nærmeste bilder er neglisjerbar.

I hvert fall Laservision-systemet kan gi en bildekvalitet som overstiger det selv gode TV-skjermer kan gi. Først ved visning på høyoppløselige monitører blir videoplaten ytt full rettferdighet. Prisen på slike monitører ligger på det mangedobbelte av vanlige fjernsyn. Selv om selve videosignalet egentlig inneholder flere linjer, regner en at et godt fjernsyn har en (vertikal) oppløsning på 250-300 linjer, og signalet på videoplaten ca. 420-430.

Utstyrskonfigurasjoner og tilleggsutstyr

En utstyrskonfigurasjon med mikromaskin og videoplatespiller vil kunne ha varierende sammensetning, avhengig av om en har én eller to skjermer. Det enkleste konseptet er det med to skjermer. Da har mikromaskinen sin egen dataskjerm, og videoplatespilleren sender sitt signal til en annen monitor (f.eks. et vanlig fjernsyn). Skal en bruke én skjerm, må de to bildesignalene på en eller annen måte blandes. Alternativene er da enten å bruke en video-mikseboks, eller å ha et eget kretskort i mikromaskinen som kan ta videosignalet fra spilleren og blande med sitt eget. Slike kort kan kjøpes for en rekke maskiner. Prisen er i størrelsesorden 10.000 kr. Har man mikromaskin med dataskjerm, og man vil bruke én skjerm, må man ofte skifte ut skjermen med en av bedre kvalitet.

Valget mellom én og to skjermer er avhengig av den aktuelle

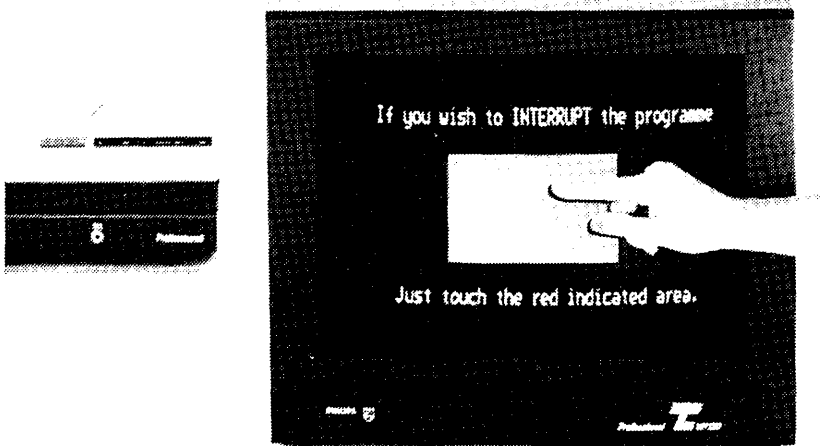
anvendelse. Generelt kan det sies at det kan være anstrengende å flytte blikket fram og tilbake mellom to skjermer, men på den annen side vil det ofte være naturlig å ha tekst og menyer på én skjerm og bilder på en annen. Et argument for én skjerm kan være ønsket om å legge grafikk fra mikromaskinen over bildene fra platen.

Hvis man vil ha tekst og enklere grafikk over bildene, kan dette løses på en annen måte, nemlig ved å anskaffe platespiller med teletekst-mulighet. (Teletekst-bokstaver og -grafikk er slikt som blir brukt i tekst-TV.) Da kan en sende koder for teletekst fra mikromaskinen til spilleren, og denne vil lage teletekst og legge over sitt eget bildesignal.

Det finnes også mulighet for å koble flere spillere i en serie, slik at bildemateriale kan hentes fra flere plate(side)r om hverandre. Kobling kan også benyttes til å få ned aksesstiden i applikasjoner der det er viktig med kontinuerlige overganger i visningen av materialet. Dette kan oppnås med to spillere med samme plate.

En del leverandører produserer integrert utstyr og selger maskin, spiller og monitor i én pakke. Det franske firmaet Thomson-CSF er i gang med å utvikle en slik pakke, komplett med læreprogrammer.

Interaksjonen med maskinen og spilleren kan foregå på andre måter



Philips berøringsfølsom skjerm.

enn via tastaturet, nemlig vha. mus eller en berøringsfølsom skjerm. Annet tilleggsutstyr er «hardcopy»-enheter for kopi av videobildet. Skal man ha en viss kvalitet, er slikt utstyr dyrt, men det finnes også rimelig utstyr på markedet. Mitsubishi produserer en videoprinter for NTSC-signal til \$390. På 15 sekunder lager den utskrifter i svart/hvitt med 16 gråtoner og en oppløsning på 280×234 punkter.

Et enkelt alternativ er å ta kopi vha. en vanlig videobåndspiller. Den har jo også den fordel at den kan kopiere levende video. Stillbilder kan f.eks. gis et halvt minutt spilletid hver.

Programvare

Programmer for interaktive videosystemer kan skrives i vanlige programmeringsspråk, og mange bruker da også i dag språk som Pascal og C. Det er imidlertid utviklet en rekke spesialspråk for interaktiv video. Flere av disse er videreføring av språk laget for datamaskinstøttet undervisning uten video. De fleste slike *forfatterspråk* (engelsk: authoring systems) er nokså kostbare, men artikkelforfatteren vil nevne et eksempel på et rimelig, nemlig Microtext. Microtext for videoplate og IBM-kompatible PC-er består av to deler – en generell del og en videoplatedel. Den generelle delen, som kan brukes til å skrive læreprogrammer uten video, koster £250 (britisk pris). Videoplatedelen kommer på et eget kretskort, som også mikser bildene fra PC-en og spilleren inn på én skjerm. Denne koster £950.

Dessverre er dagens forfatterspråk (Microtext inkludert) nokså primitive – mer på BASIC-nivå. Det foregår imidlertid interessante prosjekter ved en del amerikanske universiteter: Carnegie-Mellon, Brown (Hypermedia-prosjektet) og Stanford. Xerox har etter sigende konstruert et avansert forfatterspråk, men har støtt på arbeidspolitiske problemer (kommet inn på yrkesgruppers enemerker).

Produksjon av videoplater

Skal man ha et bilde- eller filmmateriale inn på videoplate, er det flere veier å gå. For stillbilder er det gunstig å ta materialet via 16 mm film, og så senere få filmen scannet over på masterbåndet til platen. Masterbåndet er et videobånd av høy kvalitet (én-toms). Å scanne enkeltbilder, f.eks. dias, rett inn på videobånd, er en kostbar prosess. Det er rimeligere med film siden en da kan ta hånd om avfotograferingen selv. Med et 16 mm kamera for enkeltbilder kan en fotografere både dias, papirbilder, negativ, lysbilder, dokument sider, kart, osv.

Å ta materiale direkte fra uperforert *mikrofilm* til masterbåndet kan være vanskelig. Problemet er framføringen, og at bildene kan være upresist plassert på filmen.

Til filmede sekvenser kan en bruke videokamera, men ofte er 16 mm film også her å foretrekke. Hvis video brukes, skal en være oppmerksom på følgende: Så sant en ikke har virkelig profesjonelt utstyr, blir kvaliteten lavere enn for film. I mange tilfeller er dette fullt akseptabelt. Et annet hensyn å ta er om en ønsker å kunne fryse den filmede

handlingen under avspillingen av platen. Da bør nemlig videoopptaket gjennom en spesiell prosess som kalles «correlation», ellers vil bildene flimre. Grunnen er at et videobilde egentlig består av to «halvbilder» tatt med et 50-dels sekunds mellomrom. Et halvbilde består av annenhver scanlinje i et helt bilde. Selv om en tar materiale opp på video, må også det i siste omgang over på én-toms bånd.

Dersom filmmateriale allerede eksisterer, kan det selvsagt scannes over på masterbåndet direkte fra filmen.

Enkelte hevder at en må bruke 35 mm film for å få optimal kvalitet på bildematerialet, men langt fra alle er enige i det.

Ved opptak av bildemateriale bør en være oppmerksom på at en TV-skjerm kutter bildet noe i kantene. Etter hvert som et fjernsyn blir eldre, forsterkes gjerne denne effekten. Scanneren stjeler dessuten ørlite ved overføringen til masterbåndet. Noe kan også forsvinne p.g.a. forskjellig form på bildeflatene. En fjernsynsskjerm har høyde og bredde i forholdet 3:4.

Norsk Film oppgir at pris for scanning av film til videobånd er kr. 1200 + moms pr. arbeidstime. Det lønner seg altså å gjøre mest mulig av redigeringsarbeidet på selve filmen. Arbeidstiden ved scanning av et godt redigert materiale vil ifølge Norsk Film gjennomsnittlig være lik spilletiden multiplisert med en faktor på 1,8.

Bl.a. det britiske firmaet Visual Data Systems påtar seg produksjon av plater for kunder i Norge. Framstilling av én plateside fra et masterbånd koster vel 26.000 kr. Prisen inkluderer 10 kopier. Ytterligere kopier koster fra kr. 120 til kr. 350 avhengig av antall. Inndeling av materialet i «kapitler» koster 1500-3500 kr. ekstra. Ifølge Visual Data Systems går prisene på framstilling av videoplater stadig nedover.

Man må i Europa regne med 3-4 ukers framstillingstid. I USA er det mulig å få laget en videoplate i løpet av en dag.

Det er også mulig ut fra masterbåndet å få framstilt en rimeligere versjon av platen, men da i bare én kopi. Et firma i USA tilbyr en slik tjeneste for ca. \$300, trolig pr. plateside, og ganske sikkert i NTSC-form. Det er grunn til å tro at dette gjøres med Laservision-kompatibelt write-once-utstyr som beskrevet under «Write-once videoplater».

Flere firmaer/institusjoner publiserer i dag materiale via videoplate, også digitale data.

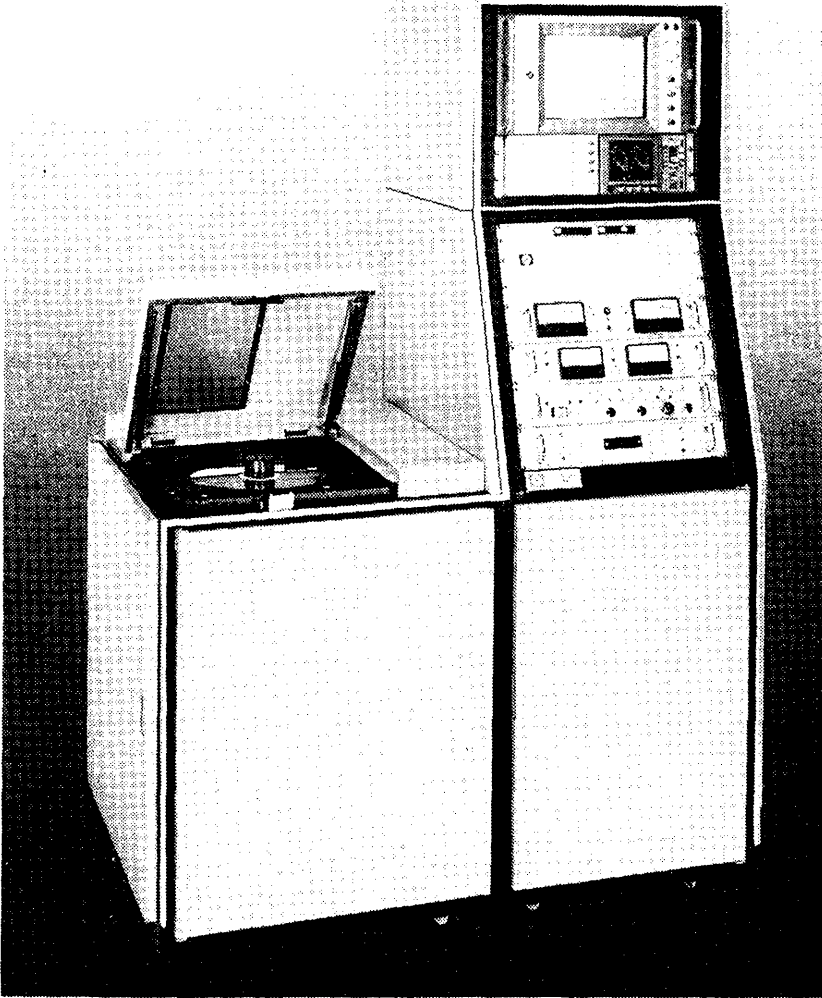
Interaktivt videobånd

Selv om denne artikkelen handler om plateteknologi, er det på sin plass å nevne et rimeligere alternativ til interaktiv video, nemlig interaktivt videobånd. Det finnes nemlig også videobåndopptakere som kan styres fra en datamaskin. Den store fordelene med bånd istedenfor plate er, i tillegg til produksjonsprisen, muligheten for å redigere og eksperimentere før det endelige resultat foreligger. Ulempene er høyere aksessetid, mer planlegging av rekkefølgen av materialet for å unngå for mye spoling, lavere bildekvalitet, slitasje og adskillig lavere kapasitet på

stillbilder (hvert bilde må vare en stund). For mange anvendelser er imidlertid interaktivt videobånd fullt ut tilfredsstillende.

Write-once videoplater

Matsushita (Panasonic) produserer write-once-utstyr i en serie med fire modeller (NTSC-standard). Den mest avanserte kan både spille inn og spille av levende video. Den neste modellen kan bare spille inn



Write-once videoplateutstyr fra Optical Disc Corporation.

stillbilder, men spille av levende video (dvs. mulighet for animasjon). De to «simpleste» modellene er kun for stillbilder. Spillerne lagrer bildene på 8-toms plater. Kapasiteten varierer fra 10.000 til 24.000 bilder pr. plate. Prisen i USA på de fire modellene er henholdsvis \$34.900, \$24.900 og \$18.900 for de to siste (kilde for prisene: Byte, juli 1984).

Optical Disc Corporation har utstyr som lager Laservision-kompatible plater (fortsatt NTSC-standard). Dermed kan platene spilles på en vanlig Laservision-spiller etterpå. Alternativt kan utstyret brukes til å produsere masterbånd for pressing av read-only-kopier.

Overskrivbare videoplater har artikkelforfatteren ikke hørt noe om.

Read-only hybridplater

Disse er allerede nevnt. Det dreier seg om analoge videoplater med digital informasjon lagt inn i videosignalet. Det finnes allerede en rekke systemer for amerikansk videosignal. I de amerikanske systemene kan opptil 480 linjer pr. «frame» brukes til å lagre opptil 25 bytes hver, dvs. totalt 12 Kbyte pr. «frame». Dette krever selvsagt en ekstra prosess i produksjonen av platen og spesielle spillere som er i stand til å dekode den digitale informasjonen igjen (f.eks. nivå 4-spillere).

Sonys spiller som er ventet på det norske marked i høst, skal kunne spille plater med digitalisert informasjon, også lyd. Philips' spiller for BBCs Domesday-prosjekt er annonsert å komme mot slutten av 1986 (Storbritannia).

De mest aktuelle anvendelsene av denne teknikken er:

- Lagring av store tekstmengder.
- Lagring av programvare. Interaktiv video-programmer kan dermed lagres sammen med materialet de bruker.
- Lagring av digitalisert lyd, f.eks. kommentarer til stillbilder. Uten denne teknikken krever en 10-sekunders kommentar til et stillbilde at bildet filmes i like mange sekunder, dvs. 250 «frames». Ved å lagre stillbildet som én «frame» og lyden i digitalisert form, kan en klare seg med 5-10 «frames».

Read-only digitale optiske plater

CD-ROM

Denne teknologien bygger direkte på Compact Disc-teknologien. Compact Disc (CD) er som kjent en slags digital grammofonplate for lyd. Den digitale informasjonen på en CD leses vha. en laser i spilleren, og følgelig er CD en digital optisk plate. Lyd lagres på platen som en sekvens av 16-sifrede binære tall (14 + 2 for feilkorrigerings). Disse tallene beskriver med stor nøyaktighet lydsvingningene. Hvert sekund lyd i hver av de to stereokanalene representeres vha. vel 44.000 slike tall, og platen rommer opptil én time lyd (egentlig er kapasiteten enda

høyere).

I prinsippet kan imidlertid den digitale informasjonen på platen være hva som helst, ikke bare lyd. CD-en kan altså benyttes som et generelt read-only datalagringsmedium. Samme teknologi som brukes i CD-spillere kan benyttes i en digital platestasjon. Navnet CD-ROM har platen fått ganske enkelt fordi den er en ROM, dvs. «Read Only Memory». (Riktignok brukes ROM vanligvis om *internt* read-only-minne i en datamaskin, mens CD-ROM mer blir å betrakte som *sekundærlager* på linje med disk og diskett.)

En CD-ROM er 120 mm (4,72 tommer) i diameter. Data legges inn i et CLV-format (se read-only videoplate), og kapasiteten er 550-650 Mbyte. På grunn av CLV-formatet er aksesstiden noe lav, nemlig ca. ett sekund. (For kapasitet målt i bilder, se under write-once digitale plater.)

Fordelene med CD-ROM er at teknologien er velp prøvd og rimelig, og at platen er pålitelig og robust. Standardiseringsarbeidet er riktignok ikke helt fullført, men ventelig vil enighet bli nådd innen et halvt års tid.

Blant firmaer som leverer CD-ROM-utstyr er Hitachi, Philips, Sony og Matsushita (Panasonic). Hitachi Sales Scandinavia oppgir en pris på ca. \$1000 for sin spiller inkludert de nødvendige kort for IBM PC, men da i et antall på over 1000 (!). Vanlig utsalgspris vil sannsynligvis være ca. kr. 15-20.000.

I tillegg vil en rekke firmaer tilby integrerte systemer. Bl.a. kommer nå Digital med CD-ROM for sin MicroVAX. Også Apple og Atari har annonsert systemer. Atari oppgir en pris på \$500 for sin spiller, og da skal den også kunne brukes i stereoanlegget.

Skal man produsere sin egen CD-ROM, må man tilrettelegge data på magnetbånd i et bestemt format. En CD-ROMs kapasitet tilsvarer ca. 15 vanlige 2400-fots magnetbånd med tetthet 1600 bpi. Produksjonstiden for en CD-ROM er for tiden ca. 30 dager. Tiden vil neppe gå særlig ned med det første. Det er et begrenset antall pressefabrikker i verden. I USA er det bare én i dag. På forespørsel til Hitachi, opplyser dette firmaet at norske kunder vil få sine plater presset i Japan. Pris for pressing ligger på kr. 50-75.000 pr. plate og kr. 50-100 pr. kopi.

CD-ROM har et stort potensiale som medium for elektronisk publisering. Platens kapasitet gjør f.eks. at innholdet kan forsynes med større og bedre indekser, noe som vil gjøre tilgangen til materialet bedre. En attraktiv tanke er også å legge data og tilhørende søkeprogram sammen på platen. Statens Rasjonaliseringsdirektorat ser muligheter i kombinasjonen SIFT og CD-ROM.

Materiale som vil egne seg for distribuering via CD-ROM er tekst, linjetegninger, grafikk, lyd og programvare.

The Library Corporation i USA publiserer nå MARC-filer med bibliografisk materiale på CD-ROM. Materialet oppdateres kvartalsvis, men det er planlagt å gå over til ukentlige oppdateringer i høst. Utstyr beregnet for lesing og bruk av MARC-platene kan kjøpes for \$3000. Utstyrskonfigurasjonen består av en IBM PC med monitor og en

Hitachi CD-ROM-stasjon. Dette er foreløpig den eneste kommersielle anvendelsen av CD-ROM som publiseringsmedium.

OROM

OROM er et konkurrerende system til CD-ROM. (OROM betyr simpelthen Optical ROM, og for ordens skyld gjøres det oppmerksom på at OROM av enkelte brukes som en generell betegnelse på all optisk ROM.) Platen har en litt større diameter (5,25 tommer). Data lagres i CAV-format, og aksesstiden er dermed presset ned i ca. ett millisekund. Dette går noe på bekostning av kapasiteten, som er ca. 400 Mbyte.

I motsetning til CD-ROM er det snakk om at det kan komme write-once- og overskrivbare utgaver senere.

3M produserer plater, men kilden til disse opplysningene kjente hverken til utstyrslleverandører, annonserte produkter eller standardiseringsarbeid. Det er aksesstiden som i mange anvendelser kan gjøre OROM attraktiv.

Et svært viktig moment er hva IBM vil satse på. Det går rykter om at IBM vil bruke digitale read-only-plater til å distribuere programvare for sine PC-er. Et annet rykte forteller at IBM har kjøpt et oppslagsverk som skal selges på plate. Det sies at IBM i hvert fall støtter OROM-standarden. Som nevnt finnes det allerede CD-ROM-utstyr som kan brukes mot IBM PC.

Andre typer digitale read-only-plater

Det finnes også andre typer plater, bla. Drexon Laser Card, som er en plate formet som et kredittkort.

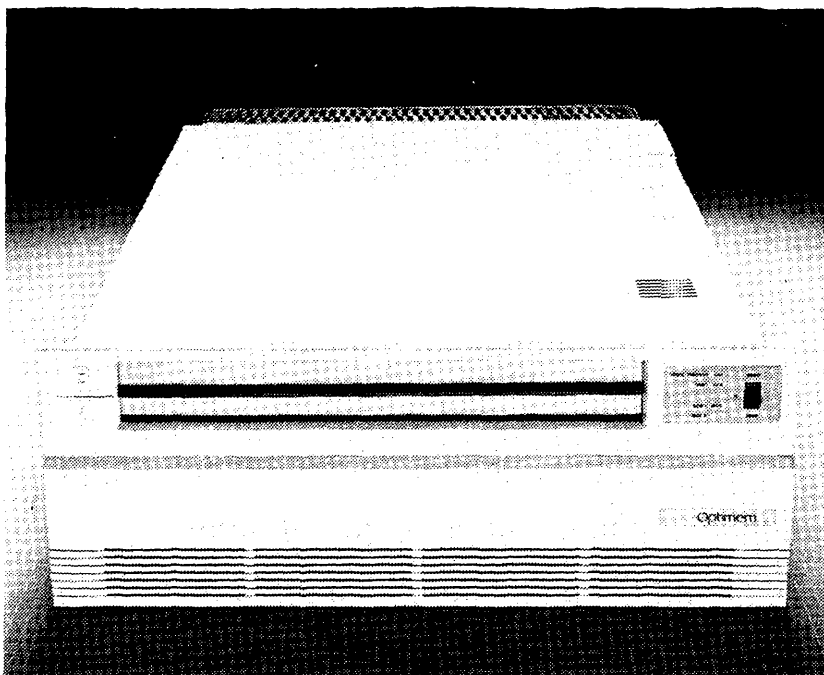
Write-once digitale optiske plater

Anvendelser som krever hurtigere oppdatering enn read-only-media, vil kunne dra mer nytte av write-once-teknologi. Digitale write-once-plater har også høy kapasitet, men er ikke egnet for mangfoldiggjøring. Distribusjon må derfor baseres på telekommunikasjon istedenfor utsendelse av kopier. Write-once-utstyr vil typisk stå som sentralt lager i nettverk, for eksempel i lokale nettverk.

Systemene som finnes, er lite standardiserte. Vi vil kunne få på markedet plater av de forskjelligste størrelser (4,75, 5,25, 8, 12 og 14 tommer), materialer (glass, aluminium, plast), osv. For 12-tommers plater er man stort sett bare blitt enige om signal-støy-forhold, festeanordningen og fasongen på hullet (!). Kommersielt tilgjengelige systemer produseres av bl.a. Thomson (Gigadisc), Philips (Megadoc) og Storage Technology. Det er imidlertid uklart om Storage Tek vil overleve.

Lagingskapasiteten på 12-tommers plater varierer fra 0,5 til 3,6 Gbyte. Størst kapasitet har utstyr fra Toshiba, med 1,8 Gbyte pr. plateside. Toshiba har allerede levert 500 systemer til prøveformål til selskaper over hele verden. Introduksjonspris: \$12.000.

Typisk 12-tommers utstyr blir produsert av Shugart (Optimem).



Write-once digital platestasjon fra Optimem.

Optimem-platen har en kapasitet på én Gbyte, men det er annonsert en ny modell med dobbel kapasitet til neste år. Aksestid er 3 millisekunder for neste spor, gjennomsnittlig 145 ms og maksimum 245 ms. Overføringshastigheten til/fra disken er 5 Mbit pr. sekund og vil gå opp til 10 Mbit («uformatert»). Det finnes interface for IBM PC (3 produsenter) og MicroVAX, og det finnes eller kommer for VAX og andre mini-maskiner.

Mange utstyrspordusenter satser på store integrerte systemer. Et eksempel er Philips med sitt Megadoc-system for kontorautomatisering. I store systemer brukes gjerne «jukebokser» for å få opp lagringskapasiteten. Da en jukeboks har mekanisk framhenting av plater, vil aksestiden for andre plater enn den som ligger i spilleren gjerne ligge på 10-20 sekunder.

I de fleste anvendelser vil det finnes data som endres raskt og ofte, om ikke annet filkataloger, hjelpeområder for diverse programmer, osv. I store integrerte systemer, og i mange anvendelser ellers, vil derfor vanlig magnetisk lager (disk/diskett) bli brukt sammen med write-once-lager. (Se ellers referat fra foredraget «Automation of a Large-Scale Storage and Retrieval System ...» fra «The 1985 International Video-disc, Optical Disk and CD-ROM Conference» annensteds i bladet.)

Et stort problem med digitale plater i dag, spesielt write-once, er mangelen på programvare til å «drive» platestasjonene. Denne programvaren må organisere filkatalogen, dvs. oversikten over hva som ligger på platen, og den må være i stand til å tolke diverse feilmeldinger og statusopplysninger sendt fra stasjonen. En kan altså ikke uten videre gå og kjøpe en digital platestasjon og koble den til som en annen disk. (For vanlige disketter ligger slik programvare allerede i operativsystemene.)

Enkelte firmaer satser på «pakker» bestående av platestasjon, kabler, interface-kort og programvare. Et eksempel er amerikanske Wyvern som selger Optimem-utstyr tilpasset VAX. Pris for utstyret er ca. \$30.000.

I Norge står Norsk Regnesentral foran innkjøp av en write-once platestasjon. Valget står mellom Thomson (Gigadisc) og Optimem. Institusjonen tenker i første rekke å bruke platen til kartmateriale.

Kapasitet for bilder

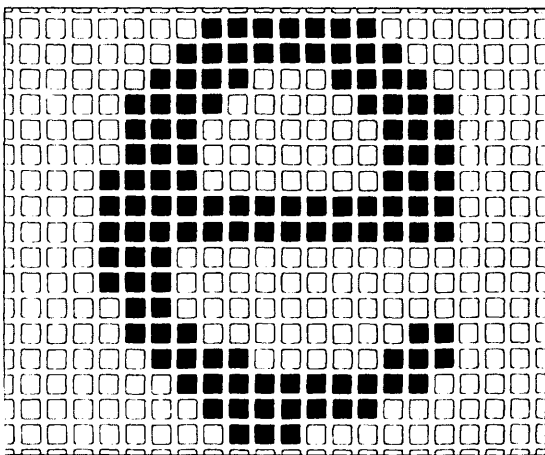
Det store spørsmål for mange er hvor mye bildemateriale en kan få inn på en digital plate. Dette vil avhenge av hvilken oppløsning en vil ha, hvor mange gråtoner eller fargetoner, og i hvilken grad de digitaliserte data lar seg komprimere.

Det nevnte Megadoc-systemet er beregnet for lagring av A4-format kontordokumenter i rent svart/hvitt, og platene har en kapasitet på ca. 1 Gbyte. Philips opplyser at Megadoc har en kapasitet på opptil 30.000 dokumentsider pr. plate. Da brukes en oppløsning på 200 punkter pr. tomme (8 punkter pr. mm), og det antas at data kan komprimeres til 6% av sin opprinnelige mengde. Hvor mye data som kan komprimeres varierer svært mye. Tettskrevne trykte sider vil kanskje bare gi en faktor på 17% mens enkle tegninger kan komme ned i 2%.

200 punkter pr. tomme er adekvat for vanlig tekst og tegninger, men enkelte anvendelser kan kreve mer. Library of Congress scanner sine kartotek kort med 480 punkter pr. tomme (19 punkter pr. mm) for å få god nok kvalitet for enkelte ikke-latinske skrifttyper. Grafisk industri trenger også en oppløsning i denne størrelsesorden (20 punkter pr. mm).

Hvis man ved digitalisering dobler antall punkter pr. tomme, øker datamengden til det firdobbelte. Går man over fra rent svart/hvitt til f.eks. 16 gråtoner, øker også datamengden til det firdobbelte. 256 gråtoner gir en åttedobling. Samme økning gjelder for samme antall fargetoner. Skal man gjengi fargebilder med god kvalitet, vil en digital optisk plate bare kunne lagre i størrelsesorden hundre stykker.

Når en kan oppnå en såpass god komprimeringsfaktor som nevnt for rent svart/hvitt, er det fordi bildene vil inneholde store ensfargede områder. En vanlig komprimeringsteknikk er istedenfor å lagre en lang sekvens av hvite punkter som en lang sekvens av f.eks. 0-ere, å representere sekvensen vha. lengden på sekvensen. En slik teknikk brukt alene krever imidlertid at platen har en god feilrate. Én bit feil i et ukomprimert bilde vil sannsynligvis ikke sees, men i et komprimert



En maskinskrevet «e» digitalisert med en oppløsning på 200 punkter pr. tomme.

bilde kan følgen bli at dekomprimeringen kommer helt ut av fase og feilen sprer seg til resten av bildet. Feilkorrigerende koding i tillegg kan motvirke dette igjen, men all komprimering og koding har sin pris i datakraft, hvis ikke spesielle kretser kan gjøre arbeidet.

Kapasiteten for lyd burde framgå av gjennomgangen under read-only videoplate. Disse kapasitetsbetraktningene gjelder selvsagt alle digitale plater, ikke bare write-once.

Innlesing av bilder

Scannere er fortsatt relativt dyre, men en ny og rimeligere teknologi er i ferd med å overta. Den nye generasjonen scannere er basert på bruk av såkalte «charge-coupled devices», forkortet til CCD. Som et eksempel på en rimelig CCD-scanner kan nevnes Microtek MS-200 som koster \$1700 FOB Taiwan. Den er beregnet brukt mot IBM PC (XT eller AT), og interfacekort koster \$100. Den kan ta formater opptil $8,5 \times 24$ tommer, dvs. et 60 cm langt ark i A4-bredde. Oppløsningen er 200 punkter pr. tomme.

Digital plate + mikrografiske medier

Mange mener at vi i tiden framover vil se mye til hybridssystemer som kombinerer digital plate og mikrografiske medier. Også i Japan, som ellers er sterkt rettet mot den nye teknologien, hevder røster dette. I slike systemer vil data gjerne gå begge veier, både fra mikrografikk til plate og fra plate til mikrografikk. Det er snakk om både eksisterende og nye mikrografiske media. Bl.a. Kodak og Canon utvikler systemer

for plate og mikrofilm. (Se referat fra «The 1985 International Video-disc, Optical Disk and CD-ROM Conference and Exhibition»).

Mens vi beveger oss i den digitale write-once-platens grenseland kan også nevnes DOC Wheel fra nederlandske DOCdata. Vha. av 100 kassetter med optisk *bånd* montert i en karusell, når dette systemet opp i en total lagringskapasitet på 600 Gbyte. Aksestiden vil selvsagt ikke være like god som for plater. En viktig anvendelse av systemet vil være som back-up-medium.

Overskrivbare digitale optiske plater

Det er usikkert når de overskrivbare platene kommer på markedet. Mange gjetter på 1987. Optimem har annonsert overskrivbart utstyr midt i 1986. Optimems system vil ha 5.25-toms plater med kapasitet på 4-500 Mbyte. 3M utvikler utstyr med samme platestørrelse og kapasitet 300-500 Mbyte. Til sammen arbeider ca. 15 firmaer med utvikling av overskrivbart utstyr.

Status og framtid for optisk lagring

Optisk plate-teknologien har nok ikke fått den utbredelse som man tidligere trodde. En kan vel si at den har lidd under det amerikanerne kaller «technology push», dvs. at forskning og utvikling har vært gjort *før* potensielle markeder er undersøkt, eller markeder skapt. Det er imidlertid ikke noen tegn til stagnasjon. Det er heller snakk om en forsinkelse.

Godt utbygde kommunikasjonsmuligheter er en forutsetning for full utnyttelse av den optiske teknologiens potensiale. Effektiv overføring av bilder over lange avstander, for ikke å snakke om film, krever en helt annen kapasitet enn telenettet kan tilby i dag. Dessverre kan ikke artikkelen gå inn på denne siden av saken. Det får være nok å nevne at løsningen på kapasitetsproblemene kan ligge i overføring via optiske fibre.

Er datalingvistikk del av anvendt språkvitenskap?

Lars Sigfred Evensen

Jeg vil forsøke å bruke dette bidraget til ANLAs symposium om datalingvistikk (Bergen 12.-13. oktober 1984) til å skape en noe klarere forståelse av den kanskje diffuse termen *anvendt språkvitenskap* slik denne brukes i datalingvistisk sammenheng. Gjennom dette ønsker jeg dels å belyse forholdet mellom grunnforskning og anvendt forskning i tilknytning til datalingvistikk og dels å skissere enkelte framtidige forskningsoppgaver i anvendt språkvitenskap.

Anvendt språkvitenskap

Begrepet 'applied linguistics' ble først tatt i alminnelig bruk i USA under 2. verdenskrig. Termen er dermed relativt ny. Dette er en av flere grunner til at det eksisterer en viss usikkerhet med hensyn til hva termen omfatter. En kanskje viktigere grunn ligger i det forhold at ikke alle forskere i anvendt språkvitenskap bruker termen på samme måte.

I det følgende skal jeg forsøke å skissere hva jeg oppfatter som to hovedretninger i synet på anvendt språkvitenskap (AS). Innenfor den ene retningen blir disiplinen AS oppfattet som i hovedsak teoridrevet (les: lingvistikkdrevet); innenfor den andre retningen blir AS oppfattet som hovedsakelig problemdrevet. Av disse tilnæringsmåtene er det den første som har lengst tradisjon og som har hatt størst påvirkningskraft i forhold til fagets profil.

AS som teoridrevet disiplin

USA sto etter Pearl Harbor overfor en situasjon hvor det var akutt behov for rask opplæring av et stort antall personer i fremmedspråk, herunder språk som amerikanere flest til dels aldri hadde hørt om (Moulton 1966). Den amerikanske regjering valgte å sette i gang et storstilt og kostbart program hvor datidas ledende lingvister i Bloomfield-tradisjonen fikk i oppdrag å delta i utvikling av intensivkurs i en lang rekke fremmedspråk.

Det var naturlig for disse forskerne å betrakte sin virksomhet som strukturalistisk lingvistisk teori omsatt i praktisk språkundervisning, hovedsakelig gjennom deskriptive, kontrastive analyser av morsmålet engelsk og de respektive målspråk. I en historisk oversikt over AS skriver Mackey (1966:247) om termen 'applied linguistics' at den ble tatt i bruk som «an effort to find practical applications for 'modern, scientific linguistics'».

Særlig rimelig var en slik sjølførståelse ut fra datidas optimistiske positivisme. For første gang i menneskehetens historie, mente man,

eksisterte det pålitelige, vitenskapelige analyser basert på språklige korpora som ble underkastet rigorøse 'discovery procedures'. Den strukturalistiske taksonomien ga gjennom sin systematikk, klarhet og etterprøvnbarhet det hittil beste utgangspunkt for et prinsipielt basert utvalg av språklig lærestoff; en progressiv ordning av dette stoffet etter strukturell kompleksitet; intensiv undervisning med utstrakt bruk av bundne drilløvinger og objektiv testing hvor lingvistens kategorier kunne inngå som 'discrete-point items'. Denne tilnæringsmåten ble opplevd som så vellykket at da det første internasjonale tidsskriftet i AS – *Language Learning* – kom i 1948, argumenterte en på lederplass med behovet for et eget tidsskrift orientert mot 'the pedagogical implications of linguistic science'.

En av hovedkildene til suksess lå i at en i det lingvistiske arbeidet avdekket strukturelle ulikheter. Dermed mente man også å oppnå systematisert analyse av problemområdene i et innlæringsforløp. I 1945 kunne således Fries si:

«The most effective materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner.» (Fries 1945:9)

Senere ble et slikt kontrastivt perspektiv lagt eksplisitt til grunn for tidligere klassikere i AS som Lados (1957) *Linguistics across cultures: applied linguistics for language teachers*.

Etter den teoridrevne modellen blir det naturlig å gi AS en underordnet rolle som en disiplin innenfor faget lingvistik. Da lingvisten Bertil Malmberg tiltrådte som redaktør av *IRAL* (The International Review of Applied Linguistics) i 1967, skrev han på lederplass om AS som «this new branch, among the numerous old ones – of linguistics».

Dette synet er også utbredt i Norge. I Hovdhaugens (1976) innføring i språkvitenskap skriver Trygve Sagen (s. 115): «På ett område har språkvitenskapen funnet en rent praktisk anvendelse, nemlig innenfor fremmedspråkundervisningen.» Videre omtales AS som «en forholdsvis vel definert disiplin innenfor språkvitenskapen.»

I en nyere innføringsbok (Bjorvand *et al* 1982:275) skriver Anne Hvenekilde at AS «bygger videre på grunnforskning innenfor språkvitenskapen og er opptatt av hvordan resultatene fra denne forskningen kan brukes som grunnlag for arbeid med praktiske oppgaver.» Også her blir AS omtalt som «egen fagdisiplin innenfor språkvitenskapen» (loc. cit.). At denne disiplinen er underordnet, går bl.a. fram av at omtalen av den først står i bokas Vedlegg 3.

Slike definisjoner tar alle utgangspunkt i et teoretisk fag – lingvistik – som så finner et bruksområde – (fremmed-) språkundervisning. Påvirkningsretningen er ensidig – fra teori til praksis. Videre er lingvistik implisitt å betrakte som tilstrekkelig betingelse for å løse oppgaver innenfor denne praksis. Uhøytidelig sammenfattet kan vi si at

de omtalte tilnæringsmåtene beskriver AS som det rørdningsnett gjennom hvilket en lingvistisk teori omsider når bakken.

Det er imidlertid problemer knyttet til en slik modell. Sol Saporta (1966:86) kritiserer den teoridrevne modellen for at en ikke tilstrekkelig klart skjelner mellom den undervisningsprosessen en elev gjennomgår og den lingvistiske strukturen i det som forhåpentligvis blir innlæringsprodukt. Dette skillet er helt sentralt også når en pedagogisk grammatikk skal skrives.

Chomsky (1966) framholder at det er lite rimelig om undervisningsproblemer skal løses med utgangspunkt i lingvistisk teori. Denne teorien er fragmentarisk og under stadig forvandling. I et moderne lingvistisk perspektiv har enkelte altså et langt mer edruelig syn på hvor langt en i lingvistisk forskning er kommet.

Et tredje problem framkommer gjennom faghistorisk analyse. Sjøl om lingvistisk teori unektelig sto sentralt i enkelte hovedaktørers bevissthet, var det faktisk slik at en hele tida hadde trukket inn også andre vitenskaper i det arbeidet som ble kalt AS – særlig (behavioristisk) psykologi, men også pedagogikk og antropologi (jf. Lados tittel ovenfor). Etter «kloakkmodellen» blir denne tverrfagligheten å betrakte som *ad hoc*. Det var også historisk sett et delvis utenomlingvistisk problem (en term som jeg vil definere som diskrepans mellom oppsatte mål og umiddelbart tilgjengelige midler) – behovet for undervisningsprogrammer – som satte fart i disiplinen utvikling.

Et fjerde problem har med forholdet mellom teoriutvikling og anvendelse å gjøre. Gjennom bruk av fagets historie som empiri blir det også klart at påvirkninga ikke er enveis, men dialektisk. Et eksempel kan belyse dette. I 60-åras USA var det utbredt bekymring over den svikt i leseferdighet en mente å kunne observere hos mange svarte storbyungdommer. Det ble satt i gang forskning om dette. En prøvde seg tidlig med såkalt kompensatorisk språktrening. Dette virket ikke (Labov 1972). En av de antatte grunnene var at en ikke kjente de språklige konvensjonene i det som var elevenes regionale sosiolekt - deres egentlige morsmål. En begynte å analysere konvensjonell språkbruk blant svarte (se f.eks. Wolfram 1969 og Labov 1969). For å gjøre en lang og spennende historie kort og kjedelig; resultatet ble sosiolingvistikkens gjennombrudd og 'variable rules' i teoretisk syntaks (Even- sen 1983).

Anvendt språkvitenskap som problemdrevet disiplin

Løsningen på de problemene som er skissert ovenfor synes å ligge i at en i større grad tar utgangspunkt i objektet for forskninga – praktiske problemer eller samfunnsoppgaver som har en språklig hovedkomponent. Allerede i 1969 sier Fraser & O'Donell i ei bok om morsmålsundervisning (s. xi) at

«It would be wrong to think of its (dvs. AS' – LSE) proper function

as being that of discovering practical applications for linguistics; this would imply that we should begin with certain theoretical answers and work back towards such subsequent questions that fitted these answers. Applied linguistics is in essence a problem-centred discipline. It starts, that is, by asking not how this or that insight into language might be employed, but rather how this or that practical language problem might be solved, whether the theoretical answers are ready to hand or not.»

Et slikt syn innebærer at AS blir mer forskningsorientert og sjølsten- dig i forhold til lingvistikk. Fraser og McDonell (loc.cit) uttrykker dette slik: «We see this discipline not as a sub-department of linguistics, but as one which, though historically and otherwise related to linguistics, exists in its own right.»

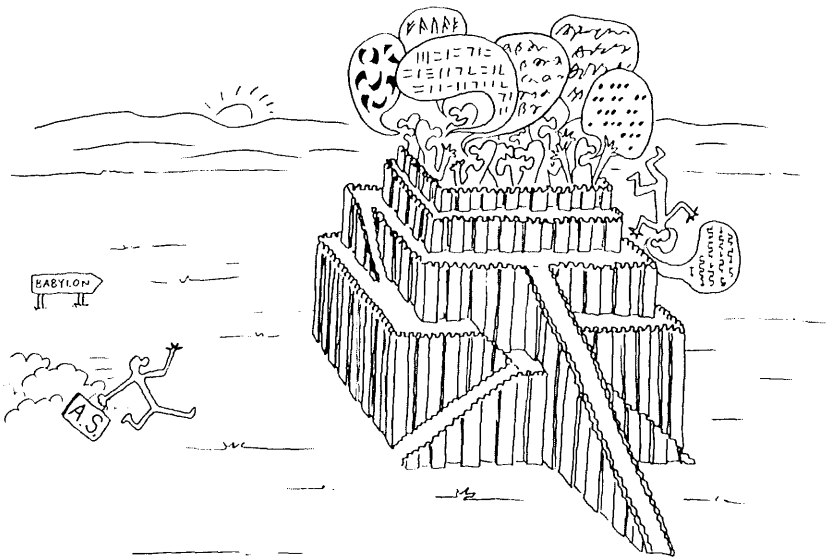
Problemorientert forskning blir svært ofte tverrfaglig. Dette skyldes at problemene (forskningsobjektet) er sammensatt av mer enn en komponent. Språkundervisning har bl.a. lingvistiske, psykologiske, pedagogiske og sosiologiske aspekter. Lingvistikkenes rolle som sentral støttedisiplin blir dermed å tilføre nødvendig, men ikke tilstrekkelig, innsikt.

