

humanistiske data 2-84



NAVF

ARTIKLER
RAPPORTER
MELDINGER
SUMMARY

NAVFs EDB-senter
for humanistisk
forskning

The Norwegian
Computing Centre
for the Humanities

SENTERETS RAPPORTSERIE

- RAPPORT nr. 1. *EDB i gjenstandsfagene*. Rapport fra en konferanse i Bergen, 18. og 19. april 1978. September 1978. 2. opptrykk november 1981. ISBN 82-7283-022-1 Pris kr. 40.
- RAPPORT nr. 2. *Et norsk datamaskinelt tekstkorpus*. Rapport fra en konferanse i Bergen, 19. og 20. oktober 1978. Februar 1979. 2. opptrykk 1981. ISBN 82-7283-016-7 Pris kr. 20.
- RAPPORT nr. 3. *Rapport fra den nasjonale konferanse om EDB i språk og litteraturforskning*, 4. og 5. januar 1979. Mars 1979. 2. opptrykk november 1981. ISBN 82-7283-024-8 Pris kr. 50.
- RAPPORT nr. 4. *Oppbygging av EDB-katalog for folkemuseu i Hordaland og kulturgeografisk registrering på Vestlandet*. April 1978. 3. opptrykk november 1981. ISBN 82-7283-000-0 Pris kr. 30.
- RAPPORT nr. 5. *Rapport fra NKKM's EDB-komite*. August 1979. ISBN 82-7283-001-9 Pris kr. 15.
- RAPPORT nr. 6. *Prøveprosjekt med EDB ved Norsk Folkemuseum*. Oktober 1979. ISBN 82-7283-002-7 Pris kr. 15.
- RAPPORT nr. 7. *Ivar Fønnes: Norsk landbruksordbok*. Prosjektreport om databehandling og tilrettelegging for trykking. September 1979. ISBN 82-7283-008-6 Pris kr. 25.
- RAPPORT nr. 8. *SEFRAK. Rapport frå prøveprosjekt for databehandling av kulturminneregisteret*. Oktober 1979. ISBN 82-7283-003-5 Pris kr. 30.
- RAPPORT nr. 9. *Jostein H. Hauge og Stebjørn Århus. Dataregistrering i humanistiske fag med vekt på optisk lesing*. August 1978. 3. opptrykk januar 1981. ISBN 82-7283-004-3 Utsolgt.
- RAPPORT nr. 10. *Roadl Skarsten: Innføring i SPSS for humanister*. November 1977. 4. opptrykk november 1981. ISBN 82-7283-005-1 Pris kr. 30.
- RAPPORT nr. 11. *Jostein H. Hauge og Knut Hoffland: Rapport fra 4 konferanser i US4 sommeren 1979*. The 17th Annual Meeting of Computational Linguistics. La Jolla Conference on Cognitive Science. The Fourth International Conference on Computers in the Humanities. Data Bases in the Humanities and Social Sciences. November 1979. ISBN 82-7283-0 07-8 Utsolgt.
- RAPPORT nr. 12. *EDB og manuskriptregistraturer*. Oktober 1977. 2. opptrykk november 1979. ISBN 82-7283-009-4 Pris kr. 20.
- 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: Søkning. 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.

Foris. 3. omslagsside.

humanistiske data 2-84

NAVFs EDB-senter for
humanistisk forskning

The Norwegian Computing
Centre for the Humanities

NAVF NORGES
ALMENVITENSKAPELIGE
FORSKNINGSRÅD

NAVFs EDB-senter for humanistisk forskning ble opprettet av Norges almenvitenskapelige 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 (33 er utkommet pr. 15.8.84).

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. 50,- for institusjoner).

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

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

Artikler, rapporter, meldinger mottas. Redaksjonen avsluttet 30. august.

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

The journal can be ordered from the address mentioned above. Contributions are welcome.

Medarbeidere fra Senteret i dette nummer:

Jostein H. Hauge, Knut Hofland, Ole Lauvskar, Kristin Natvig, Espen Ore, Øystein Reigem, Per Vestbostad.

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

Forsida: Et tibetansk tresnitt med bilde av Siddha Thang-stong rgyal-po (1385-1464). Han ble berømt som bronsesmed og oppfinner av den tibetanske kjedebrua. (Fra Detlef Ingo Lauf: Das Erbe Tibets.)

En ny metode for automatisk alfabetisering av transkribert tibetansk er blitt utviklet i Norge. Se s. 17.

Innhold

Artikler

Digitalisering av bilder. <i>Jan Wiig</i>	s. 4
Automatisk alfabetisering av transkribert tibetansk. <i>Espen Ore</i>	s. 17
Pedagogisk programvare for datastøttet undervisning. <i>Lars Vavik</i> . s.	22
PLATO og datamaskinassistert læring. <i>Rune Midtvedt</i>	s. 37
Teknikk og undervisning. Intervju med AV-sjef Lars S. Wilhelmsen, UiB. <i>Jostein H. Hauge</i>	s. 43

Rapporter

Edb-prosjekt for kunst- og kulturhistoriske museer. <i>Jon Birger Østby</i>	s. 49
Forskerstipend i automatisert arkivinformasjon. <i>Hege Brit Randsborg</i>	s. 52
Frå informasjon til kulturarv. <i>Jostein H. Hauge</i>	s. 56
Prosjektet «Norsk litteratur i Ungarn». <i>Anikó N. Balogh</i>	s. 62
Tutorial on Machine Translation. <i>Jostein H. Hauge</i>	s. 64
ALLC Louvain-la-Neuve, 2.-6. april. <i>Ole Lauvskar</i>	s. 72
Conference on Computers in English Language Research. <i>Knut Hofland</i>	s. 75
Studiereise til England 21.-25. mai 1984. <i>Per Vestbøstad</i>	s. 78
Museum Documentation Association. <i>Jon Birger Østby</i>	s. 81
Norsk Arkivseminar, Fana Folkehøgskole, 24.-26. mai 1984. <i>Anne Hals og Hege Brit Randsborg</i>	s. 84
COLING 84. <i>Knut Hofland</i>	s. 90
Nytt fra RHF/NAVF.	s. 93
Fra bokhylla.	s. 95

Meldinger	s. 96
-----------------	-------

Summary	s. 102
---------------	--------

Digitalisering av bilder

Jan Wiig

Hensikten med denne artikkelen er å henlede oppmerksomheten på en helt ny metode til å oppbevare innholdet i bilder. Jeg kommer til å bruke en del uttrykk som vil være ukjente for noen av leserne, men vel kjente for andre. For at flest mulig skal ha utbytte av artikkelen, har jeg derfor måttet forklare en del av disse uttrykkene, og ber derfor om at alle som kjenner dem fra før, hopper over de linjene.

Museer og arkiver har i mange år samlet på fotografier. Dette er samlinger av stor kulturell verdi. Hvor mange bilder det kan dreie seg om er svært vanskelig å si, men at det er mer enn én million er helt sikkert. Noe annet som også er helt sikkert, er at originalene vil forsvinne med tiden.

Bildene vi har oppbevart, er fremstilt på litt forskjellig vis. Det kan være (i rekkefølge etter oppfinnelsen av prosessen):

- Daguerreotypier på sølvplater
- Kollodium negativ på glass (våtplater)
- Tørrplate negativ på glass
- Negativ film på nitratbase
- Negativ film på acetatbase
- Positiv tørrplate på glass (for lysbilder)
- Positiv film på nitratbase (for lysbilder)
- Positiv film på acetatbase (for lysbilder)
- Farve negativ film på acetatbase
- Farve positiv film på acetatbase
- Positive sort/hvit bilder på papir
- Positive farvebilder på papir

Innenfor disse gruppene finnes det mange avarter. Det er også en del bilder som faller utenfor denne grupperingen, men det er ikke mange.

Hva gjøres i dag med denne billedmassen?

Ved de aller fleste billedsamlinger gjøres det ingen ting. Bildene blir bare liggende slik de lå da de kom inn. Dette er den sikreste og billigste måte å ødelegge samlingen på. En god del har imidlertid fornyet arkivmateriellet, dvs. lagt bildene i nye syrefrie konvolutter, anskaffet skikkelige skap, og sørget for et bedre klima for oppbevaringen. Et fåtall museer oppbevarer også sine farvebilder i fryseskap.

Et ganske lite antall bilder (i forhold til de som eksisterer), er fotografisk overført til nytt materiale. Det er også utviklet metoder til å overføre den fotografiske emulsjonen, altså selve bildet, til en ny base. Det er særlig aktuelt for negativer på nitratbase. De bildene som er mest

i faresonen, er først og fremst alle farvebilder, enten de er på film eller papir. Der er det selve bildet som blekner. Dernest er det stor fare for sort/hvit negativer på nitratbase. Der er det basen som er selvdestruerende.

Muligheten for oppbevaring for fremtiden

Det vil neppe være økonomisk mulig å avfotografere alle bildene i våre billedsamlinger. For det første er det meget tidkrevende, og for det annet vil selve materialet bli meget kostbart. Med den fotografiske teknikk vi kjenner i dag, vil vi derfor måtte se i øynene at en stor del av disse bildene vil forsvinne uten at vi kan gjøre noe ved det. Det er her spørsmålet om en helt ny teknikk kommer inn, nemlig digitalisering av bildene.

Prinsippene for digitalisering

Digitalisering vil si at vi gir noe en verdi uttrykt i form av tall. I stedet for å si at en stor haug poteter er like mye verd som en mindre haug gulerøtter sier vi at potetene koster kr 2.30 pr. kg, og gulerøttene kr 6.20 pr. kg. Vi har altså fått verdien over på en vel definerbar tallform.

På den nye laser-grammofonplaten («compact disc») er musikken digitalisert. Her har man delt inn musikken i svært små intervaller og gitt hvert intervall en verdi uttrykt ved et tall. Disse tallene er så lagret på platen. De kan oppfattes av avspilleren, som kjenner til hva disse tallene står for, og den kan omgjøre dem til musikk igjen. Fordelen med denne metode er at såfremt tallene kan leses, vil musikken bli gjengitt eksakt likt fra gang til gang. Vi kan nesten si det slik at enten virker platen helt, eller så virker den ikke i det hele tatt.

De gamle platene, eller rettere de som er i vanlig bruk i dag, er analogplater. Med det mener vi at musikken er lagret analogt. Originalen til platene er gravert med en stift som svinger i takt med musikken. Spilles det kraftig, svinger nålen kraftig ut osv. Her er det muligheter for mange feilkilder. Enten klarer ikke nålen som graverer platen å følge med, eller nålen som spiller av platen klarer det ikke. Resultatet blir forvrengning, musikken gjengis ikke slik som den opprinnelig ble spilt. Under bruk blir platen slitt ved at nålen stadig subber langs kanten på rillen. Etterhvert blir rillene flatet ut og gjengivelsen blir mer og mer forskjellig fra originalen. En datamaskin (de nye digitale spillerne er i virkeligheten datamaskiner) har ikke noe større problem med å huske 87.689 enn 13.574. Den vil altså gjengi begge verdier like godt. Heller ikke går det an å slite ned 87.689 til f. eks. 87.234.

Et vanlig fotografi er en analog registrering av virkeligheten. Til en viss intensitet i lyset svarer en bestemt svartning av filmen. En enorm mengde detaljer forsvinner allerede under opptaket. For s/hv vedkommende gjelder det bl.a. alle fargene. En rekke detaljer forsvinner fordi objektivet ikke er fullkomment. Heller ikke er filmen i stand til å

registrere alt helt riktig. Disse feilene vil bli større og større for hver gang vi kopierer bildet. Gjør vi det tilstrekkelig mange ganger, vil vi til slutt få et bilde som kan bli vanskelig å tyde.

Ved digitalisering av bilder unngår man helt forvrengning av billedkvaliteten ved kopiering gjennom mange generasjoner. Den siste generasjon vil bli helt lik den først digitaliserte, fordi tallene vil være de samme. Ved opptak av videosignaler har man helt til det siste brukt analoge signaler. Kameraet registrerer luminans (lysstyrke) og farvesammensetningen av et TV-bilde ved å dele inn bildet i 575 linjer og hver linje i 775 punkter. Det vil altså si at et TV-bilde består av 445.625 punkter. I de senere år har man gått over til å digitalisere bildesignalet visse steder underveis. Hvert enkelt punkt har blitt gitt en bestemt verdi, tilsvarende punktets luminans og farge (luminans og RGB-verdi¹). Dette signalet kunne vi tenke oss å mate inn i en datamaskin, og så lagre det der. Når vi så ville spille av programmet igjen, måtte vi kunne mate det inn i systemet på samme sted som kameraet gjør det, og på denne måte få tak i programmet igjen. Problemet er bare at det er så fantastisk store mengder med data som må lagres at alle kjente datamaskinlagre ville bli fulle ganske snart. Bare på et sekund blir det lagret 25 helbilder². Vi må altså lagre informasjon om 11 millioner punkter hvert sekund.

Fullt så stort vil ikke problemet være for lagring av enkle bilder etter det samme systemet. Det er derfor allerede konstruert en del slike apparater. Jeg skal derfor kort gjennomgå en del av det som er på markedet i dag.

Quantel DLS 6000

Dette er et digitalt bibliotekssystem for slides. Det er spesielt fremstilt for bruk i forbindelse med TV-sendinger. Systemet kan lagre bilder fra slides, eller det kan ta ut et enkelt bilde fra et film- eller videobåndprogram. Bildet lagres digitalt på en Winchester-disk (dvs. vanlig magnetplatelager), maken til den som sitter i litt større datamaskinanslegg. Alle bildene får et navn og en beskrivelse. De kan kalles opp ved hjelp av navnet eller et hvilket som helst ord i beskrivelsen. Hvis man f.eks. har et bilde av den britiske dronning som døper en ubåt og teksten er følgende: «UK dronning Elizabeth II døper en ny ubåt i Newcastle 22. januar 1984», vil maskinen finne dette bildet selv om man ikke husker hvem det var, bare ved å skrive ordet «ubåt». Ligger det andre bilder med dette ordet i teksten, vil maskinen også finne frem disse bildene. Bildene gis også et fortløpende nummer, og man kan også kalle opp et bilde dersom man kjenner nummeret. I tillegg til dette har utstyret et lite tastatur for diverse faste bilder som f.eks. «Beklager teknisk feil». Det er fullt mulig å manipulere bildet elektronisk. Bildet kan ikke forstørres, men det kan forminskes og eventuelt settes oppe i et hjørne. NRK har et slikt system i bruk daglig. NRKs system kan lagre

330 bilder.

IRIS II

Dette systemet ligner i prinsippet meget på Quantel DLS 6000. Seks brukere kan samtidig arbeide med bildene som er lagret i systemet. Antall mulige bilder er avhengig av antall disketter som er koblet til systemet. Med 3 disketter kan i alt 1125 bilder lagres.

Andre typer utstyr er Slide File fra Rank Cintel og Vist fra Logica. Alle disse er i prinsippet nokså like, og gjør stort sett samme jobben. Felles for dem er også at de er beregnet på TV-sending og ikke lagring av bilder for fremtiden.