AS får dermed også plass på linje med den en i ikke-humanistiske sammenhenger er vant til å gi anvendt forskning. Hos Ilstad, Paasche & Hovden (1977:12) beskrives anvendt forskning som den forskning «som har sitt utspring i konkrete problemer». I NAVFs årsmelding for 1979 formuleres perspektivet slik (s. 5): «Den anvendte forskning har som formål å skape ny viten med direkte betydning for praktisk arbeid.» I dette perspektivet har jeg lyst til å drøfte datalingvistikk og oppgaver innenfor denne.

Datalingvistikk

Jeg behøver ikke å gi en nærmere omtale over hvilke områder som går inn under betegnelsen 'datalingvistikk'. Her har Helge Dyvik (1984) gitt en utmerket oversikt. Det som imidlertid må presiseres i lys av foregående avsnitt er at også datalingvistikk må inndeles i grunnforskning og anvendt forskning. I datalingvistisk grunnforskning blir edb vanligvis brukt dels til utvikling eller testing av en operasjonalisert lingvistisk modell (simulering/utprøving av generativ kombinatorikk) eller til kvantitative analyser av språklige korpora (jf. Johansson 1981). I denne sammenhengen er det utvikling av lingvistisk (eller kognitiv) teori som er det primære formål for virksomheten.

I anvendt datalingvistikk blir edb hovedsakelig benyttet for å fram-skaffe forskningsbaserte løsninger på praktiske språkproblemer. I den resterende delen av denne artikkelen skal jeg forsøke å skissere enkelte datalingvistiske forskningsoppgaver av anvendt art. Det ligger da i kortene at en i ANLA-sammenheng etter mitt syn bør gi mest oppmerk-somhet til slike.



Tegning: Øystein Reigem.

Her bør det imidlertid understrekes at den dialektikken som jeg tidligere har framholdt for forholdet mellom generell og anvendt språkvitenskap innebærer at en ikke vil ha fordel av å trekke et for skarpt skille gjennom grenseflaten mellom de to. Dette blir særlig klart hvis vi tar for oss et av det moderne teknologisamfunnets hovedparadokser – teknologien som er i ferd med å stagnere på grunn av mangelfull tilgang på moderne humanistisk forskning. Mens en i hardware-sammenheng arbeider for å etablere femte-generasjons produkter, har en faktisk ikke noe etablert generasjonsbegrep knyttet til software. Dagens moderne høynivåspråk (f.eks. LISP og PROLOG) er basert på konsepter som dels er gamle (1960-åra blir vel å betrakte som en slags yngre steinalder i edb-sammenheng) og dels basert på retninger i logikken som betrakter naturlig språks flertydighet og fleksibilitet i forhold til ulike brukssammenhenger som en *svakhet*. Dette innebærer at programvaren har en rigiditet som gjør det vanskelig både å utnytte den store datakraften som ligger i moderne maskinvare og å løse de oppgavene som en vil bruke datakraft til å løse (f.eks. trekke slutninger med minst mulig forbruk av ressurser i form av algoritmer som programmene løper gjennom).

Vi har her et markert tilfelle av at edb-teknologi trenger lingvistisk kunnskap for å komme videre. Dette behovet har også et mer umarkert aspekt. Forskere i RUNIT-miljøet innrømmer f.eks. åpent at deres arbeid med spørsmål-svar-systemer er sterkt hemmet av mangel på lingvistisk innsikt.

Vi står altså i framtida overfor oppgaver av praktisk art som også betinger (data-)lingvistisk grunnforskning. Særlig er det de høyere nivåene i lingvistikens hierarki av beskrivelsesnivåer (pragmatikk, tekstnivå og semantikk), og samspillet mellom ulike nivåer, som må begripes bedre. Vi vil dermed igjen antakelig få se eksempler på at forskning omkring anvendte problemstillinger bidrar til utvikling av grunnforskning.

Dyvik (1984) har skissert flere sentrale forskningsoppgaver i anvendt datalingvistikk, f.eks. i tilknytning til maskinoversettelse (jf. EURO-TRA-prosjektet), spørsmål-svar-systemer og eksperter-systemer. Flere av disse områdene kunne omtales mer inngående. F.eks. kunne det være en interessant oppgave å utvikle eksperter-systemer til støtte for fagplanutvikling. Her ville seleksjon og ordning av spesifiserte innholdsenheter måtte gjennomføres bl.a. på bakgrunn av kunnskap om målgrupper (utviklingstrinn, motivasjon, sosial bakgrunn osv.) og behovsanalyse (hva slags kommunikasjon, i hvilke sammenhenger osv.). En kan diskutere hvorvidt en ønsker operasjonaliserte målsettinger, men på områdene ovenfor ville den presisjon som en datalingvistisk tilnæringsmåte medfører temmelig sikkert medføre hevet bevissthetsnivå omkring de beslutninger som *de facto* ligger bak enhver fagplan, også i dagens Norge.

Jeg skal likevel fare lett over de områdene som Dyvik har omtalt og supplere med oppgaver som synes sentrale sett fra mitt ståsted som forsker i anvendt språkvitenskap ved et institutt hvor språkundervisning er et sentralt felt. Jeg skal her illustrere med to oppgaver – forskning om bruk av interaktiv video i framtidens språklaboratorium og tekstlingvistiske performansanalyser av elevspråk.

I dagens forskning om språkundervisning er *kommunikativ kompetanse* det viktigste nøkkelordet i målsettingssammenheng. Dette målet innebærer at en ikke bare lærer et sett med lingvistiske enheter, strukturer og prosesser, men også hvordan slike brukes og fungerer i ulike kontekster. I et innlæringsstudio vil denne kompetansen bli lettere å utvikle hvis en har tilgang til en reell eller simulert kontekst. Her kommer video inn som et sentralt element. Dersom dette elementet skulle brukes alene, ville en imidlertid lett få rigiditet, elevpassivisering og en «mediet bestemmer»-effekt som ikke er ønskelig. I anvendt forskning om moderne språkundervisning er elevaktivitet, differensiering og elevautonomi viktige stikkord.

En av rammebetingelsene for morgendagens språkundervisning er at mange elever kommer til å ha en viss grunnleggende programmeringsferdighet. Dermed kommer interaktiv video inn som en fristende mulighet. Her kan elevene sjøl bestemme over større eller mindre deler av et undervisningsforløp.

Slike programmer fins i dag bære i svært beskjeden utstrekning. Dette skyldes ikke bare at mediet er nytt og relativt kostbart. Det er også

mange ingredienser som helt eller delvis mangler for å kunne lage vellykkete programmer: bedre kunnskap om samspillet mellom språkbruk og ikke-språklig kontekst; implementert kunnskap om hvordan en språklig kontekst er strukturert ut over setningsnivå (både for at et program skal kunne finne fram i sin egen informasjon og for at det skal kunne «forstå» og gi adekvat respons på elevens språklige aktivitet) og bedre kunnskap om innlæringsproblemer knyttet til slik undervisning, for bare å nevne noen sider ved problematikken. Her vil FoU i AS stå i skjæringsfeltet mellom lingvistikk, informatikk, filmvitenskap og språkpedagogikk.

Problematikken omkring interaktiv video peker også mot et annet problemområde som er en viktig forsknings- og utviklingsoppgave. Studiet av språklige strukturer og prosesser ut over setningsnivå er den sentrale oppgaven i tekstlingvistikk. Det tekstlingvistiske nivået er viktig både i moderne lingvistikk, i AI og i kunnskapsteknologi. Det har imidlertid også vist seg at sentrale innlæringsproblemer befinner seg på dette nivået i språkundervisning (Evensen 1985, Evensen & Lintermann Rygh under utg.). For anvendt forskning om elevspråk innebærer dette behov for tekstlingvistiske analyser.

Språklig produksjon korrelerer med en rekke variabler som dels er knyttet til den enkelte elev og dels til innlæringsssituasjonen. Videre innebærer et tekstlingvistisk studium at en er ute etter fenomener som er relativt lite hyppige. Begge disse faktorene taler for at en i slike studier bruker store og veldefinerte tekstmasser. Her kommer datalingvistikk inn som et interessant bidrag. I analysesammenheng er en imidlertid kommet kort ikke bare i studiet av elevspråk på dette nivået, men også i utviklinga av et implementerbart system av lingvistiske (konvensjonelle) regler.

Et av de problemene som elever står overfor når de skal lage en lengre språklig framstilling, er å overholde lite formaliserte konvensjoner om verbformers sekvens i løpende tekst. Det er en del av vår tekstkompetanse å vite at det umarkerte mønsteret i en fortelling er signalisert av løpende preteritum. Et slikt umarkert mønster kan imidlertid godt brytes for bestemte formål. Dersom det ikke er samsvar mellom verbal sekvens og underliggende kronologisk sekvens, kan vi f.eks. markere dette med bruk av pluskvamperfektum. Dette vil være et *motivert* brudd på temporal sekvens.

- (1) Kalle sto opp seint den dagen. Så gikk han på jobb etter at han hadde spist frokost

Elever foretar imidlertid ofte umotiverte brudd som kan ha uheldige kommunikative effekter. For den som ønsker å undersøke et slikt fenomen nærmere er situasjonen imidlertid dessverre den at en verken har god oversikt over hvor grensene går mellom motiverte og umotiverte brudd (se Lackstrom *et al* 1973 og Dijk 1977) eller programvare som er velegnet for å analysere det en faktisk *kan* formalisere på forskningas

nåværende stadium.

Sluttord

Jeg har i denne artikkelen prøvd å bidra til ANLAs arbeid med datalingvistikk ved å se datalingvistikk i forhold til en tilstrebte klargjøring av forholdet mellom generell lingvistikk og anvendt språkvitenskap. Denne drøftinga har jeg brukt til å skissere enkelte forskningsoppgaver som jeg ser som viktige i norsk anvendt datalingvistikk.

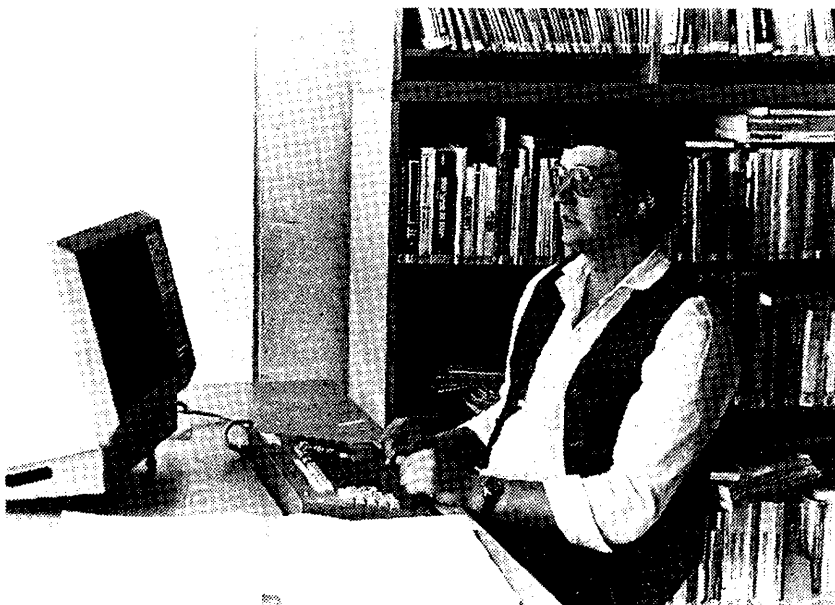
Dersom oppgavene løses på en tilfredsstillende måte, ser jeg grunn til å tro at både generell og anvendt forskning vil profitere på dette. En slik optimisme henger sammen med at jeg velger å se forholdet mellom anvendt og generell datalingvistikk som et dialektisk gjensidighetsforhold.

Jeg har også forsøkt å vise at det moderne informasjonssamfunnet utvikler motsetninger som bare kan løses gjennom opptrapping og oppvurdering av relevant humanistisk forskning. Dermed håper jeg å vise at drøftinga også har generelle implikasjoner for de humanistiske fagene. Skirbekk (1982) har framstilt humanioras antatt manglende nytteverdi som en sentral dyd. Samtidig har en i humanistiske fag lenge snakket om krise (Bruun 1983). Her tror jeg at anvendt datalingvistikk og anvendt språkvitenskap i sin alminnelighet har en viktig rolle å spille. Morgendagens humanist kan meget vel komme til å gå med hodet atskillig høyere hevet enn dagens. Det er særlig på denne bakgrunn at jeg ser det som svært positivt om ANLA tar tiltak for å styrke anvendt datalingvistikk i Norge.

Litteratur

- Bjorvand, H., Hovdhaugen, E. & Simonsen, H. G., red. 1982. Språkvitenskap. Oslo: Universitetsforlaget.
- Bruun, P. 1983. «Den humanistiska forskningens kris». I. Frängsmyr, T., red. Humanioras egenart. Oslo: Universitetsforlaget.
- Chomsky, N. 1966. «Linguistic theory.» I Lester, M., ed. Readings in applied transformational grammar. New York: Holt, Rinehart & Winston.
- Dijk, T.A. van 1977. Text and context. London: Longman.
- Dyvik, H.J. 1984. «Hva er datalingvistikk?» Humanistiske data, årg. 11, nr. 3, 4-11.
- Evensen, L.S. 1983. «David Crystal: A first dictionary of linguistics and phonetics.» Bokomtale. International Review of Applied Linguistics (IRAL), Vol. 21, No. 3, 236-239.
- Evensen, L.S. 1985. «Discourse-level performance analysis:?». I Enkvist, N.E., ed. Composition and Coherence. Åbo: Åbo Akademi.
- Evensen, L.S. & Lintermann Rygh, I. under utg. «Connecting L1 and FL in discourse-level performance analysis». Innlegg presentert på The 20th International Conference of Contrastive Linguistics, Poznan, Polen, des. 1984. Trykkes i Fisiak, J., ed. Papers and Studies in Contrastive Linguistics, Vol. 20.
- Fraser, H. & O'Donnell, W.R. 1969. Applied linguistics and the teaching of English. London: Longman.
- Fries, C.C. 1945. Teaching and learning English as a foreign language. Ann Arbor: University of Michigan Press.
- Hovdhaugen, E., red. 1976. Språkvitenskap. Oslo: Universitetsforlaget.
- Ilstad, S., Paasche, T. & Hovden, J. 1977. Survey-metoden. Trondheim: Tapir.

- Johansson, S. 1981. «Word frequencies in different types of English texts». ICAME News No. 5., 1-13.
- Labov, W. 1969. «Contraction, deletion and inherent variability in the English Copula». Language, Vol. 45.
- Labov, W. 1972. «The logic of nonstandard English.» I Giglioli, P.P., ed. Language and social context. Harmondsworth: Penguin.
- Lackstrom, J., Selinker, L. & Trimble, L. 1973. «Technical Rhetorical Principles and grammatical choice.» TESOL Quarterly, Vol. 7, No. 2, 127-136.
- Lado, R. 1957. Linguistics across cultures. Ann Arbor: University of Michigan Press.
- Mackey, W.F. 1966. «Applied Linguistics.» I Allen, J.P.B. & Corder, S.P., eds. Readings in applied linguistics. ECAL 1. Oxford: OUP.
- Moulton, W.G. 1963. «Linguistics and language teaching in the United States 1940-1960.» IRAL, Vol. 1, No. 1, 21-41.
- Saporta, S. 1966. «Applied linguistics and generative grammar.» I Valdman, A., ed. Trends in language teaching. New York: McGraw-Hill.
- Skirbekk, G. «Krise i humaniora?» Norsk Filosofisk Tidsskrift, årg. 17, nr. 2, 179-193.
- Wolfram, W. 1969. A sociolinguistic description of Detroit Negro speech. Washington D.C.: Center for Applied Linguistics.



Lars Sigfred Evensen er doktorand ved Institutt for anvendt språkvitenskap, Universitetet i Trondheim, hvor han arbeider på et NAVF-prosjekt om elevspråk.

Prosjekt i automatisk indeksering – lingvistiske metoder

Tove Fjeldvig, Anne Golden og Øystein Reigem

Bakgrunn

NAVFs EDB-senter for humanistisk forskning (Senteret) og Institutt for rettsinformatikk ved Universitetet i Oslo (IRI) har lenge arbeidet med henholdsvis design og bruk av tekstsøkesystemer og forskning på tekstsøking. Begge institusjonene har sett et behov for forbedring av tekstsøking, bl.a. ved metoder for automatisk indeksering, dvs. metoder som automatisk velger ut gode søkeord, vektet dem og bearbeider dem videre på forskjellig måte.

Ut fra felles interesser formulerte institusjonene et samarbeidsprosjekt der de ønsket å studere *enkle lingvistiske metoder* som kan brukes i automatisk indeksering. Forskningen innen automatisk indeksering har frem til i dag vært konsentrert om statistisk baserte metoder (Gerard Salton m.fl., jf. Salton/McGill 1983). Den begrensede forskning som har foregått m.h.t. lingvistiske metoder i informasjonsøking har helst skjedd utenfor Norden, og resultatene lar seg ikke uten videre overføre til de nordiske språkene.

I prosjektet ønsket vi å konsentrere oss om følgende tre metoder: automatisk gruppering av ord med felles rot (rotlemmatisering), automatisk splitting av sammensatte ord og automatisk gjenkjenning av ordklasser.

Det er to angrepsvinkler en kan bruke ved utvikling av slike metoder, enten å basere metodene på omfattende ordlister med lingvistisk informasjon, eller på mer kompakte sett med regler. Vi ville gå den regelbaserte veien.

Metoder for *automatisk rotlemmatisering* vil spare brukeren arbeidet med å holde rede på alle mulige bøyninger og avledninger av søkeord. Den vanlige metoden for å takle slike problemer i dag er (høyre)trunkering, men den er erfaringsmessig ikke tilfredsstillende. Rotlemmatisering er generelt nyttig der det er roten eller oppslagsformen til et ord som er interessant, og ikke hver enkelt bøynings- eller avledningsform. I prosjektet ville vi bygge på tidligere arbeid av prosjektgruppens medlemmer (Fjeldvig/Golden 1983 og 1984).

Metoder for *automatisk splitting av sammensatte ord* har spesiell relevans for de nordiske språk, da sammensatte ord forekommer hyppig. Splitting kan bl.a. tenkes brukt ved at søking på et sammensatt ord også gir søking på delene av ordet.

Automatisk gjenkjenning av ordklasser kan brukes på mange måter i

automatisk indeksering. Det kan f.eks. tenkes at substantiv er viktigere i søking enn de andre ordklassene, og at en ønsker å redusere søkesystemets ordbok (dvs. mengden av søkbare ord) til kun substantiv. I prosjektet vårt ville vi basere gjenkjenningen primært på en morfologisk analyse av det enkelte ord, bl.a. fordi det er en lite ressurskrevende strategi.

I januar 84 søkte vi NORDINFO om midler til et 2-års prosjekt og NORDINFO stilte seg i prinsippet velvillig til søknaden. NORDINFO så automatisk splitting av sammensatte ord som et spesielt interessant forskningsemne, og ønsket utført et forstudium av temaet. Man ønsket også at prosjektgruppen knyttet sterkere kontakter mot BDI-området og rettet prosjektsøknaden mer direkte mot det. Bevilgningen som ble gitt tilsvarte ca. 3 forskermånedsverk (dvs. FIM 47.200).

Gjennomføringen

Prosjektgruppen stilte opp en plan for et forprosjekt som skulle oppfylle NORDINFOS ønsker. Kontakt med BDI-området oppnådde gruppen ved et møte i Oslo november 84 med representanter for norske BDI-miljøer. Til møtet ble det bl.a. utarbeidet et notat for å belyse noen av sidene ved temaet for møtet.

I tillegg til NORDINFO søkte prosjektgruppen også de nordiske samarbeidsnemndene for humanistisk og samfunnsvitenskapelig forskning (NOS-H og NOS-S) om midler til prosjektforberedelser. Tanken var å bruke pengene til møter med lingvister fra Danmark, Finland og Sverige – ett møte før prosjektet kom i gang og ett noenlunde midtveis i prosjektet. Disse lingvistene ville det også bli samarbeidet med under utførelsen av prosjektet. Samarbeidet er fortsatt aktuelt i et reformulert prosjekt.

Det første møtet ble holdt i Stockholm i september 1984. I lys av NORDINFOS svar på prosjektsøknaden fant vi det naturlig for dette møtet også å trekke inn noen sentrale personer fra BDI-miljøer i Danmark og Sverige, og representanter for viktige miljøer for tekstsøking i Stockholm. Møtets første dag ble viet de lingvistiske problemer prosjektgruppen ønsket å ta opp, mens den andre dagen ble brukt til diskusjon av anvendelser og avklaring av behovene for lingvistiske hjelpemidler innen BDI.

Problemene med splitting av sammensatte ord blir behandlet spesielt i tilknytning til denne artikkelen.

Forstudium av problemer med automatisk splitting av sammensatte ord

Prosjektgruppen gjennomførte et litteraturstudium for å se om relevante forskningsresultater foreligger. Det viser seg at den eneste i Norden som har arbeidet med sammensatte ord fra en vinkel som er interessant for prosjektet, er docent *Benny Brodda* ved Institutionen för

Lingvistik/Instituttet for Rättsinformatik, Stockholms Universitet. Riktignok har en del andre forskere berørt problemområdet, men da i forbindelse med orddeling. (Selv om splitting av sammensatte ord og orddeling er beslektede problemer, er det første mer komplisert. En metode for automatisk splitting av sammensatte ord er selvsagt svært nyttig i orddeling.)

Benny Broddas betraktninger er brukt som basis for vårt forslag til arbeidsmetode, supplert med tidligere erfaringer med regelbaserte metoder (Fjeldvig/Golden 1983 og 1984).

Vi har sett på splitting av sammensatte ord som en todelt oppgave. Første delproblem er å gjenkjenne at et ord er sammensatt. Dette i seg selv er interessant i mange tilfeller. Andre delproblem er å foreta selve splittingen, dvs. finne splittingspunktet og eventuell fuge. Vi har satt opp et forslag til fremgangsmåte for hvert delproblem, og så ytterligere skilt ut en del mindre delproblemer. For detaljer, se siste del av artikkelen.

Det foreløpige teoretiske arbeidet, foreløpige tester vi har foretatt samt uttalelser fra lingvister, gir grunn til optimisme når det gjelder å finne frem til en brukbar metode for automatisk splitting av sammensatte ord.

Vi har naturlig nok ingen illusjoner om å kunne utvikle en feilfri metode. Det er imidlertid verdt å merke seg at en i et søkesystem kan benytte strategier som minsker effekten av eventuelle feil i ordsplittingen. For eksempel kan en i et system som automatisk utvider søking på et ord med søking på ordets deler, splitte tvilsomme ord *flere* steder. Dette gir selvsagt opphav til feilaktige søkeord i tillegg til de korrekte, men mange av de gale ordene vil ikke forekomme i søkematerialet. I den grad de virkelig forekommer i søkematerialet og gir tilslag på irrelevante dokumenter, vil disse (dokumentene) ofte kunne *rangeres* bort, da de som regel ikke inneholder de andre søkeordene i brukerens spørsmål. (Riktignok begrenser mange brukere seg til søking på enkle ord, men med et system som hjelper brukerne å finne på søkeord, vil dette endre seg.)

Selv om vi har sett på splitting av norske ord, er strukturen så lik i de andre skandinaviske språkene at senere forskningsresultater burde la seg overføre forholdsvis enkelt. At vi tilsynelatende lett kan bygge på Benny Broddas arbeid med svensk, styrker denne påstanden.

Resultater av kontakten med BDI-miljøene

I vår kontakt med BDI-miljøer har vi fått et klart inntrykk av at det er et stort behov for hjelpemidler for å bedre søking. De fleste hjelpemidlene er av lingvistisk art. Nedenfor følger en oppsummering av våre inntrykk fra kontakten med representanter for BDI-miljøene:

Tekstsøking og bibliografisk søking

Tradisjonelt blir biblioteksverdenen sett på som kjernen i BDI-området,

og bibliotekenes edb-bruk har til nå i hovedsak vært konvensjonell bibliografisk søking. Tekstsøking har til nå mest vært brukt innenfor fagområder som jus og humaniora. Men snart vil tekstsøking gjøre sitt inntog i bibliotekene, og metoder for forbedring av tekstsøking er dermed viktige også for dem. Mange metoder for forbedring av tekstsøking vil dessuten like gjerne kunne brukes i tradisjonell bibliografisk søking.

Hjelp til å formulere søkeargument

Det store behovet i bibliografisk søking som i tekstsøking, er for hjelpemidler som hjelper brukeren til å finne på relevante søkeord, eller mer generelt: hjelper brukeren å omforme sitt søkeproblem til en eksplisitt formulering. I tillegg til vanskene med å finne på søkeord, viser erfaringer også at mange brukere har problemer med å kombinere søkeordene på en korrekt måte (boolsk søking er vanskelig). Problemerkene vil øke sterkt med den akselererende bruk av søkesystemer en ser i dag innen BDI-området, og for så vidt i hele samfunnet.

Sentrale problemer er knyttet til

- synonymy og andre beslektede ord. Den vanlige angrepsmåten er bruk av synonymtesauri, men manuell utvikling og vedlikehold av tesauri er meget ressurskrevende.
- flerspråklighet. Flerspråklighet kan forekomme innen dokumenter, innen databaser og databaser imellom. Eventuelt kan problemet komme inn når man vil utnytte databaser i fremmedspråk. Problemene kan delvis løses med flerspråklige tesauri, men dette er også et kostbart hjelpemiddel.
- bøyninger, avledninger og sammensetninger. Disse språktrekkene utgjør et stort problem i tekstsøking, men oppleves ikke fullt så kritisk i tradisjonell bibliografisk søking. Imidlertid anser man metoder som takler bøyninger og avledninger (rotlemmatisering) også som nyttige og grunnleggende i tradisjonell bibliografisk søking. Det er generelt et stort behov for metoder som takler problemene med sammensatte ord.

Mange ønsker at søkesystemet kan hjelpe brukeren ved en eller annen form for dialog, om ikke annet fordi automatiske metoder vil gjøre feil, og brukeren i mange tilfeller bør ha det siste ord. Dialogen kan gjerne være i naturlig språk. Både i formuleringen av søkeargumentet og selve søkingen kan det være fruktbart med strategier som bygger på iterasjon. Et eksempel er metoder for automatisk reformulering av søkeargument.

Mange mener det er verdt å følge med i forskningen innen kunstig intelligens, da den kan ha nytt å tilføre konvensjonelle søkesystemer. En mulighet er å anvende kunnskapsbaserte strukturer slik en finner i ekspertsystemer, som hjelpestrukturer. I dag er tesauri de mest avanserte hjelpestrukturer som er i bruk.

Det ligger et uutnyttet potensiale i brukerens spørsmål/dialogen med brukeren, og en kan ønske seg systemer som lærer av dialogen. Et eksempel ville være et søkesystem som gir brukeren synonymforslag ut fra en tesaurus, og som så oppdaterer tesaurusen ut fra en analyse av brukerens valg samt med nye forslag fra brukeren.

Som på så mange andre områder, vil grafikk også kunne brukes i søkesystemer, spesielt i kommunikasjonen med brukeren. Grafikk vil gi nye muligheter for å visualisere strukturer og søkeresultater. Bruk av mus og vinduer, som i løpet av de siste årene har spredd seg fra avanserte arbeidsstasjoner til mikromaskiner i middels prisklasse, kan også med fordel utnyttes i søkesystemer. (Mus er en fleksibel enhet for «peking» på dataskjermen. Vindusteknikk vil si at skjermen kan deles opp, slik at mange prosesser/strukturer kan vises parallelt.)

Enkelte er opptatt av å finne alternativ til tradisjonelle søkemeter. Et eksempel er systemer hvor brukeren ikke stiller spørsmål og får mengder av dokumenter som svar, man får mulighet for å «navigere» seg gjennom dokumentmasse og hjelpestrukturer frem mot de relevante dokumenter. En slik søkemeter krever databaser med en høyere grad av intern struktur. I en bibliografisk database kan det f.eks. dreie seg om forskjellige typer referanser til andre dokumenter, til hjelpestrukturer som bibliografier, osv. Forholdene må ligge til rette for lettvent å etablere nye relasjoner. En kan også tenke seg at brukerne kan legge opp egne «spor» gjennom datamassen, som så senere kan følges. (Jf. HYPERCATalog-prosjektet ved LIBLAB, Universitetet i Linköping. Se Hjerppe/Järvelin 1984 og Hjerppe 1984.)

Tilrettelegging

I de tilfeller der datamaterialet registreres på stedet, er det behov for verktøy som kan lette tilrettelegging for databaser og som kan høyne kvaliteten på materialet og dermed på søkingen. Relevante elementer er bl.a.:

- vanlige tekstbehandlings- og skjemahåndteringsfunksjoner
- hjelpemidler for å sikre mest mulig korrekt materiale, f.eks. korrigerende av stavefeil
- hjelpemidler for indeksering
- terminologikontroll
- stilistiske hjelpemidler (jf. programmer som «Writer's Workbench»)
- metoder for håndtering av tilhørende grafikk, illustrasjoner, bilder

Også når ferdigregistrert datamateriale skal legges inn i databaser, kan slikt verktøy være aktuelt. Eksempler er korrigerende av stavefeil og terminologikontroll.

Skisse av et utvidet informasjonssøkesystem

Sammenligning med egne brukererfaringer

Prosjektet har gitt oss et grunnlag til å fortsette arbeidet med datamaskinelle informasjonssøkesystemer. Gjennom samtalene med representanter for BDI-miljøer har vi fått informasjon om svakheter ved dagens informasjonssøkesystemer og hvilke typer hjelpemidler det er størst behov for.

Erfaringene som her er kommet frem, stemmer godt overens med vår egen erfaring med tekstsøkesystemer. Systemene i dag er for lite brukervennlige, og mange finner dem vanskelige å anvende. Det største problemet er å finne frem til riktige søkeord. Et godt søkeresultat forutsetter at søkeargumentet utgjør de ordene som er brukt i dokumentene i forbindelse med det aktuelle emnet – og helst bare disse. Dette er vanskelig, og spesielt har uerfarne brukere problemer. De kjenner verken ordbruken i basen eller vet hvordan søkesystemet fungerer.

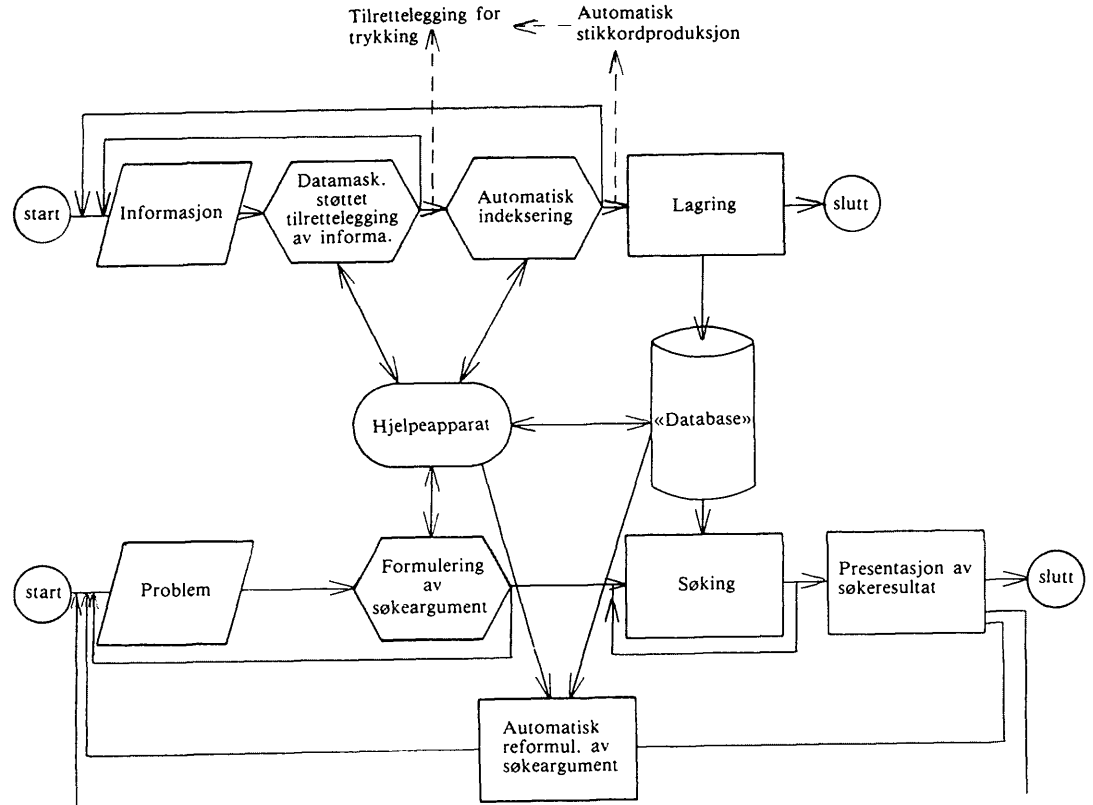
Anvender man for få søkeord, risikerer man at relevante dokumenter uteblir fra resultatlisten, og anvendes for mange, risikerer man at de relevante dokumenter nærmest forsvinner blant ikke-relevante dokumenter. Dagens tekstsøkesystemer har også begrensede muligheter til å hjelpe brukeren i formulering av søkeargumentet. De hjelpemidler som finnes, er ofte for kompliserte eller for kostbare å anvende, fordi man anvender dem ineffektivt.

Erfaringer viser at mange brukere også har problemer med å kombinere søkeordene på en slik måte at søkeargumentet gjenspeiler det informasjonsbehovet som ligger til grunn for søkingen.

De fleste søkesystemer i dag er såkalte boolske systemer, dvs. at de er basert på bruk av de boolske operatorene OG, ELLER og IKKE. Denne typen søkefunksjon dominerer, til tross for at de fleste er enige om at boolsk søking er lite egnet til bruk i tekstsøking og for vanskelig for uerfarne brukere. Spesielt kan operatorene OG være problematisk. Søkeargumenter av typen A OG B OG C OG ... N vil ved fravær av et av søkeordene (f.eks. ved at et av ordene A, B, C, osv. kun forekommer i dokumentene i en bøyingsform) føre til at ingen relevante dokumenter blir funnet. Dette kan virke uforståelig for brukeren, og det er sjelden at et søkesystem forsøker å forklare brukeren hva som kan være årsaken til et negativt resultat.

Økt kommunikasjon mellom bruker og system

På neste side har vi tegnet en grov skisse av et informasjonssøkesystem på grunnlag av de synspunkter som har fremkommet under prosjektperioden og våre egne erfaringer. Den viktigste forskjellen mellom skissen og dagens operasjonelle søkesystemer er bruk av «feedback»-informasjon – eller økt kommunikasjon mellom bruker og system. I dagens systemer er det brukeren som hele tiden må ta initiativet, og systemet svarer bare på det som det blir forespurt om. Svaret er det



samme uansett hvilken situasjon brukeren befinner seg i og hva som tidligere har skjedd i søkesesjonen. Man får ikke hjelp av systemet hvis man f.eks. står fast.

Vi har tro på at man kan løse en del av problemene ved bruk av dagens informasjonssøkesystemer ved å øke kommunikasjonen mellom brukeren og systemet og la systemet delta mer aktivt ved søkingen. Ved å initiere en dialog mellom brukeren og systemet vil f.eks. systemet kunne stille spørsmål vedrørende brukerens informasjonsbehov og brukeren vedrørende systemets informasjon om dokumentenes egenskaper. På den måten kan systemet veilede brukeren i formuleringen av søkeargumentet. Det ville også ha vært til stor hjelp for brukeren – spesielt den uerfarne bruker – om man kunne innlede søkeprosessen med et søkeargument i naturlig språk.

Skisse av et informasjonssøkesystem

Skissen er basert på den samme strukturen som dagens søkesystemer, og de muligheter den gir lar seg derfor realisere innenfor disse systemene. Kommunikasjonen (dialogen) mellom bruker og system er illustrert ved tilbakevendende piler som gir brukeren muligheten til å gå tilbake og korrigere inndata.

Det primære formålet med skissen er å vise hvilke delprosesser et informasjonssøkesystem består av, hvordan de er knyttet sammen og hvor de største utfordringene ligger med tanke på å utvikle brukervennlige og effektive søkesystemer.

Et informasjonssøkesystem har vi her latt omfatte hele prosessen fra den ønskede informasjon tilrettelegges for søking til brukeren får sitt endelige resultat presentert på skjermen. I mange brukersituasjoner har man ikke kontroll med tilretteleggelsen av dokumentene, men i en oversikt over et informasjonssøkesystem er det naturlig å ta med både tilretteleggingsdelen og søkedelen.

(1) Tilretteleggingsdelen

Denne delen av systemet sørger for å tilrettelegge informasjonen for søking. Den omfatter:

- A) Bearbeiding av informasjonen før den mates inn i søkesystemet – resultat: inndata til søkesystemet.
- B) Maskinell behandling av inndata (automatisk indeksering) – resultat: data klar for lagring.
- C) Lagring av dataene – resultat: dokumentdatabasen.

Bearbeidingen av informasjonen før den mates inn i søkesystemet avhenger av hvordan man ønsker informasjonen fremstilt av søkesystemet, hvordan den er representert før det besluttes at den skal gjøres søkbar og hva slags søkesystem vi har med å gjøre. Et tekstsøkesystem stiller normalt ikke krav til utformingen av dokumentene, men det forutsetter at det er mulig for systemet å skille ordene og dokumentene fra hverandre.

I dag er det stadig flere som velger å erstatte skrivemaskinen med tekstbehandling, og det betyr at dokumentene etter hvert vil bli registrert i maskinlesbar form alt ved førstegangsskriving. Dermed sparer man registreringskostnadene ved overgangen til et datamaskinelt informasjonssystem. I mange tilfeller er det likevel ønskelig å bearbeide dokumentene noe før de mates inn i søkesystemet, f.eks. føye til faste opplysninger eller gjøre teksten mer entydig. Det kan derfor være ønskelig med programvare som i tillegg til tekstbehandling også bidrar til å effektivisere tilretteleggelsen av dokumentene for søking. Vi tenker da på f.eks. rutiner som kan avsløre skrivefeil, sørge for konsekvent skrivemåte av forkortelser, dato, navn og ord med ulike skriftformer, og som kan gi forfatteren muligheten til å skissere dokumentformatet på skjermen slik at registrering av faste opplysninger o.l. kan foregå mer effektivt. Mye av dette finnes, men ikke alltid integrert. I mange tilfeller ville man også hatt stor nytte av et system for automatisk produksjon av sammendrag.

Etter hvert som tekstsøking blir en mer naturlig arbeidsoppgave på arbeidsplassen, vil kobling mellom tekstsøke- og tekstbehandlingsprogrammer bli sentral. Det kan være ønskelig å kalle opp tekstsøkesystemet fra tekstbehandlingsprogrammet eller tekstbehandlingsprogrammet fra tekstsøkesystemet. Her er det mange interessante arbeidsoppgaver som venter.

Den maskinelle behandling av dataene før de lagres i databasen går ut på å analysere og bearbeide dokumentene på en slik måte at de blir utnyttet effektivt i søkeprosessen. I litteraturen kalles denne prosessen for automatisk indeksering, og den består i å velge ut de ordene i dokumentet som er egnet som søkeord og tillegge dem en vekt som gjenspeiler deres betydning i forhold til innholdet i dokumentet. I automatisk indeksering inngår også teknikker for å viderebehandle ordene på forskjellig vis, f.eks. slå sammen bøyingsformer av samme ord (lemmatisering). Til tross for en del forskning på dette området, foregår denne prosessen i de fleste systemer på en svært primitiv måte. Hvert ord er i utgangspunktet søkbart, men for å spare lagringsplass sletter man i ordboken en del ord som på forhånd antas å være uegnet som søkeord (stoppord). Noen få utenlandske systemer lemmatiserer, men det er stort sett det hele.

Store utfordringer ligger i å utvikle søkesystemer som bedre kan utnytte den informasjonen som ligger i dokumentene, f.eks. kartlegge hvilke ord som

- er karakteristiske for det enkelte dokument
- beskriver samme tema
- uttrykker samme begrep (synonymer)
- har flere betydninger (homografer)

Mangelen på denne type informasjon fører lett til dårlige søkeresultat. I dag vet vi for lite om hvordan man kan innhente denne type

informasjon og hvordan den kan representeres på en effektiv måte i søkesystem. Eksempler på oppgaver i denne forbindelse er:

- rotlemmatisering (gruppere ord med felles rot)
- fastsettelse av oppslagsformen til rotlemmaet (nyttig ved presentasjon av rotlemmaet for brukeren ved formulering av søkeargumentet)
- fastsettelse av ordklassen til ordene, eventuelt som en del av en mer fullstendig syntaktisk analyse (informasjon som er nyttig ved identifisering av fraser og vektlegging av ordene i det enkelte dokument i forhold til deres viktighet i dokumentet)
- lingvistiske og statistiske analyser av teksten til bruk ved
 - homografseparering
 - etablering av lenker mellom relaterte ord
 - vektlegging av ord i forhold til det enkelte dokument
 - evt. oppdatering av et begrepsnettverk
- eliminering/merking av ord som ikke er egnede søkeord (spesielt viktig ved bruk av søkeargumenter i naturlig språk, men også ressursbesparende hvis disse ordene slettes i ordboken)

Det største problemet består i å etablere lenker mellom ord som er relatert til hverandre i en eller flere sammenhenger. Den tradisjonelle måten å etablere lenker på er manuell oppbygging av tesauri. Det er gjennomført noen få forsøk på å etablere slike lenker automatisk på grunnlag av ulike analyser av dokumentene, men ingen har gitt resultater som lar seg implementere i operative systemer (Salton 1968, Sparck Jones 1981). Den vanlige måten å etablere lenker mellom ord på, er manuell oppbygging av tesauri med utgangspunkt i viten om domenet dokumentene faller inn under. Vi er imidlertid ikke kjent med at noen har sett på muligheten av å etablere et mer avansert begrepsnettverk der domenet er beskrevet i et formelt språk.

Uansett hvordan man velger å gå frem er det viktig å få utviklet et apparat som kan gi informasjon om relaterte ord. Store forskningsoppgaver knytter seg til etablering av avanserte begrepsnettverk, eller semantiske nettverk som de også kalles.

(2) Søkedelen

Søkedelen kan vi se på som bestående av fire delprosesser:

- A) Formulering av søkeargumentet
- B) Søking
- C) Presentasjon av søkeresultatet
- D) Automatisk reformulering av søkeargumentet

Formuleringen av søkeargumentet skjer i samråd med systemet. Brukeren starter med å beskrive sitt informasjonsbehov overfor systemet, og det er en fordel om denne beskrivelsen kan gis i naturlig språk. Søkeargumentet analyseres og tolkes av søkesystemet, og deretter innledes en dialog med brukeren hvor det kan utveksles informasjon. Her diskuteres søkeargumentet, og både brukeren og systemet blir

oppfordret til å komme med forslag til nye søkeord og kommentere forslagene. Søkesystemet vil f.eks. gjøre brukeren oppmerksom på

- ord som forekommer med stor spredning i dokumentmaterialet eller på en annen måte synes å være lite egnet som søkeord
- ord som er homografer
- at enkelte sammensatte ord kan skrives om til en kombinasjon av de ordene det er sammensatt av
- ineffektiv bruk av trunkering
- andre bøyings- og avledningsformer til søkeordene
- andre ord som er relatert til søkeordene

Dialogen vil også kunne være en viktig informasjonskilde for søkesystemet. Den gir systemet muligheten til å lære på grunnlag av de forslagene og kommentarene som brukerne har kommet med. På den måten vil systemet stadig kunne forbedre dialogen med brukeren og være til større hjelp ved formulering av søkeargumentet.