Lagring av dokumentarbilder i arkiver og museer

Så vidt jeg har kunnet bringe på det rene, er det ikke laget noe utstyr som dekker vårt behov. Det vil si det er ikke noe utstyr som er komponert sammen. Men det eksisterer enkeltdele som riktig satt sammen langt på vei vil kunne dekke vårt behov. La oss først og fremst se på hva som er vårt behov, og så skissere et ideelt apparat for dette, innenfor de teknologiske og økonomiske muligheter vi har i dag eller øyner i den nærmeste fremtid.

Det vi trenger er følgende:

God oppløsning av bildet.

God farvegjengivelse av farvebilder.

Mulighet for farvebalansekontroll.

Mulighet for kontrastkontroll.

Mulighet for luminanskontroll.

Store lagringsmuligheter.

Enkel metode til å føye inn bilder i systemet.

Sikring mot at bilder blir slettet.

Lett kopierbart lagringsmedium.

Enkel metode til å finne igjen bilder.

Mulighet til å overføre innleste bilder til et annet lagringsmedium når det blir oppfunnet.

Overkommelig pris.

Jeg skal forsøke å gå gjennom disse punktene ett for ett.

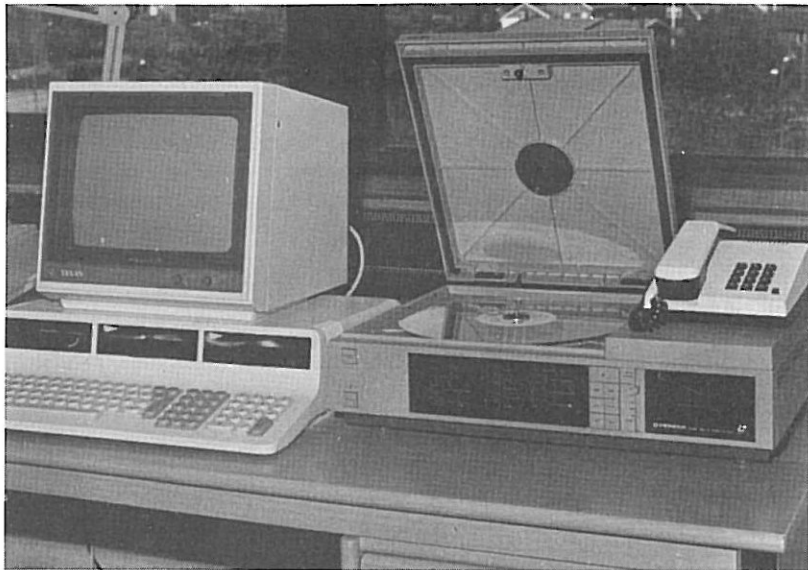
Oppløsning av bildet

Alle de eksisterende systemer er bygget opp til å holde den vanlige kringkastingsstandard. Denne kan variere litt fra land til land. De fleste europeiske landene bruker PAL-systemet. Dette er også en standard vi bør legge oss på. Det vil med andre ord si at vi får en oppløsning som helt tilsvarer det som vi kan få på et vanlig TV-apparat av god kvalitet. (I virkeligheten får vi den kvaliteten bildet har når det forlater NRK). Bildet består av det samme antall punkter som TV-skjermen, altså

445.625. Det er likegyldig hvor stort originalbildet er. Tapet i detaljrikdommen vil derfor være størst ved et stort originalbilde. Det vil være tilnærmet riktig å si at uansett originalbildets format vil detaljrikdommen i det ferdig digitaliserte bildet bli den samme.

Det er selvfølgelig teknisk mulig å øke systemets oppløsningsevne i betydelig grad, men vi må være klar over at dersom vi for eksempel ville øke dette til det dobbelte, vil det nødvendigvis ta dobbelt så stor plass på lagringsmediet. En gang i fremtiden vil det sikkert bli gjort. Allerede digitaliserte bilder må beholdes slik de er. En oppdatering vil ikke være mulig.

Ved innlesningen er det derfor viktig at bare den del av bildet som er interessant blir tatt med. Jeg skal være den første til å innrømme at det kan være meget vanskelig for en operatør å bedømme hva som er interessant i andre fotografers materiale, særlig dersom vedkommende ikke kjenner faggruppen motivet er hentet fra. Imidlertid fins det også mange tilfelle hvor en slik avmasking kan gjøres uten videre. Det vil neppe forringe bildets dokumentarverdi om man snauer litt av skyene,



Til venstre en mikromaskin (KONTIKI) for søking i billedopplysninger. Til høyre en hjemmevideoplatespiller (Pioneer). Dataskjermen brukes også til å vise bildene fra platen. Det er laget et eget kretskort til mikromaskinen for å kunne ta videosignalet fra spilleren. Kortet kan dessverre ikke sende signaler til spilleren, så bildenummer må taster inn på denne.

eller tar vekk noen gresstuser i forgrunnen. Papirbilder vil derimot som oftest være riktig beskåret av fotografen.

Forholdet mellom sidene i et TV-bilde er 3:4. Det vil si at dersom man har et 6x9 bilde, stemmer dette ikke. Her må man da enten velge avmasking, som altså gir bedre detaljrikdom i den delen av bildet som er i behold, eller man velger å ta med hele bildet med en marg oppe og nede. Det samme problemet får man ved 6x6. Da er det spørsmål om å kutte i topp eller bunn, eller å ta med marg langs kortsidene.

Margene som ikke inneholder noen informasjon, kan pakkes betydelig tettere i lagringsmediet så de ikke skal oppta så stor plass. Det samme kan faktisk skje med lite detaljrike deler av bildet. En blå himmel uten detaljer kan pakkes meget tett. Det betyr igjen at det er plass til flere bilder på samme lagringsmedium.

God farvegjengivelse av farvebilder, farvebalansekontroll

Dette er ikke lenger noe stort problem. Moderne scannere er meget gode til å registrere fargene. Dessuten er muligheten for å rette opp farvebalansen meget god. Det gjelder selvfølgelig dersom det er feil farve over hele motivet. Også her kan det innvendes at man ikke skal manipulere med motivet, men på dette punkt kan jeg ikke være enig. Det er som



Ingrid Wiig på Tiki-Data demonstrerer det nye systemet. Disken inneholder 100.000 bilder. Utstyret er analogt, og dermed bare beregnet på gjennomsyn.

oftest ikke vanskelig for en øvet fagmann å se om det er farvestikk (dvs. overvekt av én farge) i et bilde. Svært ofte vil det i et motiv være klare indikasjoner på stikket. Dersom personen på bildet er blå i ansiktet, er det neppe trolig at det var virkeligheten. Fiolett gress er heller ikke vanlig. Stort sett er det slik at dersom det grå er grått er bildet riktig når det gjelder farvebalansen.

En helt annen sak er at et farvebilde bare i begrenset grad kan brukes til å dokumentere farver. Den største fordel ved å bruke farvebilder ved dokumentasjon er at de enkelte deler av motivet skiller seg bedre ut. Naturligvis ligger det også en verdi i dette at fargene dokumenteres, men som sagt, farvenyansene dokumenteres ikke. I det hele tatt er vi ikke i stand til å dokumentere farvenyanser i dag. Med den teknikk som her er i ferd med å bli beskrevet, vil det kanskje en gang i en ikke altfor fjern fremtid kunne gjøres.

Mulighet for kontrastkontroll

Vanligvis skjer en økning i kontrasten ved reproduksjon av bilder. Denne kontrastøkningen kan helt motvirkes ved elektronisk overføring. Det er til og med mulig å heve eller holde tilbake visse deler av bildet slik at kontrasten øker eller minker, men man kan aldri få bildet så godt som et helt riktig eksponert og riktig fremkalt fotografisk bilde kan bli. Med andre ord, et feilaktig eksponert eller fremkalt bilde kan forbedres. Et riktig eksponert og fremkalt bilde blir ikke bedre.

Mulighet for luminanskontroll/korreksjon av transparens

Også på dette området er det mulighet for å rette opp et bilde. Dersom originalen er altfor tett (overeksponert), kan dette langt på vei avhjelpes. Betingelsen er selvfølgelig at det er tegninger i bildet. Er originalen så tett at detaljer er helt forsvunnet, kan de selvsagt ikke hentes opp igjen. Noe lignende er det med undereksponerte bilder. Dersom undereksponeringen ikke er for stor, kan bildene rettes opp elektronisk.

Store lagringsmuligheter

For at denne metoden skal ha noe for seg, må lageret være meget stort, og her støter vi på de største problemene i dag. Denne siden av saken er ikke godt nok utviklet ennå. Ingen av de nåværende produsenter av slikt utstyr har signalisert at de har tenkt på større lager enn 10.000 bilder. De eksisterende lagringsmedia som til enhver tid kan være on line med en slik kapasitet vil være meget kostbare (ca. kr. 150.000). 10.000 bilder vil i vår sammenheng si et meget lite antall. Bildene kan også lagres digitalt på vanlige U-matic videokassetter (U-matic er et mer profesjonelt system enn hjemmevideosystemene Beta og VHS). På en slik kassett kan man lese inn ca. 700 bilder. Ved kjøp av et større antall vil prisen ligge på under 100 kr. per stykk. Til inn- og utlesning kreves

bare en vanlig U-matic videomaskin. NB: Dette må ikke forlede noen til å tro at man kan få bildene ut på en vanlig videoskjerm ved å spille båndet av på vanlig måte. Det bildet man da får vil være helt umulig å tolke. For at det digitaliserte bilde skal bli et vanlig bilde igjen, må signalene tilbake til en digital-til-analog omformer. Deretter kan disse signalene sendes til en skjerm. En slik omformer er i dag meget dyr (ca. kr. 300.000).

Imidlertid er det tydelige signaler på det generelle datamaskinmarked om at det vil komme disketter med meget stor lagringskapasitet. Dette er de nye laserdiskene («digital optisk disk»). Hvor stor kapasitet de vil få vet man ikke nå, men det antydes svært store lagringskapasiteter. Laserdiskene er foreløpig tenkt brukt til permanente lagere. Philips har allerede et system for lagring av store mengder data på laserdisker. Det vil ikke være fysisk mulig å fjerne noe fra en slik disk. Derimot er det hele tiden mulig å gjøre tilføyelser helt til disken er full. Som man vil se er dette nærmest et ideelt lagringsmedium for vårt formål.

Enkel metode til å føye til bilder i systemet

Teoretisk sett er det meget enkelt å føye nye bilder til systemet. Det eneste man behøver å gjøre, er å laste bildet inn, og dermed er det på plass. I praksis er det ikke så lett. Som tidligere antydnet, vil det ikke være mulig at alle museer i Norge har hvert sitt innlesningsanlegg. Det vil falle altfor dyrt. Ønsker derfor et lite museum å føye 10 bilder til sin samling, må bildene sendes inn til en sentral institusjon sammen med museets lagringsmedium for å få dem oppdatert. En slik stadig oppdatering vil sannsynligvis bli både kostbar og tidkrevende. Det vil derfor være fornuftig å sende inn materiale først når det ble et visst kvantum. Et helt annet forhold er at man helt sikkert vil måtte arbeide med systemet i noen år før det bevaringsverdige materiale som allerede eksisterer, er behandlet. I mellomtiden får museene selv oppbevare sine nye opptak så godt det lar seg gjøre. Heldigvis forsvinner de jo ikke så brått. Så hurtig som utviklingen går på dette område, er det realistisk å tenke seg at dette vil bli løst før det blir et problem.

Lett kopierbart lagringsmedium

Her ligger det utvilsomt meget å vinne. Å kopiere fra disk til disk går meget hurtig. Å kopiere 1000 bilder vil bare dreie seg om få minutter. Skal man overføre fra disk til U-matic kassett vil det ta 1 time å overføre alle de informasjonene som en 1-times kassett kan lagre, ganske enkelt fordi U-matic maskinen går med vanlig hastighet. Å kopiere 700 bilder tar altså en time. Det sier seg selv at det er ikke nødvendig å passe maskinen mens den arbeider. Hvor lang tid det vil ta å kopiere en laserdisk med mange tusen bilder har jeg ikke kunnet fastslå.

Sikring mot at bildene blir slettet

Det er en ny type laserdisker like om hjørnet, og på disse diskene skjer lagringen ved at det er krystallstrukturen i materialet som blir forandret. Dette er en ikke-reversibel prosess. Ved bruk av slike diskrisikerer man altså ikke at bildene blir slettet, med mindre hele disken fysisk blir ødelagt, eller forsvinner.

På de andre lagringsmedia vil det være mulig å slette bildene. Riktignok kan det legges inn rutiner som skal forhindre dette, men helt sikker på at en feil begås skal man ikke være. Det vil derfor være aktuelt å ta én eller to sikkerhetskopier. Med den teknikk som er kjent i dag, vil sannsynligvis en overføring til magnetbånd være det mest aktuelle. Slik vil det ikke bli i fremtiden. Bildene på disse kopiene vil i alle tilfelle være helt identiske med bildene som først ble overført.

Muligheter til å overføre digitaliserte bilder til annet lagringsmedium

I den nærmeste fremtid vil det helt sikkert bli utviklet fullstendig nye lagringsmedier. Det kan derfor være nærliggende å si at da bør vi heller vente med å gjøre noe til disse er ferdig utviklet. Innenfor dette feltet er en slik tanke umulig. Dersom man hele tiden ønsker det aller siste på området, må man vente til alt er for sent. Det aktuelle spørsmål er derfor: Er det mulig å overføre allerede digitaliserte bilder til et eventuelt nytt medium som ennå ikke er oppfunnet? Svaret er ganske enkelt JA. Dersom det en gang skulle bli utviklet et lagringsmedium som ikke kan motta data fra eksisterende datamaskinutstyr, vil det neppe kunne selges, og da er jo en avgjørende faktor for at det hele skal bli suksess bortfalt. Også ved denne overføringen vil alle detaljene som var til stede ved første gangs digitalisering, forstatt være i behold. Det er jo det som er hele poenget med å oppbevare bildene i digitalisert form.

Kan bildene lett finnes igjen?

Dette er selvfølgelig et særdeles viktig spørsmål. Svaret på dette avhenger av hvilket lagringsmedium som brukes. Ligger bildene på disk, er det meget lett å finne igjen et bilde. Bildene kan nummereres fortløpende av hvert enkelt museum. Kjenner man nummeret, er det bare å be maskinen hente frem det aktuelle bildet.

Men det fins en vesentlig mer elegant metode. Til alle bildene hører det en billedtekst. Denne teksten ligger lagret sammen med bildet. Man kan søke i teksten etter de ord som ligger i den. Det vil være fornuftig å starte teksten med hovedinnholdet og så spesifisere nedover. Det vil lette arbeidet med å lage et godt program for gjenfinning av bildet. Her er et eksempel: Rammesag, tømmerdagende, brukt av van Severen & Co, Namsos. Fremstilt av J. & A. Jensen og Dahl 1932 (Myrens Verksted). Dersom vi ønsker å se på alle bilder som fins av rammesager

kan vi søke på ordet «rammesag». Da vil vi få opp alle bilder hvor teksten inneholder ordet «rammesag» uansett hvor de måtte være. Dersom vi er usikre på om det heter J. og A. Jensen og Dahl, J. & A. Jensen & Dahl, J. & A. Jensen og Dal osv. kan vi søke på Jensen ?? Da?!. Maskinen finner da frem alle bilder hvor dette navnet forekommer i teksten. Denne metode er viktig fordi det like gjerne kan være den som har skrevet inn teksten som har gjort feilskriften som den som søker.

På de større anleggene, som f.eks. det som NRK bruker, kan det vises 16 små bilder på skjermen samtidig. Man kan da bare plukke blant disse 16 det bildet man ønsker å bruke. Er man ikke fornøyd med noen av dem, er det bare å be maskinen om å gå videre med nye 16 bilder.

Dersom bildene er lagret på tape, er det straks mer tidkrevende. Sannsynligvis vil det lønne seg å dumpe teksten ned på et eget lagringsmedium som brukes bare til søkefunksjonen. Teksten tar meget liten plass i forhold til bildet. På en vanlig diskett som er omtrent så stor som en single grammofonplate kan det lagres nesten 1.000.000 bokstaver og tall. En slik plate koster kr. 30. Hvis man i gjennomsnitt skriver fire linjer om hvert bilde, får man plass til 3300 tekster på en slik diskett. Dette er det enkleste og billigste lagringsmedium. Ellers fins det nå disketter som har 640 Mbyte, dvs. det er plass til ca. 2.000.000 billedtekster på disken. Da begynner det jo å hjelpe. Fra denne disken kan vi så få opplysning om på hvilket bånd bildet ligger. De 2 millionene tekster må selvfølgelig ikke skrives inn manuelt på stordisketten. Tekstene overføres fra originalmaterialet maskinelt.

Hvordan løse problemene i praksis

Kostnadene med et anlegg som kan digitalisere og lagre bilder, vil bli meget store. Anslagsvis må vi regne med en initialkostnad på 3-5 millioner kr. I tillegg kommer så driftsutgiftene som lønninger og lokale. Til gjengjeld vil prisen pr. bilde bli meget liten. Dersom det skulle vise seg mulig å samle tilstrekkelig interesse for prosjektet, må vi regne med at det vil gå minimum 5 år 'til det kan realiseres. I mellomtiden er det temmelig sikkert at de nødvendige komponenter for lagring av så store datamengder som det her er snakk om, er utviklet.

Det kan med andre ord bare bli snakk om ett anlegg i Norge. Dette anlegget vil bestå av et farve-TV kamera med spesialutstyr for næropp-tak, montert i en optisk benk, og dessuten det spesielle digitaliseringsutstyret som kan lese bildene inn på lagringsmediet. Dette må kombineres med et tekstbehandlingsutstyr som tar seg av billedtekstene. Så langt er utstyret allerede utviklet. Det må også finnes mulighet for overføring til sikkerhetskopier.

Originalmaterialet sendes så inn til sentralen. Det vil være rimelig at man tok for seg institusjon for institusjon slik at materialet kan bli mest mulig samlet. Dersom sentralen kan gå gjennom dette materialet i løpet

av 5-10 år, må vi anse oss fornøyd. Da kan man gå løs på det som er produsert i mellomtiden og det som i første omgang ikke ble ansett som truet. Det er helt klart at det er fullt mulig å klare dette på vesentlig kortere tid dersom prosessen automatiseres, men det vil sikkert gå sterkt ut over kvaliteten.

Når bildene er ferdige, kan de sendes tilbake til eierinstitusjonen, eller de kan oppbevares av sentralen, etter eierens ønske.

Fra de digitaliserte bildene kan det så fremstilles påsiktskopier som eierinstitusjonene mottar. Dette må gjøres analogt, ellers vil utstyret bli altfor dyrt. Til dette kan man bruke en vanlig videomaskin. På et vanlig 3-timers videobånd vil det være plass til 1-2000 bilder. Ved siden av dette trengs en hjemmedatamaskin som kan styre videomaskinen slik at man kan finne igjen bildene. Teksten kan ligge både på videomaskinen og datamaskinen, eller den kan bare ligge på datamaskinen. Hvert bilde gis et fortløpende nummer. Når vi søker i teksten, kan maskinen komme med forslag til hvilket bilde som er aktuelt. Når man legger den riktige kassetten i videomaskinen, vil datamaskinen finne frem til riktig bilde. Dette er utstyr som allerede er utviklet. Et slikt anlegg vil ha en pris på ca. 10.000 kroner (med diskett, ikke disk), særlig dersom det blir kjøpt inn en del anlegg samtidig.

Det er også kommet en laserplate som lagrer TV-programmer analogt («videoplate», «analog videoplate»). Hensikten med disse platene er å selge TV-programmer til publikum. Initialomkostningene ved disse platene er ca. 25.000 kroner. Hver kopi vil bare koste et par hundre kroner. Platene vil kunne lagre ca. 100.000 bilder. De er meget holdbare, men kan ikke oppdateres, da de presses akkurat som grammofonplater. Dersom man velger dette systemet, kan i praksis alle museer og arkiver ha alle andre museers og arkivers bilder. Platene bør også kunne selges, f.eks. til forlag o.l. som til stadighet er på jakt etter bilder.

Kvaliteten på disse bildene vil ikke være god nok for reproduksjon. Dessuten kreves det spesialutstyr for å gjøre det. For en enkelt presentasjonskopi kan man selvfølgelig ta et Polaroidbilde av skjermen. Når man vil fremstille bilder til reproduksjon, er det tanken at man henvender seg til sentralen. Der har man utstyr for å lage kopier av det digitaliserte bildet direkte. Enda mer interessant er det at de nye scannerne som reproduksjonsanstaltene bruker, har en inngang for standard RGB-signal, altså den samme standard som sentralen sannsynligvis vil bruke. Det betyr igjen at det digitaliserte bildet kan overføres pr. telefon til reproanstalten, uten at det blir noen kvalitetsforringelse. Bildet overføres digitalt og omgjøres til analog form først etter at det er mottatt. Scanneren lager så separasjonsfilmene direkte.

Sammendrag

Ved digitalisering av bilder er det mulig å oppbevare bildene slik at de

ikke gjennomgår noen forandring ved stadig kopiering. Ved selve digitaliseringsprosessen blir bildene delt opp i ca. en halv million punkter uansett størrelsen på bildet. Dette vil være en merkbar kvalitetsforringing av et riktig eksponert bilde av stort format. Likevel er det å foretrekke fremfor ikke å ha bildet. Initialkostnadene ved prosessen vil være store, prisen pr. bilde svært liten sammenlignet med overføring til nytt fotografisk materiale. Digitaliseringen kan bare foregå ved en sentral som dekker hele landet. Hver enkelt institusjon kan skaffe seg rimelig utstyr for påsikt av bildene. Kopiering av lagringsmedium for slikt utstyr gjøres helt automatisk og vil være meget rimelig. Påsiktstutstyret kan også brukes i forbindelse med gjenfinning av originalmateriale som ennå er intakt. Teksten kan oppbevares sammen med bildene. Fremstilling av sikkerhetskopier vil kunne gjøres helautomatisk til små kostnader.

Det som her er skissert, vil nok endres en del underveis før en eventuell sentral kommer i drift. Min hensikt har bare vært å vekke oppmerksomheten for en interessant metode for bevaring av fotografier.

Noter

1. RGB = Rødt Grønt Blått. Fargene i et videosignal er satt sammen av grunnfargene rødt, grønt og blått. Skjermen i et farve-TV har linjer med punkter som kan gjengi disse tre fargene.
2. Man snakker gjerne om «helbilder» og «halvbilder». Hvert enkeltbilde i et TV-program kalles et helbilde, og lages på TV-skjermen ved at en elektronstråle sveiper over de 575 linjene og får punktene til å lyse opp. Av tekniske grunner foregår sveipingene i to halvbilder. Første halvbilde består av linje 1, 3, 5, ... osv., andre halvbilde av linje 2, 4, 6, ... osv.

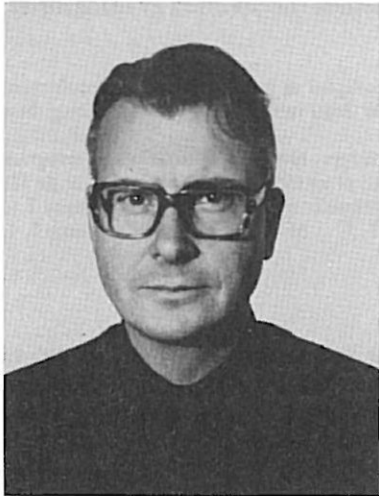
Etterskrift

Etter at dette manuskriptet ble skrevet, reiste jeg på studietur til Amerika. Ved to av museene jeg besøkte var man allerede i gang med lignende prosjekter. Lengst var man kommet ved Air and Space Museum i Washington. Der holdt de på å overføre samtlige bilder til videoplater (analoge). Hver plate inneholder ca 100.000 bilder. Foreløpig hadde de produsert to plater og var vel i gang med den tredje. Til sammen vil det bli 10 plater med 1 million bilder. De to første platene er i salg og koster \$30 pr. stk. For denne prisen kan man altså få til gjennomsyn 100.000 bilder av fly. Det fulgte også med en katalog som var ordnet alfabetisk etter flytype. Til avspøkning ble det brukt en vanlig hjemme- videoplatespiller med tilhørende tastatur for å finne frem til riktig bilde. På videoplaten er hver enkelt rute (bilde) nummerert og maskinen finner selv frem til riktig rute bare man taster inn rutenummeret.

For å overføre bildene til videoplate ble alle bildene først overført til 35 mm kinofilm i farver, uansett om originalbildet var i sort/hvitt. Dette synes jeg er noe betenkelig, fordi en 35 mm kinofilm har en begrenset holdbarhet. Selv mente de at filmen kunne holde seg i et par

hundre år dersom den ble lagret nedkjølt. Hovedårsaken til at de valgte 35 mm film var at da de startet prosjektet var det ikke mulig å finne noe lagringsmedium for så store digitale datamengder.

Ved Franklin Institution i Philadelphia hadde de derimot et prosjekt for digitalisering av bilder gående. De hadde anskaffet utstyr for meget høy oppløsning. Billedkvaliteten var meget god, og det lot seg gjøre å få utsnitt-forstørrelser på skjermen. Det var ikke bare meningen at utstyret skulle brukes til bilder, men også til tegninger og tekster. Det siste var også meget interessant, idet det var mulig å behandle maskinskrevne eller trykkete tekster som om de var skrevet inn. Maskinen kunne f.eks. finne igjen et ord på en tegning. Prosjektet var foreløpig på eksperimentstadiet, og ingen reell innlesning av bilder, tegninger eller dokumenter var gjort. Også de ventet på store digitale lagringsmuligheter. Jeg ble lovet underretning om utviklingen i forsøkene.



Jan Wiig er konservator ved Norsk Teknisk Museum, hvor han arbeider med utstillinger og restaurering av gjenstander.

Automatisk alfabetisering av transkribert tibetansk

Espen Ore

Denne artikkelen er skrevet ut fra programmeringserfaringer med indeksering o.l. av tibetanske tekster transkribert til det latinske alfabet fra originaler som finnes ved Universitetsbiblioteket i Oslo.

Tekstene blir utgitt og oversatt i et prosjekt som ledes av universitetsstipendiat *Jens Braarvig*. Undertegnede har vært programmerer ved prosjektet. Det dreier seg om en *sutra* (en buddhistisk kanonisk læretekst) som foreligger i to tibetanske versjoner (og også på kinesisk). Prosjektet har som mål en tekstkritisk utgave med oversettelse. Ut fra sanskrit-sitater og faste buddhistiske termer som finnes i tekstene, er det også meningen å gi ut en versjon av sutra'en på sanskrit.

Tekstene skrives inn med tekstbehandler på en mikromaskin. Det er da mulig å lage brukbare (evt. midlertidige) utskrifter hvor transkripsjonen ligner den man tradisjonelt finner i trykte utgaver. Indeksering- en krever imidlertid større kapasitet enn den mikromaskinen byr på. På sanskrit oppstår det et problem med sammensatte ord som til dels skal oppløses i sine enkelte bestanddeler. På tibetansk er problemet motsatt: her skrives teksten tradisjonelt oppdelt i enkelte stavelser som må føyes sammen før man får et meningsfylt indeks.

Vi har valgt å overføre tekstene til en stormaskin hvor vi lager en normal KWIC-konkordans. Konkordansen kan i utgangspunktet sorteres etter ASCII's ordningskriterier. Hver linje i konkordansen inneholder i tillegg til kontekst også en opplysning om linjenummer i den samlede tekst. Det er derfor enkelt å finne side/linje referanse.

Den opprinnelige konkordansen blir gjennomgått, og stavelser som skal høre sammen i et oppslagsord blir markert. Etter dette manuelle stadiet blir konkordansfilen brukt som inndata til et program som lager indeksreferansene (side/linje). Utdata fra dette programmet består av én forekomst av hvert oppslagsord med alle referanser samlet. Til slutt skal dette indekset alfabetiseres etter tradisjonell tibetansk sorteringsrekkefølge. (Ved først å lage indeks-filen blir mengden av data som skal alfabetiseres etter tibetansk ordning kraftig redusert.)

Vanlige sorteringsprogrammer kan ikke uten videre brukes på ord som er skrevet i vanlig transkripsjon. Man kan skrive inn tekstene slik at «samme» originaltegn får forskjellig representasjon avhengig av sorteringsrekkefølgen (se nedenfor). Her vil jeg imidlertid beskrive hvorledes man kan sortere tekster som skrives inn i standard latinsk transkripsjon.

Automatisk alfabetisering (sortering) er vanligvis trivielt. På mange datamaskiner, spesielt stormaskiner, er sorteringsprogrammer en del av

grunnutrustningen de blir levert med. Man kan kjøpe sorteringsprogrammer til de fleste andre maskiner. Som regel vil man kunne sortere etter fritt valgte ordningskriterier, man er altså uavhengig av standard alfabetisk rekkefølge. Slik kan man transkribere tekster fra det kyrilliske eller greske alfabet og få dem sortert etter originalalfabetets ordning. Sorteringen blir litt mer komplisert hvis det ikke er et én-til-én forhold mellom tegnene i det originale alfabetet og transkripsjonstegnene. De transkriberte tibetanske tekstene representerer en én-til-mange relasjon i forhold til originalen, og de byr også på flere problemer under sortering.

Tibetansk skrift ble utviklet fra et indisk alfabet i det 7. århundre. Hvis man betrakter sammensatte former som individuelle bokstaver, bruker tibetansk ca. 120 tegn. Når sammensatte tegn tas med i beregningen, skyldes det bl.a. at to tegn, X og Y, kan settes sammen på forskjellige måter: Y kan f.eks. være understilt X, eller Y kan være overstilt X. Metodene som blir beskrevet her, er blitt brukt på en tekst hvor de sammensatte tegnene er løst opp under transkriberingen slik at det i alt er ca. 35 tegn.