Mangel på gode nok søkeord i søkeargumentet er den største årsaken til effektivitetssvikt i tekstsøking. Dette har både erfaring med tekstsøking og kontrollerte forsøk vist (Fjeldvig 1980). Vi har tro på at det er mulig – også innenfor dagens tekstsøkesystemer – å bygge opp et begrepsnettverk med synonymrelasjoner ved å systematisere søkeargumentene som er stilt til systemet over tid. Vi er ikke kjent med at denne ideen har vært utforsket tidligere.

Før selve søkingen iverksettes, kan man få et inntrykk av søkeresultatet ved å la systemet beregne et estimat på gjenfinningsgrad og presisjon på grunnlag av søkeargumentet. Dette er mulig med hjelp av en modell av et tekstsøkesystem som er utviklet ved IRI (Harvold 1980). På grunnlag av disse estimatene kan brukeren avgjøre om søking skal iverksettes, søkingen avsluttes eller om søkeargumentet skal endres ytterligere.

Selve søkingen går ut på å sammenligne søkeargumentet med dokumentbeskrivelsene. Det er ulike måter å gjennomføre denne sammenligningen på, og søkefunksjoner i seg selv er et forskningsområde som er blitt viet stor oppmerksomhet. Som tidligere nevnt er de fleste operasjonelle informasjonssøkesystemer i dag basert på boolsk søking. Denne søkefunksjonen er for krevende for de fleste brukere, og dessuten gir den ikke en rangering av de funne dokumenter. Ved tekstsøking er det aldri mulig for søkesystemet å garantere at resultatlisten inneholder alle og bare de relevante dokumenter. Man får derfor et mer effektivt søkeresultat med alternative søkefunksjoner som kan gi brukeren en oversikt over de funne dokumenter rangert ut fra deres sannsynlighet for å være relevante.

I figuren er *presentasjon av søkeresultatet* fremstilt som en egen prosess. Dette er gjort for å understreke viktigheten av at søkeresultatet presenteres på en måte som gjør relevansvurderingen mest mulig effektiv. Her er det også en fordel at dokumentene rangeres, slik at

brukeren kan få silt ut de dokumenter som har størst sannsynlighet for relevans. Man slipper dermed å bla gjennom hele resultatlisten hvis man finner det man søker etter i de øverste rangsett. Det er også nyttig i forhold til relevansvurderingen å få informasjon om hvorfor et dokument er funnet og hvilken likhet det har med søkeargumentet. Dette er spesielt viktig i de tilfeller bare deler av søkeargumentet er oppfylt i de funne dokumenter, slik at man kan sile ut de som inneholder de mest sentrale søkeordene.

En effektiv relevansvurdering forutsetter også at systemet kan presentere kortere versjoner av dokumentene for brukeren. Dette er spesielt viktig når databasene er store og dermed gir tilsvarende store svarmengder. Det kan også være viktig når dokumentene i seg selv er store. Dagens systemer gir brukeren muligheten til å få se hvor søkeordene befinner seg enten ved å vise den delen av dokumentet hvor søkeordene forekommer eller ved å gi søkeordene ekstra lysstyrke. Mye kan imidlertid gjøres for å effektivisere disse rutinene, og det ville ha vært ønskelig med en metode som kunne produsere et sammendrag av dokumentet eller den delen av det som synes å være mest relevant.

Med *automatisk reformulering av søkeargumentet* sikter vi til metoder som etter et søk kan forbedre søkeargumentet ved å utnytte informasjon om hvilke av de funne dokumenter som (sannsynligvis) er relevante. Det kan være snakk om at systemet analyserer de høyest rangerte dokumentene på jakt etter flere søkeord, eventuelt kombinert med brukerens vurdering av disse dokumentene. Det er gjennomført vellykkede eksperimenter med denne type teknikker bl.a. i SMART- og RESPONSA-prosjektene (Salton/McGill 1983, Choueka 1980). De beste resultatene ble oppnådd ved å la brukeren selv avgjøre hvilke endringer som skal gjennomføres i søkeargumentet.

Aktuelle forskningsområder

På grunnlag av fremstillingen ovenfor har vi trukket frem følgende forskningsområder:

- (1) *Hjelpemidler for skribenter* til bruk ved registrering og tilrettelegging av dokumenter for søking.
- (2) *Metoder for automatisk indeksering* til bruk ved analyse og bearbeiding av dokumenter og av søkeargumenter i naturlig språk. Flere av metodene kan også anvendes ved behandling av boolske eller andre strukturerte søkeargumenter.
- (3) *Utvikling av en kunnskapsbase (eller begrepsnettverk)* på grunnlag av en analyse av det domenet som dokumentene representerer. Metoder for automatisk vedlikehold av basen.
- (4) *Effektive metoder for lagring* av dokumenter og annen informasjon som skal utnyttes i søkeprosessen.
- (5) *Systembeskrivelse av en forsats til tekstsøkesystem*. Forsatsen skal gi brukeren vesentlige bidrag for utvikling av et rikere søkeargument enn hva en typisk bruker vil kunne spesifisere uten

hjelp. Den skal ha form av en dialog mellom bruker og system. (Ved IRI foregår forskning på dette området finansiert av Norges teknisk-naturvitenskapelige forskningsråd. Prosjektet avsluttes sommeren 1985.)

- (6) *Utforming av en dialog mellom bruker og system* til bruk i søkedelen.
- (7) *Utvikling og vedlikehold av et begrepsnettverk på grunnlag av informasjon fra tidligere søk* – m.a.o. hvordan få systemet til å lære av brukerne? Begrepsnettverket kan godt være det samme eller et tillegg til kunnskapsbasen i pkt. 3.
- (8) *Metoder for effektivisering av relevansvurdering.*
- (9) *Automatisk reformulering av søkeargumentet* på grunnlag av dokumenter som alt er funnet.

Etterskrift

Etter at denne rapporten var skrevet, reformulerte prosjektgruppen sin opprinnelige prosjektplan. Den viktigste endringen var et nytt delprosjekt kalt «Oppbygging av tesaurus/kunnskapsbasert struktur ut fra brukerdialogen». Gruppen søkte NORDINFO for andre gang og fikk nå midler til delprosjektet «Automatisk splitting av sammensatte ord». I tillegg arbeider gruppen fortsatt for å få finansiering for resten av prosjektet. Arbeidet med splitting av sammensatte ord tenkes utført hovedsakelig høsten 1985.

Litteratur

- Bing J/Selmer Knut S (1980) *A Decade of Computers and Law*, Universitetsforlaget. Oslo.
- Choueka Y (1980) «Computerised Full-Text Retrieval Systems and Research in the Humanities: The RESPONSA Project»; *Computers and the Humanities* vol. 14, no. 3, 1980.
- Fjeldvig T (1980) «Controlled Experiments in Text Retrieval – an Appraisal of the Results»; Bing/Selmer 1980: 147-172.
- Fjeldvig T/Golden A (1983) «Automatisk rotlematisering», *Humanistiske Data* nr. 3 1983.
- Fjeldvig T/Golden A (1984) *Automatisk rotlematisering – et lingvistisk hjelpemiddel for tekstsøking. NORIS(64) og NORIS(58)*, COMPLEX nr. 9/84, Universitetsforlaget. Oslo.
- Harvold T (1980) «Performance of Text Retrieval Systems»; Bing/Selmer 1980: 172-219.
- Hjerpe R og Järvelin K (1984) «Prosjekt HYPERCATalog.» LiU-LIBLAB-WP:9. LIBLAB, Universitetet i Linköping.
- Hjerpe R (1984) «HYPERCATalog projektet. Några förtydliganden.» LiU-LIBLAB-WP:16. LIBLAB, Universitetet i Linköping.
- Salton G (1968) *Automatic Information Organisation and Retrieval*, McGraw-Hill.
- Salton G/McGill M J (1983) *Introduction to Modern Information Retrieval*, McGraw-Hill.
- Sparck Jones K (ed.) (1981A) *Information Retrieval Experiment*, Butterworth & Co. (Publishers) LTD.
- Sparck Jones K (1981B) «Retrieval System Tests 1958-1978»; Sparck Jones 1981A: 213-255.

Tove Fjeldvig er forskningsstipendiat ved Institutt for privatrett, avd. EDB, Universitetet i Oslo.

Anne Golden er bestyrer ved Institutt for studiet av norsk som fremmedspråk, Universitetet i Oslo.

Øystein Reigem er førstekonsulent ved NAVFs EDB-senter for humanistisk forskning.

Automatisk gjenkjenning og splitting av sammensatte ord

Anne Golden

Bakgrunn

Som svar på prosjektsøknaden til NORDINFO om midler til å studere automatisk indeksering, fikk prosjektgruppa innvilget penger til et forprosjekt som bl.a. skulle se nærmere på automatisk splitting av sammensatte ord. Spesielt skulle vi undersøke hva som er gjort i de nordiske landene på dette området.

Litteraturstudier

Et litteraturstudium viste at *Benny Brodda* var den eneste i Norden som hadde arbeidet med sammensetningers struktur ut fra et databehandlingsperspektiv. Andre forskere som arbeider med datalingvistiske metoder hadde mer berørt problematikken i forbindelse med orddeling. På møtet vi hadde i Stockholm, kom det heller ikke fram noen andre referanser til arbeider i Norden, men vi ble anbefalt å undersøke hva de hadde gjort i Tyskland på dette feltet.

Vårt forslag til arbeidsmetode bygger derfor for det meste på Benny Broddas betraktninger (se Brodda 1979 og Brodda 1982), men vi viderefører dem i forbindelse med våre egne erfaringer med regelbaserte metoder (se Fjeldvig/Golden 1984).

Når det gjelder beskrivelsen av de forskjellige ordklassenes sammensetningsmuligheter og -måter, har vi brukt F.E. Vinjes bøker (se litteraturlisten).

Alternative framgangsmåter

Automatisk splitting av sammensatte ord i en tekst kan teoretisk foregå på to forskjellige måter. Man kan

- 1) først undersøke om ordet er sammensatt, og så finne fram til de enkelte leddene.

- 2) lete direkte etter leddgrensen, og anta at ordet er sammensatt hvis man finner en slik grense.

Vi har ikke undersøkt hvilken metode som fører lengst, men vi har sett på gjenkjenning og splitting av sammensatte ord som to separate oppgaver og satt opp et forslag til framgangsmåte for hver av dem.

Automatisk gjenkjenning av sammensatte ord

For kunne skille ut de sammensatte ordene, er det nødvendig å finne fram til strukturen i de usammensatte. La oss derfor se på hvordan de ikke-sammensatte ordene ser ut:

GRUNNFORM		- PREFIKS	- AVLEDNING	- PREFIKS&AV-LEDNING
1. 1-stavelser	busk	avgå	baking	ulovlig
2. Lette 2-stavelser	jente	tiltale	krenkelse	innvielse
3. 2-stavelser av EL/EN/ER-typen	sykkel		vinteraktig/ sykling	frasykling
4. Andre 2-stavelser	nabo	medhustru	naboskap	medarbeider
5. Flerstavelser	navigere	degenerere	organisering	desavuering

Gruppe 2 (lette tostavelser) oppfører seg som gruppe 1 (enstavelsesord) så snart de får lagt til en avledningsendelse eller en bøyningendelse. Det eneste som skiller dem, er E'en i utlyd i grunnformen til de lette tostavellesordene (på norsk har vi bare -E, ikke -A som på svensk). Når vi fjerner alle affiksene (og eventuelt utlyds-E'en) til en ordform («stripper ordet»), ville begge disse gruppene bare ha en stavelse. Denne kunnskapen kan vi benytte oss av for å få en indikasjon på om en ordform er sammensatt eller ikke. Står vi igjen med bare en stavelse etter en slik stripping, er det usammensatt – står vi igjen med flere, er sjansene store for at det er en sammensetning.

Gruppe 3 har regelmessig struktur i grunnformen, formelen er:

Initialkluster + stammevokal + medialkluster + e(L,N,R)

(et kluster defineres som 0, 1, 2 eller 3 konsonanter).

Når vi legger til en avledningsendelse i denne gruppa, vil det som regel føre til at E'en faller bort, gjerne samtidig med en konsonantforenkling (eks. SYKKEL - SYKLING). Roten blir på den måten stående igjen med bare en stavelse. Det samme skjer når vi legger til noen av bøyningendelsene, men det gjelder ikke samtlige. I de tilfellene hvor en sammentrekning har funnet sted, vil vi altså stå igjen med bare en stavelse etter at affiksene er fjernet, og ordet vil bli klassifisert som ikke-sammensatt. Når vi ikke har sammentrekning, må vi altså fange opp disse ordene ved å lage regler for den spesielle strukturen de har.

Problemet blir å skille ut gruppe 4 og 5, som stort sett består av fremmedord. Ved å legge inn en liste over hyppige forstavelser i ord

som er lånt fra gresk og latin, kunne vi kjenne igjen en del av fremmedordene. Vi ville da vite at ordformen ikke nødvendigvis var sammensatt selv om den bestod av flere stavelser etter en stripping. Men for å avgjøre om de er sammensatt eller ikke, må vi kartlegge strukturen i disse ordene og sammenlikne usammensatte ord med sammensatte. Spesielt problematisk vil det bli når vi har en sammensetning hvor det ene leddet er av germansk opprinnelse, mens det andre er av latinsk eller gresk opprinnelse (eks. SELVEVALUERING).

Automatisk splitting av sammensatte ord

En sammensetning kan defineres som to eller flere frie morfemer. La oss først se på hvilke ordklasser disse morfemene tilhører:

substantiv	+ substantiv	- eks: BOKHYLLE, FYRSTIKKESKE
adjektiv	+ substantiv	- eks: GODVÆR, UNGKAR
verb	+ substantiv	- eks: KJØREBANE, SOVEROM
adverb/prep	+ substantiv	- eks: BAKSETE, OVERLEGE
tallord	+ substantiv	- eks: FIRKLØVER, HUNDREÅR

substantiv	+ verb	- eks: STØVSUGE, TALLFESTE
adjektiv	+ verb	- eks: GODSNACKE, SMÅLE
adverb/prep	+ verb	- eks: AVTA, OVERTALE, UNDERSLÅ

substantiv	+ adjektiv	- eks: HIMMELBLÅ, STARTKLAR
verb	+ adjektiv	- eks: ERTELYSTEN, SEGNEFERDIG
adjektiv	+ adjektiv	- eks: HALVFERDIG, FULLGOD

Det første leddet i en sammensetning er nesten alltid ubøyd (sett ut fra et nåtidsperspektiv), unntakene er nevnt nedenfor.

Det første leddet har flertallsform:

- FEDRELAND, MØDREHJELP (substantivet har flertallsform i noen slektskapsord).
- ÅTTETIMERSDAG, TREMÅNEDERSKURS (substantivet har noen ganger flertallsform når det innledes av et kardinaltall).
- GULERØTTER, FETEVARER (noen få adjektiv har flertallsendelse, men dette kan for så vidt regnes som en fuge -E).
- SMÅGUTT, SMÅBY (adjektivet har flertallsform, bare i forbindelser med SMÅ som førsteledd).

Andre bøyninger av førsteleddet:

- GODTFOLK, NYTTÅR, VARMTVANN (adjektiv i nøytrum).
- LILLEFINGER, LILLEBROR (adjektiv i bestemt form).
- MINDRETALL, MINSTEPRI (gradbøyd adjektiv).
- BORTEBOER, HJEMMEHJELP, UTEKLÆR (påstedsbøying av adverbet).
- BARBERMASKIN, ISOLERGLASS (E'en utelates i infinitiver

som ender på -ere).

De frie morfemene kan settes sammen på tre forskjellige måter:

- 1) førsteleddet har uendret form: FOTFESTE, INNSETTE
- 2) førsteleddet utvides med -S: DAGSREISE
- 3) førsteleddet utvides med -E: BARNEHAGE

Fugebokstaven møter man først og fremst i forbindelsen «substantiv + substantiv». I de andre tilfellene knyttes ordene direkte sammen (det finnes noen unntak).

Selv om det er sjelden at det første ordet er bøyd, vil det ofte være avledet, f.eks. FRIHETSGUDINNE, ARBEIDERPARTI.

Strukturen i sammensetninger blir slik:

1.ledd + (fuge) + 2.ledd + (bøyning).

Både 1. og 2. ledd kan egentlig selv være en sammensetning, eller de kan bestå av både prefikser og avledningsendelser. Når sammensetningen består av mer enn 2 ledd, vil sjelden det første leddet ha avledning eller fuge. Alle gruppene av ubøyde ordformer som står i skjemaet for ikke-sammensetninger er mulige som førsteledd i sammensetninger. De tre første gruppene vil ha en slik struktur:

(prefiks) + fritt morfem + (avledningsendelse) + (fuge) + (prefiks) + fritt morfem + (avledningsendelse) + (bøyningsendelse)

Når vi sammenholder dette med strukturen for de frie morfemene, får vi:

L

(prefiks) + initialkluster + stammevokal + :(medialkluster + E + N) +

R

(avledningssuffiks) + (S) + (prefiks) + initialkluster: + stammevokal +

(E)

Når vi skal splitte sammensatte ord, må vi altså lete oss fram til den sekvensen som står mellom stammevokalene i formelen ovenfor (dvs. mellom : og :), og på denne måten finne fram til hvor det ene klusteret slutter og det andre begynner. Vanskeligheten er størst når denne sekvensen inneholder få bokstaver som f.eks.

STÅ :L: AMPE

STÅ :L: ESKE

EKT :ES: ENG

HE :LS: ETNING

HA :LS: ÅPNING

Konklusjon

For å kunne skille ut de sammensatte ordene, må man altså løse følgende problemer:

1. Vi må fjerne affiksene. Til dette kan vi bruke metoden for automatisk rotlemmatisering (jf. Fjeldvig/Golden 1984), eventuelt tillemppe denne. Oppslagsformen for gruppe 3 vil være enstavet etter denne metoden. Vi testet denne metoden på ca. 2000 ulike ord hvorav ca. 40% var sammensatt. I 93% av tilfellene lyktes vi i å skille ut de sammensatte ordene. Feilene var stort sett usammensatte fremmedord som ble ansett som sammensatte.
2. Vi må kartlegge strukturen i fremmedordene og sammenlikne usammensatte fremmedord med sammensatte. Dessuten må vi undersøke i hvilken grad vi har «blandingssammensetninger» og om disse har en preferert struktur.
For å kunne splitte de sammensatte ordene blir altså oppgavene:
 1. Klustrene må kartlegges.
 2. Vi må løse problemet med overlappinger mellom klustrene. Ved å bruke statistiske hjelpemidler når klustrene overlapper hverandre, kan vi bygge opp et regelsett som prioriterer noen ordbegynnelser og noen ordslutt.
 3. Affiksene må kartlegges.
 4. Vi må kunne skille ut de «falske» affiksene, dvs. de bokstavkombinasjonene som ligner på et affiks, men er en del av roten (eks. *lillefingern*) ved å undersøke klusteret foran.

Litteratur

- Brodda, Benny: *Något om de svenska ordens fonotax och morfotax*. Papers from the Institute of Linguistics, University of Stockholm (PILUS) nr. 38, 1979.
- Brodda, Benny: *Yttre kriterier för igenkänning av sammansättningar*. Stensil 1982.
- Fjeldvig, Tove/Golden, Anne: *Automatisk rotlemmatisering - et lingvistisk hjelpemiddel for tekstsøking*. COMPLEX 9/84.
- Vinje, Finn-Erik: *Moderne norsk*. Universitetsforlaget 1976.
- Vinje, Finn-Erik: *Ord om ord*. Tapir 1973.

Edb-baserte arkivkataloger: Samkatalog for privatarkiver

Elisabeth Koren

Norsk privatarkivinstitutt arbeider for bevaring av historisk verdifulle privatarkiver slik at disse skal gjøres tilgjengelig for forskere. Et privatarkiv er kort beskrevet den samling av dokumenter som dannes av en ikke-offentlig virksomhet. Begrepet «*privat*» i denne sammenheng står i motsetning til «*offentlig*», det betyr ikke nødvendigvis «*personlig*».

Arkivverket har som oppgave å ta vare på de arkivene som blir skapt som resultat av offentlig virksomhet på statlig og kommunal sektor. Det er arkivverkets oppgave å sørge for at disse arkivene holdes sammen og ordnes slik at de er tilgjengelige både for praktiske, forvaltningsmessige formål og for forskning.

Det finnes ikke almene regler for bevaring av de arkivene som skapes som resultat av all slags ikke-offentlig virksomhet, alt fra storkonsern som Borregaard eller samfunnsfaktorer som Norsk Arbeidsgiverforening til et krisesenter for mishandlede kvinner. Det sier seg selv at slike arkiv kan ha vel så stor forskningsmessig interesse som departementsarkiver.

Selv om det ikke finnes et regelverk for bevaring av privatarkiv, så er det likevel blitt tatt vare på svært mange slike arkiv:

Universitetsbibliotekene og andre bibliotek har manuskriptsamlinger og lokale samlinger som for en stor del består av ikke-trykt materiale som bare eksisterer i det ene eksemplaret som finnes i samlingen. Likeledes har en rekke museer håndskrifter eller samlinger av dokumenter, arkivverket har tatt imot arkiver etter privatpersoner og institusjoner, og det finnes egne arkivinstitusjoner som Arbeiderbevegelsens Arkiv, som har som oppgave å ta vare på arkivene til LO og den politiske venstresiden. Privatpersoner, bedrifter og organisasjoner tar også vare på historisk verdifulle privatarkiv.

For forskere er det nesten umulig å orientere seg i denne mengden av ulike institusjoner som oppbevarer privatarkiv. Enkelte institusjoner har kataloger eller kortkartotek over sine samlinger. Andre har bare vanskelig tilgjengelige opplysninger i aksjonsprotokoller o.l. Resultatet blir at slike vanskelig tilgjengelige kilder blir lite utnyttet.

For NPI, som har som oppgave å registrere opplysninger om privatarkiv i offentlig og privat eie, har det vært en høyt prioritert oppgave å lage en felles katalog over bevarte privatarkiver, nettopp som forskerhjelpemiddel.

Instituttets forløper, Privatarkivkommisjonen, presenterte i 1976 en sensilert katalog over arkivopplysninger innsamlet gjennom 25 år,

disse var opprinnelig ordnet på katalogkort.

I 1980 fikk instituttet bevilgning fra NAVF og Riksbibliotekstjenesten til et 4-årig prosjekt som sikter på å bygge opp en edb-basert samkatalog over privatarkiv i offentlig og privat eie, i samarbeid med NAVFs EDB-senter.

Siden 1981 har datainnlesning foregått på en Altos-firebruger som NPI deler med Riksarkivet og Sekretariatet for fotoregistrering. Vi skriver inn data med programpakken Datastar, på et skjema med faste felter, og sender ferdig redigerte disketter til NAVFs EDB-senter i Bergen hvor de foretar samkjøringer på hele datamassen, hvilket vi ikke har kapasitet til på den lille maskinen vår.

Neste skritt, som ennå ikke er realisert, er publisering av katalogen, i mikrokortformat, som NAVFs EDB-senter skal stå for. Vi har valgt dette formatet fordi det vil i dag kunne gjøre katalogen lett tilgjengelig for ulike brukergrupper i alle forskningsbibliotek, i forskningsinstitusjoner og i praktisk talt alle folkebibliotek.

I løpet av 1985 får vi større datakapasitet i Riksarkivbygget. Vi kan da selv søke i basen i Riksarkivbygget, men foreløpig er ikke systemet robust nok til at basen kan være alment tilgjengelig for on-line søking utenfor huset.

NPIs Samkatalog består av tre deler. En *hovedkatalog* med summariske opplysninger om hvert arkiv, *spesialregistraturer*, det vil si opplysninger om innholdet i et arkiv på mer detaljert nivå enn det vi finner i hovedkatalogen og en *brevregistratur*, det vil si register over brevsendere og mottakere i arkivet der hvor slike er utarbeidet. *Foreløpig er bare hovedkatalogen edb-innlest, og bare deler av den er ferdig bearbeidet.*

Hovedkatalogen er delt i 2 deler: Katalogdelen (se fig. 1), som består av summariske opplysninger om hvert enkelt arkiv – arkivskaper, oppbevaringssted, arkivets omfang og innhold – og som gir brukeren en innføring i hva arkivet inneholder og registerdelen (se fig. 2) som består av fire registre som skal tjene som *innganger* til hovedkatalogen.

De registrene vi har kommet fram til er *navneregister, topografisk register, kategoriregister og emneregister.*

Arbeidet med Hovedkatalogen til nå har vært konsentrert om innlesning av opplysninger om de enkelte arkiv og utvikling av søkerprinsipper, dvs. arbeid med å lage innganger til de summariske arkivopplysningene.

I det følgende skal det kort redegjøres for grunnlaget for den maskinelle registreringen av arkivopplysninger og hvordan registrene lages på grunnlag av registreringer. Andre aspekter av registreringsarbeidet blir ikke tatt opp her.

Som tidligere nevnt foreligger det utarbeidede arkivkataloger i mangfoldiggjorte utgaver eller som kortkartotek ved enkelte arkivinstitusjoner. NPI sammendrar da disse opplysningene og registrerer dem i standardisert form i samkatalogen. Dette er tidkrevende.

SIGNATUR : ET/81
 SKJEMA : H - 08835
 NAVN : KVÆRNLIEN, ARVE
 ANDRE NAVN : Kværnlien mølle
 F./START-D./SLUTT .. : 1929 -
 YRKE : Møller - Industri, bergverk, håndverk
 BIOGR.OPPLYSN : 1872 kjøpte Henrik Pedersen Bekkemellom
 gardskverna på Søndre Røstad og bygde ny
 mølle, Kværnlien. 1905 kjøpte Anton O.
 Øverjordet Kværnlien og overdro den 1926 til
 sønnen Olaf K. som 1972 overdro mølla til
 sin sønn Arve. Mølla drives fortsatt.
 KATEGORI : Personarkiv
 TIDSRØM : 1669 - 1981
 FYLKE - KOMMUNE . : Oppland - Gjøvik
 EMNEORD : Mølle drift; Biografier
 NAVN I ARKIVINNH. : Bekkemellom, Henrik Pedersen
 ARKIVINNHOLD : Historikk om Kværnlien mølle og andre
 kverner og sagbruk i Snertingdal Opplysninger
 om Henrik Pedersen Bekkemellom (1825-)
 som bygde kvernene i Kværnlien og
 Bekkemellom
 OPPBEVARING : Eiktunet - Oppland
 UTFYLLENDE OPPL. . : 1 legg. Gave.
 NÅR INNFØRT : 1982

SIGNATUR : ET/82
 SKJEMA : H - 08836
 NAVN : AUSTDAL SKYSSTATION
 ANDRE NAVN : Austdal nordre, gnr 116, bnr 1
 F./START-D./SLUTT .. : 1895 - 1925
 BIOGR. OPPLYSN : Skysstasjonen på Austdal var før på Søndre
 Austdal, men fra 1895 ble den flyttet til
 Nordre Austdal. Stasjonen hadde to faste
 hester. De skyset til Sveen, 17 km, til
 Finden, 21 km og Gjøvik, 21 km.
 KATEGORI : Forretningsarkiv
 TIDSRØM : 1898 - 1925
 FYLKE - KOMMUNE . : Oppland - Gjøvik
 EMNEORD : Samferdsel; Reiseliv;
 ARKIVINNHOLD : Dagbøker 1898-1917, 1917-1925, 1901-1925
 OPPBEVARING : Eiktunet - Oppland
 UTFYLLENDE OPPL. . : 3 bøker.
 NÅR INNFØRT : 1982

Fig. 1: Eks. på hovedkataloginnførsel: En side fra Hovedkatalogen for Opplandsarkivet.

6.2	Detaljhandel			
	Bergseng & Co., A/S Thor	1864-1962	H-08390	135
	Bergseng, Johan	1918-1934	H-08138	54
	Biri Glasværks Landhandel	1879-1890	H-08481	176
	Bølken-Ballangrud, gnr.15. bnr.17	1876-1895	H-08568	199
	Bohmer, Johan H.	1854-1908	H-08140	55
	Engen, Christen	1820-1854	H-08157	61
	Gundersen, Christian	1845-1937	H-08141	55
	Johansen, N.	1905-1926	H-08111	44
	Knudrud, Ole H.	1877-1891	H-08145	56
	Kringlen, Ole	1861-1863	H-08150	58
	Lillehammer-forretning	1892-	H-08180	67
	Moshagen, Nils Mathias	1879-1937	H-08146	57
	Nilsen, Emil	1913-1951	H-08565	198
	Nøkleby, Ole	1877ca-1904	H-08542	279
	Olsen, Niels O.	1850-1861	H-08006	13
	Schee, Eivind	1899-1919	H-08705	246
	Skreia Handelslag	1917-1935	H-08498	263
	Vestre Tingelstad Samvirkeleg	1934-1982	H-08678	235
6.3	Hotel- og restaurant-virksomhet			
	Fjellhaug Skysstasjon	1863-1869	H-08781	264
	Grønen Skysstasjon	1859-1875	H-08500	263
	Hoff Skydsstasjon	1838-1858	H-08720	251
	Johnsens Privathotel, Chr	1893-1894	H-08184	69
	Kirkestuen, hotell	1829-1961	H-08177	66
	Klevgård, Gudbrand	1850-1930	H-08757	311
	Losnaos Hotell	1901-1967	H-08181	68
	Ødnes Hotell	1893-1945	H-08687	238
	Stavs markeds plass, Hotel	1907-1915	H-08183	68
	«Turisten», Hotel	1907-1908	H-08182	68
	Victoria hotel		H-08443	162
7.0	Transport, post og tele			
7.1	Transport og lagring			
	Austdal Skysstasjon	1898-1925	H-08836	185
	Blilie Skysstasjon	1853-1865	H-09049	294
	Böhmer, Nikolai Gunnerius	1855-1907	H-08809	149
	Eina Dampskibsselskab	1904-1906	H-09055	296
	Fjellhaug Skysstasjon	1863-1869	H-08781	264
	Grønen Skysstasjon	1859-1875	H-08500	263
	Oplandske Dampskibsselskap	1839-1883	H-08161	62
	Oplandske Dampskibsselskap	1854-1947	H-08494	262
	Strandengen Brygge	1908-1928	H-08461	168
	Wiese, Ludvig	1827-1873	H-08110	44
7.2	Post og telekommunikasjoner			
	Lillehammer/Gausdal Telefonselskab	1890-1932	H-08132	52
8.0	Bank og finansieringsvirksomhet			
8.1	Bank og finansieringsvirksomhet			
	Fron Sparebank	1863-1939	H-09007	151
	Fåberg Bygdemagasin	1831-1856	H-08158	61
	Haakenstad, Olaf	1896-1927	H-08677	234
	Totens Sparebank	1854-	H-09048	293

Fig. 2: Eks. på registerhenvisninger: En side fra Kategoriregister for Opplandsarkivet.

De fleste institusjoner har ikke utarbeidet slike kataloger og instituttet har, i et forsøk på å standardisere arkivregistreringen på landsbasis, utarbeidet egne registreringsskjemaer for hovedkataloger og spesialregistraturer til bruk i arkivinstitusjoner, hvor gjenparten av skjemaene sendes NPI.

Skjemaene er utformet med sikte på å være et hensiktsmessig redskap for den enkelte arkivinstitusjon ved katalogoppbyggingsarbeidet.

Fig. 3 illustrerer skjemaet i bruk og hvordan de opplysningene instituttet får inn om et arkiv kan være presentert fra oppbevaringsinstitusjonens side.

01 Arkivskapers navn Austdal Skysstasjon		02 Gr nr br nr 222
03 Tidligere navn Nordre Austdal		04 Gr nr br nr 116.1.
05 Annen arkivbetegnelse		
06 Start - opphør 1895-1925	07 Yrke - virksomhet Skysstasjon	
08 Uutfyllende opplysninger om arkivskaper Skysstasjonen på Austdal var før 1895 på Søndre Austdal, men fra 1895 ble den flyttet til Nordre Austdal. Stasjonen hadde to faste hester. De skysses til Sveen, 17 km Finden, 21 km og Gjøvik, 21 km.		
09 Arkivkategori Forretning		10 Tidssrom 1898-1925
11 Fylke - kommune Oppland, Gjøvik (Snertingdal)		
12 Emner Samferdsel, reiseliv		
13 Oppbevaringssted Harald Austdal	14 Signatur ET/82	15 Omfang 3 bøker
16 Klausulbestemmelse		
17 Arkivliste	18 Tidspunkt 1983	19 Kjøpt/gav/deponert
20 Arkivinnhold Dagbok 1898-1917 " 1917-1925 " 1901-1925		

Fig. 3: Eksempel på innsendte arkivopplysninger fra Opplandsarkivet.

Instituttet redigerer opplysningene og fører dem inn på et Datastarskjema på 81 felter (fig. 4), som er utgangspunkt både for det redigerte ummendraget i fig. 1 og registerdelen i fig. 2.

FIELD NUMBERS

KJEMA	H 08836		
PPBEV	Fiktupet	3	FYLKE 05 Oppland 5
AVN	Austdal Skysstation		6
NDRE NAVN	Austdal nordre (Biri 116 1 0)		7
			8
			9
			10
ØDSEL/STARTAR	189511		
ØD/SLUTTAR	192512		
ØRKE		13 14	15
		16 17	16
		19 20	21
		22 23	24
BIOGR.			
OPPLYSN.	Skysstasjonen på Austdal var før på Søndre Austdal, men fra 1895 ble den flyttet til Nordre Austdal. Stasjonen hadde to faste hester. De skyset til Sveen, 17 km, til Finden, 21 km og Gjøvik, 21 km.		
SIGNATUR	ET /82	29	
OMFANG	3 bøker	30	
FILGJENGE		31	
REGISTRATUR?	32	NAR INNFØRT 1932	
DEPONERT?	34	PRIV.OPPB.	35
KATEGORI	711 Landtransport		37
	38		39
	40		41
TIDSRUM	189842 - 1925 43		
KOMMUNE	Gjøvik	44	FYLKE 05 Oppland 46
		47	48 49
		50	51 52
		53	54 55
EMNEORD	Samferdsel	56	Reiseliv 57
		58	59
		60	61
NAVN I ARKIVINNHOLDET			62
			63
			64
			65
			66
ARKIVINNHOLD			
	Daqbøker 1898-1917, 1917-1925, 1901-1925.		67
			68
			69
			70
			71
			72
			73
			74
			75
			76
			77
			78
			79
			80
			81

Fig. 4: Eksempel på utfyllt Datastarskjema.

En sammenlikning av fig. 1 og fig. 3 viser at innførslene i katalogdelen er hentet fra forskjellige rubrikker i Datastar-skjemaet:

Oppbevaringsinstitusjonens SIGNATUR	hentes fra felt 29
NPIs interne SKJEMA-nr	fra felt 1-2
Arkivskaperens NAVN	fra felt 6
Fødsels-/stiftelsesår-døds-/sluttår	fra felt 11-12
BIOGRAFISKE OPPLYSNINGER om arkivskaper	fra felt 25-28
TIDSROM arkivet omfatter	fra felt 42-43
INNHold i arkivet	fra felt 67-81
OPPBEVARINGSSTED	fra felt 3 og 5
UTFYLLENDE OPPLYSNINGER om arkivet	fra felt 30-32 og 34
NÅR INNFØRT	fra felt 33

De øvrige rubrikkene i Datastar-skjemaet benyttes ved oppbygging av registrene. Feltene 6-10 (Arkivskaper, andre navn og betegnelser på arkivskaper) og 62-66 (egennavn omtalt i arkivet) slås sammen til et navnerregister. Feltene 44-55 (kommune og fylke arkivinnholdet omhandler) er grunnlaget for det topografiske registeret, feltene 36-41 (arkivkategori) er grunnlaget for kategoriregisteret og feltene 56-61 (emne) for emneregisteret.

Hovedinngangen til katalogdelen er navnerregisteret. Dette er et alfabetisk register som inkluderer arkivskaper, arkivskapers tidligere navn, f.eks. forening som skifter navn, annen arkivbetegnelse – enkelte institusjoner kaller et arkiv noe annet enn ved arkivskaperens navn – og navn som er omtalt i arkivinnholdet. F.eks. «Mot Dag» i Karl Evangs arkiv. Slik blir navnerregisteret inngang til Mot Dag, for en som er interessert i sosialistiske grupperinger i mellomkrigstiden, men som ikke har personkjennskap til arkivskaperen helsedirektør Evang, som hovedopplaget i katalogdelen lyder på.

Det topografiske registeret er et register over fylker, kommuner og større distrikter som er omhandlet i et arkiv. Dette er i første rekke en inngang til katalogen fra en lokalhistorisk vinkel.

Disse to registrene er uproblematisk å produsere – feltene slås sammen og sorteres etter ulike kriterier som alfabetisk orden og topografisk nivå. Man kan naturligvis diskutere alfabetiseringsregler, nivå osv., men de er greie å slå opp i:

Vet man navnet på en person eller institusjon slår man opp i navnerregisteret og ser om oppslaget finnes. Hvis så, følger man sidehenvisningen til hovedkatalogen og ser på det summariske sammendraget om det kan være aktuelt å undersøke selve arkivet. På tilsvarende måte kan en bestemt lokalitet ettersøkes i det topografiske registeret.

Dette er velkjente registertyper i mange sammenhenger. Prinsippene

for oppbyggingen av kategoriregisteret og emneregisteret er derimot spesielle for samkatalogen. Utviklingen av disse er resultatet av sorteringer og bearbeiding av det edb-registrerte materialet og hadde ikke vært mulig uten de muligheter for kryss-referanser som edb-registreringen åpnet muligheten for.

Før disse registrene presenteres, må prinsippene for arkivarbeid redegjøres for og sammenliknes med andre typer informasjonssystematisering.

Et arkiv er som sagt en samling dokumenter som er resultatet av en gitt virksomhet. Arkivarbeid tar utgangspunkt i *proveniensprinsippet*. Det vil si i den virksomheten som skaper arkivet. Materialet etter denne virksomheten holdes samlet, og adskilt fra arkiver etter all annen virksomhet. I så måte skiller arkivarbeid seg fra f.eks. tradisjonelt biblioteksarbeid hvor man nettopp prøver å systematisere og gruppere sammen bøker o.a. som hører til under samme emne. Dette kan vi kalle et realistisk ordningsprinsipp.

Mens en bok er et avsluttet hele, som kan systematiseres med utgangspunkt i innholdsfortegnelsen, er et arkiv en historisk kilde, som, vurdert som et samlet hele, forteller om arkivskaperen. Strukturen i selve arkivet beretter om utviklingen av arkivskaperen, mens arkivinnholdet belyser de arbeidsområder arkivskaperen var engasjert i. Men som en historisk kilde kan arkivet i andre sammenhenger også belyse problemstillinger som kanskje sto arkivskaperen nokså fjernt.

Et samvirkelagsarkiv belyser den forretningsmessige utviklingen av en bestemt bedrift i et bestemt lokalsamfunn. Kundeboke i dette arkivet kan belyse kredittforholdene i lokalsamfunnet, men en ernæringshistoriker kan kanskje bruke slike kundeboke som kilde til en ernæringshistorisk undersøkelse.

Arkivet, den historiske kilden, kan brukes i mange ulike sammenhenger. Det er da viktig at kildens indre orden, som beretter om den historiske sammenheng arkivet er vokst fram av, ikke ødelegges gjennom utnyttelsen av kilden, eller ved tilrettelegging av den for forskningsformål.

Hvert arkiv må derfor betraktes som en enhet, ordnes for seg, med arkivskaperen som ordningsenhet.

Forskjellige arkiv kan imidlertid ha den *indre strukturen* felles. To bedrifter i helt ulike bransjer som belyser, realistisk sett, ulike emner, kan ha helt like trekk: protokoller fra samme periode, lønningslister, statistisk materiale. Opplysninger om denne strukturen kan gi brukeren et grunnlag for å vurdere typer av problemstillinger arkivet kan tenkes å belyse.

En arkivars beskrivelse av et arkiv med utgangspunkt i proveniensprinsippet – katalogiseringsarbeidet – tar nettopp utgangspunkt i denne likheten i indre struktur som arkivene etter forskjellige arkivskapere har.

Arkivene deles opp i forskjellige kategorier. Først skilles det mellom offentlige og private arkivskapere. Gruppen private arkiv kan så deles opp i underkategorier:

- person og slektsarkiver
- gårdsarkiver
- forenings- eller organisasjonsarkiver
- bedrifts- eller forretningsarkiv
- institusjonsarkiv.

Innen hver av disse gruppene kan en så vente seg at arkivinnholdet vil ha visse felles trekk. F.eks. brev og dagbøker i et personarkiv, regnskapsmateriale i et forretningsarkiv.

Hvert enkelt arkiv er, ideelt sett, detaljert beskrevet i en spesialregistratur.

De tradisjonelle hjelpemidlene forskere får i arkivsammenheng er arkivkataloger – hovedkatalog, spesialregistraturer, arkivskaperregister (alfabetisk), kategoriregister og topografisk register. Et systematisk register, som vi kjenner fra bibliotek, ser man ikke ofte i arkivinstitusjoner. Når de forekommer gjelder de som regel helt spesielle samlinger.

Men forskere er ofte interessert i emner og problemstillinger, de vil gjerne ha henvisninger til alt om Ibsen, uavhengig av materialets proveniens. Her har bibliotekenes realistiske system store fordeler med systematiserte henvisninger på en rekke emner. Svakheten ved bibliotekenes katalogiseringssystem i arkivsammenheng er at det er vanskelig å få oversikt over et helt enkeltarkiv, hva det samlede resultatet av en gitt virksomhet var. Inngangen(e) er primært orientert mot de enkelte deler arkivet består av.

Edb-registrering av arkivopplysninger åpner muligheten for å *kombinere* disse to prinsippene for katalogisering, hvor hensyn til proveniensprinsippet ivaretas gjennom kategori-registreringen samtidig som man kan få en brukervennlig emneregistrering, uten å øve vold på prinsippet om bevaring av arkivets indre enhet.

Da NPI satte i gang med innskriving av opplysninger om de enkelte arkivene, basert på institusjonenes egne kataloger og innsendte opplysninger i 1980, forelå det lite erfaringsmateriale fra edb-arbeid i arkiv- og katalogarbeid å støtte seg på. Vi var klar over at emneregistrering krevde en thesaurus eller en annen form for kontrollert emnevokabular, som vi ikke hadde. Vi valgte å eksperimentere med å la forskjellige depotinstitusjoner som selv kjente arkivinnholdet komme med forslag til emneord og på grunnlag av opplysninger om arkivenes innhold i institusjonenes kataloger kunne vi trekke ut opplysninger om omhandlede emner.

Vi arbeidet dermed med frie emneord, og faste kategorier:

- P - personarkiv
- G - gårdsarkiv
- M - manuskript

- S - samling
- I - institusjonsarkiv, delt i 5 typer: undervisning, helse, legat og offentlig
- 3 siffer - bedriftsarkiv, delt etter SSBs næringslivsgruppering
- F - foreningsarkiv, klassifisert med utgangspunkt i inndeling i J. Moren: *Norske organisasjoner*, Oslo 1976.

«Manuskript» er ikke noen tradisjonell arkivkategori. Gjennom registreringsarbeidet har det vist seg at en rekke institusjoner ikke skiller i sine samlinger mellom avsluttede arkiver og enkeltmanuskripter av forskjellige forfattere, innsamlet for bestemte formål. NPI har derfor ved registreringen innført kategorien M - manuskript for å kunne skille ut disse og eventuelt bearbeide dem på et senere tidspunkt.

Bearbeiding av kategori- og emneregistre tok til ved at de frie emneordsforslagene ble utlistet med tanke på standardisering og gruppering av oppslagsordene.

Utlistingen viste at arkivinstitusjonene brukte emneordene til å dekke ulike typer behov for innganger til samlingene, alt etter arkivinstitusjonens art:

Noen brukte emneordene som inngang til tverrregistrering av samlingen, f.eks. «arbeiderbevegelse» på alle typer arkiv som belyser den faglige og politiske arbeiderbevegelse: fagforeninger tilsluttet LO, partier som DNA, SF, NKP, personarkiv etter veteraner, enkeltmanuskripter med arbeiderminner m.m.

Dette var en grei utnyttelse av emneordet som inngang til ulike arkivkategorier, men så dukket samme henvisning opp til et bedriftsarkiv. Her viste det seg at bruken av emneordet skiftet karakter: materialet i bedriftsarkivet - lønningslistene - ble *vurdert* som en mulig kilde til arbeiderbevegelsens historie. Denne vurderende bruk av emneordet er ikke akseptabel. Kategoribegrepet skal dekke beskrivelse av arkivets struktur. I alle bedriftsarkiv kan man i prinsippet forvente å finne lønningslister. Alle slike lønningslister kan i prinsippet være en kilde til arbeiderbevegelsens historie vidt definert.

Den andre vanlige bruken av emneordet var som stikkord til eller utfyllende opplysning om arkivinnholdet: «matoppskrifter» som henvisning til et personarkiv etter en politiker, som inneholdt en uventet opplysning om nødproviant.

Den tredje bruken av emneordene siktet på å beskrive faste trekk ved innholdet som oppbevaringsinstitusjonene var interesserte i å registrere:

Museer bruker ofte sine skriftlige kildesamlinger i illustrerende sammenheng eller til å underbygge utstillinger, og kan være primært interessert i spesielle deler av arkivet som f.eks. fotografier, lydbånd eller håndskrevne aviser.

Dette innebærer en overlappning med kategoribegrepet og kriteriet for slik bruk av emneordet må være at innholdstrekket faktisk er spesielt og ikke generelt dekket av arkivkategorien. Et emneord som «skifter»

dukket ofte opp. Det er standardinnholdet i gårdsarkiv, og vil gjelde ca. 90% av gårdsarkiv om konsekvent innført. Det er dermed normalt uinteressant som oppslagsord i arkivsammenheng.

Her er vi ved kjernen i kategoribegrepet. Det tar utgangspunkt i at man kan forvente å finne en generell felles struktur i arkivene innen en gitt kategori – nettopp skifter og adkomstpapirer i gårdsarkiv. Emneordet må derfor sikte på å beskrive trekk ved arkivinnholdet som går ut over dette.

Gjennomgangen av emneordlistene viste at brukergruppen var større enn prosjektopplegget hadde forutsatt. NPI kom fram til at målet måtte bli å utvikle kategoriregistreringen slik at den kunne gi grunnlag for å gruppere arkivene etter kriterier som gir nye innganger til arkivene – tverrregistrering – og utvikle en registerstruktur hvor emneordsinngangen supplerte kategoriregistreringen.

Emneordlistene ble sammenliknet med grupperingene som kom fram gjennom kategoriregistreringen.

Ved en kryss-gjennomgang av emneord og kategorier viste det seg at emneordene i liten grad ga utfyllende informasjon m.h.t. forretningsarkivene, som hadde allerede i utgangspunktet en detaljert kategorisystematisering (SSB's næringsgruppering).

Denne erfaringen overførte vi så til kategorien foreningsarkiv, slik disse ble findelt også ut over Morens opprinnelige inndeling. Dermed falt behovet for utfyllende emneord bort for en stor del på disse to arkivkategoriene hvor arkivene vanligvis har en vel definert struktur.

Bruken av systematisk kategoriklassifisering ble ytterligere utvidet ved at dobbelt klassifisering ble innført på edb-registreringsnivå.

Etter proveniensprinsippet er hovedkategorier entydige. Enten er et arkiv et personarkiv eller så er det et forretningsarkiv. Finner man to typer materiale blandet skilles de ut og man får to arkivskapere – personen og bedriften. Dessverre er arkiv ofte sammenblandet, bl.a. fordi virkeligheten er sammenblandet, og en konsekvent utskilling ville være å øve vold på nettopp den indre sammenhengen som proveniensprinsippet tar sikte på å bevare. Utskilling gjennom ordning fortøner seg da meningsløst. Som det går fram av gjennomgangen av hovedkategoriene for privatarkiv ovenfor har man tatt dette til følge og akseptert sammenblandede arkiv som person/slektsarkiv. Disse er som oftest en samling papirer etter flere arkivskapere.

Allerede ved den første utarbeidelsen av kategorikoder for registrering av arkivopplysninger på Datastar-skjemmet fulgte vi opp denne modifiseringen av kategoriprinsippet når det gjaldt forretningsarkiv med en findeling av kategorirubrikken, basert på bransjetilknytning, slik at vi kunne ta hensyn til at samme bedrift kunne være engasjert i en rekke ulike bransjer. Et bedriftsarkiv kunne så bli registrert med flere kategorikoder.

Etter gjennomgangen av de registrerte dataene overførte vi som nevnt

denne findelingen også til foreningsarkivene, men med visse modifikasjoner i klassifikasjonen. De fleste foreninger har *ett* formål – avhold, idrett, politikk. Virksomheten kan imidlertid gi utslag i ulike deltakende sosiale grupperinger. Forskere er ofte interessert i dette aspektet av foreningstypen – *hvem* som organiseres, såvel som arbeidsområdet – formålet.

Findelingen av foreningskategorien tar hensyn til arbeidsområdet, man kan angi ett eller flere virksomhetsområder. I tillegg gir dobbelt-kategoriklassifisering en mulighet til å angi foreningstype – kvinneorganisasjon, ungdomsorganisasjon – slik at vi her får en *tverregistrering* på kategorinivå.

Er man så interessert i kvinners generelle politiske organisering kan man skille ut arkivopplysninger som både har kategorien politisk forening og kvinneforening, og se nærmere henvisningene til slike arkiv i katalogdelen.

Vi har dermed nå etablert følgende innganger til hovedkatalogen: Navneregister – alfabetisk etter navn, etternavn på personer. Samtlige kategorier arkiv inngår i registeret.

Topografisk register – arkivskaperregister ordnet alfabetisk etter fylke, kommune/region. Samtlige arkivkategorier inngår i registeret, men det er en hovedinngang til gårdsarkivene.

Kategoriregister – arkivskaperregister over forretningsarkiv, foreningsarkiv og institusjonsarkiv ordnet etter kategorinøkkel.

På emnerregisteret gjenstår det fortsatt mye arbeid før den er endelig etablert.

Samtlige arkivkategorier inngår i dette registeret, men med de bearbeidelsene som er redegjort for, vil dette registeret primært bli en inngang til den store og sammensatte personarkivkategorien og til samlingene.

Søkingen i hovedkatalogen er basert på at man utnytter et register eller flere i kombinasjon, enten manuelt i mikrokort-utgaven, eller ved å sette opp en kombinasjon av kriterier i databasen i Riksarkivbygget.

Katalogen har ennå ikke funnet sin endelige form. Vi håper snart å kunne presentere en prototype av samkatalogen i mikrokort-format, som kan diskuteres og utprøves av en mindre krets av brukere slik at vi kan få reaksjoner og korrektiv.

Det som har vært og er utfordrende ved prosjektet er at vi er i gang med en registrering av arkivopplysninger på tvers av institusjoner og på tvers av arkivgrupper i en helt annen målestokk enn har vært forsøkt tidligere, med åpninger av nye utnyttelsesmuligheter av historisk kilde-materiale som resultat.

Elisabeth Koren er arkivar ved Norsk privatarkivinstitutt med registrering og bevaring av foreningsarkiver som særlig arbeidsområde.

Edb og arkeologi

Britt Kroepelien

I den første uken av juli i år ble det under en utgravning i den gamle bispegården i Oslo (nåv. Oslo Ladegård) startet opp et pilotprosjekt, hvor edb-baserte landmålingsteknikker forsøkes tilpasset arkeologisk feltarbeide.

Prosjektet ble vist stor oppmerksomhet. Til stede var både nordiske og klassiske arkeologer, og Aftenpostens morgenutgave den 5. juli omtalte prosjektet med et stort bilde på både første og siste side.

Dette er bare første fase av et meget omfattende prosjekt, og denne artikkelen vil derfor forsøke å gi en mer faglig detaljert beskrivelse av hele prosjektet.

Bakgrunn

I løpet av de siste årene er utviklingen innen edb gått svært raskt. Nyere forskning har muliggjort langt mer effektive teknikker innen de fleste naturvitenskapelige arbeidsfelt. Samtidig har maskinutrustning (hardware) fram til dagens relativt rimelige 16 bits mikromaskiner, åpnet for muligheter som tidligere bare store og svært kostbare datamaskiner kunne tilby. Programvare, utviklet til bruk på disse store datamaskinene, blir derfor i disse dager i stor utstrekning tilrettelagt for de små mikromaskinene.

Staten og industrien har bevilget millionbeløp til produktutvikling av avanserte dataprogrammer for landmålings- og kartindustrien. Dette har ledet til utvikling av nye teknikker, hvor også moderne målemetoder tas i bruk. Resultatet har vist seg å gi meget store fordeler, både når det gjelder hurtighet, nøyaktighet og fleksibel anvendelse av dataene. Denne programvaren blir nå i økende grad tilgjengelig for mikromaskiner.

Siden dette området er nær beslektet med det arkeologiske, tar prosjektet sikte på å finne ut hvilke gevinster arkeologien vil kunne oppnå ved å tilpasse de foreliggende resultater fra landmålings- og kartindustrien til arkeologisk metodikk. I så fall tror vi dette vil åpne for helt revolusjonerende muligheter innenfor såvel nordisk som klassisk arkeologi.

Ved arkeologiske utgravninger samles det inn et stort materiale om gjenstander og øvrige funn. Et av de store problemene i dag er å administrere denne datamengden, slik at det er mulig å trekke ut holdbare vitenskapelige konklusjoner. Det er stor fare for at oversikten forsvinner i datamengden, og det vil derfor bli langt lettere å foreta vitenskapelige undersøkelser ved en bedre dataorganisering enn den

som finnes i dag.

Prosjektet kan derfor deles inn i to hovedområder: feltdokumentasjon og arkeologisk etterarbeide.

Arkeologisk feltdokumentasjon

I dag måles og dokumenteres arkeologiske utgravninger i forhold til et rutenett. Denne metoden gir ikke eksakte koordinatverdier, men kun en referanse til hvor funnet ble gjort i hver enkelt rute. En ny målemetode vil gjøre det mulig å koordinatfeste alle funn og andre punkt av interesse i grunnriss og høyde. Dette er meget tidkrevende og gjøres i dag manuelt kun ved viktige enkeltfunn.

Sammen med de beregnede koordinater kan bygningenes konstruksjonsnummer, jordlagsnummer eller gjenstandenes funnummer registreres. Dette kan plottes umiddelbart i valgfri målestokk og kontrolleres på feltet, eller brukes i den videre planlegging av feltarbeidet. Det vil også være mulig å få tegnet ut et sett bestemte bygninger og konstruksjoner, som utfra den stratigrafiske analyse er antatt å være samtidig. Om ønskelig kan dessuten bestemte lag, f.eks. brannlag, legges på. Gjenstander kan plottes ut på plantegning over konstruksjoner, f.eks. husgeråd, håndverksavfall m.m., enten angitt ved ulike symbol eller ved påføring av funnets katalogiseringsnummer.

Arbeidet i felten vil derfor kunne skje mer effektivt, og det vil gi en høyere kvalitet hva angår nøyaktighet og datainnnsamling.

Arkeologisk etterarbeide

Den største gevinsten vil imidlertid ligge i det arkeologiske etterarbeidet. Plantegninger kan tegnes med tusj av plotter, og store deler av tegnearbeidet spares. Plotteren tegner raskt ut plantegninger ved de enkelte faser, som dermed er klare for publisering med en gang.

Kvaliteten på etterarbeidet vil heves betraktelig. Når jordlag er topp- og bunn-nivellert, kan det være aktuelt å volumberegne de forskjellige lagene, plotte de med angivelse av forskjellig tykkelse, og med statistikk over antall gjenstander i hvert lag. Denne arbeidsmetoden er allerede prøvet på gårdhaugen Soløy i Troms. Dette er imidlertid så tidkrevende at ved bygravningene og manuell innmåling av lagene vil en slik metode være utelukket.

Utvikling av programvare

Førsteantikvar *Erik Schia* tok sammen med undertegnede kontakt med rådg.ing.firma Viak A/S, fordi dette firma med støtte fra NTNF allerede har utviklet anerkjente automatiserte landmålings- og kartleggingsteknikker. Programvaren til disse foreligger i programbiter (moduler), som lar seg sette sammen, slik at forskjellige behov kan imøtekommes.

Resultatet av diskusjonene med Viak A/S er blitt et samarbeidsprosjekt, hvor Viak A/S skal utforme en interaktiv database /geografisk- og egenskapsbase/. Arkeologene vil definere datainnhold og funksjoner. Det er særdeles viktig at utformingen av egenskapsbasen blir omfattende, fordi vi da oppnår størst mulig fleksibilitet ved etterarbeidet.

Målet er ved felles anstrengelser å komme frem til en «skreddersydd» arkeologi-programpakke, som dekker de generelle arkeologiske behov, og samtidig er fleksibel nok til å imøtekomme de mer spesialiserte ønsker.

Programpakken

Programpakkens oppbygging er skissert i figuren.

Systemet deles opp i tre hoveddeler: Datafangst, lagring og presentasjon. Systemet kan i stor grad basere seg på allerede eksisterende programvare. Det er bygget opp av moduler, og de enkelte moduler vil kunne fungere delvis uavhengig av hverandre.

Datafangst

Datafangst tenkes utført ved landmåling og feltregistrering av egenskapsdata.

Totalstasjon: Dette er en kombinert teodolitt og avstandsmåler med dataregistreringsenhet. Ved å trykke på en knapp kan avstand, retningsvinkel og høydeforskjell til et punkt registreres. Operatøren kan også lagre en identitet og eventuelt en temakode sammen med observasjonene. De lagrede observasjoner kan siden spilles over til en datamaskin for beregning av koordinater i et beregningsprogram.

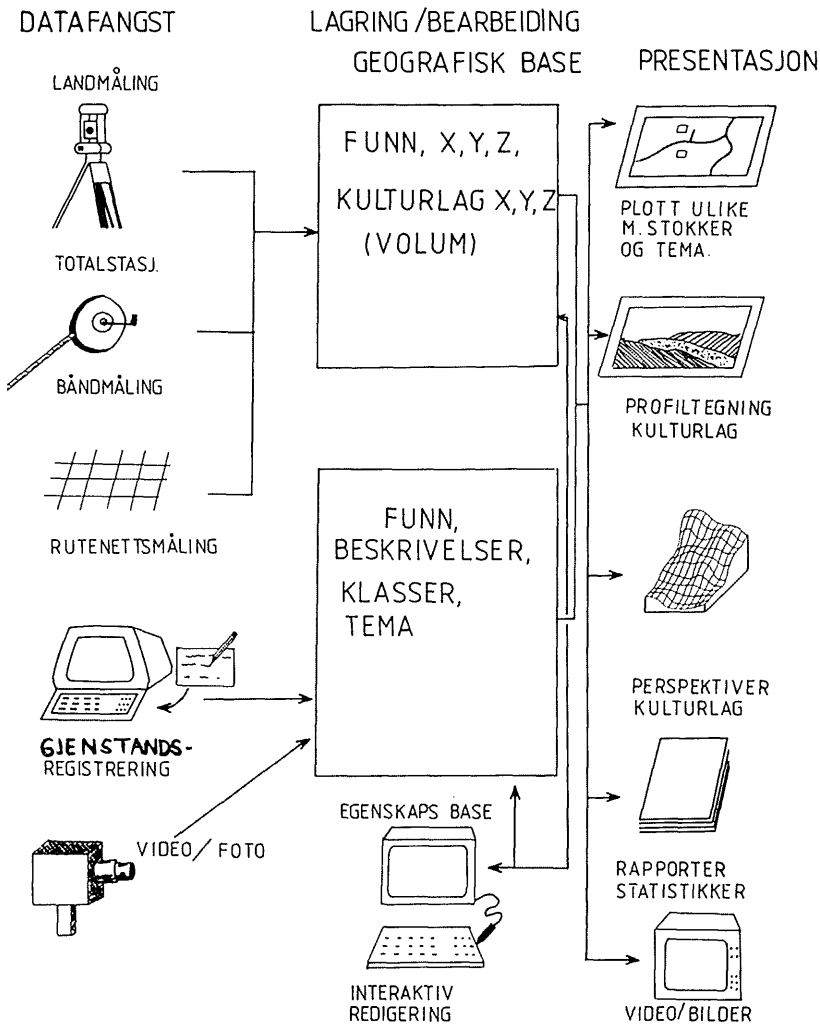
Båndmåling: Dette kan foretas ved at det opprettes enkelte kjentpunkter i og rundt utgravingen. Fra to eller flere av disse strekkes målebåndet til den gjenstand som ønskes innmålt. Høyden kan finnes ved stigningsmåler eller nivåer. Punkter skal også kunne måles inn ved hjelp av vinkelprisme og målebånd. Koordinater beregnes fra målingene i egnede beregningsprogram.

Rutenettmåling: Programmet må kunne motta koordinater fra tradisjonelle rutenettmålinger i felten.

Registrering: Egenskapsdata ved funn og kulturlag skal kunne gis inn på terminal og lagres i egenskapsdatabasen.

Video/foto: Video-opptak og fotografier av gjenstander lagres på en video-kassett, og en referanse til kassetten legges i egenskapsdatabasen.

Det skal være mulig å legge inn fotogrammetriske registreringer i databasen. En koordinatfil registrert i et fotogrammetrisk instrument leses inn i systemet. Både luftopptak og terrestriske opptak skal kunne legges inn. Fotografering og registrering kan gjøres av underleverandør



som leverer koordinatfilen. Etter fotografering sendes bildene til arkeologen som markerer de gjenstander som skal registreres før målingene i bildet gjøres. Arkeologen mottar en koordinatfil fra disse registreringene.

Koordinatberegning

Dette baseres på eksisterende programvare. Programmene kan beregne koordinater fra de nevnte innmålingsteknikker. Det eksisterer også program for å beregne koordinater fra andre typer målinger. Dette gjøres under innlesing til den geografiske databasen.

Geografisk database

Denne skal inneholde funntype, koordinater for funnet og nødvendige referanser til egenskapsdatabasen. Koordinater for punktene i kulturlaget og volum på laget skal finnes på basen. Det enkelte punkt skal ha en unik punktidentitet og en temakode. Databasen skal kunne slås sammen med andre tilsvarende databaser.

Egenskaps-database

Denne databasen skal inneholde beskrivelser av de enkelte funn og kulturlag. Disse skal kunne grupperes i tema og enkelte tema skal igjen kunne samles i klasser. Det skal være referanse til eventuelle video- og foto-opptak. Databasen kan baseres på standard databasesystemer.

Interaktiv redigering

Databasene skal kunne redigeres fra terminal. Det skal være mulig å legge til nye funn og å legge inn eller endre opplysninger om eksisterende punkter på databasen. Det kan være aktuelt å gi mulighet for redigering og tegning av databasene på grafisk skjerm.

Presentasjon

Data skal kunne presenteres fra databasen som kart, profiler, perspektivskisser, statistikk og video/fotoopptak.

Kart: Det skal kunne tegnes kart i forskjellige målestokker, og det skal være valgfritt hvilke temaer og/eller klasser som skal tegnes. Det skal kunne brukes forskjellige symboler for de enkelte tema. Dette kan delvis gjøres med eksisterende program.

Profiler: Det opprettes en digital terrengmodell fra den geografiske databasen som kan generere profiler for de enkelte kulturlag. Profilets retning og plassering i terrenget skal kunne velges fritt. Det skal kunne velges målestokk i høyde uavhengig av målestokk i grunnriss.

Perspektivskisser: Dette kan tegnes av de enkelte kulturlag fra terrengmodellen beskrevet over. Skissen skal ha valgfri målestokk og betrakningsretning. Det finnes program for digital terrengmodell som kan gjøre dette i dag.

Statistikk: Fra egenskapsdatabasen skal det kunne listes statistiske data som f.eks. antall funn av forskjellig type i et gitt kulturlag, antall funn pr. volum i et lag, liste over alle funn med bestemte egenskaper etc. Det skal være mulig å generere rapporter fra databasen.

Video/foto opptak: Fra terminal skal det kunne kalles fram fotografier og filmsekvenser på skjerm fra en videospiller. Disse skal kunne vises både under redigering av basen og som en egen presentasjon av type: «vis alle kniver funnet i et bestemt kulturlag og område».



Gamle og moderne måle- og tegneinstrumenter side om side. Til venstre på bildet: manuell måling og tegning. I forgrunnen til høyre innmåling ved hjelp av en tegne maskin (pantograf). I forgrunnen midt på bildet et edb-basert landmålingsinstrument som via en reflektor på målestedet (på stangen i bakgrunnen) gir X-, Y- og Z-koordinater ved å trykke på en knapp. Bygningen i bakgrunnen er Oslo Ladegård.
Foto: Knut Snare, Aftenposten.

Fremdriftsplan

Prosjektet består av 3 hoveddeler. Del 1 er en egen selvstendig del, som kan fungere uavhengig av delene 2 og 3. Dersom økonomien gjør det vanskelig å finansiere det hele, vil det uten tvil være et viktig skritt fremover for arkeologien om prosjektets del 1 kan gjennomføres.

Prosjektets del 1:

Ved det allerede omtalte pilotprosjektet i den gml. bispegården i Oslo, ble moderne landmålingsinstrument prøvet for registrering av konstruksjoner, jordlag og funn. Bearbeiding av måledataene vil skje utover høsten 1985. Med utgangspunkt i de erfaringene vi her har vunnet, vil programpakken for feltregistrering utarbeides våren 1986. Planen er å prøvekjøre denne på en ny, stor utgravning i Gamlebyen (utgravnings-sesong mai-oktober).

Prosjektets del 2:

Arbeidet vil i denne delen av prosjektet konsentreres om å tilpasse den nye programpakken de allerede utviklede systemer for katalogiseringen av gjenstander, spesielt i Gamlebyen. Disse er tidligere blitt utarbeidet med midler fra NAVF.

Det vil bli arbeidet med tilretteleggelse av gjenstandsmaterialet for video, slik at verbalbeskrivelsen kan suppleres med et bilde.

Undertegnede vil tilrettelegge programpakken for klassisk arkeologi.

Prosjektets del 3:

Det vil i samarbeid med NAVFs EDB-senter arbeides med tilpasningsmuligheter til ulike programpakker for statistisk databehandling av arkeologiske funn. Spesielt gjelder det programpakken STAR, som er utarbeidet av prof. *Stig Welinder* for midler fra NAVF.*

Økonomiske aspekter

Et prosjekt med disse dimensjoner har selvfølgelig økonomiske aspekter.

For det første er det enormt kostbart å utvikle det. Her har vi håp om at NTNf vil yte støtte til utvikling av programvaren. Videre håper vi at NAVF vil gi støtte til forutgående metodeforskning.

Et annet økonomisk aspekt er at ved å innføre det nye system, vil kostbar forskertid kunne frigjøres for viktigere oppgaver. Dermed vil de midler samfunnet stiller til disposisjon for de humanistiske disipliner, bli mest mulig effektivt utnyttet. Det var egentlig dette som inspirerte oss til å gå i gang med prosjektet.

Utgravninger i middelalderbyene er kostbare og tidkrevende. Prisen pr. utgravd m³ er i dag ca. kr. 4000. Etter lov om kulturminner av 9.

* Se HD 1-83 (Red.anm.)

juni 1978 plikter grunneieren å betale eventuelle arkeologiske utgravninger før utbygging kan starte. En kostnadsbesparelse vil derfor også komme grunneieren til gode i tillegg til den tidsbesparelse han oppnår ved den nye metoden. I 1984 betalte private byggherrer 7 millioner kroner for Riksantikvarens arkeologiske utgravninger ved byggeprosjekter bare i middelalderbyene Oslo, Tønsberg, Bergen og Trondheim.

Etter at vi kom i gang, har vi dessuten forstått at vårt prosjekt på grunn av sin oppbygging også kan ha interesse ut over Norges grenser, og her kommer det tredje økonomiske aspektet. Greier vi å bygge opp et system etter den målsetting som er beskrevet foran, vil det ikke bare kunne effektivisere vårt eget arbeid. Det vil også f.eks. på royalty-basis kunne stilles til disposisjon for andre lands forskningsinstitusjoner. De humanistiske vitenskaper vil derved få et verdensomspennende felles-sprog, som vil bidra til et samarbeid om ivaretagelsen av den felles kulturarv – et samarbeid som vi i dag bare aner konturene av.



*Britt Kroepelien er stud.mag.art.
i kunsthistorie ved Universitetet i
Bergen.*

DUS-forfatter-systemet: ComCats

(ComCats – The Complete Computer Assisted Training Systems)

Erik Meistrup

Baggrund

DUS – Datamatformidlet UndervisningsSystem – er et dansk edb-forfatter system, som oprindeligt er udviklet i et samarbejde mellem edb-folk og pædagoger på Århus Tandlægehøjskole og Jydsk Telefon.

Dette fællesprojekt havde som målsætning at tilvejebringe et edb-værktøj for undervisere på alle niveauer, hvormed disse kunne skabe og producere egne undervisningsprogrammer på mikrodatamater uden at have et større kendskab til edb-programmering i øvrigt.

DUS-sproget

Projektet resulterede også i udformningen af et særligt høj-høj-niveausprog DUS-sproget, som er kendetegnet ved at betjene sig af ca. 100 maskinordrer, og ved at næsten alle ordre er danske.

Et højniveausprog er et programmeringssprog, som forenkler maskinprogrammeringen uden væsentligt at begrænse programudviklingsmulighederne.

DUS-filosofien

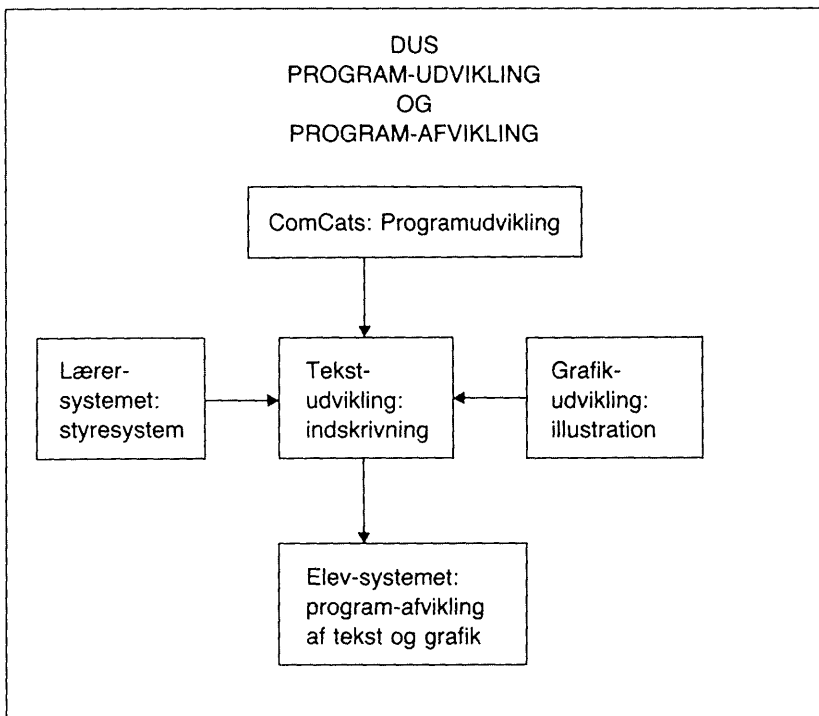
Den bærende filosofi har været at skabe et alsidigt værktøj, som kunne imødekomme både den uøvede og den professionelle programforfatter. Systemet er derfor udformet som en dynamisk og åben pædagogisk struktur, hvor også sprogbugen er koncentreret om at skabe maksimal pædagogisk kontrol med teknikken.

En anden meget væsentlig filosofi har været at skabe et edb-værktøj, der lægger op til at tænke og arbejde i helheder – i modsætning til at arbejde i enkeltlinier indenfor det enkelte skærmbillede som f.eks. Basic og COMAL er baseret på. Dette indebærer, at det aller vigtigste programudviklingsarbejde finder sted uden for datamaskinen, hvor man arbejder med målformuleringsproblematikker (pædagogisk hensigt), programstrukturudformning og skærmdesign.

Et tredje princip har været at koncentrere sig om de «personlige» datamater i modsætning til de store edb-anlæg, der giver mulighed for fjernundervisning. Man opdagede, at undervisning bliver nedprioriteret, hvis anlægget også bruges til andre formål, hvilket kan give urimelige ventetider i afviklingsfasen.

Endelig kan man fremhæve, at et fjernundervisningsideal også indebærer den implicitte forudsætning, at man lettere arbejder med centraliserede målformuleringer. Centraliserede målsætninger lægger op til stordrift løsninger (billiggørelse) ved programudviklingen og programbrugen. Man kan altså frygte, at en stigende brug af fjernundervisningsideen kan medvirke - yderligere - til at fremme et mekanistisk menneskesyn.

DUS DIALOGBASERET FORFATTERSYSTEM



DUS-systemet består af 2 dele, nemlig en programudviklingsdel og en programafviklingsdel.

Programudviklingsdelen

Denne del kaldes også *forfatterdelen*, og den sætter brugeren i stand til at skabe forskellige slags programmer til forskellige former for undervisning.

På grund af den omfattende indbyggede vejledning er det ikke bare let at lære at betjene sig af systemet og dets faciliteter, men også meget

let at bevare sine evner til – og muligheder for – at lave nye programmer, selvom man ikke arbejder med programudvikling hele tiden. Andre mere traditionelle programmeringssprog fordrer, at man regelmæssigt arbejder i det pågældende sprog for at kunne bevare sine programmeringsevner på et brugbart stade. I forfatterdelen er det muligt at arbejde i 3 forskellige former for indskrivning (også kaldet editorer):

En *blanketform*, der kræver et absolut minimum af kendskab til programmeringssprog, og derfor kan betjenes af de fleste, der ønsker at arbejde som programforfattere.

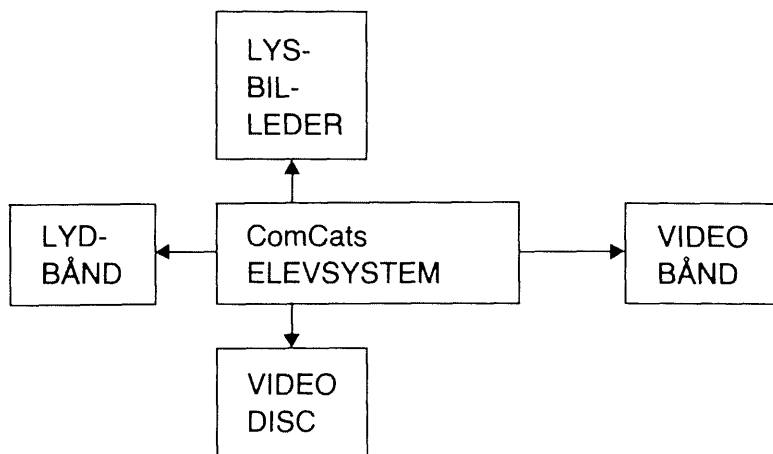
En *fri form*, der fordrer kendskab til DUS-sprogets ca. 100 programmeringsordrer, men som til gengæld giver en væsentligt forøget frihed i programopbygningsfasen.

En *grafik del*, hvor det er muligt uden programmering via en række fastlagte funktionskald at lave forskellige grafiske illustrationer til brug for tekstbehandlingen. Disse illustrationer kobles bagefter sammen med tekstindskrivningen.

Programafviklingsdelen

Denne del benævnes også ELEVSYSTEMET. Dette system indeholder en fortolker, der muliggør, at et program kan anvendes af en bruger (elev).

Elevsystemet giver mulighed for at bruge andre audiovisuelle medier som illustration til programindholdet.



Programudviklings måder

DUS-systemet er på én gang både et *menuorienteret* og *ordreorienteret* system.

Menuorienteringen betyder, at man som programforfatter præsenteres for en hovedliste (hovedmenu), og ud fra den vælger sig ind til andre lister (undermenuer). Menusystemet er altså et hierarkisk system, hvor man vælger ordrer ved at bevæge sig fra ét niveau til et andet. Skiftet fra den overordnede menu til de underordnede menuer sker ved valg af ét af de ordrebogstaver, som den overordnede menu stiller til rådighed.

Ordreorienteringen betyder, at man kan kombinere flere på hinanden følgende ordrevalg i en og samme arbejdsgang, og at man fra et vilkårligt sted i menustrukturen kan aktivere en ordresekvens, som også involverer hovedmenuens ordretilbud.

DUS-hovedmenu

The screenshot shows a terminal window with a main menu. The menu is organized into three main sections: Programming, Printing, and Help functions. Each section lists specific actions and their corresponding keyboard shortcuts. At the bottom, there are instructions for selecting menu items and exiting the program.

DELEHNAVN	EMNAVN	DREV
Programmering	O: Oprette (ønske, delemsne) V: Vælg (ønske, delemsne, drev) I: Indskrive/rette program K: Klargøre program til slævbrug A: Afprøve slævprogram	C
Udskrifter	U: Udskrift (programtekst) L: Lister (strukturinformation)	
Hjælpfunktioner	F: Fjerne (ønsker, delemsner, rammer) N: Navnændring (ønsker, delemsner, rammer) C: Kopiere (ønsker, delemsner, rammer)	

Vælg menupunkt

F1: Hjælp F9: Afslut DUS

Vælg menupunkt og tryk RETUR

Ideen bag udformningen af hovedmenuen er at samle programarbejdet i tre hovedområder: «Programmering», «Udskrifter» (af eksisterende programmer) og «Hjælpfunktioner» (til eksisterende programmer).

Programudviklingsblokken indeholder igen tre områder:

1) «Oprette» en emnestruktur eller «Vælg» sig ind på en allerede

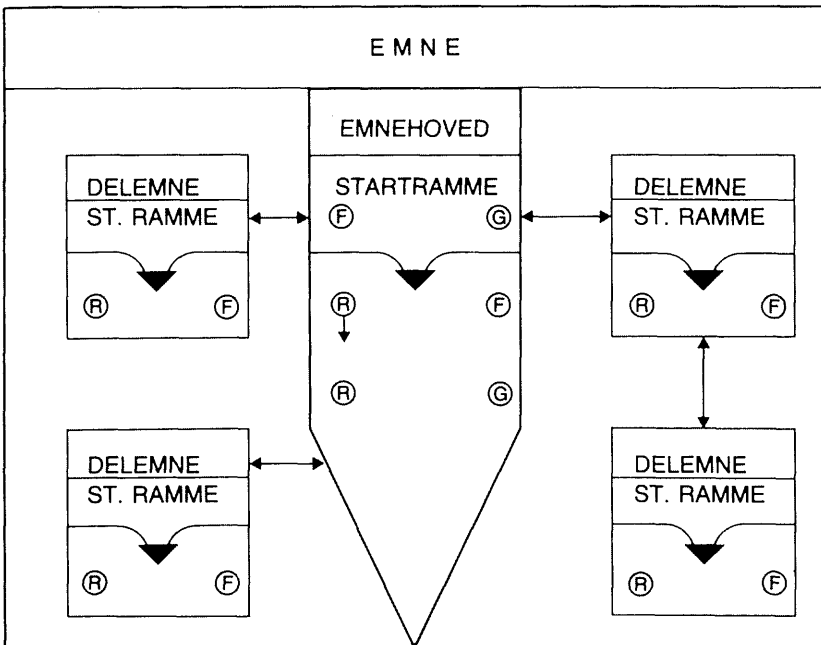
oprettet struktur.

- 2) «Indskrive» sit programindhold eller «rette» i et tidligere indskrevet program, samt efterfølgende «Klargøre» det udviklede program, dvs. oversætte sin struktur og sit indhold til et program, og endelig
- 3) muligheden for at «Afprøve» det færdige program, således som eleven vil opleve det.

Det er yderligere muligt at benytte sig af området med hjælpefunktioner under programarbejdet, idet man kan lave forskellige former for navneændringer og kopiering. Man kan altså genbruge allerede oprettede programmer eller dele heraf.

Det første, man gør, når man vil indskrive et program, er at *oprette* et

DUS-SPROGETS HOVEDBEGREBER



- Ⓜ rammer
- Ⓧ funktioner
- Ⓨ grafiske billeder

emne. Man taster bogstaven «O» ind i svarfeltet ud for «vælg menu-punkt» efterfulgt af tryk på returtasten (↵). Dernæst navngiver man emnet (dvs. reserverer arbejdsplads på disketten) og skriver den titel, eleven skal møde og vælge ud fra senere hen. Emnet skal ikke omfangsbestemmes fra starten. Systemet reserverer ny plads i takt med, at emnet vokser ved indskrivning.

I ovenstående model er DUS-programmets strukturelle hovedbegreber vist.

Emne/delemner

Når man som programforfatter laver et undervisningsprogram, placerer man det i en hovedstruktur, der kaldes et **emne**. Emnet gives et *navn* (max. 8 tegn), som dels kan fortælle noget om indholdets art, dels muliggør at systemet altid kan finde programmet ved senere brug.

I ethvert emne er der altid mindst ét **delemne**. Dette faste delemne kaldes **emnehoved**. Ud over dette obligatoriske delemne kan man oprette/navngive op til 23 delemner i hvert emne.

Rammer/funktioner

Emner og delemner danner den overordnede struktur, men selve programindholdet – tekster og grafiske billeder – anbringes i RAMMER eller FUNKTIONER. Der kan maksimalt være 66 rammer og/eller funktioner i hvert delemne.

Rammen er systemets mindste enhed og består af op til 132 linier i blanketformen og op til 528 linier i fri form.

Alle rammer og funktioner skal *navngives*, så systemet kan finde dem igen (max. 10 tegn og begynde med et bogstav). Den *første ramme* i ethvert delemne skal altid hedde STARTRAMME. Navnene skal være forskellige fra hinanden – to rammer eller funktioner i et delemne må altså ikke hedde det samme – men man må godt bruge de samme ramme-/funktionsnavne i forskellige delemner.

Rammerne kædes sammen på den måde, at man fra hver ramme henviser til den/de næste ramme(r) i den oprettede kæde eller grenstruktur.

Grafik-billeder er også rammer. De har egne navne og er oprettelsesmæssigt placeret i emnehovedet. Der kan maksimalt oprettes 65 billeder i hvert emne.

Billederne kan kaldes ind i alle delemnerne efter behov.

Funktioner bygges programmeringsmæssigt og indholdsmæssigt op som rammer, men de bruges anderledes.

Fra rammerne kan man kalde en funktion til f.eks. belysning af noget af en rammes indhold. Når så funktionen er afviklet/udført, vender man tilbage til det pågældende sted i rammen og fortsætter programafviklingen.

Det er muligt at kalde en funktion fra en anden funktion.

Funktioner, der er placeret i emnehovedet, kan bruges overalt i emnet, mens funktioner, der er oprettet i andre delemner, kun kan aktiveres i disse delemner.

Hvis man skal sammenligne begreberne med noget mere kendt, kan man overforenklet sige følgende:

DUS/DFU-programmet	svarer til en bog
EMNET	svarer til bogens titel
EMNEHOVEDET	svarer til indledningen
DELEMNERNE	svarer til kapitlerne
RAMMERNE	svarer til siderne
FUNKTIONERNE	svarer til noteapparatet og/eller appendix.

DUS/ComCats forfatter-systemfaciliteter

Blanket-form

Indskrivning (editering):

- dynamisk blanketstruktur
- programrelaterede hjælpesider
- definition af og skrivning i vinduer 1)
- tekstfremhævningsmuligheder:
farveskift med 8 karakterfarver og 8 baggrundsfarver, highlight, tekstblink, omvendt video, understregning
- tekstflytning:
indsætte, slette enkelttegn og linier og automatisk justering af restteksten, kopiere slettet tekst
- styring af svartid
- nøgleordssøgning 2) (begrænset udnyttelse)
- standardfunktioner (slet billede, vent, tildeling/oprettelse af variable, forgrening, udførelse af delemne, m.v.)
- indsættelse af grafiske illustrationer

Strukturmuligheder:

- fast elevmenu eller egen elevprogramforside
- lineær opbygning
- fri forgrening (bl.a. for informationssøgningsprogram)

Fri form (har yderligere disse muligheder)

Indskrivning (editering):

- menuopbyggede hjælpesider
- oprettelse af egne forfatterdefinerende funktioner
- tekstpositionering via koordinatbestemmelse
- udvidet nøgleordssøgning og mulighed for at blande svar bestående af tal og tekst
- animering af tekst og grafiske billeder
- tilfældighedsbearbejdning: opgaver, svarreaktioner osv.
- parallelprocesoprettelse og styring 3)

- skabe programstrukturer, der løbende giver mulighed for selvstændig elevbearbejdelse med nye/ukendte data – sammenkæde forskellige programmer

Strukturmuligheder:

- simulering (parallelle forløb/handlinger)

1) Vinduer

At oprette vinduer betyder, at man definerer et særligt (lukket) område af skærmen til placering af tekst og/eller grafiske billeder. Man har mulighed for at oprette flere vinduer, ja endda vinduer i vinduer, således at man på sit skærmbillede har separate felter, som kan fungere uafhængigt af hinanden. Det betyder f.eks., at man kan beholde et billede i et vindue, mens man har en tekst i et andet vindue som udskiftes løbende.

Der er mulighed for i alt at aktivere 16 vinduer (og selv frit at definere de 15).

2) Nøgleordssøgning

Med den eksisterende teknologi og de eksisterende programmeringsprog ligger et afgørende kriterium for anvendeligheden af et forfatter-system i, a) hvor nøjagtigt systemet kan vurdere – dvs. læse – et frit elevsvar, b) hvilke krav til elevsvaret, systemet stiller og endelig, c) hvor differentieret en programforfatter kan lade sit program reagere på et givet elevsvar.

Man kan eksempelvis stille et spørgsmål, hvor man bevidst ønsker at undgå at tage stilling til f.eks. stave- eller syntaxfejl. Det er indholdet, der vægtes i den pågældende situation. Man kan også bestemme sig for en høj grad af præcision, hvor man ser efter et helt bestemt ord eller enkelte bogstaver i ord for at teste lige netop dem. Den samme mulighed er til en vis grad til stede ved talsvar og ved blanding af tal og tekstvar. I disse og andre situationer kan man i sin programmering let sikre sig, at systemet kun aflæser de dele af et svar, man er interesseret i i den pågældende situation.