I transkripsjon av tibetansk (og sanskrit) representeres enkelte originaltegn som flere tegn fra det latinske alfabet, f.eks. «kh», og andre som latinske bokstaver med tilleggsmarkeringer, f.eks. «s'». Før noen som helst sortering kan finne sted, er det derfor nødvendig å lage en intern kode med én-til-én representasjon mellom originaltegn og kode-tegn. Dette blir gjort av en egen programmodul som stort sett er identisk med en jeg har brukt for sortering av sanskrit (det er delvis samme transkripsjon som er brukt). Det ligger fortsatt noe kunnskap om transkripsjonssystemet innebygget i programmet, men det er meningen at den skal fjernes. På den måten vil brukeren kunne definere sine egne transkripsjonskoder, f.eks. i en fil, og så oppgi den som parameter til programmet. Slik vil programmet bli mer generelt. (Programmet for tibetansk alfabetisering er allerede utviklet et stykke i den retning, se nedenfor.)

Etter å ha «filtrert» transkripsjonen slik at det er én-til-én forhold mellom originalbokstaver og intern kode (men med sammensatte tegn fortsatt oppløst), kommer vi til de sorteringsproblemene som er spesielle for tibetansk. Alle vanlige sorteringsprogrammer forutsetter at man begynner i én ende av nøkkelfeltet (f.eks. venstre, dvs. første bokstav på norsk) og så finsorterer etter bokstavene i retning av den andre enden (på norsk mot høyre). Dette gjelder ikke for de tibetanske ordene. Tibetanske ord (eller riktigere: stavelser) kan ha en foranstilt bokstav, men de trenger ikke ha det. Videre *mā* de ha en hovedkonsonant (med evt. over- eller understilte bokstaver i sammensatte former) som er oppslagsbokstaven. Når sammensatte former blir oppløst under transkriberingen, blir overstilt bokstav plassert foran hovedkonsonanten, mens den understilte blir plassert etter. De *mā* også ha en vokal, men enkelte ganger kan hovedkonsonant og vokal smelte sammen. Til slutt

kan det komme én eller to bokstaver samt evt. bøyningssendelse.

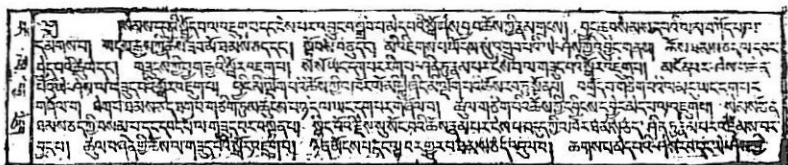
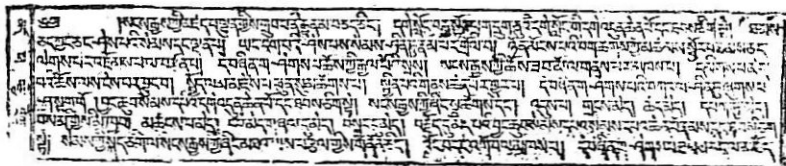
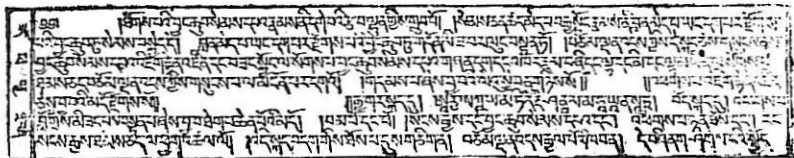
Et ord som har alle elementer kan se slik ut i transkripsjon:

b s g r u b s
1 2 3 4 5 6 7

Tallene hører ikke med i transkripsjonen. De er brukt her for å forenkle fremstillingen. 1 er foranstilt bokstav, 2 er overstilt, 3 er hovedkonsonant, 4 er understilt, 5 er vokal og 6 og 7 er de avsluttende bokstaver som ikke byr på noen sorteringsmessige problemer. På tibetansk ville de forskjellige bokstavene bli plassert slik, rent fysisk:

s
b g u b s
r

Man vil da med én gang se hvilken gruppe hver enkelt bokstav tilhører.



Begynnelsen på originalteksten som finnes på Universitetsbiblioteket i Oslo. Teksten er sentral i Mahayana-buddhismen og omtaler bodhisattva'enes etikk. (I Mahayana er bodhisattava'ene de som har nådd frem til terskelen av Nirvana men ikke går over denne før alle levende vesener er frelst.)

I eksemplet ovenfor var alle bokstavgrupper representert (jeg vil heretter kalle dem 1,2,3 osv.). I kodefeltet som skal konstrueres fra ordet og brukes som nøkkelfelt under sorteringen blir rekkefølgen slik: 3 2 1 4 5 6 7. Før kodingen leser programmet 7 filer som inneholder de tegnene som kan forekomme i gruppene 1-7 og som også gir tegnenes innbyrdes sorteringsrekkefølge i de forskjellige gruppene. På denne måten er man til en viss grad uavhengig av opprinnelig ortografi (hvis tekstene ikke er skrevet inn med standard tibetansk ortografi) og transkripsjonssystem.

Siden alle bokstavgruppene er representert i "bsgrubs" er det ikke noen problemer forbundet med å lage kodefeltet. Det er like lett hvis vi har et ord som "pa". Her må første bokstav være fra gruppe 3 og andre fra gruppe 5. Kodefeltet blir da "3---5--" ("-" = blank).

Men det finnes ord som f.eks. "rnams" eller "sbyin". Alle bokstaver som er mulige gruppe 1-, 2- eller 4-medlemmer er også mulige gruppe 3-medlemmer. På empirisk grunnlag har vi laget følgende mengder for mulig gruppe-tilhørighet. (Bokstavene blir gjengitt i den formen de får når ordene skal sorteres på stormaskin. På mikromaskin hvor tekstene som skal utgis, blir skrevet, brukes litt andre koder for å gi pen utskrift.)

Gruppe 1: g, d, b, m

Gruppe 2: r, l, s

Gruppe 3: k, kh, g-, g, n≠, c, ch, j, n~, t, th, d, n, p, ph,
b, m, ts, tsh, dz, w, ž, z, 'y, r, l, š, s, h, a

Gruppe 4: y, r, l

Gruppe 5: a, i, u, e, o

Gruppe 6: g, n≠, d, n, b, m, 'r, l, s

Gruppe 7: s, a, i, u, e, o

Algoritme

Programmet finner først hvilken plass vokalen har i ordet. Er det plass nr. 2 eller 5, har vi et av de to enkle tilfellene nevnt ovenfor. Hvis vokalen har en annen plassering, velges det mellom to analyser avhengig av om vokalen er på plass 3 eller 4.

Hvis vokalen har plass nr. 3, testes det om første bokstav kan tilhøre gruppe 1. Kan den det, har vi enten 1 3 eller 3 4 i begynnelsen av ordet siden gruppe 1 og 2 ikke har noen felles elementer. Hvis første bokstav er eller andre bokstav ikke er element i gruppe 4, har vi 1 3 (dette bygger igjen på empiri). I motsatt fall har vi 3 4. Hvis derimot første bokstav ikke er element i gruppe 1, testes det om andre bokstav er med i gruppe 4. Er den det, har vi rekkefølgen 3 4. Er den det ikke, har vi 2 3.

Når vokalen har plass 4, starter programmet også med å teste om første bokstav er element i gruppe 1. Er den ikke det, er saken grei: rekkefølgen må være 2 3 4. Hvis første bokstav er med i gruppe 1, har vi enten 1 2 3 eller 1 3 4. Hvis tredje bokstav er element i gruppe 4, er det

siste riktig, ellers har vi 1 2 3.

Sorteringsrekkefølgen er slik (forutsatt samme hovedkonsonant): Først kommer ord som bare har 3 og 5, så (i rekkefølge) 3 4 5, 1 3 5, 1 3 4 5, 2 3 5, 2 3 4 5, 1 2 3 5 og til slutt 1 2 3 4 5. Siden fordelingen 2 3 4 5 kommer før 1 2 3 5, og kodefeltet har rekkefølgen 3 2 1 4 5 (6 7), er det laget et ekstra flagg-felt som viser når vi har rekkefølgen 1 2 3 5. Det fullstendige kodefeltet ser altså slik ut: 3 F 2 1 4 5 6 7.

Andre sorteringsmetoder

Peter Nancarrow har beskrevet en sorteringsalgoritme som bygger på et annet prinsipp¹. I stedet for å basere seg på transkriberte tekster bruker han en innskrivningsmetode direkte fra tibetansk som ikke løser opp de sammensatte bokstavene. I tillegg gir han de foranstilte (vår gruppe 1) en egen verdi under innskriving som er forskjellig fra den verdi de har når de opptrer som hovedkonsonant. Denne metoden forenkler sorteringen. Men den forutsetter at tekstene skrives direkte inn fra tibetansk av personer med fagkunnskap på området. Nancarrow velger dessuten å bruke ordboksformer i et program som automatisk gjør om det innskrevne til transkribert tekst. I ordboksformer skrives hovedkonsonanten ofte som en versal. I de tekstene som vår metode er anvendt på, har det vært ønskelig å reservere versalene til første (hoved-)bokstav i egennavn, i setninger osv. Vårt system kan også brukes på tekster som skrives (eller leses ved hjelp av OCR) direkte fra allerede eksisterende transkriberte utgaver. Dette vil sannsynligvis være en fordel der man ønsker å kunne behandle store tekstmengder så fort som mulig og ikke har tilstrekkelig med tibetansk-kyndig arbeidskraft til å skrive tekstene inn direkte fra tibetansk.

Noter

1. Peter Nancarrow: «A System for Processing Tibetan Texts in their Original Orthography» i ALLC Journal, Vol. 1, No. 1, 1980, pp. 19-24.

Pedagogisk programvare for datastøttet undervisning

Ulike problemstillinger knytta til forskjellige bruksmåter*

Lars Vavik

1. Innledning

Oppsamlinger av forskningsrapporter om forskjellige former for datastøttet undervisning gir oss etter hvert mulighet til å reflektere over dagens praksis, men gir også næring til spekulasjon om framtida. Med datastøtte i undervisningen mener jeg pedagogisk programvare som kan brukes som et hjelpemiddel i undervisningen i forskjellige fag og som redskap i løsningen av tverrfaglige problem.

Vi spør oss om dette hjelpemiddelet kan utvide den formen for veiledning som f.eks. gis gjennom trykte læremidler. Om vi kan lettere etterligne systemer i natur og samfunnet, som ellers er komplisert å få innsikt i. Om maskinen kan hjelpe oss i å holde orden på og oversikt over en stadig voksende informasjonsmengde.

Vi kan sammenligne datamaskinen med mange forskjellige undervisningsmetodiske hjelpemidler, f.eks. film, dias, video, kassettbånd, bøker. Flere av de egenskapene som gjør disse nyttige å bruke, finnes nå i ett og samme instrument. Datamaskinen gir mulighet til å behandle og fremstille tekst, til å fremstille og kontrollere bildebevegelser, til å konstruere og høre lyd (tale, musikk) og selvfølgelig til å lagre eller bearbeide numeriske data. Disse egenskapene vil, når de integreres i ett og samme medium, legge grunnlaget for en ny undervisningsmessig dimensjon vi vanskelig kan forutse effekten av nå.

I mye integrasjonsproblematikk er regelen den at effekten er mer enn summen av delene. Dette gjelder trolig også for datamaskinen i forhold til andre audio-visuelle hjelpemidler. Forutsetningen er at datamaskinens spesielle muligheter utnyttes. Vi kan selvfølgelig legge lærebøkens tekst og illustrasjoner inn på datalager og «bla i boken» ved hjelp av datamaskinen. Det utbyttet vi kan ha av slike program, blir mer et spørsmål om økonomi enn pedagogikk; papirtekst kontra elektronisk tekst. Vurdering av datastøttet undervisning har i første rekke med nye pedagogiske muligheter å gjøre, deretter kommer kostnadsoverslagene.

2. Klassifikasjon av pedagogisk programvare

Datastøttet undervisning er ikke et entydig begrep. Vi snakker om

*Utdrag fra denne artikkelen står i tidsskriftet «DATATID», juni 1984.

mange bruksmåter og trenger derfor et grupperingssystem. Vi kan ordne pedagogiske program i forhold til fag og i forhold til pedagogisk metodikk. Jeg har i denne oversikten lagt mest vekt på den metodiske inndelingen fordi den både er teoretisk interessant og praktisk nyttig.

Undervisningsmetoder kan klassifiseres i grader av frihet, alt etter hvordan de gir elevene mulighet til å ta initiativ og bestemme hva de skal arbeide med og hvordan. Dette er et ordningsprinsipp av generell pedagogisk interesse. Undervisningen i det ene ytterpunktet preges av sterkt strukturerte metoder. Dette kommer særlig godt fram i det vi kaller programmert undervisning. Den andre ytterligheten inneholder «åpen», «oppdagende» undervisning eller selvgenererende læring. Mellom disse finnes flere overgangsformer. Datastøttet undervisning vil kunne spenne over hele repertoaret. Fire hovedkategorier peker seg allikevel ut. Disse har jeg valgt å kalle: datastøttet instruksjon, simulering, modellbygging og informasjonsbehandling.

2.1 Datastøttet instruksjon

I midten av 60-åra ble flere nye systemer for datastøttet undervisning prøvd ut. De fleste var sterkt farget av behaviouristiske læringsteorier og organiseringsprinsipper. Stoffet ble trinnvis bygd opp med økende grader av vanskelighet. Rask tilbakeføring er et viktig psykologisk prinsipp i disse systemene. Av slike program finnes to typer, de såkalte lineære og forgrenede programmer.

Den første typen ble først konstruert av psykologen *B.F. Skinner* og er brukt uavhengig av datateknologisk utstyr. Programmene kalles lineære fordi alle brukerne går gjennom de samme trinn i et bestemt stoff. *N. Crowders* system har fått navnet forgreinet fordi elevene ikke nødvendigvis går gjennom de samme programmene i et bestemt stoff, men møter materiale av ulik vanskelighetsgrad. Skinners metode skulle gi mulighet for tempodifferensiering, mens Crowders system skulle resultere både i tempo- og nivåddifferensiering.

Oppleggene ble kritisert på tre viktige punkt. Det første gikk på stofforganiseringen. Det finnes flere oppfatninger om hva som er lett eller vanskelig når stoffet skal ordnes. Vi må derfor ha mange innlæringsveier. Det andre ankepunktet har å gjøre med undervisningsmetoden, som bygger på spørsmål og elevsvar som den eneste mulighet. Avhengig av om svaret er rett eller galt må en enten repetere eller gå tilbake til «enklere» oppgaver. Det er kun absolutt riktige svar som aksepteres. Den tredje innvendingen er en mulig konsekvens av de to første: motivasjonen for å lære gjennom slike systemer er vanskelig å holde ved like.