Man kan generelt sige, at man med DUS-nøgleordssøgning ikke pålægges logiske begrænsninger i sit arbejde. De systembegrænsninger, der trods alt er til stede, er af teknisk art og afhænger af den brugte programkompleksitet.

Eksempel på begrænset brug af nøgleordssøgning – instruktion fra den indbyggede vejledning i Blanketformen:

INSTRUERENDE EDITOR - RIGTIGT TEKSTSVAR

Systemet accepterer som rigtigt svar kun, hvad du instruerer det om. Du kan f.eks. konstruere dit svarudtryk ud fra følgende eksempler:

- b! ? rigtigt svar er bogstaverne b og l skrevet med stort eller lille, uanset den sammenhæng, de indgår i.
- b! rigtigt svar er bogstaverne b og l skrevet med småt.
- b! ! rigtigt svar er kun bogstaverne b og l og intet andet.
- b og l rigtigt svar er både b og l, skrevet med små bogstaver.
- b eller l rigtigt svar er enten b eller l, skrevet med små bogstaver.
- b? og ikke l rigtigt svar er b skrevet med småt eller stort, eller stort evt. kombineret med andre. Dog må l ikke forekomme.

Principper for brugen af nøgleordssøgning:

- 1) Eleven svarer inden for et forfatterbestemt omfang med tekst og/eller tal (efter forfatterens anvisninger).
- 2) Svaret vurderes af systemet efter programforfatterens anvisninger med hensyn til omfang af evt. ønskelig korrekthed eller anden udtryksmåde.
- 3) Svaret placeres af systemet (igen efter forfatterens anvisninger) i en såkaldt «betingelsesstruktur», dvs. en bestemmelse af, hvad der skal ske med det enkelte svar.
- 4) Eleven præsenteres for forfatterens reaktioner i form af f.eks. tekst og grafik eller brug af andre (ydre) illustrationer. Reaktionen bliver altså forsøgt afpasset efter elevens mulige svarudtryk.

Det er denne «dialog» mellem elev og program (der repræsenterer forfatteren), der betegnes som interaktion (samspil).

3) (Parallelle) Processer

DUS-systemet indeholder en helt anden mulighed for programudvikling end den, hvor man har struktureret sit program (rammer/funktioner) til at følge bestemte rækkefølger og dermed ordnet alt i tid. Denne mulighed – PROCES – åbner op for programtyper, hvor de enkelte dele i programmet afvikles i en nok på forhånd fastlagt rækkefølge, men *uafhængigt* af hinanden (forskelligt indhold) og med forskellige hastigheder.

Når flere processer afvikles samtidigt taler man om parallelle processer, og det er muligt at lade op til 10 processer arbejde uafhængigt af hinanden og med forskellige hastigheder.

DUS-PROCES-begrebet indeholder yderligere den mulighed at lade processerne i større eller mindre udstrækning samarbejde dvs. udveksle data (via fælles datastruktur). Denne dataudveksling er det muligt at regulere (synkronisere), idet DUS-systemet indeholder nogle synkroniseringsordre, der gør det muligt for den enkelte proces at få eksklusiv adgang til denne datastruktur.

Processerne kan skabes og iværksættes fra rammer, funktioner eller

andre processer og kan derved indgå som dele af «normale» programmer og samtidig være selvstændige programmer. Disse sidste vil typisk simulere hændelsesforløb, hvor flere handlinger er i gang samtidigt. Man kan yderligere udforme programmet således, at brugeren (eleven) får mulighed for at være en aktiv deltager i forløbet med ret til at foretage ændringer og dermed ændre processens forløb.

Her er et eksempel, der simulerer et kontrolpanel i en fabrikationsproces. De enkelte processers resultater er anbragt i vinduer og forinden på skærmen er det angivet, at man har to muligheder for indgriben: Standse forløbet eller bede om mulighed for at ændre på én eller flere af de bagvedliggende data og derved ændre på procesresultatet.

Segment nr:	Setpkt.temp	Produkttemp.	Kædøl temp.	RPM
1	20.2°C	11.84 °C	12.85 °C	80
Drej.moment	Vakuüm	Varmetilf.	Segmenttid	Procestid
6	20%	100%	0.3 hrs	1.7 hrs

_____ Kommandoer : stop og rette _____

Operatør kommando :

Pædagogiske struktureringsmuligheder i og med DUS

De programudviklingsmåder, som DUS-systemets forgreningsmuligheder indebærer, kan skitseres som følgende hovedområder.

DatamatFormidlet Undervisning (DFU) – dvs. hvor programmet styrer forløbet:

- Lineære programmer med en enkel fastlagt struktur uden mulighed for et frit elevvalg med hensyn til gennemløbets form.
- Forgrenede programtyper, der åbner op for elevvalg af forskellig art. Det kan være opgavetyper, det faglige niveau, afviklingsmåden eller andre muligheder stillet til rådighed af programforfatteren.

DatamatStøttet Undervisning (DSU) - dvs. hvor brugeren styrer forløbet:

- Informationssøgningssystemer, hvor man som bruger kan søge rundt i og frit bruge de indlagte oplysninger. Altså en slags mindre eller større elektroniske biblioteker.
- Simuleringsprogrammer, hvor man som bruger kan eksperimentere med forskellige former for modeller af eller om virkeligheden.

Kort fortalt kan man altså sige, at DFU er programtyper, der fremmer en elevholdning, mens DSU er programformer, der mere fremmer en (edb)brugerholdning.

Det er et filosofisk-pædagogisk valg, hvilken type man benytter i den pågældende situation. Der er selvfølgelig en mere glidende overgang mellem DFU og DSU, end her er antydnet, idet forgrenede programtyper kan være skabt med henblik på at være en videns- og træningsbank til fri afbenyttelse.

Erik Meistrup arbejder som frilanser med specialisering i undervisning/pedagogikk. Han har tidligere været ansatt som utdanningskonsulent hos DUS INTERNATIONAL A/S og har skrevet en rekke artikler om bruk av datamaskiner i undervisningssammenheng.

RAPPORTER

BBCs Domesday-prosjekt

Øystein Reigem

Øystein Reigem besøkte i juni BBCs kontorer i Ealing, London. Der ble han mottatt av Michael Tibbets, som ga ham en grundig orientering om BBCs Domesday-prosjekt. I tillegg demonstrerte Tibbets en pilotversjon av Domesday-systemet.

I år 1085 beordret Vilhelm Erobreren at det skulle utarbeides en oversikt over grunneiendommene i England. Eiendomsforhold, bruk og verdi av jorda skulle registreres, samt opplysninger om innbyggere og husdyr. Resultatet – den såkalte Domesday Book – ble fullført høsten 1086. Den dekker England med unntak av de nordligste deler og en del byer. Den opprinnelige utgaven finnes den dag i dag og oppbevares på The Public Record Office, som er en nasjonal institusjon med mye historisk materiale. I tillegg ble verket trykt og utgitt et par ganger på 1700- og 1800-tallet. Verket er et meget verdifullt kildeskrift.

Domesdayboka har 900-års-jubileum neste år, og i den anledning har BBC startet et stort anlagt prosjekt med sikte på å produsere en ny utgave. Det siste innen teknologi vil bli tatt i bruk i prosjektet. 1986-versjonen vil nemlig foreligge på videoplate, og tilgangen til informasjonen vil være via mikromaskin. Hele Storbritannia vil nå bli dekket.

Domesday-prosjektet er sannsynligvis det største og mest ambisiøse videoplateprosjekt i verden. Mye av kildematerialet til platene skal samles inn av skolebarn over hele landet. Til nå har 13-14.000 skoler meldt seg på, dvs. nærmere halvparten av alle primary og secondary schools i Storbritannia. Også andre organisasjoner, som f.eks. speidere og naturverngrupper, vil bidra sammen med skolene. Til sammen vil ca. en million mennesker involveres i prosjektet på denne måten.

Ytterligere materiale skal samles inn fra nasjonale kilder. Totalt vil de to platene som prosjektet skal frambringe, romme data tilsvarende to millioner tekst-TV skjermesider, 20.000 kartsider og nærmere 120.000 bilder.

Viktig støtte til prosjektet vil bli gitt av de lokale undervisningsmyndigheter, en rekke universitetsmiljøer og kartoppmålingsvesenet (Ordnance Survey).

Til å gjennomføre dette prosjektet, har BBC opprettet et eget, kommersielt drevet firma. Dette firmaet vil kunne få oppgaver ut over Domesday-prosjektet, i form av salg av konsulenttjenester eller produksjon av interaktive videoplater for andre. Det er også tenkt at Domesday-prosjektet skal følges av en hel serie videoplateprosjekter for BBC.

Domesday-prosjektet har et budsjett på £2,5 millioner. BBC Enterprises, det britiske handels- og industridepartementet og Philips bidrar med en halv million hver. I tillegg investerer Philips én million i utviklingen av en ny videoplatespiller for Domesday-platene. Handels- og industridepartementet ser på prosjektet som en katalysator for Storbritannias engasjement innen områder som video-programvare og informasjonsteknologi generelt. Prosjektet kan bli det endelige gjennombruddet for videoplateteknologien.

Platene

Den første av Domesday-prosjektets to plater er den såkalte *lokale* platen. Den skal bestå av informasjon om Storbritannia på lokal basis – data, tekst og bilder samlet inn av de deltakende skolene og andre lokale organisasjoner, samt et omfattende kartmateriale. Plate nummer to – den *nasjonale* platen – skal inneholde informasjon fra viktige nasjonale datakilder. Disse data vil være grunnlagsmateriale for diverse statistiske framstillinger. Den nasjonale platen skal også inneholde om lag 40.000 bilder fra museer, arkiver, m.m. Det kan i tillegg tenkes at denne platen vil inneholde en innføring i bruken av platene. I tillegg til selve bilde-, tekst- og datamaterialet skal de to platene også inneholde alle indekser og all programvare som brukerne trenger for å finne, kombinere, sammenligne og presentere Domesday-materialet på dataskjermen.

Den lokale platen

Basert på Ordnance Survey-kart er Storbritannia delt i blokker på 4 × 3 km, såkalte D-blokker. (Fasongen på D-blokkene er tilpasset dataskjermen.) Omtrent 23.000 D-blokker dekker hele Storbritannia. Opplysninger om hver D-blokk skal samles inn av skoleelever og andre lokale frivillige. Hver gruppe som deltar i innsamlingen av materiale, får seg tildelt én eller flere D-blokker. I tillegg til det gruppene samler inn, skal materiale om spesielt interessante ting, som kjente bygninger og minnesmerker, samles inn separat.

På platen vil kartmaterialet foreligge med forskjellig oppløsning, fra et oversiktskart over hele Storbritannia til de enkelte D-blokker. Når man bruker platen, vil dette gjøre det mulig å «zoome inn» på interessante områder. Et tilhørende register over stedsnavn m.m. på kartene vil gi en annen inngang til materialet. D-blokkene vil bli



THE
DOMESDAY
PROJECT

BBC TV

avfotografert fra kart i målestokk 1:50.000.

De lokale gruppene skal samle inn følgende informasjon om hver D-blokk:

- Opplysninger om bruken av området. Det er blitt utarbeidet 20 kategorier for å beskrive landområder (1 = boligstrøk, 2 = handel/kontor, 3 = olje/kjemisk/kraftverk, osv.). Innsamlingen skal gjøres ved å dele D-blokken i 12 kilometerstore kvadratruter, og så bestemme hovedkategorien for hver rute. Om nødvendig tillates to tilleggskategorier for hver rute. Siden det i dag ikke finnes noen fullstendig oversikt av denne typen for Storbritannia, er dette viktige nye data som kan bli nyttige i mange sammenhenger – ikke bare i undervisning.
- Opptelling av en rekke forskjellige tjenester og aktiviteter i hver kvadratkilometer rute, 64 i tallet. Her skal det telles alt fra postkontor til parker, fra kirker til kloakkanlegg, fra snackbarer til slakterier, fra hagesenter til høyskoler og fra bussavganger til bryggerier.)
- Bilder med motiv som er representative for området. Det skal

sendes inn fire dias, hvorav tre blir brukt på platen (det siste er reserve). Minst ett av bildene bør være et oversiktsbilde eller på annen måte gi et generelt inntrykk av området. Det er også ønskelig at de andre bildene formidler noe av livet og virksomheten på stedet, og at det er mennesker med. Til bildene skal det settes én lang og én kort tittel, samt indeksord for framsøking.

- Fri tekst som reflekterer det gruppen og den lokale befolkning synes er viktig og relevant for området. Teksten kan fylle inntil tjue skjermesider med 22 linjer à 40 tegn (tilsvarer ca. 140 engelske ord pr. skjermeside). Den første av disse sidene må være en presis, generell oversikt over D-blokken som helhet. For resten av teksten står innsamlerne nokså fritt, bortsett fra inndeling i skjermesider med korte overskrifter for hver side. For å muliggjøre søking i teksten, skal teksten suppleres med fire indeksord pr. side, hvorav minst ett skal hentes fra et kontrollert vokabular (37 termer).

Bortsett fra bildene skal alt D-blokk-materialet samles inn via mikromaskin. Dette er mulig da det er en god utbredelse av mikromaskiner i de britiske skoler og hjem. (Ifølge Daily Telegraph november 1984 hadde 23.000 skoler på det tidspunkt en BBC mikromaskin.) Dataregistreringen skal gjøres vha. spesiell programvare som er utviklet for formålet.

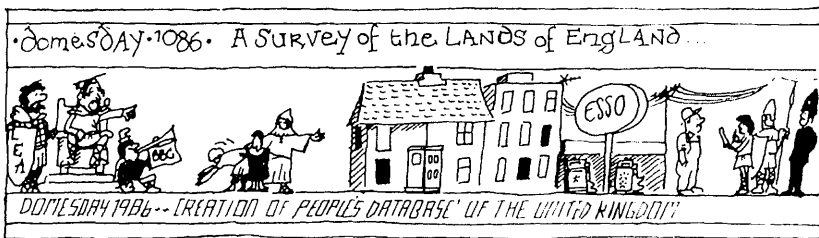
I tillegg til at innsamlings- og bearbeidingsfasen vil bli et meget interessant prosjekt for den enkelte skole som deltar i Domesday-prosjektet, vil denne fasen også kunne brukes som utgangspunkt for en rekke andre aktiviteter. Som eksempler kan nevnes:

- Elevene kan arrangere utstillinger av innsamlet materiale (fotos, undersøkesskjema, tekst, m.m.) på skolen, i kjøpesentra, i biblioteker, osv.
- De enkelte elever kan gi en muntlig framføring av sin egen innsats i prosjektet, kanskje illustrert med dias.
- Elevene kan overtale en lokalavis til å skrive artikler om skolens prosjektdeltakelse, eventuelt om resultater av datainnsamlingen. Lokale radio- og fjernsynsstasjoner kan også være interessert i denne typen stoff.

Den lokale platen i bruk

En sesjon med bruk av den lokale platen kan tenkes foregå på følgende måte: Foran deg står en BBC mikromaskin med fargeskjerm, tilkoblet en videoplatespiller. Du setter deg, starter mikromaskinen og legger platen i spilleren. Programvaren som er nødvendig for å styre platespilleren og presentere materialet lastes nå automatisk over fra platen til maskinen. (Det trengs dermed ikke diskettstasjon.) Opp på skjermen kommer et oversiktskart over Storbritannia.

Du er nå interessert i å vite mer om et bestemt sted som du kjenner



Tegning: Jonathan Hall.

beliggenheten av. Med fingeren peker du på riktig sted på oversiktskartet, og straks kommer et mer detaljert kart over et mindre område til syne. (I ditt system har du en berøringfølsom skjerm, men utstyret kan også betjenes med en mus eller vanlige piltaster.) På nytt peker du på skjermen, og det «zoomes inn» til en blokk på 40×30 kilometer. En siste peking henter fram D-blokken på 4×3 kilometer. Nå har du tilgang til de opplysningene som er lagt inn om D-blokken. Du blar deg gjennom tekstsidene og kikker på bildene.

Du bestemmer deg så for å ta en tur i nabolaget til dette stedet. Ved å hente fram igjen D-blokk-kartet kan du nå spasere fra D-blokk til D-blokk i de fire himmelretninger. Større sprang over landskapet foretas lett ved å gå et nivå opp til 40×30 -kilometer-kartet.

På et av D-blokk-kartene lyser det en gul firkant. Du skjønner at det er noe av spesiell interesse på dette stedet. Du peker på den gule firkanten, og et nytt, enda mer detaljert kart springer fram. Nå ser du tydelig at en kjent katedral er avmerket. Ny peking henter fram et foto og en plantegning av katedralen.

Steder du ikke vet den nøyaktige posisjonen til, finner du raskt ved å gå via kart-indeksen. Alle navn på kartene er lagt inn i indeksen, og kan dermed søkes på. Og etter hvert som du har besøkt de stedene du kjente navnet på, går du også over til å lete opp interessante steder ved å søke på ord i indeksen til teksten og bildene.

En ting du ergrer deg litt over, er at platespilleren ikke kan spille mer enn én plateside om gangen. Det lokale materialet er nemlig for stort til å rommes på én side. Landet er derfor delt i en nordlig og en sørlig del som har fått hver sin side av platen. Platen må tas ut og snus når du krysser grensen mellom nord og sør. Heldigvis er selve *kartdelen* duplisert, slik at all orientering vha. kart kan gjøres uten snuing.

Du lar deg imidlertid fascinere av de grafiske funksjonene som kan brukes på kartmaterialet. Ved å trekke linjer mellom punkter kan du måle avstander i luftlinje. Ved å trekke serier av små linjestykker langs veier og jernbanelinjer kan du måle kjøreavstander. Ved å «omringe» områder kan du beregne arealer.

For å hjelpe deg i bruken av platen forteller en enkel menylinje hvilke

muligheter du har til enhver tid. Mer omfattende hjelp er også tilgjengelig via menyen. Menyene kan slås av og på etter behov. Menyene og tekster blir vist i pene og tydelige teletekst-bokstaver (dvs. som på tekst-TV).

Den nasjonale platen

Den nasjonale platen skal inneholde så mye som mulig av nasjonale data hentet fra eksisterende kilder, og på en form som passer i interaktiv video-sammenheng. Bl.a. følgende institusjoner skal veilede prosjektgruppen i utvelgelsen av dette materialet: Institute of Terrestrial Ecology, Centre for Urban and Regional Studies ved University of Newcastle-upon-Tyne, Birbeck College ved London University og University of Essex. Ved sistnevnte institusjon ligger National Data Archive, som besitter maskinleselige data fra de fleste større nasjonale undersøkelser.

Den nasjonale platen skal inneholde informasjon om nasjonale ressurser, transport, befolkningsstatistikk, demografiske data, sosiale aktiviteter, kunst og kunsthåndverk, undervisning, underholdning, helse, flora, fauna og mye annet.

Til å anskueliggjøre og viderebehandle de statistiske opplysningene vil det bli brukt datamaskingrafikk, både tradisjonelle framstillingsformer som søyle- og kakediagram og andre mer avanserte metoder. F.eks. vil områderelatert informasjon kunne vises lagt oppå kart. (Disse kartene er spesiallagd for dette formålet, da Ordnance Survey-kartene viser for mye detaljinformasjon.) På platen skal altså ligge *statistisk grunnlagsmateriale* i form av tabeller og *programmer* for å manipulere/presentere dette. All *grafikk* (unntatt selve kartene) vil bli *generert* av mikromaskinen ut fra grunnlagsmaterialet.

I tillegg til statistiske data og kart, skal platen inneholde 40.000 «nasjonale» bilder. Disse skal omhyggelig velges fra spesialarkiver og bildesamlinger, og vil gi et visuelt inntrykk av livet i 80-års Storbritannia. Eksempler på tema er arkitektur, mote, familieliv og arbeidsliv. Resultatene av en nasjonal fotokonkurranse kommer også med på platen.

Bildematerialet vil også gjøre det mulig å foreta simulerte reiser gjennom det britiske landskap. Langs rutene vil interessante steder være illustrert med bilder og tilhørende tekst. Brukeren vil kunne få både nærbilder fra stedene og utsiktspbilder i alle retninger. Og hvis man vil holde seg fullstendig innendørs, vil et tilsvarende materiale med opptak fra en serie britiske hjem gi muligheten for det.

Sannsynligvis vil baksiden av den nasjonale platen inneholde opplæring i bruken av utstyret og begge platene.

Den nasjonale platen i bruk

For å illustrere de statistiske mulighetene ved den nasjonale platen kan vi fortsette den tenkte sesjonen ved mikromaskinen og videoplatespilleren. Du tar nå ut den lokale platen og legger inn den nasjonale. Da du er interessert i mediasituasjonen i Storbritannia, ber du om materiale om de største avisene i landet. Lettvint veksler du mellom data for forskjellige distrikter og får dem presentert som histogrammer. For hvert distrikt skifter du raskt fra år til år, og får på denne måten et fritt visuelt inntrykk av hvordan de enkelte avisers leserkrets varierer over tid. Med et tastetrykk skifter du presentasjonsform fra søyler til kake. Også mer avanserte statistiske programmer står til din rådighet, og du oppdager at det er lagt inn sikkerhetsforanstaltninger for å hindre at du trekker konklusjoner på sviktende datagrunnlag.

Du skifter så til folketellingsdata. Du henter fram kart over det området du er interessert i, maskinen ruter dette opp og fyller så i fargene avhengig av folketettheten i den enkelte rute. Du oppdager at du ønsker mer detaljert informasjon om de minst folketette områdene. Kanskje skilles det nå mellom 1-100, 101-200, 201-500, osv. innbyggere pr. rute. Uten store problemer redefinerer du grensene for de forskjellige fargene, slik at det nå skilles mellom f.eks. 1-5, 5-20, 21-50 osv. innbyggere pr. rute.



Engelske skolebarn samler inn opplysninger om sin D-blokk. (Foto utlånt av BBC).

Utstyr og teknologi

For å oppsummere: Presentasjonen på skjermen baserer seg i Domesday-prosjektet på følgende tre elementer, som kan forekomme alene eller i kombinasjoner med hverandre:

- Video enkeltbilder og levende videosekvenser (levende sekvenser muligens bare i opplæringsdelen på baksiden av den nasjonale platen)
- Teletekst
- Grafikk

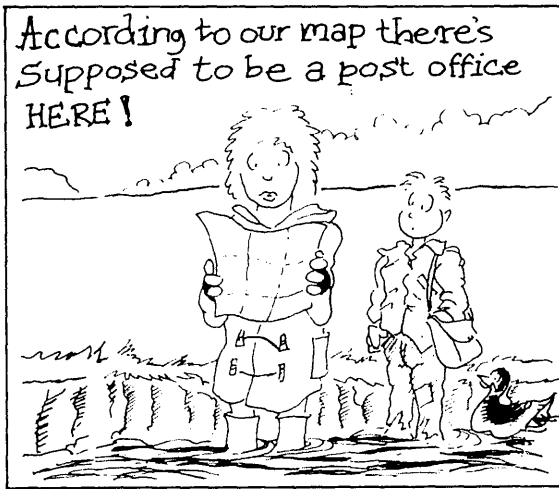
Tradisjonelt videoplateutstyr er fullt i stand til å ivareta det første elementet. Den vanlige videoplaten er beregnet på lagring av et vanlig videosignal. Videoplaten har dermed mulighet for å lagre levende sekvenser, og - siden levende video består av 25 enkeltbilder pr. sekund - også enkeltbilder. Kapasiteten for en videoplate av Laservision-typen, som er det beste og mest utbredte systemet, er ca. 54.000 enkeltbilder pr. side, eller ca. 36 minutter levende film. (Dette er for plater i det såkalte CAV-formatet. For spillefilm brukes CLV-formatet, som kan ta en hel time film pr. side, men ikke enkeltbilder.)

I utgangspunktet er ikke kapasiteten så god for tekst på videoplaten (i motsetning til *digitale* optiske plater). I Domesday-prosjektet hadde man imidlertid bruk for også å lagre store mengder tekst. Å ta bilder av tekst og lagre som video enkeltbilder er plassmessig uøkonomisk. Tekst-TV-delen av videosignalet har heller ikke god nok kapasitet for tekst. Løsningen i Domesday-prosjektet ble å omforme digital tekst (dvs. inntastet og maskinlagret tekst) til et lydsignal og lagre i *lydsporet* (dvs. det frekvensområdet i videosignalet som brukes til lyd). Også *programmene* er digital informasjon og vil kunne lagres med denne teknikken. Opptil 324 Mbyte (324 millioner tegn) kan da lagres på hver plateside.

For å kunne dekode den digitale informasjonen til data og programmer igjen, kreves ekstra elektronikk i spilleren. Philips skal framstille en slik spiller. Denne spilleren skal også være i stand til å legge et eksternt videosignal (dvs. i praksis grafikk fra mikromaskinen) over sitt eget signal.

Det andre elementet, teletekst, er den typen bokstaver og enkel grafikk som vi kan se på tekst-TV. Teletekst gir klare og tydelige bokstaver og mulighet for farger. Men som en selv lett kan se på tekst-TV, er mengden tekst på en skjerm side mindre enn for vanlig dataskjermtekst. (Det er like mange linjer (24), men bare 40 mot 80 tegn pr. linje.)

Når det gjelder det tredje elementet, grafikken, er det naturlig at *mikromaskinen* genererer denne. Man fant at de grafikkmulighetene som ligger i vanlig teletekst ikke var gode nok. BBC-maskinen har noe av den beste grafikk blant maskiner i sin prisklasse, og ble i denne sammenheng funnet ideell for Domesday-prosjektet. En annen god



Tegning: Jonathan Hall.

egenskap ved BBC-maskinen er at videosignalet den selv genererer er av høy kvalitet (den har ikke egen skjerm, og er da avhengig av å sende et videosignal til en monitor, f.eks. et TV). Dette er slett ikke tilfelle for alle dens konkurrenter.

Forståelig nok trengs mye *programvare* til Domesday-prosjektet. En fant imidlertid at det eksisterende programtilbud var for dårlig. Acorn Computers, som produserer BBC-maskinen, utvikler nå programvare for prosjektet sammen med BBCs partnere i det såkalte Computer Literacy Project. Programvaren består blant annet av et nytt operativsystem, protokoller for interaktiv video, samt en omfattende program-pakke for gjenfinning, behandling og presentasjon av Domesday-materialet. Mye av programvaren vil kunne brukes i andre prosjekter.

Selv om BBC naturlig nok holder på BBC-maskinen som den mikromaskin som passer best i Domesday-systemet, vil det utvikles programvare tilpasset andre utbredte mikromaskinmerker. En tenker seg at Domesday-platene kan inneholde versjoner av programvare for flere maskiner.

Utgangen som er beregnet brukt fra mikromaskinen mot videoplate-spilleren er dessverre hverken vanlig seriell RS-232 eller parallell Centronix, men såkalt SCSI (Small Computer Systems Interface - som da selvsagt BBC-maskinen har). Hvis mikromaskinen en benytter ikke har SCSI-utgang, kan spilleren også styres fra en RS-232-utgang, men kommunikasjonen vil gå langsommere, og denne løsningen kan muligens by på tekniske problemer. Alternativet er da å utstyre maskinen med et ekstra kort for SCSI-utgang.

Domesday-spilleren skal være tilgjengelig samtidig med platene, nemlig mot slutten av 1986. Ventelig vil den etablere en ny standard for Laservision-spillere til institusjonsbruk, i undervisning, og til og med i kommersiell sammenheng. Den nye spilleren vil være i stand til å spille gamle videoplater, men i praksis vil ikke gamle spillere spille de nye platene.

For den som er interessert i å skaffe seg Domesday-spilleren og platene når den tid kommer, kan det opplyses at prisen vil ligge på om lag £1000 for spilleren og £100 for platene. Komplette utstyr vil koste ca. £1500. Norske priser ligger som kjent over britiske, så det er klokt å gange med en faktor på 1,5-2.

Hva med et slikt prosjekt i Norge?

Som enkelte vil kjenne til, er NAVFs EDB-senter i gang med sitt eget videoplateprosjekt. Ca. 20 institusjoner (museer, arkiver, universitetsbibliotek, humanistiske institutt ved universiteter o.a.) vil bidra med bilde- og datamateriale. På KUDs seminar om pedagogisk programvare i Grimstad i sommer, formulerte en gruppe enda et norsk prosjekt med en videoplatekomponent, «Data Heimsins». Data Heimsins har flere likhetstrekk med Domesday-prosjektet. Blant deltakerne i prosjektgruppen er personer fra Stord Lærarhøgskule, som også deltar i Senterets videoplateprosjekt. Gjennom Senterets prosjekt får gruppen en sjanse til å produsere et pilotmateriale på videoplate.

Det er å håpe at et nasjonalt prosjekt kan gjennomføres. En kan vinne mye ved å bygge på Domesday-prosjektets erfaringer, uten at dette betyr å vente til 1987 før en setter i gang. Det finnes etter hvert en del kompetanse i Norge, både for videoplate teknologi og pedagogisk programvare. Det er på tide at våre myndigheter, både undervisningsmyndigheter og andre, får øynene opp for den nye teknologiens muligheter.

Samisk database

Anders Løøv

Pr. i dag er det i den samiske databasen ved Universitetsbiblioteket i Trondheim registrert ca. 2000 bibliografiske enheter. I hovedsak er det litteratur publisert i Norge etter 1950. Av alt det som er skrevet om samene eller om samiske forhold er dette en liten del. Etter mine egne beregninger er den totale mengden av slik litteratur ca. 20.000 bibliografiske enheter. Og den vokser stadig. I de 20.000 er også det som er utgitt

i andre land tatt med. Den årlige tilveksten er det vanskelig å gi sikre tall for. Etter det som er registrert i databasen var den i 1980 i Norge 200 enheter. I Sverige, Finland og Sovjetsamveldet, de andre landene med en samisk befolkning, blir det publisert noe mindre, men det blir også skrevet en god del om samene i andre land. Den samlede utgivelsen av nevnt litteratur er derfor antagelig mellom 700 og 800 pr. år.

Å foreta en egen registrering for en samisk database kan kanskje synes overflødig. Ville det ikke være tilstrekkelig at det ble registrert nasjonalbibliografiene og i tidsskriftindeksene?

Det er det ikke.

Om en i databasen søker etter det som ble gitt ut i Norge i 1980 får en 205 treff. En kontroll mot Norsk bokfortegnelse: årskatalog 1980 og Norske tidsskriftartikler: årskatalog 1980 viser at ikke mer enn 62 av disse er å finne der.

For databasen er informasjonssystemet POLYDOC brukt. Det er utviklet av Norsk Senter for Informatikk, Oslo. I POLYDOC-systemet er fleksibilitet en av de ting det er lagt stor vekt på. Rent alment kan en ikke si annet enn at det har gått greit å tilpasse det til den samiske databasens behov.

Inputformatet i databasen ser slik ut:

feltnavn	kode	søkbar	"P	"P TRIAL	"P FUL
Registreringsnr	RN	*			*
Forfatter	AU	*	*	*	*
Medforfatter	CA	*			*
Tittel	TI	D	*	*	*
Utgave	RE			*	*
Trykkangivelse	PU			*	*
Kollasje	NP			*	*
Navn på serie	SL			*	*
I:	ST			*	*
Bibliografi	BI			*	*
Bi-personer	OP	*			*
Kommentar	CO			*	*
Stikkord	KW	*			*
Språk	LA	*			*
System.nr.	SN	*			*
Geografisk inform.	GN	*			*
Kilde	SO	*			*
Utgivelsesår	PY	*			*
Utgivelses-land	PC	*			*
Finnes i	SI				*
Ekstrafelt	XF				

For hver enkelt innførsel kan en benytte seg av et varierende antall datafelt. Hvert felt kan deles inn i flere delfelt. Feltene blir spesifisert med en feltkode bestående av to bokstaver. Den som foretar registreringen i databasen bestemmer hvilke felter som skal bli brukt for hver enkelt innførsel som blir lagt inn. I oppsettet ovenfor kan en også se hvilke felter som blir tatt med i de tre standardformatene for online output fra basen. "p vil gi bare forfatter og tittel. "P vil gi samme informasjon som en normal bibliografisk innførsel. "p full gir all den informasjon som er registrert i basen. Hvordan en slik "p full ser ut kan en se av følgende eksempel.

RN : 10867
AU : Eidheim, Harald
TI : Aspects of the Lappish minority situation. -
RE : 3.ed. -
PU : Oslo: Universitetsforlaget, 1977. -
NP : 86 s. -
SL : (Scandinavian university books)
KW : minoritetsforhold
LA : engelsk
SN : 13c ; 14a
SO : eo
PY : 1977
PC : Norge

Online dialogen ligger nært opp til søkespråket Orbit. For de brukere som er kjent med det systemet er det lett å sette seg inn i søketeknikken i den samiske databasen.

Kommunikasjonen med databasen er alltid en av følgende:

brukeren gir systemet en kommando
brukeren svarer på spørsmål fra systemet
brukeren gir ett eller flere søkeord eller bygger opp søkestrategier
ut fra søk som allerede er foretatt

Søketermene er av to hovedtyper: ord i tittel og «andre». Distinksjonen er viktig siden systemet vil anta at et søkeord er et ord i tittel om det ikke blir fortalt at det ikke er det. Det blir det fortalt gjennom at ordet blir fulgt av en parentes med feltkoden. For forfatterfeltet er det eksempelvis (AU). Søk kan foretas på alle felt som i feltoversikten ovenfor er markert * eller D. Søk er videre ikke bare mulig på hvert delfelt men også på hvert enkelt ord og på deler av (trunkerte) ord.

Her følger et enkelt eksempel på hvordan et søk kan bli foretatt:

Søkeord og felt som det blir søkt på	søk nr.	antall «treff»	spørsmålet gjentatt
engelsk(LA)	1	56	ENGELSK(LA)
minorit:(TI)	2	111	MINORIT:(TI)
1 og 2	3	10	1 = 2
"P 1			

- 1 -

AU : Eidheim, Harald

TI : Aspects of the Lappish minority situation.-

Her blir det søkt bare i to felter, språk og tittel, som deretter blir kombinert med hverandre. Bare en av de 10 treffene ble ønsket skrevet ut. For å dekke inn alle mulige sammensetninger med minorit- ble ord trunkert. Det er mulig å bygge opp et søk til å omfatte så mange som : forskjellige søketermer som hver og en kan bestå av ett eller flere ord. Det er bare å velge ord som er relevante for det en måtte være interessert i å finne litteratur om, og så å kombinere dem med de logiske operatorene OG, IKKE, ELLER.

Bruken av datateknikk i arbeidet med en samisk bibliografi ved Universitetet i Trondheim har ikke bare tatt sikte på å lage en online database. Virkeligheten er den at det ennå er svært få bibliotek som har muligheter til å foreta søk online i databaser. Som database kan den bli benyttet av alle som har muligheten å kommunisere med RUNI dataanlegg i Trondheim.

Dataregistreringen av materialet er også tenkt som et hjelpemiddel for å produsere trykte bibliografier om samiske emner.

Som den første av slike publikasjoner ble det i 1984 gitt ut «Samisk bibliografia = Samisk bibliografi: utgivelser i Norge 1966-1980». Valget av denne avgrensingen var begrunnet av at dette var det som langt var dårligst dekket av allerede eksisterende bibliografier. Om det kan leses mere i en artikkel av meg i Heimlen 1980, s. 534-537.

Som det framgår av beskrivelsen ovenfor er det ingen grenser for hvor en kan søke etter. En bibliografi kan som et eksempel bli definert til omfatte:

- litteratur publisert
- i et bestemt land
- på et bestemt språk
- i en bestemt tidsperiode
- om et bestemt emne

Et eksempel: «Litteratur publisert i Sverige på finsk 1920-1970 om samisk kunsthåndverk : en bibliografi.»

De fleste slike bibliografier vil ikke bli særlig omfangsrike, men

meget anvendelige for forskerne.

En personbibliografi kan tjene som eksempel på en slik bibliografi. Den produseres i løpet av få minutter. Her blir bare et utsnitt fra en slik bibliografi vist. Den er redigert slik som det vil bli gjort med alle bibliografier som blir tatt ut over printer. Innførlenes utforming kan med letthet forandres slik at den svarer til det som er ønskelig i enhver sammenheng.

PERSONALHISTORIE

= = = = =

AARSETH, BJØRN

Rikets kulturpris til en samisk foregangsmann. -

S. 5-9 : ill.

I: Ottar 88(1976)

Om Hans J. Henriksen

KULTURHISTORIE

= = = = =

AARSETH, BJØRN

Røros-prosjektet : et museumspedagogisk samarbeidstiltak.

- S.102-107: ill.

I: Ottar 116-117(1979)

I bibliografien er det forskjellige overskrifter. De blir produsert automatisk og satt inn på rett plass i samsvar med de klassifikasjonsnummer som blir gitt enhver innførelse i basen.

Til en slik bibliografi er det videre mulig å produsere enhver form for register: forfatternavn, stedsnavn, stikkord osv. Det kan også produseres register med alfabetisk ordnede stikkord som etterfølges av en tekst, såkalte KWAC (keyword and context) register. Etter samme prinsipp kan det også lages KWIC (keyword in context) og KWOC (keyword out of context) register. Det er da bare redigeringen som er forskjellig. Videre er det muligheter for rotasjonsregister og frekvensordlister. Registerne kan lages for ett eller flere av registreringsfeltene.

Den fortsatte registreringen i basen

Av forskjellige grunner er det ikke foretatt noen nyregistrering i basen etter 1. juli 1984.

Fra høsten 1985 av vil arbeidet med en løpende registrering fortsette.

Framtidige arbeidsoppgaver er det ellers nok av. Selve registreringen er bare én.

Kvantitativt utgjør det som har med samiske forhold å gjøre en liten del av det som er med i nasjonalbibliografiene. De spesielle krav som må stilles til en slik bibliografi gjør imidlertid at det å lage en slik

bibliografi er en oppgave som i en eller annen form bør finne sin løsnir på permanent basis.

Den viktigste oppgaven nå er å få til et fast samarbeid med bibliotel Sverige, Finland og Sovjetunionen som kan ta seg av registreringen i det som blir gitt ut i de respektive land. Utvekslingen av data med dis vil ikke by på nevneverdige problemer. Ved Universitetsbiblioteket Trondheim er det utviklet et eget registreringsprogram for registrering POLYDOC-format for mikromaskin. Men et eget registreringsprogra har egentlig ikke mange fordeler å by, slik som registreringsformatet utformet. Hvilken maskin en har adgang til har lite å si. Det tilstrekkelig at en har et tekstbehandlingssystem som kan produse rene ASCII-filer. Med hjelp av et kommunikasjonsprogram er det c ikke noen vanskeligheter å overføre de registrerte dataene til database. Med dette oppnår en også betydelige besparelser. All korrekturlesing c retting kan fullføres uten kostbar tilknytning til hovedmaskinen.

Et vesentlig moment å få avklart er også hvorvidt det vil være noe besparelse i å benytte seg av det som blir registrert i nasjonalbibliogr fiene og i tidsskriftindeksene. Som ovenfor nevnt er det helt utilstrekkl lig i forhold til det som er ønskelig for den samiske databasen. At 198 er det eneste år som det fins sammenlignbart materiale fra for Norge kanskje litt uheldig. Det var det første året som det igjen ble gitt ut e norsk tidsskriftindeks etter en «pause» på femten år og dekningsgrade for årene deretter er sannsynligvis bedre. Å hente informasjon fra d som er registrert der byr ikke på nevneverdige problemer. Men tilføyen sen av forskjellige tilleggspopplysninger som er nødvendige for å få til c ovenfor nevnte sorterings- og søkemulighetene, vil kanskje ta så mye t at besparelsene med dette går tapt. At det ene tar sikte på å væ generell og det andre på å være spesiell vanskeliggjør selvsagt en sl samkjøring. Den kan virke opplagt for den som ikke har sansen for d som er av interesse for den som arbeider med det spesielle. Denne på s side kan ikke akseptere at mye informasjon går tapt fordi et gener system ikke kan ivareta det spesielles behov.



Anders Løøv er universitetsbibliotekar ved Universitetet i Trondheim, Museet. For tida har han forskerstipend fra NAVF for å arbeide med sørsamisk historie.

Nytt om Registreringsentral for historiske data, Tromsø

Jostein H. Hauge

Registreringsentral for historiske data (RHD) er også omtalt i flere tidligere nummer av Humanistiske Data (jf. nr. 1-82, 3-83).

RHD er organisert som en faglig enhet ved Universitetet i Tromsø og med en tilknyttet registreringsentral i Moen i Målselv, for tiden bemannet med 8 tidligere televerkansatte. Når det nye Statsarkivet i Tromsø er ferdig i 1987 vil RHD få lokaler der. Statsarkivet vil bli bygget på universitetsområdet i Brevikva.

Ved Universitetet i Tromsø er det følgende faste stillinger fra 1985 av: daglig leder (for tiden *Fredrik Fagertun*), edb-konsulent, 1/2 driftsassistent og 1 kontorassistent. Det arbeides for tiden for å få gjort stillingene på Moen permanente.

RHD har som hovedoppgave å skrive av og databehandle historiske kilder (særlig folketellinger og kirkebøker) med opplysninger om mennesker som levde i forrige hundreår og tidligere. Hele landet er arbeidsområde for RHD som også tar imot henvendelser fra og gir service-ytelser til en lang rekke brukergrupper.

Materialet kan brukes av:

- *historikere* til å studere vanlige menneskers historie
- *demografer* for å granske befolkningsutviklingen
- *samfunnsvitere* til forskning i befolkningens sammensetning
- *navnegranskere* for å finne fram til gamle person- og stedsnavn
- *lokalhistorikere* som kildegrunnlag for bygdebøker, gårds- og slektshistorie
- også forskere fra andre fagfelt og skoleverket har sagt seg interessert i det materialet som tilrettelegges ved RHD

Hittil har RHD konsentrert seg om folketellingene fra 1865, 1875 og 1900. Det endelige siktemålet er å dekke hele Norge når det gjelder kirkebøker og folketellinger fra 1700- og 1800-tallet.

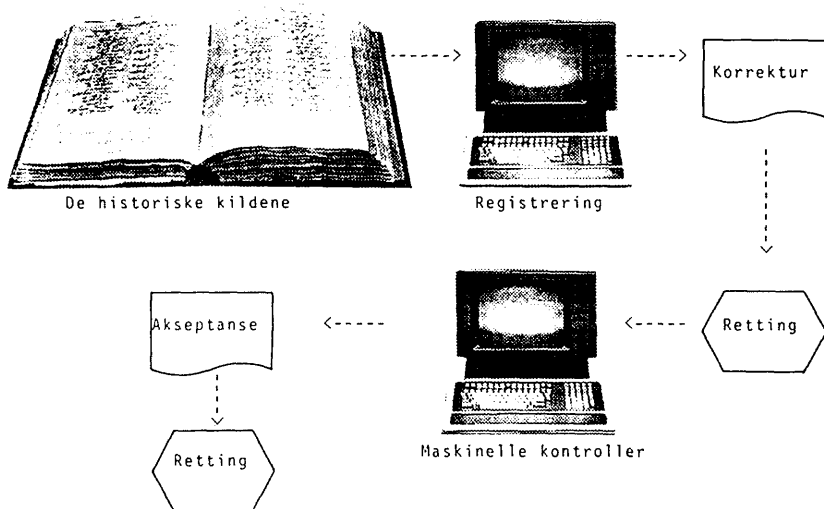
I tillegg til arbeidet med kilde tilretteleggingen som foregår i et samarbeid mellom Tromsø og Moen (jf. figuren) er det også satt i gang flere prosjekter knyttet til RHDs arbeid.

I *skoledataprojektet* samarbeides det med Tromsø lærerhøgskole om pedagogisk bruk av primærkildene. Det er nå utviklet programvare for mikromaskin for fremfinning, utlisting og optelling av folketellingsdata. Et annet prosjekt gjelder utvikling av metoder for automatisk å kunne identifisere de samme personer i ulike historiske kilder. Dette kalles ofte historisk lenking.

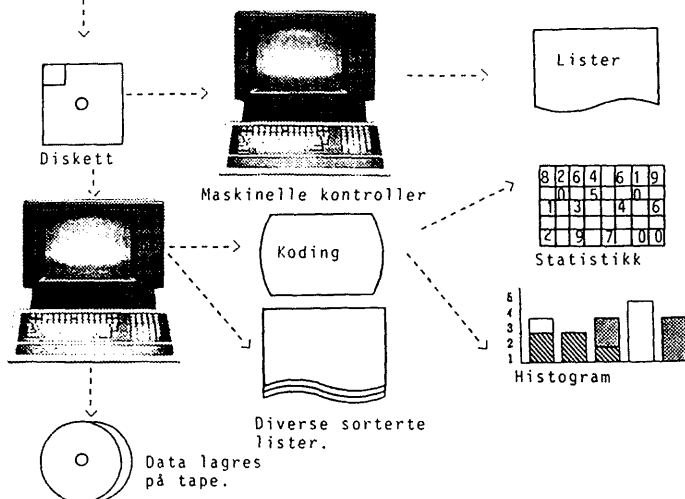
Det er også foreslått at Universitetet i Tromsø oppretter et semesteremne i humanistiske og/eller samfunnsvitenskapelige fag hvor bruken

FLYTSKJEMA FOR KILDEBEHANDLING

MOEN



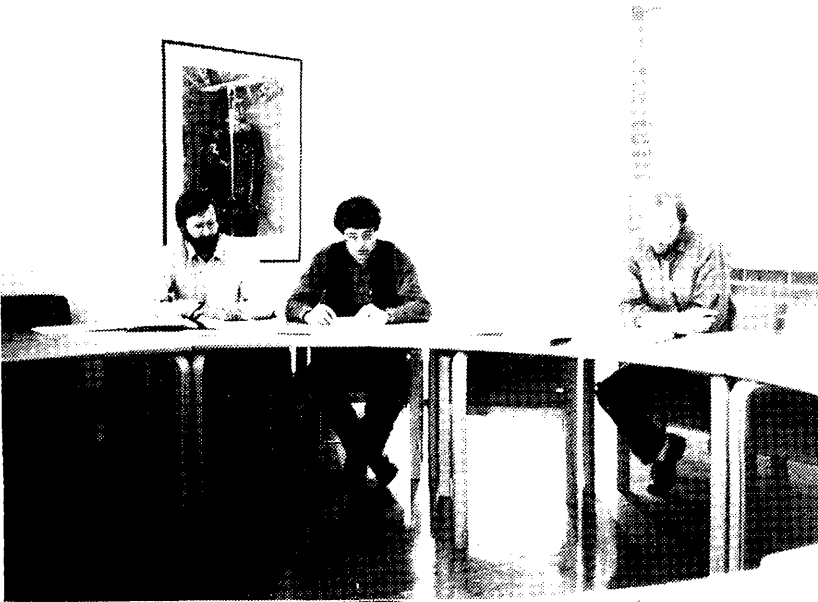
TROMSØ



av lokalhistoriske data i pedagogisk sammenheng vil få en sentral plass.

Det tredje rådmøtet for RHD ble holdt i Tromsø 10. mai 1985. Rådet er sammensatt av representanter for de faghistoriske miljøene, arkivverket, Troms fylkeskommune og NAVF.

På møtet ble det bl.a. orientert om de planer som foreligger i Troms og i andre deler av landet for å etablere registreringssentraler for historiske kilder og museumsmateriale. I Nord-Troms er det etablert en



F.v. daglig leder (perm.) Gunnar Thorvaldsen, formann i rådet for RHD Einar Arne Drivenes og daglig leder (vik.) Fredrik Fagertun.

midlertidig sentral med 17 arbeidstakere. Det arbeides med lignende planer i Trøndelag og også i Sør-Norge. For tiden er det usikkert om sysselsettingsmidlene i fremtiden vil kunne brukes til å etablere varige arbeidsplasser for historisk registreringsarbeid, men det vil være gode muligheter for å kunne opprette midlertidige sentraler dersom sentralene organiseres som ledd i opplæringstiltak rettet mot ordinære arbeidsplasser i offentlig virksomhet og næringslivet.

RHD har vært trukket sentralt inn i planleggingen av flere knoppskytingstiltak og er innstilt på å gi konsulenthjelp og inngå faglig samarbeid innenfor sin kapasitet.

En vesentlig sak på rådsmøtet 10. mai var drøfting av de prinsipper for kildetilrettelegging som RHD bør benytte i årene som kommer. Saken er aktualisert ved et sterkt ønske om at RHD påtar seg den komplette registrering av folketellingsmaterialet fra Oslo år 1900. Dette kildematerialet har stor lokal – og nasjonal – betydning og vil bl.a. kunne utnyttes i arbeidet med Oslo bys historie, men vil samtidig båndlegge en betydelig del av arbeidskraften ved RHD. Rådsmøtet vil komme tilbake til denne saken når mer bakgrunnsmateriale foreligger.

En praktisk viktig sak for RHD og arkivverket er spørsmålet om hvilket avskriftsmateriale som kan/bør benyttes i arbeidet, dvs. valget mellom originalmaterialet og kopier. Arkivverket har betenkeligheter med å utlåne originalkilder i store mengder, mens RHD finner det

slitsomt og vanskelig å arbeide med mikrofilm og Xerox-kopier. Det ble vedtatt å søke fleksible løsninger på problemet ettersom begge parter er meget interessert i at kildeoverføringen kan skje med høy kvalitet og størst mulig hurtighet.

For dem som er interessert i å vite hvilke kilder som finnes og hva de kan brukes til, kan vi henvise til informasjonsmaterialet som er utarbeidet av RHD, bl.a. en trykt rapport fra mars 1985. Det er også laget et skolefjernsynsprogram om arbeidet i RHD, som lederen av RHD, Fredrik Fagertun, også kan gi nærmere orientering om.

RHDs adresse er: *Registreringsentral for historiske data, Universitetet i Tromsø, Boks 678, 9001 Tromsø.*

«High Technology and Computers in Education»

Lise Opdahl

Dette var navnet på en utstilling som ble holdt i London 23.-26. januar år i Barbican Centre.

Britene ligger for tiden i fremste rekke i Europa når det gjelder utvikling av datamaskiner og bruk av disse i undervisningssammenheng. I britiske skoler er det nå f.eks. mer enn 100.000 mikromaskiner flere pr. elev enn i noe annet land. Det er også tatt et nasjonalt initiativ til å utvikle programvare for undervisning på universitetsnivå på mange forskjellige fagfelt. Denne utstillingen var derfor en god anledning til å få et visst overblikk over status på dette område.

Utstillingen var den første i sitt slag, og favnet vidt. Praktisk talt alt høyteknologi og alle datamaskiner kan jo brukes i en eller annen utdanningssammenheng, og dette ble en utstilling der man fikk se de siste innen maskinvare og roboter. I tillegg kom selvsagt alt som nå finnes på markedet av utstyr for input og output: tastaturer, mus, berøringstavler, tegnebrett og «voice input», samt printere av alle slag. Det som ikke alltid kom like klart frem, var hvilken funksjon alt dette kunne ha i undervisningssammenheng. Det hindret selvsagt ikke interesserte i å få ideer til egen undervisning på egen hånd.

Ordet «education» var tolket vidt, noe som var naturlig på en slik første utstilling. Til sammen dekket utstillerne svært mange typer utdanning for svært mange behov: fra utviklende spill og leker for barn i 1 1/2-2-årsalderen(!) til opplæring av psykisk utviklingshemmede, fra almenutdanning på grunnskolenivå til meget spesialiserte kursopplegg for universitetene, fra avansert teknisk yrkesopplæring til «hjemmeopplæringsprogram» i slanking og cocktailmiksing. Arrangørene hadde tatt sine forholdsregler: publikum var begrenset til personer over 18 år

med yrker relatert til undervisning.

Utstillingen gav altså mulighet til oversikt over status for utviklingen innen maskinvare, og flere firma benyttet anledningen til å lansere nye produkter på dette området. For en humanist med undervisningsoppgaver og uten spesiell teknisk eller datafaglig bakgrunn, ble imidlertid hovedinteressen ved utstillingen å prøve å få en oversikt over hva som fantes av programvare til undervisningsbruk, samt bøker og tidsskrifter på området. Og utstillingen syntes å gi et godt bilde av situasjonen, særlig hvis man tok rikelig for seg av brosjyrer – naturlig nok kunne bare et fåtall program bli demonstrert. Hovedinntrykket var imidlertid følgende: det er fremdeles alle slags spill som dominerer markedet – selv om det i denne sammenheng ble argumentert med at disse var «educational», fordi de appellerer til fantasi og kreativitet. Forskjellige typer lekeprogram for førskolebarn var det relativt mange av: stove- og tallprogram, øvelser i form og farve, dyreprogram o.l., der små tavler med figurer til å peke på ble brukt til å gi inndata. Logo-skilpadden var også på plass.

Når det gjelder programvare for vanlige skolefag var det, som ventet, realfagene som dominerte: enkelte brosjyrer hadde mange sider med presentasjon og omtale av program til bruk i undervisning i matematikk, fysikk, kjemi og biologi på flere nivåer – også for universitetene. For samfunnsfag som økonomi og geografi fantes det også programvare, mens tilbudet innen humaniora foreløpig var meget tynt: de fleste produsenter/forlag hadde ikke noe på dette området. Et fag som historie burde f.eks. kunne by på mange seriøse programmuligheter for et så stort marked som det britiske, men med ett unntak fantes det bare «adventure»-program og program av spørreboktypen. Unntaket var «Town and Country», et opplegg basert på folketellingen av 1861 – i prinsippet ikke ulikt de opplegg som er kommet fra NSD.

Med undervisningsoppgaver innen engelsk som fremmedspråk var jeg særlig interessert i hva som fantes av program innen engelskfaget, samt innen fremmedspråksopplæring generelt. For engelskopplæring i norsk grunnskole kan antakelig en del av de mange «language games» og «spelling programs» laget for elever med engelsk som morsmål ha en viss interesse som supplement til undervisningen, særlig fordi mange var laget slik at de kan redigeres av den enkelte lærer. For videregående skole og universitetsnivå kan det ha interesse at et forlag er gått i gang med de engelske litterære klassikere, med program som ifølge vaske-seddelen skal passe til selvstudium eller repetisjon. Enkelte Shakespeare-skuespill er allerede på markedet, mens andre kjente verk vil følge i løpet av våren. I sommer kan derfor den som vil, bruke datamaskinen til å studere/repetere såvel fem Shakespeare-skuespill som andre klassikere som Great Expectations, Pride and Prejudice, Jane Eyre, Lord of the Flies og Animal Farm – bare for å nevne noen. Om dette er en fornuftig måte å bruke tid og penger på, gjenstår å se. For videregående skole og for postgymnasialt begynnernivå kan noen



av språkprogrammene innen staving og vokabular produsert for engelsk morsmålsundervisning ha verdi.

Med den store bredde i programvare som utstillingen tross alt viste var det påfallende at området engelsk som fremmedspråk ikke var representert i det hele tatt. Dersom utstillingen var representativ – og det er det ingen grunn til å tro at den ikke var – er det altså fritt frem for alle som vil prøve seg på å lage program på et område med stort behov.

For fransk og tysk som fremmedspråk fantes det noen få program for O-level (passende for ungdomsskole/videregående skole), og flere kommer i løpet av våren.

For den som eventuelt gjerne vil se nærmere på programvarene, er selvsagt formatet og kompatibiliteten viktig. Mange av programmen på utstillingen var fremdeles på kassett, særlig de av spill-typen, men det var tydelig at diskettene nå kom for fullt, særlig da i 5 1/2" format til bruk for BBC modell B, som ved siden av Research Machines er den dominerende mikromaskin i engelsk skole i dag. Ellers var det programvare fra Commodore, Sinclair og Electron som dominerte.

Presentert på utstillingen var en god del av den stadig voksende litteratur om bruk av datamaskiner generelt og om deres funksjon i undervisningssammenheng. Man merket seg også at det i løpet av våren vil komme flere nye tidsskrift på området datastøttet læring/undervisning. Dette betyr at mange nå oppfatter datastøttet læring som et nytt og viktig område.

Utstillingen hadde sin store verdi ved at den som helhet gav et godt inntrykk av status på et noe uoversiktlig område i stadig forandring og vekst. Samtidig viste den relativt klart en del problemer og svakheter ved dagens situasjon.

Fordi utstillingen i utpreget grad var kommersiell, gav den meget få opplysninger om hva som foregår av programutvikling ved de enkelte undervisningsinstitusjoner, eller om pedagogiske forsøk og/eller erfaringer med datastøttet læring. Med de forskjellige firma i fokus var det heller ikke så enkelt å orientere seg om eksisterende programvare på eget fagfelt. En annen organisering av utstillingen, f.eks. etter skole-type, nivå eller fagområde, ville gjort den mer brukervennlig for det store flertall av besøkende, for hvem tilbudet innen programvare nok var det viktigste på en slik utstilling. Men kanskje må kompatibilitetsproblemene bli mindre før dette kan fungere tilfredsstillende. – Fremtidige utstillere vil nok i stadig sterkere grad merke hvor viktig det er at god maskinvare følges opp av god og rikelig programvare. En avansert mikromaskin som ble lansert på utstillingen, Research Machines' «Nimbus», vakte f.eks. liten interesse, fordi svært lite programvare er utviklet for den.

Bare noen få britiske forlag var å finne på utstillingen, noe som er representativt for situasjonen: hittil har få av de tradisjonelle britiske lærebokforlag engasjert seg i utvikling av programvare til undervisningsbruk. Hvis kvalitetsforlag som f.eks. Longman og Macmillan etter hvert tar opp utfordringen, vil dette forhåpentligvis hjelpe på den faglige standard.

Alt i alt kan man vel si at utstillingen bekreftet inntrykket av at mens man når det gjelder datamaskiner nå regner med å være i 4. generasjon, er man på programvaresiden for undervisning nok fortsatt i 1. generasjon. Dette gjør det desto viktigere å lage programvare som holder faglige mål.

Det er klart at produksjon av faglig tilfredsstillende programvare krever høy faglig kompetanse og solid pedagogisk erfaring. Mange humanistiske fagområder burde ha stoff til å lage minst like interessante program for undervisning på forskjellige nivåer som en del av det som ble presentert på utstillingen for realfagene. Britiske universiteter har fått bevilgninger til dette, og er begynt – når begynner vi?

Lise Opdahl er universitetslektor i engelsk språk ved Universitetet i Bergen.

ANTEM III, Paris 12.-15. mars 1985

Knut Hofland

ANTEM, hvis offisielle tittel er «Nye teknikker i undervisning, profesjonell opplæring og kultur», er et av 18 prosjekter innen det internasjonale utviklingsprogrammet «Teknologi, utvikling og sysselsetting», som ble satt i gang i 1983 etter initiativ av François Mitterand. ANTEM er knyttet til CESTA (Centre d'Études des Systèmes et des Technologies Avancées) i Paris. CESTA har eksistert i 3 år og har prosjekt innen tre hovedområder 1) produksjon, automatisering og robotteknikk, 2) informasjon, kommunikasjon og media og 3) bioteknologi. For nærmere opplysninger om ANTEM se HD 1-85 s. 65.

Denne konferansen var den tredje i rekken og var delt inn i 3 hovedsesjoner:

- 1) Nye teknikker og opplæring innen servicesektoren
- 2) Pedagogisk anvendelse av nye kommunikasjonsteknikker
- 3) Pedagogisk anvendelse av bilder

I tilknytning til konferansen var det en utstilling av pedagogisk programvare og utstyr, i hovedsak fra forskjellige firma i Canada. Det ble også holdt fire arbeidsgrupper under konferansen.

I det følgende gis en omtale av noen av foredragene.

Guy Latge, CASSIE, Paris, presenterte et interaktivt videodisk opplæringsprogram om bruk og vedlikehold av og feilsøking i bankautomater. Det ble brukt en Philips Laservision platespiller. Det er behov for å lære opp 750 personer pr. år og opplæringstiden er redusert fra 2.5 dag til 1.5 dag ved innføring av dette opplæringsprogrammet. Utviklingskostnadene fordelte seg med ca. 20% til programmering og teknisk produksjon av plate og resten til analyse, opplegg og regi av de forskjellige sekvenser på platen. Det ble brukt skuespillere og musikk under produksjonen.

Trevor Bentley, CNJ Systems, London, orienterte om planene for et innebygget opplæringsystem til et programsystem, f.eks. et databaseprogram. En prøver her å smelte sammen opplæring og hjelpefunksjoner til programsystemet i en integrert enhet som er tilgjengelig sammen med systemet. Læringsystemet skal tilpasses den enkeltes nivå og behov ved at bruken av systemet blir registrert i en brukerhistorie. Systemet gjør også nytte av lyd-kassetter og trykkt opplæringsprogram. Brukeren starter med lyd-kassetten og det trykte materialet og går deretter over til opplæringsmodulen i systemet. Etter hvert begynner brukeren å kjøre systemet med hjelpefunksjonene innkoblet. Disse kobles ut etter hvert som brukeren blir fortrolig med systemet. I Storbritannia er det 25 millioner arbeidstakere og i løpet av få år vil

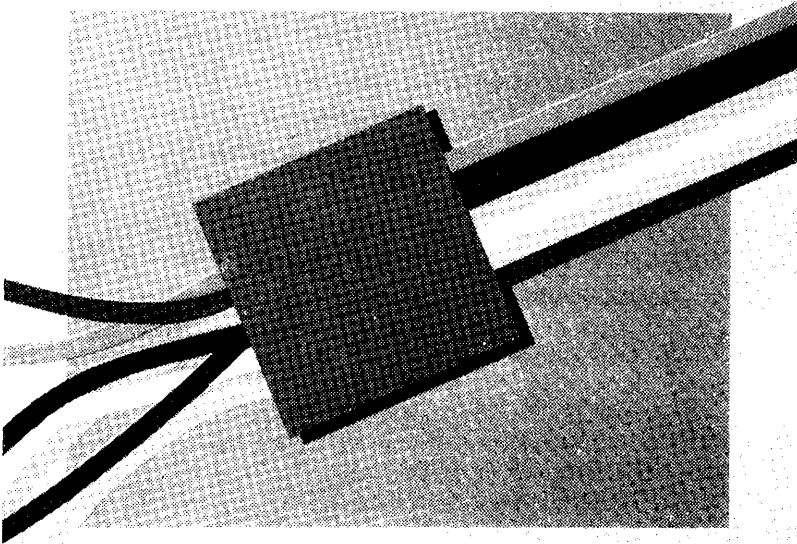
tallet på personer som gjør bruk av edb være steget fra 2 til 10 millioner. Opplæringsbehovet er således stort. Prosjektet er statlig finansiert med 8 mill. kr.

Georges Cailleteau, Agence pour le Développement de l'Informatique, Frankrike, ga en presentasjon av et mikronettverk for skoler. Det er i dag ca. 100 slike i drift og en ny utgave vil trolig øke utbredelsen. Nettverket knytter sammen billige arbeidsstasjoner og IBM-kompatible maskiner via vanlig telefonkabel og med en hastighet på 300 Kbit. En overbygging av de enkelte operativsystem (CP/M, MS-DOS og Thompson-DOS), gjør systemet enkelt og uniformt i bruk. Lærerne har et eget program for å legge inn og tilpasse nye programmer og data til systemet.

Christopher Humphries, Council for Educational Technology, England, orienterte om «telesoftware», distribusjon av programvare via elektroniske nettverk. I Storbritannia fins det i dag ca. 10.000 programmer som kan brukes i de ca. 30.000 skoler som fins der. Det ble fra 1981 til 1984 gjort forsøk med distribusjon av programvare via Prestel teledatanett. 98% av skolene har adgang til Prestel via telefon med lokaltakst. Formatet som ble brukt er senere blitt Prestels standardformat og er også brukt i flere private teledataselskap. Prosjektet tok for seg tekniske, organisatoriske og økonomiske faktorer i distribusjon av pedagogisk programvare. Det viste seg at systemet virket best for mindre programmer og programmer med begrenset dokumentasjon samt for demonstrasjonsutgaver av program og eksempler på data. Kostnadene ved distribusjon ble redusert med 40% ved et opplag på 500 kopier. Fra høsten 1985 skal det startes et kurs innen Open University i markedsføring i eksportnæringen der telesoftware sammen med videobånd, kassetter og trykket materiale utgjør kursmaterialet.

Bertrand Retaillion, APRIM, Paris, representerte et firma som har laget flere videoplater, bl.a. en som skal brukes i videreutdanning av 36.000 tannleger i Frankrike. Tidligere ble det holdt 300 kurs lokalt og en stor samling med lysbilder ble transportert landet rundt. For ca. 50.000 kr. får nå tannlegene en IBM kompatibel maskin, monitor og videoplatespiller. Dette utstyret kan både brukes til administrative oppgaver og til opplæring. Opplæringssystemet består av tre deler: 1) læreprogram med presentasjon av tilfeller, diagnoser og behandling, 2) et dentalt leksikon og 3) en kommunikasjonsmodul. Tannlegen kan ringe opp en lærer som har samme type utstyr og begge har mulighet til å styre hverandres videospiller for å trekke frem aktuelle bilder/filmsekvenser. I tillegg kan verbal kommunikasjon foregå via tastatur/skjerm eller separat telefonlinje.

Vincent Thompson, Butler Cox & Partners, London, presenterte et gjør-det-selv teledatasystem basert på mikromaskiner. Systemet gjør det mulig for lærere og elever å legge opp egne teledatabaser. Informasjon fra eksisterende teledatabaser kan lastes inn i systemet og brukes videre der. Det ble demonstrert en database over fiske som noen elever



hadde laget som resultat av en hjelpeklasseundervisning. Elevene hadde samlet inn opplysninger i lokalmiljøet om fiskeplasser, utstyrsforretninger o.l. Systemet gjør det mulig å editere både tekst og grafikk

Gordon Kerr, British Telcom Research Laboratory, Ipswich, holdt et

meget interessant foredrag om interaktiv video via kabelnett. Dette er et prøveprosjekt i en del av London som vil blir kommersielt tilgjengelig høsten 1985. En gjør her bruk av et bredbånds kabelnett i stjerneform. Forbindelsene mellom hovednodene er optiske fibre og ut til forbrukerne er det en vanlig koaksialkabel. I tillegg til enveis kabeltjenester som fjernsyn og radio kan en tilby forskjellige interaktive tjenester. Det eneste tilleggsutstyret utenom et vanlig fjernsyn er tilkoblingsboksen til nettet med en egen fjernkontroll. Denne kan ha et rent numerisk tastatur eller et komplett alfanumerisk tastatur. De interaktive tjenestene er av to typer: teledata og et videobibliotek. Teledataformen brukes som informasjonstjeneste i nettet. I tillegg fins det mulighet for lokal eller nasjonal teledata. Den lokale teledata kan inneholde fotografisk materiale som åpner nye muligheter for denne type tjeneste, f.eks. innen salg av eiendommer o.l. En nasjonal teledatasentral som kan vises, er f.eks. Prestel. Ved den interaktive videotjenesten kobles brukeren opp mot en videokilde og kan gi kontrollsignaler tilbake via tastaturet. Videokilden kan være en grafisk skjerm, videobånd eller videoplate. Tjenester som er aktuelle er videoleksikon (Longman har allerede en plate om kroppen), visuelle databaser (kart, diagrammer o.l.), bilder fra kunstgallerier, forskjellige typer opplæring fra faglige kurs til gjør-det-selv håndbøker. Brukeren har full kontroll over videospilleren m.h.t. stopp, tilbakespøling, sakte kino o.l., som om den står i stua. Videospilleren som er brukt er en enkel Philips Laservision spiller som er utstyrt med interface som er laget av British Telecom. En tenker seg 300 videospillere med mekanisk arm og et bibliotek på 1000 titler.

Jean-Luis Dessalles, Den franske telehøgskolen (ENST), presenterte SAVANT 2: en videotjeneste for undervisning. Dette systemet er ment som et tilleggshjelpemiddel i undervisningen, spesielt for repetisjon. En lærer ikke noe nytt stoff, og det er ingen spørsmål. SAVANT, som er et rent teledata læreprogram, er blitt brukt i 3 år og er delt inn i 14 deler. I SAVANT 2 er dette utvidet med et bibliotek av videosekvenser, ca. 1000 i tallet. Disse varer fra noen sekunder til 1-2 min. I tillegg kan en ha en egen lydkilde. Disse tre kildene, som er sentralt plassert, blir distribuert via et stjernenettverk og blir blandet i to omganger (via en matrisekobling) før de når brukerterminalene. Det er kun brukt kommersielt tilgjengelige produkter, bl.a. optiske fibre.

Pascal Levay, CCETT, demonstrerte et system for høykvalitets tredimensjonale bilder/animasjoner. Kjernen er en spesialmaskin for tredimensjonal grafikk som kan ha flere prosessorer og der mikrokoden kan tilpasses den aktuelle bruk. Maskinen kan brukes som en terminal mot VAX, men senere skal det være mulig å knytte den til en IBM-PC. En har som mål å komme ned fra en kostnad på 20-30.000 franc pr. sekund animasjon til 1000-3000 franc. Det ble vist flere eksempler, de fleste reklameinnslag, men også et som var en rekonstruksjon av et historisk gravkammer med utsiktspunkt fra forskjellige steder og



En av utstillerne på ANTEM-konferansen demonstrerer «Mimi», et program for barn i alderen 2-8 år.

forflytting gjennom gangene.

Bob Salkeld, BBC, orienterte om BBC Domesday Project og viste er video slik som systemet ville virke for brukeren. Prosjektet er beskrevet i en egen artikkel i dette HD og omtales derfor ikke videre her.

Christopher Jones, Edinburgh, gav et foredrag om bruk av interaktiv video i utdanning av døve. Foredragsholderen er selv döv og fulgte konferansen via døvetolk! Tre prosjekter ble beskrevet. Det første omhandler lesing og forståelse og bruker tegnefilm med undertekster i tekst-tv. Filmsekvensene blir stoppet slik at elevene får lest tekstene ferdig. I tillegg blir det stilt kontrollspørsmål. Det andre prosjektet gjør bruk av tegnefilm i opplæring av verbets tider og passive former av verb. Det siste prosjektet er produksjon av en videoplate som skal brukes som en interaktiv ordbok/leksikon. Foredragsholderen konkluderte med at bruk av video vil være et effektivt hjelpemiddel til å høyne leseferdigheten hos døve. 55% har i dag en leseferdighet lavere enn 9-årsstadiet og 75% lavere enn 13-årsstadiet.

To av utstillerne hadde produkter innen videoplateteknologien. Den ene var et fransk firma som har laget programvare for styring av videospillere fra en IBM-PC og oppbygging av en database over materialet på platen. Den andre var et kanadisk firma, Interactive

Image Technologies, som er en veteran innen utvikling av interaktive videosystemer. Leder er *Joseph Koenig* og han har produsert 12 videoplater siden 1981. En av disse omhandler en simulering av en pionerfamilies liv på 1780-tallet i Canada.

Et sjarmerende og prisbelønnet program kalt «Mimi» beregnet på barn helt ned til 2 år, blir vist på utstillingen. Det er skrevet til en Commodore 64 og består av et tyvetall tegnefilmsekvenser med klassiske musikkstykker som bakgrunnsmusikk. Hver sekvens kan kalles frem med en bokstav på tastaturet og barnet kan programmere sin egen historie ved å sette sammen sekvenser i en serie.

Konferansen hadde en serie med demonstrasjoner som viste bruk av datamaskin, videobånd og platespiller, modem o.l. Demonstrasjonene foreløp uten problemer – unntatt for foredragsholderen som hadde glemt passordet for å komme inn på et transatlantisk nettverk. Det var simultanoversetting til engelsk, fransk og delvis italiensk. Konferansen blir dyktig administrert av Anne-Françoise Tester. Antem IV vil bli arrangert i mars 1986 i tilknytning til en konferanse om datamaskinassistert undervisning.

Referanse

Preprints ANTEM III, CESTA, 1 rue Descartes, 75005 Paris

Humanioradagene 1985 ved Universitetet i Oslo

«Humanister i dataalderen» 18.-22. mars

Tellef Kvifte

Etter to år med «Humaniorauker» som satte hele fakultetet på ende med åpne hus og arrangementer ved praktisk talt alle institutter og avdelinger, hadde det Historisk-filosofiske fakultet ved Universitetet i Oslo i år satsset på et litt mindre format og mer veldefinert tema. Under tittelen «Humanister i dataalderen» ble det avviklet seks foredrag/seminarer på formiddagstid og fjorten mer lokale instituttarrangementer.

Årets arrangement var mer rettet mot universitetets og fakultetets ansatte og studenter enn tidligere, og noe av hensikten var å bevisstgjøre humanistene om hva slags bidrag fakultetet kan ha både i forståelsen av den nye teknologien og i utviklingen av den. Det forhindret ikke at mye av publikum også i år var andre folk enn fakultetets egne.

Programmet spente over et vidt spekter, fra filosofiske implikasjoner

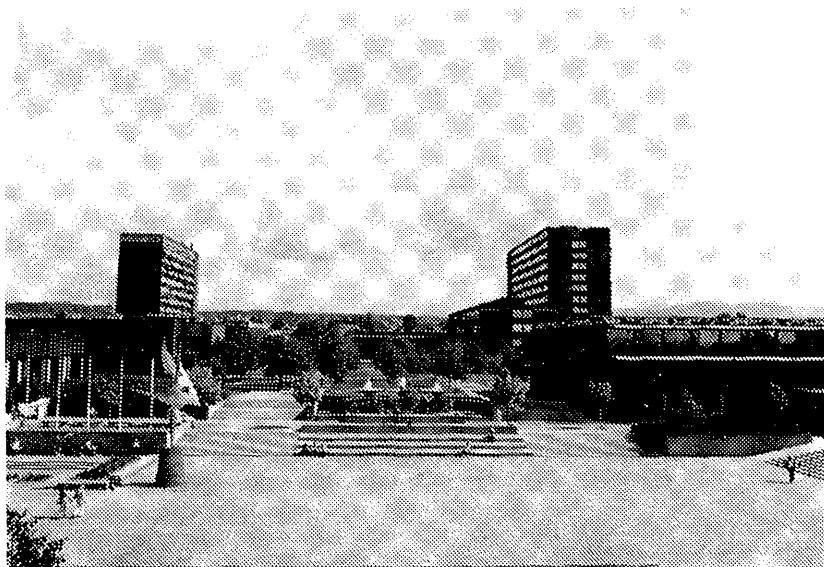
av kunstig-intelligensforskningen til tekstbehandling og edb-faget i skolen, og debatten ble til tider riktig frisk etter fordedragene. Det ble også arrangert utstillinger; Bibliotekjentesten stilte ut bøker innen fagområdet, og en del mikromaskinleverandører viste maskiner og programvare. Det må sies at denne utstillingen ikke var særlig godt besøkt – delvis så det ut til at maskinleverandørene ikke oppfattet HF som et interessant marked og derfor ikke gjorde den helt store innsatsen, og delvis virket det som folk enten ikke var interessert i maskiner, eller hadde skaffet seg den informasjonen de trengte annetsteds.

Først ute i hovedarrangementene var *Astrid de Vibe* og *Tor Edvin Dahl* med foredrag omkring spill av «adventure»-typen og perspektiver på interaktive bøker. Astrid de Vibe ga oss et innblikk i dataspillenes mannsdominerte kultur, og viste forbindelser mellom spillene og folkeeventyr og eldre heltediktning. Tor Edvin Dahl var nok noe skeptisk til fenomenet interaktive bøker, der leseren kan «gripe inn» i handlingen; blant annet fordi han mente at forfattere er interessert i å vise en bestemt utvikling når de først skriver, og at dette ikke lar seg forene med valg som leseren foretar.

Tirsdagens arrangement hadde den megetsigende tittelen «EDB er svaret, men hva er spørsmålet». *Christian Beck* fra Pedagogisk forskningsinstitutt gav en utmerket begrunnelse for hvorfor humanister er nøkkelpersoner i edb-undervisningen i skolen. Kort fortalt argumenterte han med at edb-undervisning ikke bør handle så mye om programmering som om hvordan bruke programsystemer, og at effekten av «datarevolusjonen» blant annet blir at enorme datamengder blir lett tilgjengelig. Den viktige oppgaven blir dermed å gi elevene redskap til å orientere seg i datamengdene, og det viktigste redskapet til dette er det vi gjerne kaller «kulturell bakgrunn». Med seg hadde Beck to hovedfagstudenter, *Lene Nyhus* og *Signe Jensen*, som la fram erfaringer fra edb-undervisning i grunnskole og ungdomsskole, der edb-verktøy hadde gått inn i prosjektarbeid i andre fag, der elevene blant annet lærte å bruke database.

Onsdag var det to arrangementer. På det ene fortalte *Helge Dyvik* og *Jens Erik Fenstad* om datalingvistikk; på det andre fortalte *Arvid Staupe* om departementets planer for edb i skolen framover. Han ble fulgt av *Kåre Andersen* fra HF-data, som presenterte sitt program RUBREG og hvordan dette sammen med for eksempel folketellingsmateriale kan brukes i et integrert undervisningsopplegg i historieundervisningen i skolen.

Torsdagens arrangement var et av de best besøkte. *Andrew Jones* og *Dagfinn Føllesdal* tok opp fundamentale spørsmål i tilknytning til vårt forhold til maskiner, og gikk i dybden på problemene med hva vi mener med «kommunikasjon» og «intelligens». Gir det mening å bruke disse begrepene om det som skjer mellom mennesker og maskiner, og om det som skjer inne i maskinene? Innbudte kommentatorer var *Stein Bråten*



Blindern, Universitetet i Oslo. (Foto: Universitetet i Oslo).

og *Kristen Nygård*, og med god hjelp fra salen fikk de i gang en diskusjon som ble riktig frisk.

Det avsluttende arrangementet på fredag var viet kildebehandling. *Asbjørn Brændeland*, *Trond Berg Eriksen* og *Tellef Kvifte* innledet, og hevdet synspunkter som spente fra begeistring over store besparelser i edb-støttet materialbehandling og -gjenfinning til dyp mistillit til gjen-nomføring av den type formalisme vår omgang med maskinene nødvendigvis må være preget av.

På ettermiddagstid var det Bibliotekstjenesten som hadde de mest omfattende arrangementene, og demonstrerte litteratursøking ved edb alle dagene. Det ble også vist datautstyr for blinde ved Romansk institutt, og edb-baserte hjelpemidler til notetegning ved Musikkvitenskapelig institutt i samarbeid med Institutt for informatikk. Institutt for nordisk språk og litteratur ga oversikt over hva de har brukt edb til, og på Britisk institutt delte nye tekstbehandlingsbrukere sine erfaringer med andre. Her holdt også *Michael Benskin* innlegget «How computery is changing the character of research in the humanities», som omhandlet såvel edb som redskap som de sosiale endringer forskningsprosessen er utsatt for.

Institutt for etnologi hadde fått *Jon Birger Østby* til å fortelle om edb-virkomheten innenfor kunst- og kulturhistoriske museer, og Institutt for namnegransking tok opp både praktiske og prinsipielle sider ved edb-bruk i namnegransking.

Arrangementet som helhet viste at det er stor bredde i de edb-faglige aktivitetene ved fakultetet. Aktiviteten er stigende, og stadig flere institutter begynner å interessere seg for området, enten med tanke på rent praktisk nytteverdi, eller på mer teoretisk og faglig grunnlag. Dette vil etter hvert også gjenspeile seg i studietilbudet. I øyeblikket er to nye grunnfag under planlegging i kjølvannet på denne utviklingen: Institutt for filosofi vil tilby et grunnfag i kommunikasjonsteori og semantikk*, mens HF-data vil utvide det eksisterende semesteremnet i edb for humanister til grunnfag.

*Se HD1-85 (red. anm.)

Tellef Kvifte er amanuensis ved Norsk Folkemusikksamling, Universitetet i Oslo, hvor han bl.a. har laget et arkivopplegg for edb.

Computers and Art History Workshop

University College, London 2.-3. april 1985

Espen Ore

Ideen om denne workshop kom frem blant enkelte deltagere på konferansen om automatisk behandling av kunsthistoriedata og -dokumenter i Pisa, september -84. På denne konferansen var det siden av plenumssesjonene med paneldebatter, også parallelle demonstrasjoner. Erfaringene fra disse demonstrasjonene ga mersmak, og mulighetene for et noe mindre arrangement basert på demonstrasjoner og diskusjoner ble luftet. *Will Vaughan* og *Dave Guppy*, begge fra University College, London arbeidet videre med planene og kunne ønske ca. 60 deltagere velkommen. De aller fleste kom fra Storbritannia, men det var også representanter fra Italia, Norge, Polen og USA.

Første dag begynte med foredrag og diskusjoner om nytten av å bruke allerede eksisterende programmer, fremfor å utvikle spesialprogrammer. Innlederne viste bl.a. hvorledes vanlige editorer og standard sorteringskommandoer på operativsystemer kan være til stor nytte, at man ikke har behov for forseggjorte databasesystemer. Blant innlederne kan nevnes *Tim Benton* fra Open University som arbeider med arkitekttegninger. Han viste hvor mye vanskeligere denne typen forskningsobjekter er m.h.p. sortering og gruppering enn f.eks. oppslagsord i et leksikon. Tegningene kan grupperes etter internt løpenummer, men det er lite hensiktsmessig i en forskningssammenheng. Man kan også

sortere dem etter type, skisser kontra ferdige byggetegninger osv. Benton tok i sitt arbeid utgangspunkt i det enkelte bygg, altså hvilket fysiske objekt tegningene representerte. Dessverre byr også dette utgangspunktet på problemer: han viste et eksempel på en nesten ferdig tegning av ett hus som også inneholdt en skisse av et annet.

Benton konkluderte ang. sesjonens hovedtema at han fant det mest hensiktsmessig å bruke mikromaskin med vanlige, kommersielt tilgjengelige programmer (f.eks. DBaseII) til dataregistrering. Deretter kan data overføres til en stormaskin for mer ressurskrevende arbeid som sortering o.l.

Michael Greenhalgh fra Leicester University tok opp tråden og argumenterte sterkt for størst mulig bruk av mikromaskiner. Han mente at de har særlig fire fordeler:

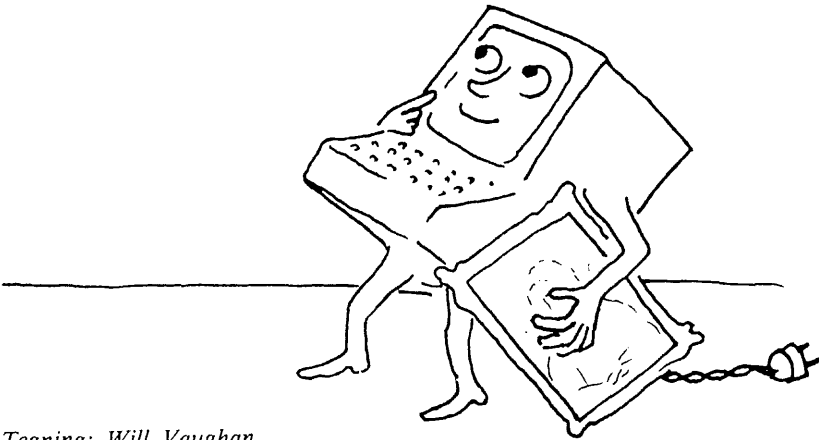
- De er ikke opptatt p.g.a. annen bruk
- De har nye og brukervennlige programmer
- De har mye bedre tekstbehandlingsmuligheter enn stormaskiner
- De er billige

Når en bruker mikroer har man også muligheten til å finne fritt tilgjengelige («public domain») programmer gjennom kontakt med brukergupper. Likevel, mente Greenhalgh, gjenstår det ett vesentlig problem for forskere i kunsthistorie: billedkvalitet. Siden visuell informasjon er så vesentlig innen kunsthistorie, stiller man langt høyere krav til oppløsning og fargegjengivelse enn f.eks. næringslivet gjør til forretningsgrafikk. I Greenhalgh's øyne var f.eks. en Macintosh ikke å anse som mer enn et glorifisert leketøy. Han mente at kunsthistorikere kan få det de trenger dels ved at høykvalitets grafiske arbeidsstasjoner blir billigere og dels gjennom utvikling av videoteknologi. Men også Greenhalgh konkluderte med at kunsthistorikere bør satse på å benytte rimelige og eksisterende programmer så langt dette er mulig.

Senere samme dag innledet Will Vaughan. Han argumenterte for nytten av å lære seg et programmeringsspråk. Som et eksempel viste han til egne eksperimenter der han hadde laget programmer som simulerte Mondrian-komposisjoner. P.g.a. Mondrians strengt rektangulære stil var det mulig å nøye seg med relativt enkel datagrafikk. Slike programmer fant Vaughan nyttige til f.eks. analyse av utviklingen av Mondrians stil.

Den andre dagen fordelte deltagerne seg på grupper som tok for seg spesielle temaer: dataregistrering og -organisering, statistikk for kunsthistorikere og billedprosessering og -lagring. Det ble gitt demonstrasjoner og deretter diskutert i gruppene. Undertegnede demonstrerte et program som blir brukt i prosjektet Litera (se HD1-85), og jeg vil benytte anledningen til å takke Dave Guppy og personalet ved Computer Centre og Department of Computer Science, University College, for all hjelp jeg fikk med å implementere det på en GEC mini under operativsystemet EUCLID.

COMPUTERS AND ART HISTORY



Tegning: Will Vaughan.

Gruppen som tok for seg billedprosessering og -lagring, dro først på en ekskursjon til The Remote Sensing Unit of the Geography Department, University College. Her har man spesialisert seg på å behandle satelittbilder som kan inneholde data både fra synlige og ikke-synlige bølgelengder. For å analysere bildene har The Remote Sensing Unit tatt i bruk flere metoder som kan være av interesse for kunsthistorikere.

Bildene blir vist på en grafisk skjerm med en oppløsning på 512×512 bildelementer (pixler). Det er fullt mulig å zoome inn på detaljer, og kvaliteten på gjengivelsen er kun avhengig av oppløsningen i det lagrede bilde – skjermens oppløsning berører altså ikke mulighetene for forstørring.

Satelittbildene vi ble vist, var i utgangspunktet svært uklare. De var tatt på overskyete og tåkete dager, og bare trange belter av spektrum var representert. Ved å «strekke» ut disse beltene slik at de spente over hele den synlige del av spektrum, kom alle fine nyanser meget tydeligere frem, og man kunne se detaljer som ellers var fullstendig skjulte. Siden satelittenes sensorer også registrerte ikke-synlige bølgelengder, inneholdt satelittbildene også informasjon om f.eks. infrarød stråling. Ved å omforme disse bølgelengdene, kunne man se dette direkte på bildet. I behandlingen av satelittfotografier kan slik informasjon brukes til studier av vegetasjon, forurensning osv., men også innen kunsthistorie er slike muligheter interessante.