Mange datastøttede instruksjonsprogram er konstruert for innøving av kunnskap som regnes for å være nyttig. Slike program ble i praksis avgrenset til emner det er lett å trene på. Drilløvinger i matematikk og

K L A S S I F I K A S J O N A V

	TYPE	BESKRIVELSE
I DATASTØTTET INSTRUKSJON	1. Programmert undervisning	Enkle innlærings- sekvenser. (Bit for bit) (Behaviouristisk undervisning)
	2. Forgreinet programmert undervisning	Flere læringsveier
	3. Øvelsesprogram	Innøving av ferdigheter, begreper med eller uten kunstige effekter
	4. Tilpasset undervisning	Programmet inneholder en representasjon av a) eleven og/eller b) emnet som skal læres og/eller c) undervisningsmåter
II DATASTØTTET SIMULERING	5. Simuleringer	Eleven utfører eksperimen- ter på en datamaskin- modell av virkeligheten
	6. Rettlednings- systemer	Programmet gir oppgave- spesifikke råd gjennom simuleringen
	7. Modellbygginger	Elevene bruker et modell- verktøy for å skape sin egen matematikk, grafikk, musikk eller fortelling
III DATASTØTTET MODELLBYGGING	8. Modellstyring	Elevene bruker et program for å styre instrumenter
	9. Informasjons- behandling	Et system som gir brukeren mulighet for å lagre og/eller hente og bearbeide informasjoner
IV DATASTØTTET INFORMASJONS BEHANDLING		

D A T A S T Ø T T E T U N D E R V I S N I N G S P R O G R A M

EKSEMPLER	BEGRENSNINGER	MULIG FORBEDRING
Skinners lærermaskiner	Liten mulighet for individualisert læring!	
Noen PLATO program	a)Programmet tar liten hensyn til elevenes forskjellige måter å lære på b)Fiksert undervisningsstrategi c)Restriktiv dialog	Tilpasset D.B.I. (Databasert instruksjon)
Dominerer mikromaskinmarkedet fram til 1984	Effektene kan ta oppmerksomheten bort fra fagmålsettingen	Veiledningsprogram "Initiativesprogram" "Kulissanalyse"
Buggy Scholar	Kompliserte program innenfor avgrensede fagområder	Generelle program "Ekspert systemer" med undervisningsveiledning
Quadratic Tutor		
Fysikksimuleringer (A. Bork, Irvine)	Det gis ingen kommentarer på utførelsen. Det er vanskelig å forandre forutsetningene i systemet	Rettledningssystemer Modellforandringsmuligheter
West	Komplekse program	Generelle rettledningsprogram
LOGO Songwriter Melody Maker Dr. Draw Story Maker Thing Lab	Kan bli for generelle. Læringen kan mangle fokus og retning	Informasjon og kunnskapssystemer kombinert med rettledningsprogram
LOGO som robotstyrer	Vi har svært liten erfaring med nytten av slik undervisning	
Microquery Visial	Ingen undervisning	Informasjon og kunnskapsbaser kombineres med rettledningssystemer

språk er typiske eksempler. Datamaskinen fungerer som en oppgavegenerator og velger øvinger etter tilfeldighetsprinsippet, uavhengig av teorier om stofforganisering. Forskjellige kunstige effekter bygges inn i drilløvingene for å motvirke motivasjonstapet under trening. Disse kan ha tilknytning til treningsoppgaven. I andre tilfeller er det ingen sammenheng mellom øvingen og «belønningseffekten».

Hvor mange oppgaver vil du ha? 10, 15 eller 20?

E: **15**

Hvor vanskelig skal de være? Vanskelig, Middels, Lett?

E: **Middels**

6 x 5

E: **30**

Riktig. 9 x 6

E: **54**

Flott. 3 x 8

E: **26**

Nei. Forsøk igjen. 3 x 8

E: **23**

Her har du visst problemer. Be læreren din om å hjelpe deg.

Fig. 1. Enkel programmert instruksjon/øving i matematikk.

.....

Å sove

E: **schlawf**

Nei. Forsøk igjen.

E: **schlafen**

Riktig.

.....

Fig. 2. Programmert instruksjon/øving i språk.

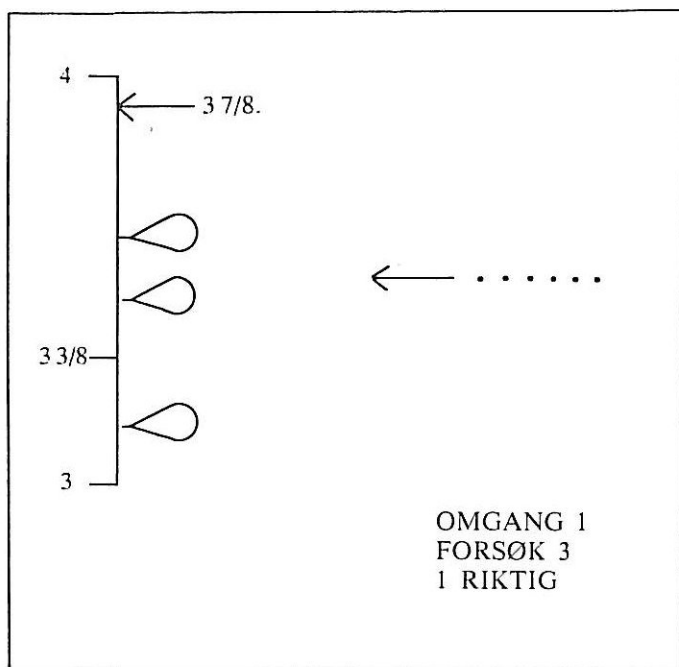


Fig. 3. Øvingsoppgave med motivasjonseffekter. (Pilkast på tallinja.)

Utvikling av øvelsesprogram med ulike motivasjonsskapende effekter dominerer i dag det pedagogiske programvaretilbudet. Poeng for riktig svar, lyd- og fargesignaler eller dramatiske hendelser brukes for å holde oppmerksomheten fanget.

Kunstig kontra naturlig motivering er et gammelt stridsspørsmål i pedagogisk psykologi. Det er bare i kommersiell interesse å bruke effekter for å øke etterspørselen uten at vi samtidig vurderer det faglige utbyttet. Vi vet mye om hva slags effekter som gjør et program spenningsfullt, men mindre om dette samtidig tar oppmerksomheten bort fra eller øker de faglige målsettingene.

Det vil bli en viktig oppgave å vise hvilke sammenhenger det er mellom kulisseeffekter, læringsmotivasjon og faglig utbytte. Vi vet også lite om hva slags innvirkning slike pedagogiske spillprogram har på undervisningen i sin alminnelighet.

Så langt kan vi si at programmerte undervisningssystemer fra 60-årene bygde på et ensidig og spinkelt teoretisk grunnlag og representerte for mange en psykologisk retning med et sett verdier som var uakseptable. Drilløvinger var mer konstruert ut fra et behov for

treningsoppgaver enn ut fra læringsteoretiske nyvinninger.

Intens forskning og utviklingsarbeide innenfor feltet kognitiv psykologi har i dag åpnet nye muligheter for datastøttet instruksjon. Undervisningsprogrammene har i engelskspråklig litteratur blitt kalt «Intelligent Tutoring Systems».

Begrepet «intelligent» er på mange måter uheldig fordi vi lett kan tilegne maskinen menneskelige egenskaper. Datastøttet instruksjon skal ikke forveksles med lærerens ansvarsområde. Systemet er ikke intelligent, men avansert både i forhold til lærebokveiledninger og behavioristiske læringsteorier.

Utviklingsarbeidet har vært styrt av målsettingen om å:

- utforme en «dialog» med brukeren i mest mulig naturlig språk.
- bygge inn i programmet så mye informasjon om det emnet som er tatt opp at det er mulig å analysere uventede svar.
- gi muligheter for å tolke brukernes misoppfatninger.
- gi muligheter for å profitere på brukererfaringer og eksperimenterere med undervisningsstrategier.

En konsekvens av dette arbeidet er at mye av initiativet til f.eks. å formulere problemstillinger, hypoteser og velge presentasjonsformer kan bli skjøvet over til brukeren. Veiledningssystemene kan allikevel variere fra hyppig intervensjon med fullstendige forklaringer og rettledning til å gi veiledning uten «å si noe». I det siste tilfelle er det mer snakk om å tilrettelegge problemene og på den måten stimulere læringsprosessen. Mellom disse ytterpunktene finner vi igjen flere mellomposisjoner. Fig. 4 nedenfor gjengir dialogen mellom bruker og et program som har noen flere muligheter innebygd. De best utbygde systemene som f.eks. «BUGGY», «SOPHIE» eller «MYCIN» er det vanskelig å få inntrykk av gjennom et skjerm bilde (se klassifikasjonsskjemaet).

Fig. 4 nedenfor gjengir dialogen mellom bruker og et program med noen få flere tolkningsmuligheter.

Fig. 4. Program med mer utbygd tolking av elevsvar.

Prøv å løse likningen $x \cdot x - 18 = x \cdot 9$ Finn svarene.

E: Hva skjer hvis jeg ikke klarer det?

Bare skriv et tall eller skriv **STOPP** hvis du vil stoppe.

E: 1,2

Nei

E: 3

Det er riktig. Klarer du å finne det andre svaret?

E: 4,5

Nei

E: 6

Bra. Nå har du klart begge svarene.

Prøv å løse likningen $x \cdot x + 18 = x \cdot 19$ Finn svarene.

E: 3

Nei. Forsøk igjen.

E: 4,5,6,7,8,9,10

Du forsøker ikke å finne noen regel som kan hjelpe deg.

Vi tar et annet eksempel.

Forsøk å finne en regel.

Prøv å løse likningen $x \cdot x + 48 = x \cdot 19$ Finn svarene.

E: 1,18,12,48,19,64,9,2

Nei

E: 3

Det er riktig. Klarer du å finne det andre svaret.

E: 8

Nei. Forsøk igjen.

E: 3

Ja, men det har du funnet tidligere.

E: 15,18,9,1

Nei.

Det var mye gjetting.

La oss forsøke et annet eksempel.

De riktige svarene var 3 og 6.

Et lite utsnitt fra programmet BUGGY gjengitt nedenfor viser prinsippet med å finne fram til systematiske feil i f.eks. enkle subtraksjonsoppgaver.

Oppgaver og riktige svar:

45	40	139	500	312
<u>-23</u>	<u>-30</u>	<u>-43</u>	<u>-65</u>	<u>-243</u>
22	10	96	435	69

Elev svar:

+	+	+	565	149
*	*	*	***	***

- *** BUGGY forutser elevens gale svar
- * BUGGY forutser elevens riktige svar
- + Eleven svarer rett

I dette eksemplet er det enkelt å stille «diagnosen». Et avansert analysesystem trer i kraft for å skille mellom systematiske og tilfeldige feil når feilsøkingen blir mer komplisert.

Vi har liten erfaring fra skolen med det jeg har kalt «tilpasset undervisningsprogram». Dette har hovedsakelig en praktisk forklaring, det er få program som er lagt over på skolemaskiner.

2.2 Simuleringer

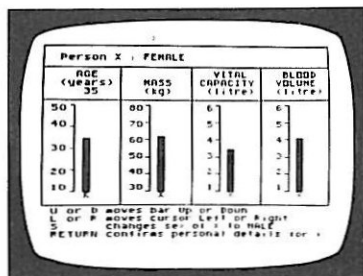
Begrepet «simulering» blir brukt når datamaskinen er programmert som modell for et system. Gjennom manipulering med det simulerte systemet, skal brukeren finne ut hva slags sammenheng det er mellom forskjellige variabler. Denne metoden gir muligheten for å eksperimentere med ting som ellers ville være for farlige, kostbare, tidkrevende eller umulig å gjennomføre av andre årsaker.

Forskjellige simuleringssystemer finnes særlig innenfor naturfag og samfunnsfag. Bølgelære, næringskjeder, planetenes bevegelse i solsystemet og økonomiske modeller for næringsdrift er noen få eksempler på hyppig valgte temaer for simuleringssystemer.

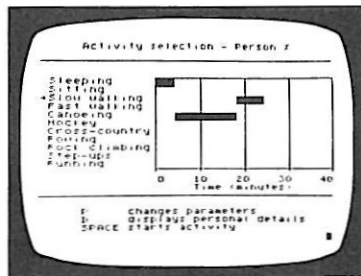
Det er et pedagogisk problem at simulerte systemer lett kan sammenblandes med virkelighet. Det bør gå klart fram av modellen hva slags forutsetninger systemet bygger på. Dette gjelder særlig samfunnsvitenskapelige modeller. Dersom modellen bygger på bestemte verdier, så vil konklusjonen også være verdibestemt. Derfor må forutsetningene også være mulig å forandre. Denne betenkeligheten gjelder også lærebøkene.

Det er tidkrevende å konstruere simuleringssystemer, men dette kan oppveies i de tilfeller en har å gjøre med vanskelig tilgjengelig lærestoff

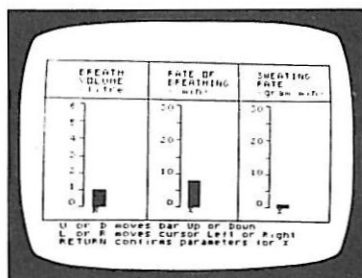
eller når målsettingen for undervisningen krever et program som ligger så nært opp til virkeligheten som mulig.



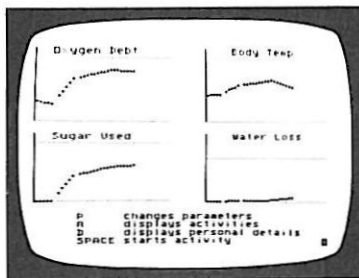
a) Variabelvalg



b) Aktivitetsvalg



c) Parametervalg



d) Resultat

Fig. 5. Simulering i fysiologi.

2.3. Modellbygging

Det teoretiske grunnlaget for denne formen for datastøttet undervisning bygger på reformpedagogiske ideer fra f.eks. Piaget, Dewey og Montessori. Konstruktørene viser også med all tydelighet at ideer fra disse pedagogene er grunnlaget for å skape verktøy for selvgenererende læring. Utformingen av modellverktøyet LOGO ved Massachusetts Institute of Technology er et særlig godt eksempel på at datateknologi kan brukes i reformpedagogisk tjeneste. Vi hører igjen de samme argumentene mot formell strukturert undervisning/kunnskapsformidling som vi tidligere er kjent med. Denne gangen kommer kritikken fra pedagoger som ser muligheten for å realisere aktivitetspedagogiske prinsipper med nye hjelpemidler.

En av de fremste forskerne fra MIT, S. Papert formulerer sitt syn på den tradisjonelle skole slik: «Modellen for vellykket læring er måten

barnet lærer å snakke på – en prosess som finner sted uten bevisst og organisert undervisning. Jeg ser klasserommet som et kunstig og ineffektivt læringsmiljø som samfunnet har blitt tvunget til å oppfinne fordi de uformelle læringsmiljøene var ufullstendige for mange formål som skriving, grammatikk og matematikk. Jeg tror datamaskiner vil gjøre oss i stand til å endre læringsmiljøet. Mye av den kunnskap skolen for tiden forsøker å formidle med mye slit og begrenset hell kommer til å plukkes opp av elevene slik de lærer å snakke. Det jeg sier innebærer at skolen slik vi kjenner den i dag ikke får noen plass i fremtiden. Men det er et åpent spørsmål om den klarer å tilpasse seg ved å bli til noe nytt eller kommer til å smuldre vekk og bli erstattet av noe annet». (S. Papert, «Dialog med datamaskinen», Cappelen 1983, s. 15).