Etter demonstrasjonen ble det diskutert hvorledes slike metoder som vi hadde sett, kunne brukes i bl.a. restaureringsarbeid. Ved å registrere et bilde digitalt også m.h.p. ikke-synlige bølgelengder, vil man ha en langt mer eksakt viten om dets tilstand på et gitt tidspunkt enn man kan få gjennom f.eks. slides. Ved å gjenta slik registrering hvert 10. eller 50. år, kan man få empiriske data som støtte til å regne ut hvorledes pigmenter har endret seg siden bildet ble skapt. I dag brukes bl.a. kjemisk viten til slike utregninger, men det vil være svært nyttig å kunne kople denne sammen med rent empiriske data.

Workshop'en sluttet med en plenumssesjon der *Katarzyna Zawaisa-Staniszevska* fra Adam Mickiewicz Universitetet i Poznan redegjorde for status quo i Polen. Hun gikk først og fremst inn på arbeidet som blir gjort i Poznan med å kartlegge det kunstneriske miljø fra 1890-1930. Alle data om kunstnere og kunstverk så som korrespondanse, utstillingskataloger, anmeldelser osv. blir registrert. Siden tekstenes informasjonsverdi varierer sterkt (det er forskjell på et maleris omtale i en salgskatalog og i en anmeldelse), blir tekstene tagget.

Etter hennes innlegg ble erfaringene fra workshop'en og mulige tiltak fremover diskutert. Arrangørene påtok seg å starte utgivelsen av et meldingsblad om bruk av datamaskiner i kunsthistorie (se HD1-85). Det ble også bestemt å arrangere éndagsseminarer et par ganger i året, men man regnet realistisk med at disse ville få nasjonal (britisk) karakter. Ettersom man også mente at det var et klart behov for internasjonale sammenkomster, ble det besluttet at man skulle forsøke å arrangere tiltak som denne workshop årlig.

Proceedings fra workshop'en vil muligens komme ut i løpet av høsten -85.

Utviklingsseminaret på Ustaoset 17.-19. april 1985

Espen Brodin

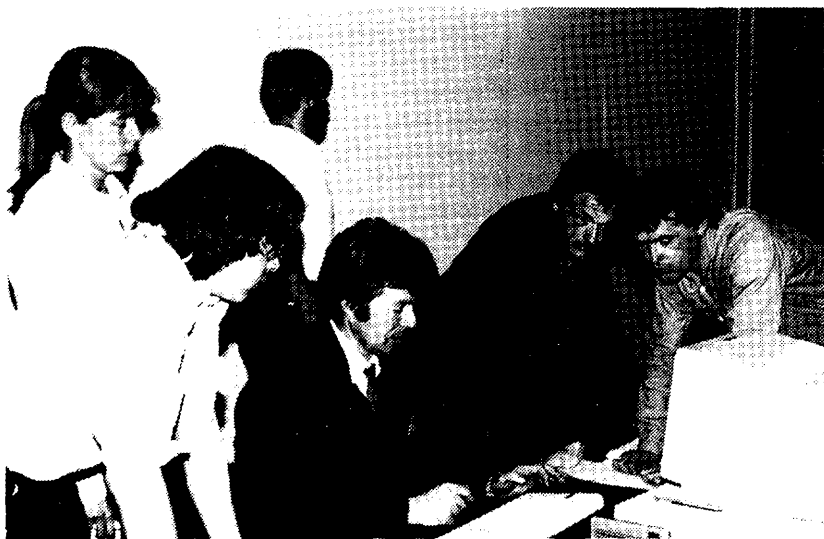
Temaet for seminaret var ekspertsystemer eller kunnskapsbehandlings-systemer. Disse såkalte fjerde og femte generasjons datamaskinsys-temer gjør datateknologien tilgjengelig på en helt annen måte enn tidligere. Tidligere systemer, som blant annet er kjennetegnet ved møysommelig programmering, erstattes med et system som kan forholde seg aktivt til kommandoer på dagligdags språk, resonnere utifra disse og supplere med ny informasjon. Kombinert med en, eller flere databaser, får vi et kraftig pedagogisk verktøy som i aller høyeste grad er et humanistisk anliggende: utviklingsmessig såvel som i anvendelse.

Årets seminar var begrenset til 30 deltakere og basert på en kombina-sjon av foredrag, demonstrasjon og øvelser.

Den første dagen demonstrerte *Asle Hjellbrekke K-mann* (Knowl-edge manager), et fjerde generasjonsverktøy som også deltakerne fikk testet ut. Systemet var brukervennlig og svært fleksibelt på demonstra-sjonsmaterialet. Registrering og definering av forskjellige enheter, statistiske bearbeidelser og utskrift av data er uproblematisk. Integre-ring av K-manns moduler synes også å være oversiktlig og enkelt. Men for at systemet skal kunne anbefales brukermiljøene, kreves videre bearbeidelse og testing, for å kunne finne ut hvor tilpasningsdyktig det er. Det bør utføres i forbindelse med enkeltprosjekter hvor også andre systemer kan testes og sammenlignes.

Uwe Hein introduserte forskjellige ekspertsystem og hvordan blant annet LISP kan brukes til oppbygging av et system. Sammen med *Einar Dehli's* presentasjon av LOOPS, et overbygg til LISP, ble det tydelig demonstrert hvilke muligheter fjerdegenerasjons-verktøy gir, men Dehli viste ikke overbevisende nok at LOOPS var verdt sin pris i tillegg til LISP. Det er derfor behov for konkrete prosjekt som kan utforske de muligheter som ligger i denne type teknologi, og i anvendelse av eksisterende ekspertsystemskall. Det kan ikke betviles at de humanistiske brukermiljøer vil være tjent med å ta i bruk fjerde- og femte-generasjonsverktøy. Derfor bør disse miljøene bygge opp kompetanse, både når det gjelder utvikling av teknikken og undervisning i den. Humanister besitter særkompetanse innen bl.a. lingvistikk og logikk som kan bidra til å utvikle den generelle teknikk. Skal humanistisk forskning i Norge kunne følge med i den internasjonale forskningsfront i fremtiden, må økte bevilgninger fremskaffes slik at de kommer opp på det nivå teknisk forskning tilføres.

Tore Amble gav en innføring i PROLOG som er det programmerings-språk japanerne valgte som grunnlag for sitt ambisiøse femte-



Tore Amble fra RUNIT demonstrerer PROLOG for deltakere ved Utviklingsseminaret 1985.

generasjonsprosjekt. Dette språket har klare fordeler når en skal bygge opp binære systemer med veldefinerte avgjørelsesprosedyrer, og en kan stille spørsmål av typen:

«Which country bordering the Mediterranean borders a country that is bordered by a country whose population exceeds the population of India.»

Både LISP og PROLOG egner seg godt til analyse av naturlige og kunstige språk og vil være til enorm hjelp for lingvister, men disse språkenes muligheter til å bygge opp ekspertsystemer vil være av vel så stor interesse for andre humanistiske fag: *Erik Fjornes* orienterte om hvordan dette kan gjøres for faget kunsthistorie. Sentrale begreper for analyse av kunstverk presenteres verbalt og vises frem ved hjelp av piler og streker på et bilde som fremkommer på skjerm. Deretter kan systemet kommunisere interaktivt for en videre forståelse og bearbeidelse av studentens ferdigheter gjennom å stille spørsmål, kommentere svar osv.

På seminaret kom det frem klare ønsker om at NAVFs EDB-senter snarest bør arbeide videre med ekspertsystemer, både ved å bygge opp basissystemer og sette i gang anvendelsesorienterte prosjekter slik at de humanistiske fag ved universitetene kan få et kompetent utviklingsmiljø. Det ble foreslått å nedsette en ad hoc komité som kan fordele arbeidsoppgaver og initiere et større forskningsprogram. Senteret har etter konferansen innledende drøftet hvordan en bør legge an arbeidet, i

første rekke med å få fram en utredning om kunnskapsbaserte system anvendt i de humanistiske fag.

Seminar om pedagogisk informatikk

Stord lærarhøgskule 27.4.85

Per Vestbøstad

Stord lærarhøgskule har nyleg teke opp undervising i pedagogisk informatikk for å gje kandidatane innsyn i læreprogram og datamaskinstøtta undervising. Dei legg særleg vekt på å vurdera slik undervising og tilknytt programvare ut frå ein pedagogisk synsstad.

Som ei utviding av eit internt seminar om dette emnet 26.4. hadde dei bede inn professor *Tim O'Shea* frå Institute for Educational Technology ved The Open University i England. Han heldt to trippel-forelesingar laurdag 27.4.:

1. «The Future of Computer Assisted Learning»
2. «Artificial Intelligence in Computer Assisted Learning»

Eg skal her referera nokre sentrale synspunkt frå forelesingane.

1. Om framtida for datamaskinstøtta læring

Professor O'Shea tok utgangspunkt i skilnaden mellom tradisjonelle lærebøker og datamaskinstøtta undervising: Skal vi bruka edb i undervisinga, må det godtgjerast at dette er klart betre enn å bruka bøker, som jo er mykje billigare både å laga og å spreia.

Viktigaste fortrinnet for datamaskina er at ho kan presentera og simulera prosessar, medan boka berre kan presentera fakta. Dette meinte prof. O'Shea kunne få like store konsekvensar som oppfinninga av prentekunsten. Elles har datamaskina m.a. desse viktige fortrinna framfor bøker:

- Ho er interaktiv og kan såleis gje elevane snøgg tilbakemelding på prestasjonane deira,
- ho kan hugsa elevhandlingar og kan tilpassa læringa etter elevnivå, og ho kan kommunisera på mange måtar: Tale inn/ut, bilete inn/ut, styringa kan gjerast ved peiking beint på skjermen eller via skjermlos («mus»). Dette siste kan gje mange funksjonshemma tilgang til undervising dei elles kan henda ikkje ville kunna få.

Vidare peikte prof. O'Shea på nokre sentrale problem ved edb-bruk i

undervisinga:

- Systemet skjønar ikkje naturleg språk, og det er truleg langt fram før ein kan rekna med noko slikt.
- Undervisningsprogramma kan ikkje sjølv det emnet som dei skal undervisa eleven i. Dette bør vera eit krav til alle undervisningsprogram. Difor vil kunstig intelligens måtta stå svært sentralt i pedagogisk informatikk i framtida.
- Programma skjønar ikkje *kvifor* eleven gjer feil eller misforstår, og dei har inga kjensle av kor godt eleven har oppfatta stoffet (om han «er med»).

Dersom læraren likevel vil satsa på datamaskinstøtta undervising, møter han mange avgrensingar/hindringar:

Den fysiske tilgangen på maskiner er oftast knapp, og der finst lite programvare av høg kvalitet. Skal elevane kunna gjera seg nytte av utstyr og programvare, må dei læra seg å bruka tastatur, operativsystem og andre utanomfaglege «knep». Og endeleg kan det vera krevjande å innpassa programvaren i kurset/undervisinga elles.

For å laga gode læreprogram trengst det eit godt samspel mellom fleire faggrupper: Lærarar, undervisingsteknologar, programmerarar og folk med kompetanse i kunstig intelligens. Alle vellukka system er laga av arbeidsgrupper – solo-prestasjonar duger ikkje.

Kostnadane ved produksjon av læreprogram stig, sjølv om maskinkostnadane vert halverte annakvart år. O'Shea rekna at det kostar i snitt 200 timeverk å laga ein times læreprogram, og dette talet vil veksa med tida, fordi folk vert meir kresne etter kvart. Utviklinga vil gå mot snøggare program, bruk av fargeskjerm, rettskrivingskontroll, setningsanalyse og intelligente programreaksjonar.

Prof. O'Shea skisserte til sist ulike måtar læreprogram kan spreiest på. Spreiing via diskettar er det vanlegaste i dag, medan program som er fast innbrent på såkalla ROM-brikker vil verta meir brukt i framtida. Overføring via telefon- eller kringkasting (radio, fjernsyn) har vore prøvd, men set store krav til linjer og overføringsutstyr. Spreiing via videoplater vil verta vanleg i framtida.

2. Om kunstig intelligens i pedagogisk informatikk

Kunstig intelligens finn sin praktiske bruk i ekspertsystema. Som døme på slike peika Tim O'Shea på eit av dei første, MYCIN. Det var laga for å diagnostisera infeksjonar i blodet. I tillegg til å koma med framlegg om diagnose, kunne MYCIN gjera greie for eigne slutningar og generera kunstige pasientar for øvingsformål.

Sjølv om læring er det sentrale omgrepet i kunstig-intelligensforskning, har ein også arbeidd mykje for å få systema til å skjønna naturleg språk. Det mest kompliserte brukar-grensesnittet som er laga til no, er den amerikanske marinen sitt øvingssystem (kalla «Sophie»)

for å finna feilen ved straumstans på båtane deira. Det kostar e årsverk pr. time læreprogram, og likevel maktar Sophie berre analysere setningar som gjeld straumproblem.

Om val av programmeringsspråk

Prof. O'Shea gjekk grundig inn på spørsmålet om korleis ein sku velja programmeringsspråk. Han sette opp ei liste over dei faktora ein då bør vurderer: Syntaks, semantikk, verkemåtar, dokumentasjon og kulturelt opphav. Språk som er meint å skulla dekkja det beste f fleire andre programmeringsspråk, er sjeldan vellukka.

Han gav vidare ei grundig orientering om tre viktige programmeringsspråk innom kunstig-intelligens-verda: LOGO, PROLOG og SMALL TALK. (Han tok ikkje tid til å kommentera LISP, då dette var bet kjent og vel etablert innom fagfeltet):

LOGO var opphavelig laga for fjernskrivar-terminalar og var på v ut då mikromaskinene kom og gav det ein renessanse. Språk dekkjer i dag delar av «storebror» SMALLTALKs aksjonfelt, og velegna til problemløysing, prosess-simulering, geometri og algebra men ueigna for deklarativ uttrykk. Forsøk har synt at LOGO-øvelevar ikkje vart betre problemløysarar enn andre, men dei had mykje betre verbal uttrykksevne!

PROLOG er eit deklarativt språk og (difor?) svært vanskeleg å læ bort. Viktigaste argumentet for å satsa på PROLOG er at japanarar gjer det i sine femtegenerasjons datasystem. O'Shea refererte forsi som synt at programmering i deklarativ språk gjev klårt bet resultat enn programmering i dei vanlege prosedyre-orienterte språk SMALLTALK er laga på maskinutstyr i millionklassen. Det einaste fullblods grafikk-språket, og høver godt til simulering og animasjon. Men syntaksen er kompleks («barokk») og semantikk problematisk, så det er svært vanskeleg å læra. Prof. O'Shea v imponert over å høyra at her i landet fanst 4-5 aktive SMALLTALK programmerarar - i Storbritannia hadde dei 2-3!

Alle tre språka vil sikkert utviklast vidare, men vi kan ikkje rek med å finna nokon av dei i aktiv bruk om 20 år. På spørsmål frå sal svara Tim O'Shea at LOGO truleg er språket som vil døy ut før

Dei vel 50 deltakarane på seminaret kom frå universitet, lærarhøgskular og tekniske høgskular, hovudsakleg på Vestlandet. Dei tykte vera godt nøgde med professor O'Sheas presentasjon av eit felt rivande utvikling.

Computers in English Language Research

ICAME 6th 1985, Röstånga 19-22 May 1985

Stig Johansson

During the last few years there have been yearly conferences on computers in English language research arranged within the framework of ICAME (International Computer Archive of Modern English). The sixth and latest was arranged by Professor Jan Svartvik, University of Lund, at Röstånga, a small village some 35 kilometers north of Lund. About 50 participants from 13 countries (14 if Scotland is counted separately!) enjoyed three beautiful days at Röstånga Gästgiveri at the foot of Söderåsen. There was a packed programme with papers and panel discussions.

About half of the projects were devoted to reports from major projects at Birmingham (John Sinclair, Antoinette Renouf), Lancaster (Geoffrey Leech, Susan Blackwell, Stephen Elliott, Gerry Knowles), Nijmegen (Jan Aarts, Nelleke Oostdijk), Amsterdam (Willem Meijs, Eric Akkerman, Pieter Masereeuw), and Lund (Jan Svartvik, Bengt Altenberg, Anna-Brita Stenström, Mats Eeg-Olofsson). The Birmingham collection of texts is still growing and is being used for lexical analysis. The Lancaster team continues its work on the automatic parsing of the LOB Corpus and on a context-sensitive spelling checker. A new development is the study of automatic intonation assignment (Gerry Knowles). The Nijmegen group is occupied with parsing, using an Extended Affix Grammar. In the course of the next few years the aim is to produce a million-word linguistic data base permitting a variety of ways of displaying analysed trees. The Amsterdam project aims at the development of a lexical data base and analysing system making use of the *Longman Dictionary of Contemporary English*. Current work includes the development of a morphological component which can recognise forms which are the result of the productive rules of English morphology. The new project at Lund is concerned with describing the rules that govern the natural prosodic segmentation of continuous discourse, using the London-Lund Corpus as the basis of study. Reports were given on pauses and on the segmentation into tone units.

Some of the participants reported on studies based on computer corpora. Karin Aijmer (University of Lund) gave a paper on «Conversational Phrases in *A Corpus of English Conversation*». Stig Johansson (University of Oslo) reported on the distribution of word classes in the tagged LOB Corpus as compared with the Brown Corpus (a study

undertaken in collaboration with Knut Hofland, Norwegian Computing Centre for the Humanities, Bergen). Göran Kjellmer (University of Gothenburg), in a paper on «'The Lesser Man': On the Role of Women in Modern English», spoke about the distribution of masculine and feminine pronouns in the Brown Corpus and the LOB Corpus. Douglas Biber and Ed Finegan (University of Southern California) contributed a highly interesting paper on textual similarities and differences among spoken and written text types, based on a study of the Brown Corpus, the LOB Corpus, and the London-Lund Corpus. Factor analysis was used to analyse the co-occurrence of 40 linguistic features in the corpora «identifying three underlying textual 'dimensions' ... that can be used to chart the relations among spoken and written text types in English: (Interactive versus Edited Text, Abstract versus Situated Content, and Reported versus Immediate Style). Biber and Finegan showed how «factor scores» can be used to characterize texts.

Several participants had used the computer-tape version of the *Longman Dictionary of Contemporary English* and presented reports of their work: Dirk Geens (Economische Hogeschool Limburg), Willem Meijs (University of Amsterdam), and André Moulin (Université de Liège). Other papers were: Eric Steven Atwell (University of Leeds),



Dr. Jan Svartvik, University of Lund and Dr. Nelson Francis, Brown University at the 6th ICAME conference.

«Software Tools for English Language Analysis»; Nina Devons (University of Jerusalem), «Observations on the Lexicographical Treatment of 'One' and the Approach adopted in the FREQSUCON»; Petr Sgall (Charles University, Prague), «Experiments with Machine Translation and Computational Question-Answering Systems in Prague»; Valerie Jones (University of Sterling), «A Query Language for Semantic Network Data Bases»; Henry Kucera (Brown University), «The Analysis of the English Verbal Group»; Cheng Yumin, (Fudan University, Shanghai), «An Attempt at Analysing Linguistic Style».

Wallace Chafe (University of California at Berkeley) gave a talk on «Options for the Archiving of Spoken and Written Data», including an account of spoken and written material held by members of staff at Berkeley. This provided a good starting-point for a panel discussion, planned and chaired by Gunnel Tottie (University of Uppsala), on the need for a corpus of spoken American English. John Algeo (University of Georgia) suggested that a new corpus could make use of material gathered for the *Dictionary of American Regional English* (DARE). The material includes narrative and descriptive speech as well as conversational passages. The general feeling seemed to be that it is practical to make use of already existing material. Comparability with the London-Lund Corpus is not a major priority. The question of delicacy of transcription was also touched on. Gunnel Tottie will further explore the possibilities of assembling a spoken American corpus and will investigate the practicability of different methods of transcription.

In a concluding discussion on «Habeus corpus. Quo vadis?» the participants seemed to be in favour of keeping the informal organisation of ICAME. It was decided to include summaries of the conference papers in the next issue of *ICAME News* (to appear in the spring of 1986). Preliminary suggestions for the location of the next conferences were Brown University or Oslo/Bergen (1986), Amsterdam (1987), and Helsinki (1988). The general impression is that interest in ICAME is increasing and is reaching well beyond the original «founding nations».

Stig Johansson is Professor of English Linguistics at the University of Oslo. His main fields of work include computer-assisted linguistics and automatic grammatical analysis. He is editor of Computer Corpora in English Language Research and co-author (with Knut Hofland) of Word Frequencies in British and American English.

The 1985 International Videodisc, Optical Disk and CD-ROM Conference and Exhibition

Øystein Reigem

Denne konferansen ble holdt i London 30. og 31. mai i år, med tilhørende «workshops» den 29. Arrangører var Meckler Publishing som er utgivere av tidsskriftet *Videodisc and Optical Disc*. Meckler Publishing vil arrangere slike konferanser hvert halvår framover annenhver gang i USA og England. Neste konferanse blir i Philadelphia 9.-12. desember.

Workshop'ene var svært dårlig lagt opp (referenten deltok på én i dem), men konferansen ga et mer solid inntrykk. At en del av foredragene var nokså sterkt markedsorientert, var i og for seg positivt. Hvilket tilbud det vil bli av utstyr, programvare og tjenester innen optisk plate-teknologi i årene framover, vil, enten vi ønsker det eller ikke, avhenge av markedsmekanismene. (Dette har vi nettopp sett et tydelig eksempel på når det gjelder personlige datamaskiner. Takket være sin markedsdominans generelt, førte IBMs inntreden på personlig datamaskin-markedet til at IBM-kompatible PC-er i stor grad er blitt standard.)

Utstillingen bød også på en del interessante ting, selv om en savnet et så sentral teknologi som CD-ROM.

Referenten har her valgt å gjengi innholdet av tre av foredragene på konferansen. Alle disse foredragene handler om konkrete prosjekter om anvendelser av optisk plate-teknologien. I stedet for å referere fra foredrag som omhandlet selve teknologien, hvilket utstyr som er tilgjengelig osv., er slikt stoff nedfelt i en egen artikkel annensteds på bladet («Optisk plate-teknologi»).

Det første foredraget som refereres, omhandler videoplate brukt i undervisning innen legestudiet. Selv om medisin ikke er et humanistisk fag, er såkalt interaktiv video et hjelpemiddel med stort potensiale også innen humaniora. Det andre foredraget er viet håndtering av en stor samling tegninger vha. mikrografiske medier og digital optisk plate. Mange humanistiske institusjoner forvalter store dokumentsamlinger både tegninger og skriftlige dokumenter, som kan tenkes lagret på tilsvarende måte. Det siste foredraget forteller om spesialutgivelser av spillefilm på videoplate, utgivelser som er mer rettet mot filmhistorikerne enn vanlig publikum.

Highlights of Videodisc Applications in the Medical and Health Care Sciences in the United States and Canada

I konferanseprogrammet sto som foredragsholder *Scott Stewart*, u

giver og redaktør av MedicalDisc Reporter – et spesialtidsskrift viet videoplate brukt innen området medisin. Stewart arbeider også ved Human Resources Research Organization (HumRRO), Virginia, USA. HumRRO er et ikke-kommersielt firma som produserer interaktive videoplater for undervisningsformål, bl.a. for medisinske anvendelser. Den som holdt foredraget var imidlertid *Carol Hargan* fra samme firma.

Hargan konsentrerte seg om HumRROs virksomhet, og knyttet mye av stoffet til en demonstrasjon av en videoplate med tilhørende programvare fra firmaets egen produksjon.

Temaet for platen som ble vist, er behandlingen av en pasient. Når systemet settes i gang, starter handlingen umiddelbart med følgende dramatiske scene: Pasienten, en ung mann med et stygt skuddsår i magen, blir blodig og hulkende ført inn på intensivavdelingen av to pleiere. Legestudenten som bruker platen skal så «undersøke» og «behandle» pasienten ved å velge fra menyer på en berøringsfølsom skjerm. I den filmede handlingen på platen er det en ung lege som studenten skal identifisere seg med (konvensjonelt nok spilt av en mann).

Platen er lagt opp rundt en rekke relativt korte handlingssekvenser – ca. 50 scener à 20 sekunder. (Enkelte scener er i tillegg filmet i flere versjoner for å sikre kontinuitet med det som har skjedd før.) Etter at en scene er avspilt, stopper platen, hvorefter studenten skal gi sin respons på situasjonen. Til sin rådighet har han en rekke laboratorietester, behandlingsmåter, medisiner, røntgen, osv. Hvilken ny scene som blir avspilt er avhengig av hans respons.

For å gjøre simuleringen mer realistisk, er det lagt inn en tidsdimensjon. F.eks. tar det tid å utføre de enkelte testene, og resultatene er dermed ikke øyeblikkelig tilgjengelige. En klokke viser kontinuerlig hvor lang tid som er gått. Klokken står riktignok stille hver gang studenten skal gjøre sine menyvalg, men nøler han for lenge, blir handlingen automatisk satt i gang igjen, gjerne med en innledende kommentar fra en av de andre personene i scenen. Eksempelvis vil nøling etter den innledende scenen føre til følgende kommentar fra en pleier: «Synes du ikke vi burde *undersøke* pasienten, doktor?»

Når studenten er ferdig med behandlingen (enten pasienten overlever eller ikke), får han en «eksperts» oppsummering av det hele. Oppsummeringen inneholder både faste elementer og elementer som er avhengig av det aktuelle hendelsesforløpet i behandlingen. Spesielt blir de feil studenten har gjort underveis gjennomgått.

Med det store antall responsmuligheter studenten har for hver scene på denne platen, og når handlingen i tillegg må være avhengig av hva som har skjedd tidligere, vil selvsagt ikke platen kunne dekke alle mulige veier handlingen teoretisk kan ta. Det er derfor brukt diverse metoder for å begrense handlingsmønstrene. F.eks. blir mange sannsynlige feiltagelser korrigert med en kort kommentar fra en av personene i

handlingen. Andre feiltagelser fører til en skriftlig kommentar på skjermen. Ifølge Hargan er den totale handlingsstrukturen likevel for komplisert til å tegne ut i et diagram. Systemeringen av platen og de tilhørende programmene ble derfor gjort med utgangspunkt i hver enkelt scene/meny.

Det er ikke uvanlig at strukturen i et interaktiv-video-system blir kompleks. Det er derfor forståelig at et nytt system ikke alltid fungerer tilfredsstillende når det kommer rett fra produksjonsprosessen. HumRROs erfaring var at man i enkelte tilfeller måtte ta nye videoopptak og presse platen på nytt. Oftest var det imidlertid tilstrekkelig å gjøre endringer i programvaren. Det er altså i dette tilfellet en fordel ikke å ha programvaren lagret på videoplaten (se artikkel om BBCs Domesday-prosjekt annensteds i dette bladet for videoplater *med* programvare).

I den utstyrskonfigurasjonen HumRRO bruker nå, inngår to skjerm-er – én for den filmede handlingen fra videoplaten, og én for menyer og annen tekst. Firmaet arbeider nå med å legge alt over på én skjerm, uten at Hargan vil ha sagt at det betyr noen pedagogiske fordeler i det aktuelle anvendelsesområdet.

I sin videoplateproduksjon til nå hadde HumRRO gjort alle opptak på 16 mm film og så overført til 1-toms videobånd, som er det som kreves for et masterbånd. (Dette er ikke uvanlig for å gi god kvalitet på opptakene.) Firmaet planlegger nå å gå over fra 16 mm til opptak på Sonys nye Betacam-kamera og 1/2-toms bånd, noe som ifølge Hargan gir nesten 1-toms kvalitet.

Kostnadene forbundet med utvikling av en slik videoplate som ble vist anslo Hargan til ca. \$150.000.

Robert J. Kalthoff: Automation of a Large-Scale Storage and Retrieval System for Engineering Drawings in a Major Communications Utility

Robert J. Kalthoff er president i Access Corporation, Cincinnati, Ohio, USA. Access Corporation er en av en håndfull veletablerte firma som selger/utvikler integrerte dokument-baserte systemer. De hevder å ha installert de fleste av tungvektene blant automatiserte dokument-baserte systemer i drift i dag. Dokumentene det dreier seg om kan være både tekst og tegninger.

Kalthoffs foredrag var viet en installasjon firmaet nå gjør for New York Telephone Company (her kalt NYTEL). Etter å ha blitt skilt ut fra AT&T, satt NYTEL i januar 1984 plutselig selv med ansvaret for å administrere en og en halv million tekniske tegninger på mikrokort. Et ekstra problem lå i at NYTEL i tillegg til et sentralt hovedkontor har seks underavdelinger spredt over hele staten New York. I dag besitter disse underkontorene selv over en halv million kort, hvorav noen er duplikater av kort fra hovedkontoret og noen er egne. Dette referatet vil

imidlertid ikke gå inn på de problemer som eksisterte og de løsninger som ble valgt for kommunikasjonen mellom avdelingene.

NYTEL bruker sitt tegning-materiale hyppig. Antall forespørsler ligger på over 1000 tegninger daglig, og det er ventet at antallet vil stige til det tredobbelte. I tillegg blir daglig 500 tegninger endret, og 50-100 nye lagt inn.

Til det nye systemet måtte det stilles følgende krav: Sentralisert kontroll med materialet, automatisering av kopier/utskrift, mulighet for digitalisering og ikke minst en *gradvis* overgang fra det eksisterende systemet.

I NYTELS system ville en total overgang til digitaliserte tegninger føre til en rekke problemer: Et fullstendig nytt system ville for det første kreve en meget stor omstilling av bedriftens ansatte. For det andre ville systemet måtte gå gjennom en fase med uttesting og feilfinning, der den daglige driften ble forstyrret. Uttesting og feilfinning ville bli adskillig enklere dersom utskiftingen ble begrenset til én del av systemet om gangen.

Andre problemer ville oppstå på den tekniske siden. Med digitaliserte tegninger ville f.eks. all utskrift måtte gå til laserskriver eller annen elektronisk skriver. Med mikrokortene kan kopier gjøres på vanlig fotografisk vis, og kapasiteten for produksjon av kopier er stor. En annen konsekvens av overgangen ville være at alle forespørsler til systemet ville gå gjennom datamaskinens «santidskø», uansett om forespørslene krevde umiddelbar respons eller ikke.

Kalthoff understreket at dette var et generelt problem når man konstruerte slike systemer. Flaskehalsen vil nemlig oftest komme på prosesserings- og kommunikasjonssiden. Han advarte mot å stirre seg blind på aksesstider for disketter og hastigheter for scanning av dokumenter.

Løsningen for NYTEL ble å beholde mikrokortene, og så bruke optisk og vanlig magnetisk lager som et supplement. Det ble bestemt at alle tegninger skulle eksistere i kortform, mens et varierende utvalg skulle finnes digitalisert.

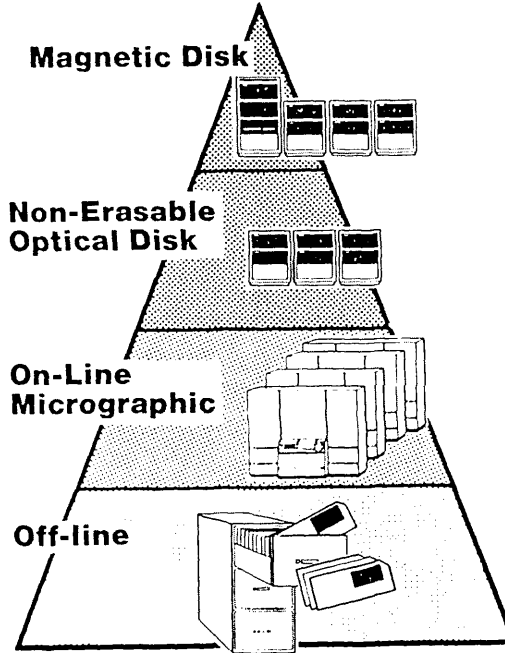
Kalthoff mente generelt at de mikrografiske media langt fra ville bli utkonkurrert av optiske plater, i det minste ikke med det første. Også andre er kommet til samme konklusjon, hevdet han. Flere og flere ser for seg lagringssystemer som *kombinerer* mikrografiske medier og digitale optiske plater. Slike systemer kan gjerne tenkes som en pyramide av medier med stigende tilgjengelighet mot toppen (se figuren). For NYTELS system (som ikke er komplett ennå) kan pyramiden tenkes slik:

- Bunn av pyramiden utgjøres av et off-line mikrografisk medium, nemlig mikrokort i arkivskap. Her fins tegninger som sjelden er i bruk. Ved behov kan disse tegningene hentes, legges i scanner, digitaliseres og legges inn på et av platemediene på toppen av

IMAGE TRANSMISSION SYSTEM

access®

Storage Capacities



- 3,720 C Size Drawings on Two 380Mb Magnetic Disks
(Maximum Capacity Over 14,000 Drawings)
- Optional Optical Disk Storage for 5,000 C Size Drawings per 1Gb Optical Disk
(Maximum Capacity 16 Disks per Juke Box)
- 640,000 Aperture Cards in Four Aperture Card Retrieval Modules
- 800,000 Aperture Cards in Off-line Storage

- pyramiden. Om ønskelig kan kort flyttes over på neste nivå ...
- ... som er samme medium *on-line*. Her er framhenting og scanningen av kortene automatisert. Kortene er plassert i spesielle framhentingsmoduler med egen scanner. I NYTELS system vil tid for framhenting og digitalisering typisk ligge på rundt 10-15 sekunder. Selve scanningen tar 2-3 sekunder. Det er da snakk om A2-tegninger med en oppløsning på 200 punkter pr. tomme (8 punkter pr. mm).
 - Nivå 3 er et digitalt optisk platelager, muligens med flere plater i en jukeboks. Dette brukes til å lagre et stort antall tegninger som en ønsker raskere tilgjengelig. Tid for lagring eller framhenting av en A2-tegning vil være på 0.5-2 sekunder, men hvis det må skiftes plate (jukeboks), vil tiden øke til 6-12 sekunder. Ved en jukeboks-løsning vil det altså lønne seg å prøve å gruppere tegningene fornuftig. Det kan dreie seg om å holde tegninger fra samme, aktive prosjekt på samme plate eller å holde de mest aktive tegninger på én plate. Uansett jukeboks eller ikke, kan man med jevne mellomrom reorganisere ved å produsere nye plater med de mest aktive tegningene.
 - På toppen av pyramiden befinner seg en vanlig magnetisk disk. Den har følgende oppgaver: Være raskt temporært lager slik at en kan få data hurtig inn og ut av systemet, håndtere forespørsler fra dag til dag, og lagre de aller mest brukte tegningene. Tegninger legges inn på en først-inn-først-ut-basis, hvor de mest aktive tegningene flyttes til toppen av køen. Tegninger som ikke blir brukt, vil dermed etter en tid bli skjøvet ut av disken. I tillegg vil en del tegninger kunne lagres mer permanent. Tid for lagring eller framhenting av en A2-tegning vil være på 0,2-1 sekund.

Når de *overskrivbare* optiske platene kommer, vil de kunne tenkes å erstatte begge de to øverste nivåene.

En vanskelighet man støtte på under utvikling av scannerdelen i NYTELS system, var at bildeplanets avstand til scannerlinsen varierte (mikrofilmen forekom montert på begge sider av kortet). En annen var at både sølvnitrat- og diazofilm var brukt. Det siste problemet måtte løses ved at scanneren dynamisk satte grenseverdier mellom det den oppfattet som svart og som hvitt.

Originalene til NYTELS tegninger varierer i format fra A4 til A0. Oppløsningen som ble valgt for digitaliseringen var 200 punkter pr. tomme for A4, A3 og A2. For A1 og A0 gikk man ned til henholdsvis 132 og 106 punkter da tegningene i disse formatene som regel hadde færre detaljer. (200 punkter pr. tomme = 8 punkter pr. mm, 132 ppt = ca. 5 ppm, 106 ppt = ca. 4 ppm.) Man fant at de digitaliserte data kunne komprimeres til ca 10%. Likevel vil en A2-tegning kreve 187,5 Kbyte lagerplass.

Selv om datakomprimering er plassøkonomisk gunstig, kan komprimering og dekomprimering være en stor belastning på en generell

datamaskin. Disse funksjonene kunne med fordel vært overført fra programvare til maskinvare. Kalthoff etterlyste en komprimeringsbrikke («-chip») som i lengre tid har vært «like rundt hjørnet». Han nevnte også at en større brikke-produsent planlegger å selge sine brikker spesialtilpasset komprimering/dekomprimering vha. av «firmware», men at produktet ennå ikke er ferdig.

Peter Crown: King Kong and Citizen Kane: Old Media, New Message

Dette foredraget brakte både gode og dårlige nyheter. Den gode nyheten var at filmklassikerne Citizen Kane (1941) og King Kong (1943) foreligger på videoplate i CAV-format. CAV-format tillater seeren å stoppe platene på hvert enkelt bilde, slow motion, hopp til et vilkårlig bilde, osv. Med plater i vanlig spillefilmformat (CLV), har man kun mulighetene for vanlig avspilling, rask scanning og å hoppe til «kapitler» (hvis platen er inndelt i sådanne). Spillefilmer utgis ellers kun i CLV-format.

Den dårlige nyheten var at platene er produsert i USA, og dermed følger amerikansk videosal-standart (NTSC). For å kunne spille platene, trenger en dermed en amerikansk videoplatespiller og en monitor som kan vise et NTSC-signal. Det siste kan man leie, men videoplatespilleren er det verre med.

Firmaet som har produsert platene heter Criterion og holder til i California. Firmaet har lagt svært mye arbeid ned i produksjonen, og *Peter Crown* (som tilsynelatende ikke var tilknyttet Criterion) mente innsatsen må ha vært «a labour of love».

Et spesielt problem som Criterion hadde lagt svært mye arbeid i å løse, var det såkalte «3-2-pulldown»-problemet. Når film, som har 24 enkeltbilder pr. sekund, skal overføres til NTSC-video, som har 30 enkeltbilder, må nye bilder lages av de gamle. Ved den vanlige måten å overføre fra film til NTSC-video på, vil to av fem enkeltbilder flimre, og det er lite tilfredsstillende for en filmhistoriker/cineast som ønsker å fryse scener. Teknikken som brukes for å unngå flimring gjør at overføringen blir svært tidkrevende.

Crown viste utdrag fra Citizen Kane-platene. Kvaliteten på filmmaterialet var forbausende god. I tillegg til selve filmen, inneholdt platene mye filmhistorisk interessant materiale. Blant dette var scriptet (med endringer), en lenge forsvunnet trailer, stillbilder av plakater og et visuelt essay om produksjonen av filmen.

King Kong- og Citizen Kane-platene er tenkt brukt i et vanlig nivå 1-system, som bare er et fint ord for en videoplatespiller med fjernkontroll. Referenten mener imidlertid at disse platene må være et ypperlig grunnlag for et mikromaskinstyrt system (dvs. nivå 3), f.eks. for undervisning i faget filmhistorie. Bruk i undervisning ville bli spesielt interessant hvis det skulle komme flere CAV-produksjoner fra Crite-

rion. Da ville det kanskje bli verdt å investere i en amerikansk videoplatespiller likevel? At vanlige videoplateprodusenter skal produsere CAV-plater er ikke å vente, da platekapasiteten synker fra én time til 36 minutter pr. side når en går over fra CLV- til CAV-plater.

Hvis noen ønsker å skrive til Criterion, er adressen: *P.O. Box 2310, Santa Monica CA 90406, USA.*

Nytt fra RHF/NAVF

Vitenskapelig utstyr til humaniora

Etter initiativ fra Rådet for humanistisk forskning (RHF) og Rådet for samfunnsvitenskapelig forskning (RSF) har NAVFs utredningsinstitutt laget utredningen «Utstyr for forskning. Situasjonen innen humaniora og samfunnsvitenskap».

Utredningen foretar en gjennomgang av utstyrssituasjonen innen humanistiske og samfunnsvitenskapelige fag. RHF drøftet utredningen på sitt møte 07.06.85. Rådet fant da at den dokumenterer et stykke på vei hvordan det i løpet av de siste 5-10 år har foregått en markert endring i bruken av vitenskapelig utstyr innen de humanistiske fag. For en rekke av fagene er spesielt edb-utstyr i ferd med å bli en nøkkelressurs for å drive avansert forskning. Tilgangen til audio- og videoutstyr er dessuten avgjørende for muligheten til å drive aktuell forskning innenfor mange fagfelt.

Utredningen påviser en klar diskrepans mellom tilgangen til og behovet for vitenskapelig utstyr innen humaniora i Norge. Samtidig som den humanistiske forskningen i et helt annet omfang enn tidligere er i ferd med å bli utstyrsbasert, står den altså i en situasjon der mangel på utstyr fører til klart påvisbare skadevirkninger.

- Manglende tilgang på utstyr styrer forskningen i retning av mindre utstyrskrevede aktiviteter. Det settes grenser for den faglige utvikling, grenser som i stor grad avgjøres av utstyrstilgangen og bare i annen rekke er bestemt av faglige kriterier.
- Forsknings samarbeid mellom forskjellige fag og institusjoner vanskeliggjøres på en rekke felter.
- Flere forskningsmiljøer blir forbigått av den internasjonale utvikling i sine fag p.g.a. manglende utstyrstilgang.
- Fagmiljøer hindres i å utføre viktige samfunnsmessige og kulturelle oppgaver som flere humanistiske fag har når det gjelder utviklingen innen telekommunikasjon, datautstyr, media og kulturminneforvaltning.

Utredningen konkluderer med å anslå et umiddelbart udekket behov for vitenskapelig utstyr til humaniora og samfunnsvitenskap samlet på

mellom 30 og 60 millioner kr. Beregningen bygger imidlertid på et noe mangelfullt grunnlag, og RHF mener det er rimelig grunn til å anta at det reelle behovet derfor er vesentlig større.

NAVFs er i løpet av 1980-årene gitt to store ekstraordinære tildelinger til vitenskapelig utstyr. Begge ganger er disse midlene i svært liten grad kommet humanistisk forskning til gode. I 1981 fikk dette fagrådet en del midler til edb-utstyr, og i 1984 fikk humaniora 0,5 mill. kr til fordeling. Da midlene for 1984 ble utlyst, kom det inn søknader for til sammen 3,3 mill. kr.

For 1985 er det gitt nok en ekstraordinær utstyrsbevilgning til NAVF. Den er på til sammen 12,5 mill. kr. Av dette har RHF fått tildelt 1 mill. Midlene er utlyst og frist for innlevering av søknad er 6. september 1985. De er ment å bidra til å bedre utstyrssituasjonen ved forskningsinstitusjonene. Det vil fortrinnsvis bli gitt bevilgning til vitenskapelig utstyr som har tilknytning til bestemte forskningsprosjekter, men det kan også søkes om generell støtte til innkjøp av utstyr. RHF vil foreta sin utdeling i oktober.

Leif T. Andressen
Konsulent, RHF

Styrking av RHF's informasjonsvirksomhet

NAVFs styre har fattet vedtak om at informasjonsarbeidet i NAVF skal legges om. Den nåværende informasjonsavd. er besluttet omorganisert slik at deler av den legges under direktørens sekretariat (Fellesadministrasjonen) samt at det opprettes en egen redaksjon for bladet «Forskningnytt». I denne forbindelse er det også besluttet at de faglige rådene skal ivareta sine egne informasjons- og formidlingsoppgaver.

RHF mener at slike oppgaver direkte knyttet til rådenes virksomhet løses best ved at ansvaret legges til hvert enkelt råd og dets administrasjon og har gitt sin tilslutning til den omorganisering som nå er skjedd. Rådet utreder for tiden sitt framtidige arbeids- og innsatsområde og vurderer i den forbindelse bl.a. nye arbeidsstrategier. Alt nå er det imidlertid klart at RHF er innstilt på å satse på en mer utadrettet profil enn tidligere, overfor de forskningsutførende institusjonene, de bevilgende myndigheter, spesielle brukergrupper og den brede allmennhet.

I denne forbindelse mener RHF at det vil bli særdeles viktig å spre informasjon om humanistisk forskning og det viktige fag- og forskningspolitiske arbeidet rådet gjør. Dette er bakgrunnen for at RHF på sitt møte 19.04.85 vedtok å foreslå for NAVFs styre at det opprettes en stilling som informasjonskonsulent tilknyttet rådets administrasjon.

RHF ønsker å satse på såvel informasjon og PR-virksomhet om rådets virkefelt og beslutninger generelt som samfunnsrettet formidling av det faglige innhold og resultater av RHF-støttede og andre prosjekter

innen humanistisk forskning.

En viktig arbeidsoppgave for den nytilsatte informasjonskonsulenten vil bl.a. bli å bygge opp et kontaktnett til dagspresse/radio/TV og andre media med interesse for forskningsstoff. RHF har dessuten planer om å starte et eget informasjonsblad. Samarbeid med NAVFs informasjonstjeneste for humanistiske forskningsprosjekter i Bergen vil også stå sentralt.

Elisabeth Johnsen
Avdelingsleder, RHF

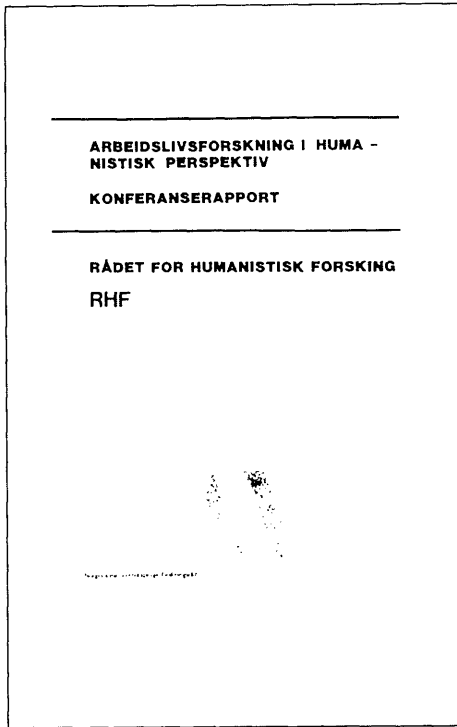
Arbeidslivsforskning i humanistisk perspektiv

Dette er tittelen på en konferanserapport fra Rådet for humanistisk forskning (RHF) som ble gitt ut i begynnelsen av 1985. Rapporten inneholder i bearbeidet form innlegg som ble holdt på en konferanse på Lindås utenfor Bergen i 1984, arrangert av prosjektet Humanistisk orientert arbeidslivsforskning. Dette prosjektet, som er satt i gang av RHF, har som målsetting å kartlegge de verdier, holdninger og opplevelser som er knyttet til arbeid og å undersøke historisk de forandringer som har foregått på området. Forandringene blir sett i lys av de endringer som ellers har skjedd i arbeidslivet, endringer i arbeidsprosesser, i maktforhold på mikro- og makroplan, når det gjelder organisasjonsmessige forhold osv. Det er konsentrert om utviklingen i Norge i vårt århundre og opererer med et relativt vidt arbeidsbegrep. Medarbeiderne i prosjektet kommer fra fagene etnologi, historie og filosofi.

Nå finnes det ingen systematisk bearbeiding av dette forskningsområdet. Det var på denne bakgrunn at prosjektet tok initiativ til konferansen. Den samlet en rekke deltagere fra humanistiske fag, men hadde også et betydelig innslag av forskere fra samfunnsvitenskapene. Konferansen hadde til hensikt å få fram hva de enkelte humanistiske fagene har vært opptatt av i forbindelse med arbeidslivsforskning og å drøfte metodiske og vitenskapsteoretiske problemer knyttet til slik forskning.

Flere av bidragene i konferanserapporten tar opp sentrale og prinsipielle spørsmål innen arbeidslivsforskningen. Arbeidsbegrepet blir nøye undersøkt (Ida Blom), likeså relasjonen mellom arbeid og kulturell identitet, stikkord: bedriftskultur (Guttorm Fløistad) og betydningen av lønnsforholdet under ulike lønssystemer (Knut Vennesslan). På det mer innholdsmessige plan blir det gått inn på arbeid og arbeidsvilkår innen jordbruket (Bjørn Fjellheim), på industri og arbeidslivshistorie med utgangspunkt i skipsbyggingsindustrien (Håkon With Andersen) og på særtrekk ved utviklingen av norsk arbeidsliv (Ståle Seierstad).

Videre bør nevnes at Rolf Braadland i sitt innlegg retter oppmerksomheten mot kunst- og kulturspørsmål i forbindelse med jernarbeidernes



historie i Bergen som for tiden er under skriving, mens Ragnhild Bjerre Finnestad trekker fram det flerkulturelle arbeidsliv og drøfter relevansen av religionshistorisk forskning i forhold til dette temaet. Jostein Gripsrud tar for seg hvorledes litteraturvitenskapen kan være egnet til å belyse sider ved arbeidslivets utvikling som ofte blir oversett. I tillegg inneholder rapporten en presentasjon av prosjektet Humanistisk orientert arbeidslivsforskning som skisserer dets forskningspolitiske sammenhenger og hovedtrekk ved forskningsområdet.

På bakgrunn av konferansen er det berettiget å hevde at det finnes en betydelig interesse for arbeidslivsspørsmål innen humanistiske fagmiljøer. Spesielt bør bredden i tilnæringsmåter framheves. Interesserte kan få tilsendt konferanserapporten ved å henvende seg til RHF eller direkte til prosjektet. Adressen til prosjektet er:

Humanistisk orientert arbeidslivsforskning
Filosofisk institutt
Universitetet i Bergen
5000 Bergen

Knut Venneslan

Førstelektor, Filosofisk institutt, Universitetet i Bergen

MELDINGER

Office of Scholarly Communication and Technology

Ved University of Leicester, England ble Office for Humanities Communication opprettet i 1982 (se HD 3-84, s. 86). Høsten 1984 etablerte The American Council of Learned Societies (ACLS) the Office of Scholarly Communication and Technology i Washington D.C. Formålet er å oppmuntre vitenskapelig personale til å bidra til aktiviteter knyttet til vitenskapelig kommunikasjon og å utvikle et program med følgende målsettinger:

- å fremme forståelse og samarbeid mellom vitenskapelig personale, forleggere, bibliotekarer, administratorer o.a. som arbeider med humanistiske fag
- å formidle informasjon om viktige endringer i det vitenskapelige kommunikasjonssystemet
- å stå for studier av kvaliteten til det vitenskapelige kommunikasjonssystemet
- å utforske hvordan teknologisk endring virker inn på måten vitenskapelig personale betrakter og utfører arbeidet sitt

Aktiviteter som finner sted ved Office of Scholarly Communication and Technology omfatter utgivelse av meldingsbladet *Scholarly Communication*; en spørreundersøkelse blant vitenskapelig personale i ulike fag vedr. synspunkter på publisering av bøker og tidsskrifter, tilgang på informasjon og bruk av ny teknologi; og en serie «workshops» ved universiteter, som samler bibliotekarer, informasjonsmedarbeidere, administratorer, edb-personale o.a. for å diskutere saker av felles interesse.

En av de første oppgavene for Office of Scholarly Communication and Technology har vært å definere nærmere feltet «vitenskapelig kommunikasjon». Et hjelpemiddel i dette arbeidet er produksjonen av en annotert bibliografi over 100 bøker og artikler innen feltet.

Adresse: *Office of Scholarly Communication and Technology, 1717 Massachusetts Ave, N.W., Washington D.C., USA.*

Se for øvrig anmeldelsen av *Recent initiatives in communication in the humanities* i HD 1-85 for flere opplysninger om vitenskapelig kommunikasjon i USA og Storbritannia.

ICAME NEWS no.9

ICAME NEWS no. 9 appeared in May. This issue contains the following articles:

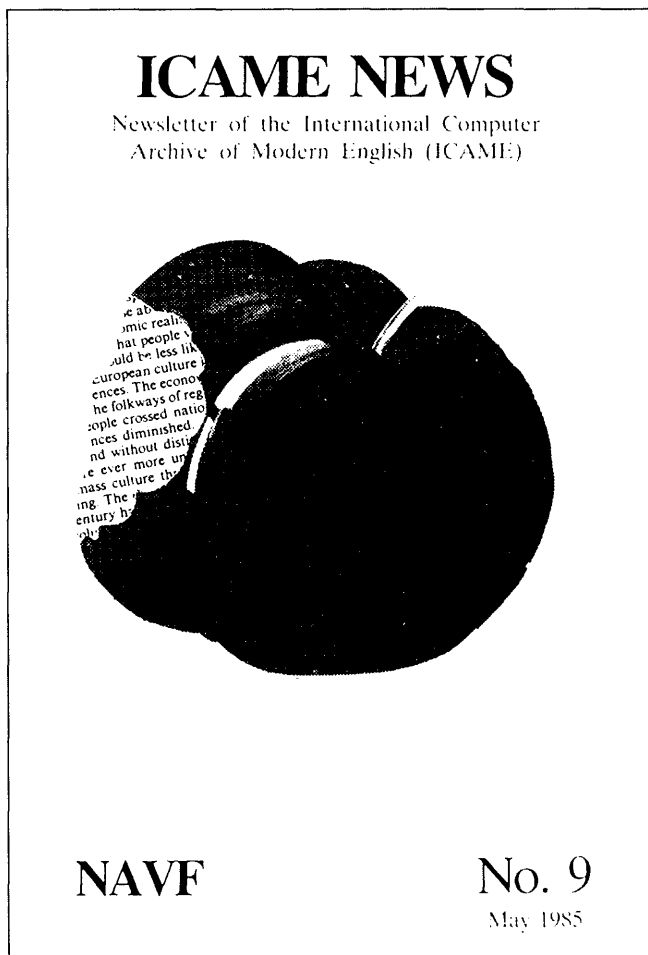
S.A. Blackwell: A Survey of Computer-Based English Language Research

B.M. Booth: Revising CLAWS

Jean-Marie Baïssus: G.E.R.A.S.

S.V. Shastri: Word Frequencies in Indian English: A Preliminary Report

ICAME NEWS is distributed free of charge by the Norwegian Computing Centre for the Humanities.



Kommunikasjonsteori og semantikk

Nytt grunnfag ved Inst. for filosofi, UiO fra høstsemesteret 1985*

Stadig nye områder av menneskelig virksomhet blir i vår tid sterkt endret som en følge av revolusjonen i moderne informasjons- og kommunikasjonsteknologi. Det er en viktig oppgave for de humanistiske fag å klargjøre perspektiver som kan gi en dypere forståelse av disse forhold. Studiet i KOMMUNIKASJONSTEORI og SEMANTIKK (heretter KS) søker å imøtekomme dette behovet fra en filosofisk synsvinkel. Det gir en grundig innføring i de områder av moderne logikk og filosofi som er mest relevante for analysen av en rekke sentrale begreper på dette feltet: beregnbarhet, informasjon, mening, kommunikasjon, intelligens, kognitive systemer.

Studiet av naturlige språk og språkbruk representerer et meget viktig forskningsfelt i forhold til utviklingen av informasjonsteknologien. Arbeid innenfor logikk, filosofisk logikk og språkfilosofi har bidratt, og kommer fortsatt til å bidra, svært mye til denne forskningen. Studenter som tar KS skal, bl.a., få en innføring i elementære formallogiske metoder, moderne formalsemantikk og talehandlings-teori. Kunstig intelligens (AI) og dens betydning for studiet av mennesket blir også et sentralt tema.

Undervisningen i KS er delt i fire kurs:

- (Høstsemesteret) KS1 Logikk og beregnbarhet
- KS3 Kommunikasjon og intensjon
- (Vårsemesteret) KS2 Semantikk
- KS4 Informasjon og erkjennelse

Studiet er lagt opp slik at det kan være av interesse for studenter fra alle fag og fakulteter, men vil kanskje være mest relevant for studenter med bakgrunn i informatikk, matematikk, psykologi, lingvistik eller filosofi.

Studenter fra Det matematisk-naturvitenskapelige fakultet og andre institusjoner som bruker en studieordning basert på vekttall, kan ta 5, 10 eller 20 (= hele grunnfaget) vekttall innenfor kurssets ramme. Tillatte kurskombinasjoner for 10 vekttall beskrives i studieplanen.)

Videre detaljer fås ved henvendelse til *Dr. Andrew J.I. Jones*, eller *inst. sekretær Ola Hole*, Inst. for filosofi, Niels Treschows hus, 9. etg., P.b. 1024 Blindern, 0315 Oslo 3. Tel. (02) 456963.

Det forutsettes Departementets godkjenning.

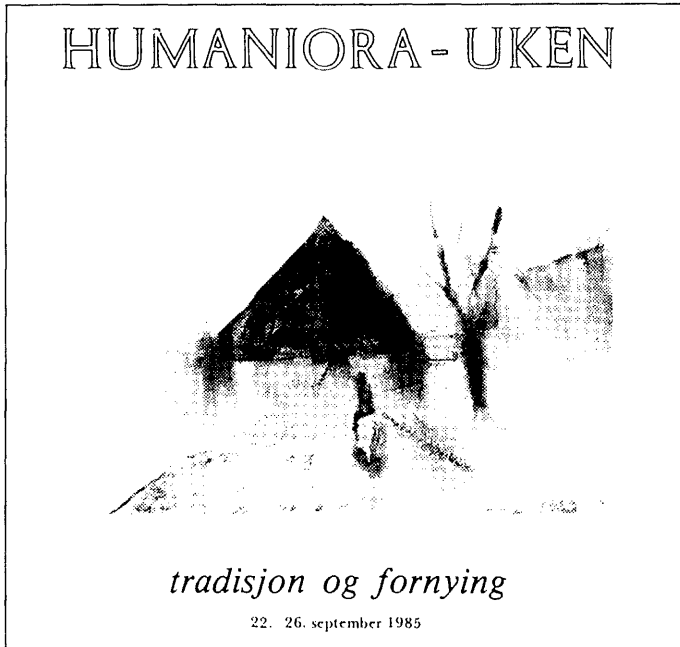
Humaniora-uke ved Universitetet i Bergen

Det historisk-filosofiske fakultet ved Universitetet i Bergen skal arrangere en «humaniora-uke» for første gang. Hovedtema for uken, som skal finne sted 22.-26. september, er «Tradisjon og fornying». Emnet for en av sesjonene er humaniora i datalderen. I denne sesjonen vil følgende foredrag bli holdt:

Edb i humaniora, humaniora i edb. Jan Oldervoll, Historisk institutt
Maskinen og ordene. Lars Otto Grundt, Romansk institutt
Maskinen og språkopplæring. Magnar Brekke, Engelsk institutt
Maskinen og språkforståelse. Helge Dyvik, Institutt for fonetikk og lingvistikk.

Det skal i tillegg arrangeres et edb-verksted hvor edb-anvendelser i de humanistiske fagene blir demonstrert. Blant disse er talesyntese, data-lingvistikk og læreprogrammer i engelsk. NAVFs EDB-senter for humanistisk forskning skal bidra med demonstrasjoner av videoplate-teknologi, informasjonssøking og edb-metoder for tekstrekonstruksjon.

Programheftet for humaniora-uken kan fås ved henvendelse til: *Ulf Lie, Engelsk institutt, Universitetet i Bergen, Sydnesplass 9, 5000 Bergen.*



Aktuelle tidsskrifter

SPECTRA

Meldingsbladet SPECTRA utgis kvartalsvis av Museum Computer Network, Inc., en internasjonal organisasjon av museer som har som formål å forbedre museumsdokumentasjon og adgangen til den ved hjelp av moderne informasjonsteknologi. SPECTRA inneholder meldinger og korte rapporter og fås gratis ved henvendelse til: *Museum Computer Network, Inc., Center for Contemporary Arts and Letters, ECC, Building 26, State University of New York at Stony Brook, Stony Brook, New York 11794, USA.*

Archaeological Computing Newsletter

Første nummer av det nye meldingsbladet Archaeological Computing Newsletter utkom i desember 1984. Bladet koster £2 pr. år for fire nummer. Adresse: *Archaeological Computing Newsletter, Department of Computing, North Staffs Polytechnic, Blackheath Lane, Stafford ST18 0AD, England.*

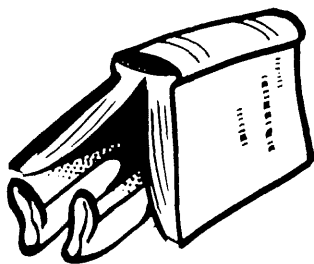
Praktisk Lingvistik

Praktisk Lingvistik er en skriftserie utgitt ved Institutionen för Lingvistik, Lunds universitet. Innholdet i nr. 8-83 består av to artikler om COMMENTATOR, et mikromaskinbasert forskningsredskap for lingvister. Interesserte kan henvende seg til: *Lunds universitet, Institutionen för lingvistik, Helgonabacken 12, 223 62 Lund, Sverige.*

UNESCO ALSSED-LSP Newsletter

LSP Centre (LSP = Language for Specific Purposes) ved Handelshøjskolen i København utgir et meldingsblad to ganger pr. år. Meldingsbladet, som er flerspråklig, inneholder rapporter både om pågående LSP-forskning og fra konferanser/seminarer, bokanmeldelser, konferanseinnbydelser og småmeldinger. En del av stoffet omhandler edb-metoder. Adresse: *UNESCO ALSSED-LSP Network and Newsletter, The LSP Centre, Handelshøjskolen i København, Fabrikvej 7, DK-2000 København F., Danmark.*

Nytt i biblioteket



Språk

- Akkerman, E., Masereeuw, P., Mejs, W.: Designing a computerized lexicon for linguistic purposes. Amsterdam, 1985. 80 s.
- Fjeldvig, Tove/Golden, Anne: Automatisk rotlematisering - et lingvistisk hjelpemiddel for tekstsøking. NORIS (64) og NORIS (58). Universitetsforlaget, 1984. 99 s.
- Källgren, Gunnel: Automatisk excerpering av substantiv ur löpande text. Ett möjligt hjälpmedel vid datoriserad indexering. Stockholm, 1984. 22 s.
- Nishida, Toyooki: Studies on the application of formal semantics to English-Japanese machine translation. Kyoto University, 1983. 255 s.
- Norevik, Bjarne (red.): Datamaskinen og språket. Universitetsforlaget, 1985. 190 s.

Edb-emner

- Allan, Boris: LOGO - en innføring. Universitetsforlaget, 1984. 111 s.
- Austvik, Lars m.fl.: VIIN - Videodisk og informasjonshåndtering. Trondheim, 1984. 97 s.
- Goldberg, Adele: Smalltalk. The interactive programming environment. Addison-Wesley, 1984. 516 s.

Edb i undervisning

- Landa, Ruth Kaplan: Creating courseware. A beginner's guide. New York, 1984. 295 s.
- Self, John: Microcomputers in education. A critical evaluation of educational software. The Harvester Press, 1985. 184 s.

Diverse

- Bunch, Allen: The basics of information work. London, 1984. 159 s.
- Behovet for datautdanning. NAVFs utredningsinstitut, Oslo 1985. 142 s.
- Jones, Kevin P. (ed.): Informatics 7. Cambridge 1983. Intelligent information retrieval. Proceedings. London, 1983. 149 s.

- Knappskog, Karl: Automatisering av gjenfinningssystem for avisartikler. Bergen, 1983. 181 s.
- Knappskog, Karl/Syvvertsen, Trine: Pressedatabankar – arbeidsverktøy, salsvare eller offentlig gode? Bergen, 1983. 41 s.
- Laan, van der A./Winters, A.A. (eds.): The use of information in a changing world. 42nd FID Congress. Hague 1984. Proceedings. Amsterdam, 1984. 469 s.
- Online Information, 8th international meeting. London 1984. Oxford. 555 s.
- Revue informatique et statistique dans les sciences humaines. 19 année 1983 nos 1 à 4. Universitetet i Liège, 1985. 251 s.
- Rich, Elaine: Artificial intelligence. McGraw-Hill, 1983. 436 s.
- Syvvertsen, Trine/Knappskog, Karl: Avisarkiv, yrkesroller og ny teknologi. Bergen, 1983. 219 s.

Konferanser

The First International Expert Systems Conference

Denne konferansen arrangeres av Learned Information (Europe), Ltd i London 1.-3. oktober 1985. Konferansen blir fire-delt: plenumsesjoner, produktpresentasjoner, seminarer og en utstilling. Én av konferansens målsettinger er å fungere som en brobygger mellom kunstig intelligens-teori og brukerne av denne teoriens sluttprodukter. I foredragene vil det bli lagt vekt på praktisk bruk av ekspertsystemer og potensielle anvendelser.

Flere opplysninger kan fås fra: *Learned Information (Europe), Ltd, Besselsleigh Road, Abingdon, Oxford OX13 6LG, England.*

New Technologies in Training

Den andre internasjonale konferansen om «New Technologies in Training» skal finne sted i London 2.-4. oktober 1985. Konferansen vil bli delt opp i plenumsforedrag og sesjoner for deltakere med forhåndskunnskaper på ulike nivåer. Blant emnene som vil bli tatt opp, er interaktiv læring, kunstig intelligens og ekspertsystemer. Det blir også arrangert produktdemonstrasjoner.

Flere opplysninger fra: *Lynn Brook, Conference Organiser, Queensdale Exhibitions & Conferences, Blenheim House, 137 Blenheim Crescent, London W11 2EQ, England.*

Translating and the Computer

Aslib, the Aslib Technical Translation Group og the Translators' Guild arrangerer den 7. konferansen om edb og oversettelse i London 14.-15. november 1985. Programmet skal bestå av foredrag holdt av representanter fra USA, Øst-Europa og EF-landene. En utstilling skal arrangeres i tilknytning til konferansen.

For flere opplysninger, skriv til: *Sherry Jespersen, Manager, Professional Development Group, Aslib, The Association for Information Management, Information House, 26-27 Boswell Street, London WC1N 3JZ, England.*

9th International Online Information Meeting

Denne konferansen skal holdes i London 3.-5. desember 1985. Programmet er ikke fastsatt ennå, men blant foredragsemnene blir å finne: elektronisk publisering, programpakker, utvikling av brukervennlige systemer, databaser, informasjonssøking, brukerevaluering og elektroniske bibliotek. Det skal dessuten arrangeres seminar om spesielle emner, produktpresentasjoner og en utstilling av programvare.

Flere opplysninger fra: *Learned Information (Europe), Ltd, Besselsleigh Road, Abingdon, Oxford OX13 6LG, England.*

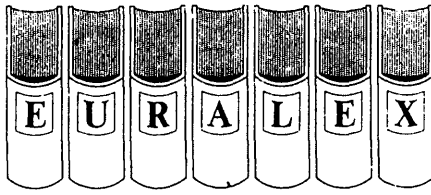
ICAMCI 2

The Second International Conference on the Application of Micro-Computers in Information, Documentation and Libraries skal finne sted i Baden-Baden 17.-21. mars 1986. Konferansen skal skape et forum for internasjonal utveksling av erfaringer på området. Det skal også legges vekt på en diskusjon av tendenser i utviklingen og virkningene av mikromaskinapplikasjoner både i globale, regionale og lokale nettverk og som egne systemer. Noen av hovedtemaene blir beskrivelser av eksisterende systemer, krav til programvare, applikasjoner i bibliotek og informasjonssentre, spesielle problemer i utviklingsland og mikromaskiner i elektronisk publisering. Det skal også arrangeres symposier om spesielle emner og en utstilling av operative systemer.

Flere opplysninger fra: *Secretariat, Deutsche Gesellschaft für Dokumentation e.V. (DGD), Westendstrasse 19, D-6000 Frankfurt am Main 1, Federal Republic of Germany.*

EURALEX Second International Conference

Denne konferansen skal finne sted i Zürich 9.-14. september 1986. Konferansens siktemål er å skape et forum for alle som arbeider med ordboksproduksjon. Foredrag ønskes om alle aspekter av leksikografi. Et av temaene det skal legges spesiell vekt på, er datamaskinens rolle i leksikografi.



EUROPEAN ASSOCIATION FOR LEXICOGRAPHY

Konferansens struktur blir symposier i plenum etterfulgt av parallelle sesjoner som skal bestå av spesialiserte foredrag sammen med enten diskusjoner eller workshops.

Sammendrag av foredrag (800-1000 ord) ønskes tilsendt innen 1. november 1985 til: *PD Mary Snell-Hornby, Dekanat der Philosophischen, Fakultät I (EURALEX-Tagung), Universität Zürich, Rämistrasse 71, CH-8006 Zürich, Switzerland.*

SUMMARY

Optisk plate-teknologi

Optical disc technology

In this article Senior Computing Officer Øystein Reigem of the Centre gives an overview of optical disc technology at present. Analog and digital optical discs are described and compared. Detailed information is given on read-only, write-once and erasable discs. The three «levels» of interactive videodisc systems are described, as are the different videodisc standards and the two systems VHD and Laservision. A comparison is made between Laservision's two formats, CLV and CAV, and between domestic and professional players for this system. Ways of combining videodisc players with microcomputers are outlined along with special programming languages for interactive video.

Information is given on how to produce one's own videodiscs according to type of disc. The advantages, disadvantages and potential of the different types are outlined, including the two read-only digital optical disc systems CD-ROM and OROM.

In this connection write-once digital optical discs are dealt with in detail, also concerning picture capacity, scanning of pictures and the combination digital discs/micrographic media.

Finally, the present status and the future of optical storage is discussed.

Er datalingsvistik del av anvendt språkvitenskap?

Is computational linguistics part of applied linguistics?

Research Fellow Lars Sigfred Evensen, University of Trondheim, discusses the role of computational linguistics in ANLA's (The Norwegian Association of Applied Linguistics) future work. In studying the historical development of the term «applied linguistics» it is possible to distinguish between two approaches – one is theory-driven and the other problem-driven. The theory-driven approach uses established linguistics for practical purposes. In the problem-driven approach, practical language problems are the starting-point for interdisciplinary research. In this approach, whose case Evensen argues, linguistics is a necessary but not sufficient condition for problem solution.

A distinction must be made between theoretical and applied computational linguistics. The most important applied task lies in the discrepancy between very advanced hardware and not so advanced software. Logical formalisms may be too rigid for tomorrow's high-level programming. Natural language, with its *underdetermined meaning* and context-dependence, offers a much more flexible potential. The exploitation of this potential, however, requires a better understanding of pragmatics and text linguistics.

The development of software for interactive video in language teaching is another important task in applied computational linguistics. Successful R & D here requires the interaction of linguistics, informatics, film science and pedagogics.

Prosjekt i automatisk indøksering – lingvistiske metoder

Project in automatic indexing – linguistic methods

The Centre and the Norwegian Research Centre for Computers and Law, University of Oslo, have for a long time worked on, respectively, design and use of information retrieval systems and research in information retrieval. Both institutions have seen the need for improving information retrieval, by e.g. automatic indexing.

The two institutions formulated a joint project aiming at studying simple linguistic methods that can be applied to automatic indexing. The project's co-workers – Tove Fjeldvig, Anne Golden and Øystein Reigem – wanted to concentrate on three methods: automatic grouping of words with joint roots (root lemmatization), automatic splitting of compound words and automatic recognition of parts of speech. They chose to base their methods on a compact set of grammatical rules, rather than on the method of dictionary look-up.

The project group received in 1984 a grant from NORDINFO (The Nordic Council for Scientific Information and Research Libraries) for preliminary studies of automatic splitting of compound words and for establishing contacts with relevant projects within the field of libraries, documentation and information. The splitting of compound words is dealt with separately in connection with this article. As for the latter, an urgent need has been uncovered for aids to information retrieval, particularly of a linguistic nature. These include assistance in formulating search arguments and tools that can ease the preparation or raise the quality of material intended for data bases.

The project group gives a run-down of the disadvantages of existing information retrieval systems and presents an outline of an extended system. The most important feature of their system is the incorporation of a higher level of communication between the user and the system.

This system also encompasses the preparation phase of information to be retrieved – processing prior to entry, automatic indexing and storage. These three steps are described in detail, along with the four processes that constitute retrieval according to this system: formulation of queries, retrieval, presentation of the information retrieved, and automatic reformulation of queries.

On the basis of the outlined information retrieval system the project group points out a number of possible areas of research.

Anne Golden outlines the work the project group has carried out on automatic splitting of compound words. This includes studying literature on the subject and investigation of alternative methods. The group regards recognition and splitting of compound words as two separate tasks and have set up procedures for each of them. These procedures are described in detail. The group has been awarded a grant from NORDINFO for further work in this field in 1985/86.

Edb-baserte arkivkataloger: Samkatalog for privatarkiver

Computer-based archive catalogues: Joint catalogue for private archives

In 1980 the Norwegian Institute of Private Records started a four-year project aiming at the creation of a computer-based joint catalogue of private archives in public and private possession. This project is being carried out in co-operation with the Centre, which will publish the catalogue in the form of microfiches.

Archivist Elisabeth Koren gives a description of this catalogue and the work that has taken place on it so far. The catalogue consists of three parts: a main catalogue, special inventories of archives and letter inventories. The main catalogue is divided into two: a catalogue section, and a register section (containing names, topographical information, categories and subject terms) that serves as an entrance into the catalogue.

An account is also given of the basis of the data entry of archive information and how the registers are constructed. Archive information is contributed by each institution on special forms developed by the Institute.

The principles behind the establishment of the category and subject registers are a result of the computer processing of the material itself, which among other things, has made employment of cross-references possible.

Edb og arkeologi

ADP and archaeology

In July a pilot project was started in Oslo that aims at adapting computer-based surveying techniques to archaeological field work. This is the first phase of a comprehensive project which is described in detail by MA student Britt Kroepelien.

The project can be divided into two main areas: field documentation and archaeological analyses. It is planned as a joint venture between the University of Oslo and a firm of engineering consultants, which will design an interactive data base (for geographical information and characterizing information on the finds). The final goal of the project is an archaeological program package, consisting of modules for recording, storage and presentation of archaeological data.

The recording of data will be executed via surveying and field registration of characterizing data. Existing computer programs can calculate coordinates via the surveying techniques. A geographical data base will contain types of find, coordinates for the finds and necessary references to the characterizing data base. This will consist of descriptions of each find and cultural layers. Interactive editing of the data bases will be possible. Data can be presented in the form of maps, profiles, perspective sketches, statistics and video/photos.

DUS-forfatter-systemet: ComCats

The DUS authoring system: ComCats

In this article Erik Meistrup gives an account of the Danish computer-assisted training system DUS. This system allows teachers at all levels to create their own teaching programs on microcomputers without extensive knowledge of computer programming. It is based on the special high-level programming language DUS, which consists of about 100 instructions, almost all in Danish.

The system is designed as a dynamic and open pedagogical structure. The programming language aims at creating maximum pedagogical control over the technology. The most important work connected with program development takes place outside of the computer, i.e. the setting of pedagogical goals, program structure development and screen design.

DUS consists of a system for teachers and another for students. Meistrup describes the principles of both of these systems in detail.

BBCs Domesday-prosjekt

The BBC's Domesday Project

In June Senior Computing Officer Øystein Reigem of the Centre visited the BBC's offices in London, where Michael Tibbets gave him a thorough account of the BBC's Domesday project.

In order to mark the 900th anniversary of the Domesday Survey of England, the BBC has started the production of a new version on videodisc. Access to the information will take place via microcomputer. This project is probably the biggest and most ambitious of its kind. Most of the information is being collected by school pupils and members of organizations – one million people in all. The resulting two videodiscs will contain data equivalent to two million text-TV screen pages, 20,000 pages of maps and close to 12,000 pictures.

In order to carry out this project, the BBC has established its own commercially run company. Philips is investing £1 million in the development of a new videodisc player for the discs.

The Domesday Project's two videodiscs are a *local* disc, consisting of information on Great Britain on a local basis, and a *national* disc, which will contain information from important national data sources – the basis for various statistical surveys – and an introduction to the use of the discs. Both discs will contain indices and software for manipulating the material. Reigem gives a detailed account of the contents, structure and use of these discs.

In the Domesday project screen presentation is based on three elements: single video pictures and live video sequences, teletext, and graphics. Reigem discusses the technology problems raised by these forms of presentation and the software and hardware that is being developed in order to overcome the difficulties.

Samisk database

Lappish data base

At the University Library in Trondheim an on-line Lappish data base has been established that contains 2000 bibliographical units. University Librarian Anders Løøv gives a detailed account, including examples, of this data base. Information is given on input format, communication and search words.

The computer entry of this material is also intended as an aid to the production of printed bibliographies on Lappish subjects. A bibliography of literature published in Norway 1966-1980 appeared in 1984.

An important future task is to establish routines for exchanging data with Swedish, Finnish and Soviet libraries and a system for entry of this data into the data base.

Nytt om Registreringsentral for historiske data, Tromsø

News from the Norwegian Historical Data Archives, Tromsø

Director Jostein H. Hauge of the Centre comments on recent developments at the Norwegian Historical Data Archives (RHD). RHD's main task is to prepare and data process historical sources (in particular censuses and church registers) from the whole country containing information on people who lived in the 18th and 19th centuries. RHD services (local) historians, demographers, social science scholars, name researchers and other interested parties.

Projects initiated in connection with RHD's work include pedagogical use of source materials and record linkage (automatic identification of the same individuals in different historical sources).

At RHD's annual meeting in May a presentation was given of plans for establishing similar archives in other parts of Norway. At the meeting a discussion also took place of the principles for source preparation that RHD should follow in coming years. Another discussion was held on the pros and cons regarding the use of original versus copied source material at the data preparation stage.

«High Technology and Computers in Education»

An exhibition, «High Technology and Computers in Education», held in London's Barbican Centre in January, was attended by University Lecturer Lise Opdahl, The English Department, University of Bergen. The exhibition was the first of its kind in Britain, and the word «education» was applied to the use of computers in the broadest sense, covering areas from games for 2-year-olds to university courses. The exhibition gave a relatively good overview of the present state of the art for hardware and commercial software. The visitor's main interest was software for education in the humanities, but the development in this field seems still to be in its infancy. The general impression was that while computers are moving into the 5th generation, educational software for the humanities remains in the first generation. There is thus ample scope for initiative in this field.

ANTEM III, 12.-15. mars 1985

ANTEM - «New Technologies Applied to Education, Professional Training and Culture» - is one of 18 projects within the international program «Technology, Growth, Employment». Senior Computing Officer Knut Hofland of the Centre took part in the third ANTEM conference in Paris, which was divided into three main sessions: New Technologies and Training in the Tertiary Sector, Pedagogical Applications of New Communication Technologies, and Pedagogical Applications of the Image.

Hofland reports on some of the talks given at the conference. These dealt with: an interactive videodisc training program on the use and maintenance of bank automats, an expert system built into an information system to adapt training to each user's needs, a micro-network for schools, telesoftware for teaching and training, a videodisc for training dentists, a do-it-yourself videotex system for schools, interactive video via a cable network, SAVANT 2: a video

service for the training of telecommunications engineering students, a system for high quality three dimensional synthetic imaging, the British Domesday project, and interactive video in the training of the deaf.

In connection with the conference an exhibition was shown of pedagogical software and hardware, including videodisc technology and a program for children from 2 up.

Humanioradagene 1985 ved Universitetet i Oslo

The «Humanities Week» 1985 at the University of Oslo

The title of this year's «Humanities Week» at the University of Oslo was «Humanists in the Data Age». University Lecturer Tellef Kvifte reports that six lectures/seminars were held in addition to 14 local arrangements at various departments. Part of the aim of the arrangement was to heighten humanists' awareness of how the Faculty of Arts can contribute to the understanding and development of new technology.

The lectures given dealt with the following subjects: «adventure games», interactive books, ADP training in schools, computational linguistics, man-machine communication, data processing of source materials, ADP in museums of art and art history, and ADP in name research.

Microcomputers, software and books on computing were exhibited. Demonstrations included literature searching via ADP, computing equipment for the blind and ADP-based aids for the drawing of musical notes.

In Kvifte's view, the arrangement demonstrated a wide range of and an increasing interest in computing activities at the Faculty of Arts.

Computers and Art History Workshop

This workshop was arranged in April at University College, London for 60 participants, mainly from Great Britain but also from other European countries.

Computing Officer Espen Ore of the Centre reports that the first day started with lectures and discussions on using already existing software rather than developing special programs. Ore gives details of these lectures.

On the second day the participants split up into groups in order to discuss special subjects: data entry and organization, statistics for art historians, and image processing and storage. Demonstrations of programs were also given.

The latter group visited The Remote Sensing Unit of the Geography Department, University College, where processing of satellite pictures was demonstrated. Afterwards a discussion took place on how the demonstrated methods can be used by art historians in restoration work.

The workshop ended with a plenary session where an outline was given of the status quo of computing in art history in Poland and future initiatives were discussed. As a result of the latter, a newsletter on computing in art history is to be started (see HD 1-85). Attempts will be made to arrange one-day seminars twice a year and an annual international workshop.

Utviklingsseminaret på Ustaoset 17.-19. april 1985

«Developmental seminar» at Ustaoset 17-19 April 1985

The theme of this year's «Developmental seminar» was expert systems. Former Computing Officer Espen Brodin of the Centre reports that the seminar was attended by 30 people and was based on a combination of lectures, demonstrations and exercises.

The program package K-man, the combination LISP/LOOPS and PROLOG were presented. Both LISP and PROLOG are well-suited to the analysis of natural and artificial languages, and can be of enormous help to linguists as well as in constructing expert systems in other humanistic subjects. A demonstration was given on how this can be achieved in the field of art history.

At the seminar it was suggested that the Centre continue work on expert systems. Since then preliminary plans have been made for a report on the application of knowledge-based systems in the humanities.

Seminar om pedagogisk informatikk

Seminar on pedagogical informatics

Stord Teachers' Training College has recently started a course in pedagogical informatics. In April Professor Tim O'Shea from the Institute for Educational Technology, The Open University, England was invited to speak on this subject. Dr. O'Shea held two triple lectures: «The Future of Computer Assisted Learning» and «Artificial Intelligence in Computer Assisted Learning». Senior Computing Officer Per Vestbøstad gives an account of the main points in these lectures.

In Dr. O'Shea's view, computer programs should only be used in education if they are demonstrably better than books. He pointed out the advantages of using a computer, the most important being that it can present and simulate processes and not merely facts. He also drew attention to the disadvantages of today's computer programs and how these can be remedied.

In addition, Dr. O'Shea discussed the question of how to choose a programming language. Factors that should be evaluated are syntax, semantics, modes of operation, documentation and cultural origins. Finally, thorough analyses of the programming languages LOGO, PROLOG and SMALLTALK were presented.

The 1985 International Videodisc, Optical Disk and CD-ROM Conference and Exhibition

This conference was arranged in London in May by Meckler Publishing, publishers of the journal *Videodisc and Optical Disc*. Senior Computing Officer Øystein Reigem of the Centre gives summaries of three of the speeches made at the conference on concrete projects. These dealt with: an interactive videodisc used in the training of medical students; a system for handling a large collection of drawings with the aid of micrographic media and digital optical discs; and special editions of film classics on videodisc. Reigem points out the relevance of the first two projects to humanistic disciplines.

The next conference of this kind will be arranged in Philadelphia in December.

Nytt fra RHF/NAVF

News from the Council for Research in the Humanities

The Institute for Studies in Research and Higher Education has published a report on research equipment in the humanities and social sciences. According to this report computing equipment has in particular become a key resource for carrying out research. However, there is a clear discrepancy between access to and the need for scientific equipment within the humanities. In order to help remedy this situation, the Council for Research in the Humanities has been awarded an extra grant of 1 mill. Nkr for 1985.

The Council for Research in the Humanities has applied for the establishment of a new position as Information Officer. An important task for a new Information Officer will be to establish contacts with the media. The Council also wishes to start an information bulletin. Cooperation will take place between NAVF's Information service for Humanistic Research Projects in Bergen.

The Council for Research in the Humanities has published a conference report on work life research in a humanistic perspective. The conference was arranged at Lindås near Bergen in 1984 by the project «Humanistic orientated work life research», initiated by the Council. The goal of this project is to chart values, attitudes and experiences connected with work and to investigate historically the changes that have taken place in this area. The aim of the conference was to uncover what each humanistic discipline has been concerned with in connection with work life research and to discuss methodical and theoretical problems connected with research of this kind.

Meldinger

News

In the autumn of 1984 the American Council of Learned Societies established the Office of Scholarly Communication and Technology in Washington D.C. One of the Office's aims is to encourage the participation of scholars in activities related to scholarly communication. The Office publishes the bulletin *Scholarly Communication* and has started the production of an annotated bibliography of 100 books and articles in the field.

A new course - «Communication Theory and Semantics» - will be offered at the Department of Philosophy, University of Oslo from this autumn. The course will give a thorough introduction to the areas of modern logics and philosophy that are the most relevant to the analysis of a number of central terms that will be dealt with in the course: calculability, information, meaning, communication, intelligence and cognitive systems. Students will also be instructed in elementary formal logic methods, modern formal semantics, speech act theory and artificial intelligence.

In September the University of Bergen will arrange a «Humanities Week» for the first time. The theme of one of the sessions will be the humanities in the data

age. This session will include a number of lectures on and demonstrations of applications of computers in the humanities. The Centre will contribute with demonstrations of videodisc technology, information retrieval and computing methods for text reconstruction.

Several English-language journals are presented. *UNESCO ALSED-LSP Newsletter* is published twice annually and contains material on computing methods. The newsletter *SPECTRA* is published quarterly by Museum Computer Network, Inc. The first issue of the quarterly *Archaeological Computing Newsletter* appeared in December 1984.

Forthcoming conferences:

The First International Expert Systems Conference - London, 1-3 October

New Technologies in Training - London, 2-4 October

Translating and the Computer - London, 14-15 November

9th International Online Information Meeting - London, 3-5 December

The Second International Conference on the Application of Micro-Computers in Information, Documentation and Libraries - Baden-Baden, 17-21 March 1986

EURALEX Second International Conference - Zürich, 9-14 September 1986.

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.

Av innholdet:

Øystein Reigem: Optisk plateteknologi

Lars Sigfred Evensen: Er datalingvistikk del av
anvendt språkvitenskap?

Tove Fjeldvig, Anne Golden og Øystein Reigem:
Prosjekt i automatisk indeksering – lingvis-
tiske metoder

Elisabeth Koren: Edb-baserte arkivkataloger:
Samkatalog for privatarkiv

Britt Kroepelien: Edb og arkeologi

Erik Meistrup: DUS-forfatter-systemet: Com-
Cats

Returadresse:

NAVFs EDB-senter for humanistisk forskning

Boks 53