Datastøttet undervisning vil i denne retningen bety en bestemt måte å bruke maskinen på. Modellbygging viser tilbake til datamaskinen som verktøy for aktiv læring. Det vil stå i skarp motsetning til at maskinen kan brukes til å lære riktige framgangsmåter ved hjelp av øvelser med passende vanskegrad. Når datamaskinen brukes som instrument for drill, representerer dette et syn fra den «gamle pedagogikk».

S. Papert står for det revolusjonære standpunkt i verkstedpedagogisk sammenheng. Datamaskinen innleder et studium i pedagogikkens historie hvor radikale forandringer er mulig, etter hans mening. Det revolusjonære vil være å innpasse ideer inspirert av teorier om «learning through discovery», «learning by doing» eller læring gjennom modellbygging (programmering) i skolens undervisning.

«Forfatterverktøy» i musikk og språk faller inn under den samme filosofien. De er foreløpig lite utprøvd i skolene.

Det er mange uklare sider i aktivitetspedagogisk undervisningslære. De blir ikke borte selv om vi får et nytt dynamisk hjelpemiddel inn i skolen. Noen av de mest generelle spørsmål er f.eks.: Hva er det vi antar at elevene skal kunne «oppdage»? Er det generaliseringer, faktakunnskap, relasjoner eller arbeidsmetoder? Hva skal vi gjøre hvis elevene kommer fram til slutninger som utvilsomt er uriktige? Når skal læreren gripe inn? Hvordan skal vi stimulere sjølstyrt læring?

Vi må finne svar på disse spørsmålene gjennom forsøks- og utviklingsarbeid med dataverkstedpedagogikk i skolen.

2.4. Datastøttet informasjonsbehandling

Databehandlingsprogram gir brukeren mulighet til å behandle en større mengde informasjon enn det var praktisk mulig å gjøre tidligere. Det gjøres samtidig med en hastighet som frigjør mer tid til å studere og diskutere betydningen av informasjonen. Undervisningsstøtte vil derfor her si økt informasjonstilgjengelighet i form av ny informasjon, rask fremkalling og interaktiv presentasjonsform.

Det finnes i dag mange forskjellige informasjonsbehandlingspakker på markedet. Mange er laget for spesielle formål i industri og handel.

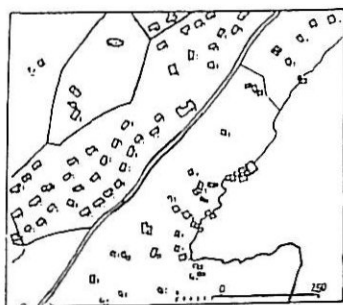
Andre har mer generell utforming og kan brukes som pedagogiske verktøy. Det er to sider ved datastøttet informasjonsbehandling som har særlig pedagogisk betydning. Det første gjelder utvelgning av informasjon. Det andre gjelder presentasjonsformen.

Hva slags informasjon bør ligge på «lager»? Vi spør egentlig om hva slags kunnskap som bør være tilgjengelig i undervisningen. Det er avhengig av faglige målsettinger. Målene har satt ramme omkring stoff og arbeidsmåter. Men samtidig åpner informasjonsteknologien opp andre muligheter som ikke er skrevet inn i planen fordi de var vanskelige å realisere.

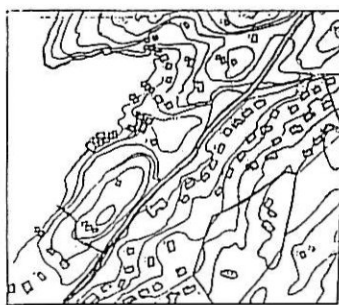
Orienteringsfagene i grunnskolen vil f.eks. være et område som lettest kunne utnytte mulighetene. De didaktiske hovedspørsmålene om f.eks. global orientering kontra lokale studier, om breddekunnskap kontra fordypning i sentrale begreper, vil bestemme innholdet i informasjonslageret (jfr. forskjellen mellom danske og norske læreplaner).

Hvordan skal data fra informasjonslageret presenteres? Vi har svært lite forskning å bygge på når vi skal finne fram til presentasjonsformer som gir best informasjon for brukerne. Typen av informasjon er med på å bestemme hva slags form den kan framstilles i.

Aggregerbare data er lettere tilgjengelige i grafisk eller kartografisk form. Kartografisk og grafisk kommunikasjonsteori er derfor nyttige verktøy i konstruksjon av presentasjonsprogrammet. Det ligger store pedagogiske fordeler i presentasjonsformen dersom brukeren har kontroll over hvilke data som skal presenteres, hvilken rekkefølge de skal komme i og hvilke grafiske symboler som bør velges. Har en ideografiske data i lageret, er det mer snakk om å utvikle fleksible søkeverktøy som gjør at en kommer raskt fram til de opplysningene en ønsker.



5.valg (bygninger)



6.valg (høydekurver)

Fig. 6. Ulik lokal informasjon presentert kartografisk.

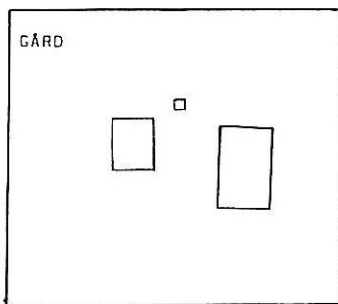
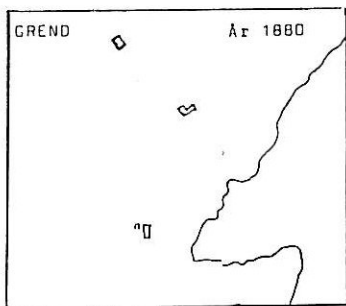


Fig. 7. Historiske kart (1880).

HUSSTAND	ÅR 1880
F	Anton Larsen
M	Louise Larsen
S	Leif
S	Harald
D	Ida
T	Petra

INDIVID
Navn.....
Født.....
Gift.....
Yrke.....
Flytting.....
Nattverd.....
Anmerkning.....
.....
Død.....
Barn.....

Fig. 8. Historisk informasjon.

3. Hovedtrekk ved programvareutviklingen fram til i dag

Datastøttet instruksjon, simulering, modellbygging og informasjonsbehandling kan ideelt sett hver for seg flytte grenser for hva som tidligere var pedagogisk mulig. Datastøttet instruksjon utvider lærebøkene muligheter, simuleringer etterligner systemer vi tidligere hadde liten mulighet for å arbeide med i skolen, modellbygging skaper rammer rundt sjølgenererende læring og informasjonsbehandling øker tilgangen på nye opplysninger. Det må allikevel være riktig å si at det er stor avstand mellom de mulighetene edb-teknologien gir og kvaliteten på mye av det som markedsføres under tittelen «pedagogisk programvare». En påfallende stor del bygger på enkle prinsipper om stimulus, respons og feedback. Om dette nødvendigvis reflekterer en bevisst pedagogikk eller om det tilfeldigvis har blitt slik, er vanskelig å avgjøre. Mye tyder på at teknologien fram til de siste årene har satt så store krav

til programtekniske ferdigheter at det ikke har blitt plass til å studere prinsipper for pedagogisk redigeringsarbeid. Dette har og vil fortsette å bedre seg ved at maskinene får større kapasitet, samtidig som det stadig utvikles bedre og lettere tilgjengelig «forfatterverktøy». I mange tilfeller vil god programvare være knytta til krav om interaktivitet, fleksibilitet eller brukerkontroll. Dette gjør oppgaven vanskeligere og behovet for pedagogisk innsikt større.

Tilfeldighetsloven gjelder også for hvilket emne det var hensiktsmessig å gi datastøtte i. Ingen skolefag har unngått å bli stilt overfor tilbudet om såkalte pedagogiske program. Dette skjer uten at datamaskinens spesielle muligheter vurderes i forhold til målsettingen i faget. Det er nødvendig med en grenseoppgang mellom hva som er teoretisk interessant å utforske med datateknologi i forhold til faglige spørsmål, hva som er praktisk mulig å videreutvikle med det vi nå har av hjelpemidler og på hvilke fagområder datateknologien har lite å bidra med. En analyse av hva datastøtte kan gi i forhold til grunnskolens undervisningsoppgaver, har gitt dette avgrensingsforslaget.

GRUNNSKOLENIVÅ:

Type	Fag						
	Samf.	Naturf.	Mate.	Språk	Musikk	Forming	
Instruksjon	1	T	T	T	T	T	T
	2	T	T	T	T	T	T
	3	T	T	T	T	T	T
	4	U	U	U	T	T	T
Simulering	5	I	I	I	T	T	T
	6	I	I	I	T	T	T
Modellbygging	7	I	I	I	U	U	U
	8	I	I	I	T	T	T
Informasjonsbehandling	9	I	I	I	T	T	T

I Interessant utviklingsfelt

U Bør utredes bedre

T Tvisomt bruksområde

I språkfagene, musikk og forming vil det være lite relevant med utstrakt bruk av datastøttet undervisning. Generering av bilder, musikk og tale er fortsatt på utforskningsstadiet. De programmene som finnes for mikromaskiner, er mer å se på som kuriositeter enn praktisk nyttige

verktøy. Selv om datamaskinen *kan*, er det ikke sikkert at den *skal*. Vi kan tenke oss nye tekniske forbedringer, uten at dette automatisk gir større plass for datamaskinen i disse fagene.

Dette går på tvers av mange anbefalinger i internasjonale interesseorganisasjoner omkring «computers in language and second language teaching, computers in fine art and design, computers in music education». Konklusjonen gjelder undervisningsoppgaver på grunnskolenivå. Mange av «drillprogrammene» som i dag dominerer markedet, vil trolig gjøre liten skade med begrenset bruk, men det er vanskelig å se at de kan ha noen betydning. Hvilken rolle «komposisjonsverktøy» i språk, forming og musikk får, er det vanskelig å si noe om uten at dette er utprøvd bedre. At elever på de laveste klassetrinnene kan lære å bruke slike redskap, er det liten tvil om.

Det er lettere å argumentere for forskjellige former for datastøtte i orienteringsfagene. De mest interessante utviklingsoppgavene ligger i simuleringer, informasjonsbehandling og muligheten for å knytte disse til veiledningssystemer.

Herbert Simon, en kjent forsker i informasjonsvitenskap, har pekt på at det blir viktigere i framtida å vite hvordan en skal få tak i informasjon enn å være i besittelse av den. Det kan tolkes slik at arbeidsmåten er viktigere enn stoffets innhold. Dette er det metodeformalistiske utdanningsprinsipp som vi forsøkte å bygge Normalplanen for folkeskolen fra 1939 på: arbeidsmåten skulle være pensum. Det er vanskelig og betenkelig å skille stoff og metode. Datastøttet undervisning kan også bli oppfattet slik at det er viktigere å lære å betjene de tekniske instrumentene enn å vurdere hva slags kunnskap vi kan formidle. Utviklingen har i hovedsak vært drevet av de store framskrittene på den teknologiske siden, mens de undervisningsmessige mulighetene har blitt hengende etter. Den programvaretypen vil stå sterkest som både kan være et aktivum i fagundervisningen og samtidig vise datamaskinens muligheter og begrensninger.

Lars Vavik er lektor ved Stord Lærerhøgskule, hvor han underviser og foretar utviklingsarbeid innen feltet pedagogikk og informatikk.

PLATO og datamaskinassistert læring

Rune Midtvedt

Control Data's PLATO er et av de eldste og mest velkjente systemer innen datamaskinassistert læring. I USA brukes systemet på alle trinn i undervisningssystemet, samtidig som en lang rekke bedrifter bruker det i den bedriftsinterne opplæring.

Bruken av datamaskinen som hjelpemiddel i undervisning og opplæring er et av de virkelig spennende felt innen edb i dag. Tanken er forsåvidt ikke ny. De første forsøkene begynte for over 25 år siden. Men det er først med den store utbredelsen av mikromaskiner at datamaskinassistert læring er blitt aktuell i stor skala, både innen skoleverket og ellers.

De viktigste egenskapene ved denne form for opplæring er:

- læringen blir en aktiv prosess. Eleven blir stadig stilt overfor spørsmål eller beslutningsvalg som han må ta stilling til for å komme videre.
- individualisert opplæring. Programmene kan lagres slik at den videre gangen gjennom materialet blir gjort avhengig av det/de svarene eleven allerede har gitt. På den måten kan en elev som har problemer få en utdyping av et bestemt felt, mens en annen kan gå direkte videre på neste.
- sjølstyrt tempo. Eleven bestemmer selv i hvilket tempo han vil gå gjennom materialet.
- presentasjonsmuligheter. Skjermterminalen gir helt nye muligheter for presentasjon av materialet. Grafikk og tekst kan kombineres fritt. Man har et ubegrenset antall «sider» tilgjengelig. Det er derfor ikke nødvendig å fylle hver enkelt side med masse tekst, slik som i en bok. En detalj av et bilde, en enkelt setning eller et bestemt ord kan få stå helt alene på skjermen om det er ønskelig. Visuelle effekter som blinking, inners video, farger, osv. gjør at man kan framheve bestemte deler av et bilde eller en tekst.

Hva er så PLATO?

For det første er PLATO en omfattende samling av eksisterende kursvarer. I PLATO-biblioteket finnes det tilgjengelig hundrevis av kurs innen en rekke områder. Disse kan benyttes enten via terminal og oppringt linje, eller på mikromaskin.

For det andre er PLATO en måte å gjennomføre kursene på som benytter seg av alle datamaskinens fortrinn. En utstrakt bruk av grafikk og animasjon («tegnefilm») gjør at stoffet presenteres på en grei og oversiktlig måte. Kombinert med bruk av trykkfølsom skjerm gjør dette PLATO spesielt velegnet til simulering. (Trykkfølsom skjerm vil si at man kan gi maskinen kommandoer ved å trykke på skjermen i stedet

for å bruke tastaturet). Ved hjelp av simulering kan man få øvd inn nye ferdigheter som er kostbare, farlige eller tidkrevende å gjøre i virkeligheten. Eller man kan friske opp gamle kunnskaper. Et eksempel er regulering av trykket i oljebørner for å forhindre utblåsing. Etter å ha gått gjennom teorien og testet at han forstår denne, kan eleven få prøve seg i «praksis». På skjermen kan han nå få en kopi av måleinstrumenter og kontrollbord. «Måleinstrumentene» viser hvordan trykket endrer seg i den tenkte oljebørnen. Operatøren må korrigere ved å trykke på de rette «kontrollknappene», på den trykkfølsomme skjermen. Avhengig av om eleven har gjort dette rett, vil målerne vise at trykket har stabilisert seg. Eventuelt at man har fått en utblåsning!

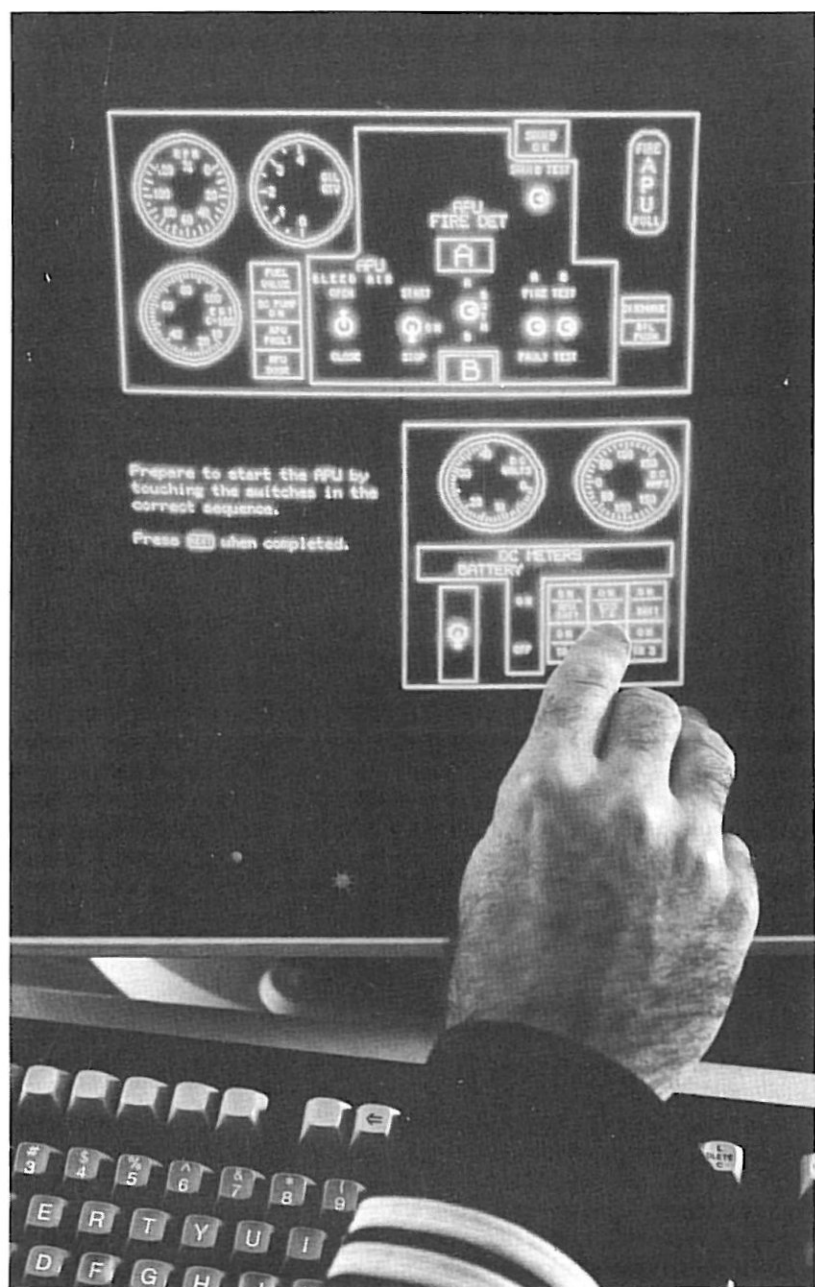
PLATO gjør det også enkelt å legge inn forklarende hjelpetekster/bilder. Dermed kan eleven når som helst få hjelp dersom han blir stående fast eller vil ha ytterligere opplysninger. Det er også mulig å kople flere terminaler sammen i et nett og la en lærer ha kontrollen med det som skjer. Læreren vil da kunne sjekke hvordan det går med de forskjellige elevene og eventuelt gi sine kommentarer. Altså omtrent som i et språklaboratorium.

For det tredje består PLATO av forskjellige «forfatterspråk» til å utvikle program. Inntil ganske nylig var det kun mulig å utvikle læreprogrammer på stormaskin, selv om programmene skulle kjøres på mikromaskiner. Etter hvert har det imidlertid kommet forskjellige «modeller» for utvikling på mikromaskiner (også IBM PC). Disse må foreløpig sies å ha relativt begrensede muligheter. Men det er lovet en bedre mikroversjon. Felles for alle «språkene» er at man, i hvert fall i teorien, skal kunne bruke disse uten forutgående kjennskap til programmering. For å konstruere f.eks. en sirkel, går du først inn i grafisk modus, plasserer markøren der du vil ha sentrum og trykker på 'o'. Programmet ber deg da om å spesifisere radien. Du angir denne ved å trykke på skjermen for å plassere markøren i et punkt som går gjennom sirkelperiferien. Du trykker på ENTER. Og det er det hele. Sirkelen blir tegnet ut på skjermen. Like enkelt er det å tegne linjer, firkanter, halvsirkler, vektorer m.m.

Videre kan man bruke forskjellige skrifttyper, lage egne tegn/figurer osv. Det vil føre for langt å gå i detalj her, men mulighetene er mange.

Selv om en trenger en relativt kraftig mikromaskin for å utvikle PLATO-kurs, vil en kunne overføre selve kursene til disketter som kan brukes på en lang rekke billigere maskiner, f.eks. Commodore 64. Utpå høsten skal noen av utviklingsmodellene komme på norsk.

For tiden er standard PLATO-terminaler uten farger fordi dette gir best kvalitet og oppløsning. Men det arbeides med å få til brukbare fargeskjermer som tilfredsstillir Control Data sine krav. På Hannover-messa i år viste Control Data Tyskland en farge- og trykkfølsom skjerm. Denne var også koblet mot en videodisk, slik at en fikk både bilder og tekst på samme skjerm. En kunne i tillegg styre videodisken ved å trykke på skjermen.



Del av et PLATO-kurs demonstrert på en trykkfølsom skjerm.

PLATO - video-stasjon

En ny kombinasjon av kjente hjelpemidler i undervisningen

Datamaskinassistert undervisning er i mange tilfeller en overlegen metode for å overføre kunnskap. Datamaskinens grafikk har imidlertid en begrensning i det at mange har vanskelig for å visualisere en skjematisk framstilling og overføre dette til en konkret situasjon. Det riktige er først å se en virkelig situasjon, for deretter å få en grafisk framstilling renset for eventuelt unødvendige detaljer.

Kombinasjonen video og datamaskin

PLATO's video-stasjon har gitt den maskinassisterte undervisningen en ny dimensjon. På PLATO's nye fargeskjerm kan en nå framstille konstruksjoner og hendingsforløp på en realistisk måte med alle de muligheter for lyd og levende bilder som videoteknikken gir. PLATO kombinerer på samme skjerm video, grafikk, tekst og beregninger i en programmert og elevstyrt framstilling. I tillegg gir også PLATO-terminalen mulighet for en dialog med læreprogrammet ved berøring av skjermen.

Bak denne annonseringen av PLATO's video-stasjon ligger en mangeårig erfaring med oppbygging av læreprogrammer og forfatterspråk for datamaskinassistert undervisning som nå også er kombinert med de visuelle mulighetene som ligger i filmteknikken. Billedframstillingen fra videoplaten blir hele tiden styrt av læreprogrammet gjennom datamaskinen.

Hvilke muligheter gir PLATO's video-stasjon?

Kombinasjonen av video og datamaskin gir muligheter som langt overgår det som disse teknikkene kunne gi hver for seg. Nå kan en bruke stille og bevegelige bilder, tidsavsnitt og zoom-effekter. En ønsket bildesekvens kan bli kallet opp fra program. Den kan ledsages av lydeffekter, tekst eller tale og musikk i hi-fi-kvalitet. Bilder kan forstørres eller forminskes, filmavsnitt kan kjøres med redusert fart. Tekst lagret i datamaskinen kan koordineres og skrives over bilder fra video, og f.eks. markere navn på deler som er vist. Dette kan igjen styre en dialog med systemet ved berøring på skjermen på de markerte felter.

Praktiske anvendelser

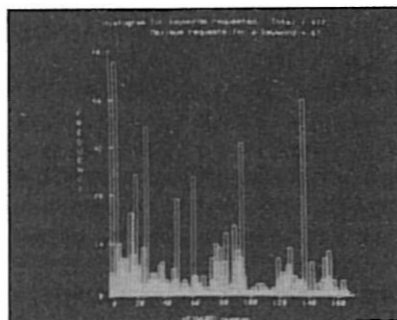
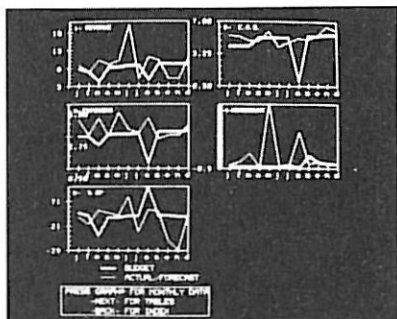
Ved demonstrasjoner

Datamaskinstyrte interaktive video-stasjoner har blitt installert i demo-rom for biler. Slike installasjoner gir en kunde mulighet til sjøl å undersøke deler av en bil, finne tekniske spesifikasjoner osv., uten å måtte bry en selger med spørsmål. Eller selgeren kan sjøl finne svar på

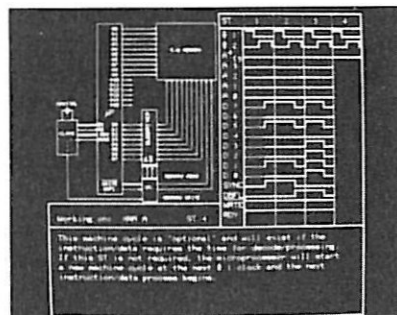
vanskelige spørsmål som kunden stiller.

Som et referanseverk

Stadig mer komplisert utstyr i industrien stiller stadig større krav til opplæring av de som skal bruke det. En interaktiv video-stasjon gjør det mulig å bygge et edb-styrt referanseverk som kan holdes à jour og dessuten lett kan brukes i ordinær opplæring. Teknikken kan brukes som en kombinert brukerhåndbok og opplæringsressurs for ansatte i bedrifter der en har et stort antall rutiner som skal beherskes, eller viktige sikkerhets- eller beredskapsprosedyrer som hele tiden må holdes ved like.



Eksempler på PLATO-grafikk.



Datamaskinassistert læring

Innen området datamaskinassistert læring bruker en tekst og grafikk, gjerne i kombinasjon med videosekvenser på separat utstyr. Denne kombinasjonen vil sette videoteknikken inn i en bedre pedagogisk ramme ved at den nå kan gjøres interaktiv og eleven kan dermed få en aktiv rolle i utnyttningen av mediet.



Cand. mag. Rune Midtvedt er systemkonsulent ved LOGOS A/S hvor han har PLATO som hovedbeskjeftigelse.

EDB OG HUMANIORA

Teknikk og undervisning

Intervju med AV-sjef Lars S. Wilhelmsen, Universitetet i Bergen

Jostein H. Hauge

Audiovisuell avdeling, Universitetet i Bergen ble startet i 1972 som en felles serviceavdeling for Universitetet. Avdelingen har i dag 6 medarbeidere. Lars Skjold Wilhelmsen har vært AV-sjef siden avdelingen ble opprettet. Han har tidligere arbeidet som rektor ved Studentersamfunnets Fri Undervisning i Bergen og undervisningskonsulent ved Det medisinske og Det odontologiske fakultet, Universitetet i Bergen. Wilhelmsen har vært en av de drivende krefter bak arbeidet med universitetspedagogikk i Norge, og han er for tiden redaktør av tidsskriftet UNIPED.

Kan du si litt om hvilke oppgaver AV-avdelingen har?

Vår oppgave er to-delt: For det første skal vi være en serviceinstitusjon for produksjon av læremidler, informasjonsprogrammer osv. ved hjelp av video, foto og lyd. Dessuten har vi en viktig oppgave i å arrangere kurs i bruk av denne type teknologi for pedagogiske formål og for formidlingsarbeid i det hele. Vi har sett det som vesentlig ikke bare å være en «inn- og utlånsentral», men å sette bruken av programmer og utstyr inn i en **pedagogisk ramme**.

Er det forskjell på bruken av AV-tjenester mellom fakultetene?

Det må vel sies at de fagene som bygger på naturvitenskapelig metodikk til nå har vært de største brukerne – kanskje fordi en del av deres lærestoff umiddelbart egner seg for visualisering. Forskerne i disse fagområdene har vært vant til å bruke bildestoff og annen visualisering, tenk bare på bruken av menneskekroppen i medisinsk undervisning.

Men har ikke også de humanistiske fag behov for visualisering og for formidling via lyd?

Det er klart. Jeg kan her f.eks. vise til språklaboratoriene. Men ellers er det vel rett å si at HF- og SV-fagene til i dag i hovedsak har formidlet kunnskap gjennom ord. Men det foreligger også et overbevisende erfaringsmateriale som viser at f.eks. bruk av video kan være et hjelpemiddel i språkopplæringen i fremmedspråk. Her i Bergen er det



Lars Skjold Wilhelmsen

bl.a. laget studentproduserte opptak som skal forestille hverdagssituasjoner i et tysktalende land som et utgangspunkt for samtale og språktrening i kollokvegrupper. Effekten er dobbel: studentene har språklig utbytte både av produksjonen og av drøftingene knyttet til visningen av videofilmen.

Hvilke teknologiske hjelpemiddel vil bli særlig interessante i årene som kommer for humanistisk fag- og forskningsformidling?

Spørsmålet er galt formulert, slik jeg ser det. Det er lite formålstjenlig å liste opp hvilket utstyr som for tiden egner seg best for humanistisk fagformidling. Et riktigere utgangspunkt er å spørre hva formidlingsoppgavene i humanistisk forskning består i. Hva ønsker en å formidle? Hvordan er strukturen i lærestoffet? Hvilke funksjoner skal lærestoffet tjene? Skal studentene bli i stand til å reprodusere fakta eller utvikle ferdigheter til problemløsning innenfor et felt? Hvilken interaksjon er nødvendig mellom lærer og studenter? Når slike spørsmål er besvart, kan vi komme inn på de pedagogiske og teknologiske virkemidler.

Men la oss likevel – med de forbehold du har tatt – komme litt nærmere inn på teknologien. Hvor og hvordan kan edb-teknikken komme inn i det universitetspedagogiske arbeid?

For å forstå det nye som edb-teknikken bringer inn, må vi huske på at tradisjonell, teknisk basert formidling har vært énveis kommunikasjon. Læreprogrammet fortalte deg alltid i detalj hva du skulle gjøre ved en godt strukturert og faseinndelt oppbygging og alltid i en lineær rekkefølge av spørsmål-svar-sekvenser. Det nye som bruk av edb-teknikk kan føre til, er at du kan programmere læreprogrammet til å reagere på forskjellige måter alt etter hvilke svaralternativ som studentene kan tenkes å gi. Tradisjonelle læreprogrammer har i så måte vært lukkede systemer.

Med edb aner vi muligheten av å skape i alle fall en viss grad av åpenhet. Da begynner vi å nærme oss lærerens suverene fortrinn i en undervisningssituasjon. Hans eller hennes undervisning er åpen i den forstand at den i alle fall i prinsippet kan gi rom for spontane reaksjoner tilpasset det som skjer i læresituasjonen.

Er det edb-baserte læreprogram med forgreinete læringsveier som representerer de nye muligheter?

Nei, egentlig ikke. Edb-baserte system med forgreining er også i en viss forstand lukkede systemer. Riktignok kan du bli ledet gjennom læreprogrammet på flere ulike måter avhengig av de svar du gir på ulike steg. Men likevel arbeides det innenfor et begrenset sett av strukturelle alternativ. Et åpent læreprogram skal mest mulig være identisk med den problemløsning som foregår når mennesker møter og søker å løse et problem, f.eks. i en faglig sammenheng. F.eks. kan vi tenke oss et undervisningsprogram der en ved hjelp av en rekke ulike edb-hjelpemidler (f.eks. primærkildedata) studerer et historisk spørsmål. Det viktigste blir da ved hjelp av informasjonssøking på edb å velge ut, sammenstille og bearbeide intellektuelt og kvantitativt det materialet som er til rådighet. Et slikt undervisningsopplegg kan si oss mye om studentenes arbeidssett, problemløsningsstrategier og evnen til å arbeide metodisk. Samtidig som akkurat denne type læring er en god forberedelse til det kommende, såkalte informasjonssamfunnet.

Men vil det studentene lærer i dette tilfelle, være direkte overførbart til bruk i f.eks. en yrkesmessig sammenheng?

Nei, det er klart at virkeligheten gjennomgående er langt mer kompleks enn vi klarer å reprodusere i et lærings-program, men typiske trekk av problemløsning vil kunne «simuleres» om vi kan uttrykke det slik, og være en meget viktig stimulans til den metodeundervisning som gis i de fleste fag.

Vi kan nå registrere at stadig flere ønsker å lage læringsprogrammer for skole- og universitetsbruk, f.eks. for mikrodatamaskiner. Hva synes du om denne utviklingen?

Ekspimentering er en god ting, men jeg er bekymret for den retning arbeidet kan ta. I dag sitter mange fagfolk i ensomhet og pønser ut læreprogrammer. Jeg tror at pedagogisk velfunderte læringsprogram vil forutsette en høy grad av teamarbeid og tverrfaglig innsikt. Det hadde vært en meget god idé å skape fora på universitetsnivå for edb-støttet opplæring der både emnespesifikk innsikt, pedagogisk kunnskap og edb-teknisk ekspertise kunne være representert. Det er denne tre-enigheten som skal til for at et læreprogram kan bli godt. Jeg er også litt skeptisk til den vekt det synes å legges på programmeringsferdigheter. Den vanlige edb-bruker i fremtiden vil ikke drive med programmering.

Edb-teknologien er i stadig utvikling, og nå er den optiske videoplaten i kombinasjon med mikrodatamaskin lansert som et nytt verktøy også for læring.

Ja, slik jeg ser det, åpner den optiske videoplate i kombinasjon med databehandling for helt nye muligheter i en læringsituasjon. Den kompakte lagringsform og den hurtige fremsøking av informasjon som her tilbys, kan bidra til å bryte ned den lineære form som konvensjonelle læringsprogrammer har. Samtidig gjør videoplaten det mulig også å bruke billedstoff og film i kombinasjon med datateknikk. For humanister bør relevansen av denne teknologi være åpenbar både for fag- og formidlingsarbeid. Bruken av bildemateriale er fundamental i undervisning i de estetiske fagene og i en rekke av de øvrige humanistiske fag. I språkfagene kan videoplate bety at du får et samspill mellom det visuelle og det auditive inntrykk. Her kan jeg vise til det eksemplet som er gitt i Elin Solstrands artikkel om videoplaten i *Humanistiske Data* 3-83, der det gis eksempler på bruk av denne teknologien i et opplæringsprogram i spanskundervisning.

Men for å lage slike, vi kan kalle dem «halvåpne» systemer, må det investeres store ressurser?

Ja, utvilsomt. Og det er også et argument for å utvikle teamarbeid i fremstilling av slike læreprogrammer. La meg nevne at vi selv har planer om å eksperimentere med videobånd og videoplate i forbindelse med utvikling av undervisningsopplegg i kunsthistorie. Her håper vi også på et samarbeid med NAVFs EDB-senter for humanistisk forskning når det gjelder det edb-baserte læreprogrammet og koplingen mellom læreprogram og videomedier.

Etter hvert vil teknisk utstyr, inklusive datautstyr, bli langt vanligere i skoleverket enn tidligere. Vil dette få konsekvenser for universitetsundervisningen?

Ja, dette spørsmålet opptar meg for tiden. Hva skjer når «data-barna» om en tid melder seg på universitetet? Da vil de som starter på universitetsnivået som den naturligste ting av verden betjene terminaler for en rekke formål. Hvordan skal vi utnytte disse ferdighetene i vår undervisning? Det er lite trolig at vi i det teknologisk varierte mediesamfunnet som nå er i ferd med å realiseres, kan fortsette å basere oss på den tradisjonelle kateterundervisning på universitetet. En klok politikk fra universitetets side ville være snarest mulig å gjennomdrøfte denne situasjonen praktisk og pedagogisk for å forberede oss på den. Skal universitetet på noen måte kunne forberede for senere yrkespraksis, må vi snarest mulig omstille oss til arbeidsmåter som vil bli vanlige, ja selvsagte i samfunnet i fremtiden.

Hvordan har etterspørselen vært etter de tjenester som AV-avdelingen kan tilby, f.eks. videoproduksjon?

Her er vi inne på et litt forunderlig forhold. Vi har lenge hatt utstyr for ulike typer bildefremstilling, men det er først i det siste – etter at videoteknikken har spredd seg til de tusen hjem, og utleiebutikkene for videofilm har skutt opp på de underligste steder – at vi har fått en plutselig og markant stigning i etterspørsel etter denne type tjenester. Nå kommer f.eks. tekstbehandling og bruk av mikrodatamaskiner for

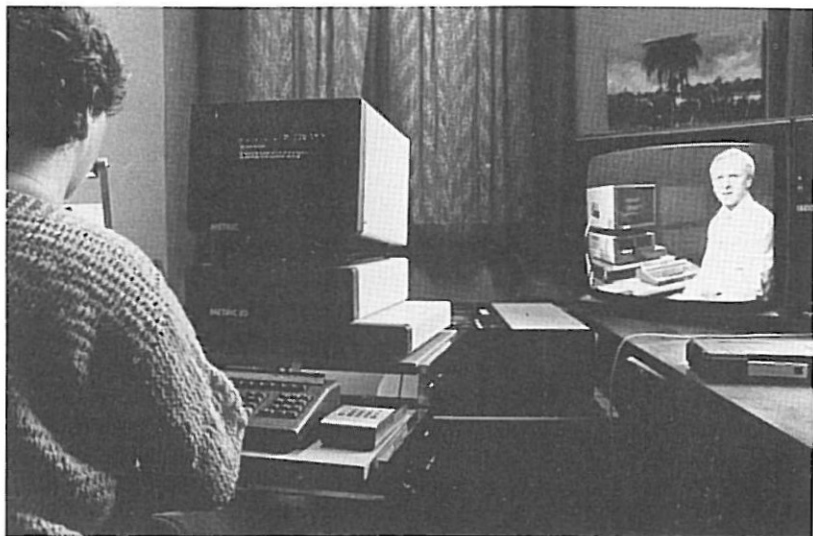
fullt også innen de akademiske kretser. Vi venter derfor også nå en kraftig etterspørsel etter edb-baserte læremidler og ikke minst pedagogiske strategier for fremstillingen.

Men har Universitetet – og AV-avdelingen for den del – vært offensive nok i å tilby tjenester?

Handlingsplaner er tydeligvis ikke nok. Det virker som om det er påvirkningen fra samfunnet utenfor som preger etterspørselen etter AV-tjenester ved Universitetet. På den annen side tas det også nye initiativ innenfor Universitetet. Her i Bergen kan vi nevne faget «Medie- og massekommunikasjon» som i det siste er vokst frem som et fellesfag mellom HF- og SV-fakultetene. Kvalifisering i formidling via nye medier vil være et viktig innslag. Det nye tilbudet er for øvrig typisk for utviklingen. Det er i grenseflatene mellom flere fag at det faglig spennende skjer i våre dager.

Men satser AV-avdelingen selv aktivt for å følge med f.eks. når det gjelder de muligheter edb frembyr i undervisningssammenheng?

Ja, etter fattig evne gjør vi det. I 1982 arrangerte vi f.eks. et seminar om edb-orientert undervisning. Vår hovedoppgave vil for øvrig i hovedsak være å gi miljøene pedagogisk stimulans. Vi har ikke kapasitet til å drive edb-teknisk utviklingsarbeid. Men det er også klart at det i vårt universitetssystem ligger tydelige begrensninger for hvor langt vi kan nå.



Utprøving av et selvinstruerende fjernsynsprogram i tekstbehandling laget for HF-fakultetets EDB-seksjon.

Hva tenker du spesielt på?

Stikkordsmessig kan jeg kalle det belønningssystemet på universitetet. I dag har ikke pedagogisk utviklingsarbeid samme status som forskningsvirksomheten på universitetet. Så lenge meriteringssystemet vesentlig er bygd opp omkring det å kunne legge på bordet lærde avhandlinger, vil vi ha problemer med å fremme våre ønsker. Vi har mange eksempler på at universitetspersonale gjennom våre universitetspedagogiske kurs blir entusiastiske og motiverte for å sette i gang en mer systematisk og bedre formidlings- og undervisningsvirksomhet. Men det er ikke skapt tilstrekkelig positiv holdning til denne virksomhet rundt på instituttene til at interessen kan holde seg når forskerne kommer tilbake til den normale «institutt dagen».

Det akademiske kollegium kan her gi et viktig bidrag, både ved sin faglige tyngde og gjennom muligheten til å sette av midler til pedagogisk utviklingsarbeid og etablere permisjonsordninger. Slike tiltak vil i forskersamfunnet bli tolket som en signalisering om at denne virksomhet også er meget viktig. Det vil være en legalisering, så å si.

For øvrig gleder det oss som arbeider på dette feltet at Rådet for humanistisk forskning i NAVF har vært opptatt av formidlingsproblemer innenfor humanistisk forskning. Rådet har holdt en konferanse om bruk av formidling via film og video, og har nylig etablert en stipendordning for humanistiske fagmedarbeidere som ønsker å gjennomføre et formidlingsprosjekt. Dette er nye signaler, og vi håper på flere slike fra andre hold i fremtiden.

RAPPORTER

Edb-prosjekt for kunst- og kulturhistoriske museer

Jon Birger Østby

Bakgrunn

I samarbeid med NAVFs EDB-senter for humanistisk forskning har museene drevet prøvevirksomhet med edb siden midten av 70-tallet. Dette arbeidet har vesentlig dreiet seg om registrering av gjenstandsdata og katalogproduksjon. Norske Kunst- og Kulturhistoriske Museer (NKKM) har hatt en egen edb-komite som har hatt ansvar for å tilrettelegge bruk av edb, og komiteen har utarbeidet et katalogkort for gjenstander. Kortet kan benyttes for optisk lesing, men det kan også med fordel benyttes av museer som ikke har planer om å bruke edb. Kortet er i dag tatt i bruk av ca. 40 institusjoner, men det er bare 10 av disse som har lagt inn data på edb.

Språksituasjonen her i landet med to skriftspråk og alle dialektale synonymymer og varianter av ord, skaper spesielle problemer ved bruk av edb. For å ta fatt på dette problemet har det vært drevet et treårig prøveprosjekt for nomenklatur for etnologisk gjenstandsmateriale, og det arbeides nå for å få en fast stilling til å følge opp dette arbeidet.

Databehandlingen har i vesentlig grad skjedd ved EDB-senteret, men i de siste årene har flere museer anskaffet egne mikromaskiner. Dette har gitt muligheter for databehandling i egen regi, og denne utviklingen har skapt behov for nye støtte- og samarbeidstiltak. Det er utsendt et spørreskjema til de kunst- og kulturhistoriske museene om bruk og planer for bruk av edb. Pr. 4. juni er det kommet 131 svar. 9 kulturhistoriske museer opplyser at de har tatt edb i bruk til gjenstandsregistrering. Disse museene har til sammen lagt inn data for 37.000 gjenstander, mens de har en samlet gjenstandsmasse på ca. 300.000 nummer. 50 museer oppgir at de har planer om å ta i bruk edb til registrering av gjenstander og/eller foto, men for de fleste av disse museene er tidsrammen for gjennomføring av planene noe diffus. Den

vesentligste årsaken til dette ser ut til å være problemer med å skaffe midler til innkjøp av utstyr, og noen museer svarer at de vil vente på opplegget fra dette edb-prosjektet. Men 17 museer har mer konkrete planer og mener at de skal komme i gang med registrering på edb i 1984 eller 1985. (De arkeologiske museene er ikke tatt med i denne oversikten.)

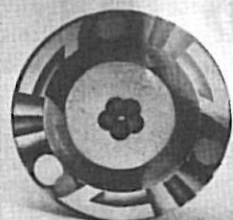
Med øket bruk av edb rundt om i museene er det nå et sterkt behov for å utvikle et standardopplegg for registrering, sortering og utlising av data på mikromaskiner, og etter søknad fra NKKM har Rådet for humanistisk forskning, NAVF, opprettet et toårig forskerstipend for dette området. Stipendet er knyttet til NAVFs EDB-senter.

Målsetting

Hovedmålsettingen for prosjektet er å videreutvikle felles opplegg og rutiner for bruk av edb ved museene. Det er et mål å samordne edb-oppleggene slik at museene i størst mulig grad kan samle seg om felles løsninger både med tanke på selve datateknikken og det maskinlagrede kildematerialet. Med felles edb-opplegg vil vi lettere kunne gjennomføre felles opplæring, det vil være enklere å utveksle data og vi vil kunne dra større nytte av hverandres erfaringer på edb-sektoren.

Det er planlagt at prosjektet skal ta opp følgende arbeidsoppgaver: 1. Utarbeide oversikt over museenes behov for datatjenester og gjøre opp status for bruk av edb ved museene i dag. I denne forbindelse er det

NORSK FOLKEMUSELANS SAMLING					
01	SLIKETID	Asjett, 6	04	ÅR	1979
02	SLIKETID	Fruktasjett ?	05	ÅR	1979.03
03	SLIKETID	Fruktasjett ?	06	ÅR	1979
04	SLIKETID	1935 ca.	07	ÅR	NF.1981-75/80
05	SLIKETID	(01) Oslo	08	ÅR	264
06	SLIKETID		09	ÅR	
07	SLIKETID	(21) Rosander, Gerd Aarland, museumslektor, HF.	10	ÅR	
08	SLIKETID	(11) Vu BD	11	ÅR	
09	SLIKETID	(41) Andreassen, Aslaug, sykepleier, Alders Hvile, Cicinengt. 1, Fredrikstad	12	ÅR	
10	SLIKETID		13	ÅR	
11	SLIKETID		14	ÅR	
12	SLIKETID	Stentøy	15	ÅR	
13	SLIKETID	Presset	16	ÅR	
14	SLIKETID	OVERF: Blyglasur, gul begjning	17	ÅR	
15	SLIKETID	Rund, flat	18	ÅR	
16	SLIKETID	Gul med sort dekor	19	ÅR	
17	SLIKETID	Sprøytet	20	ÅR	
18	SLIKETID	Funkis	21	ÅR	
19	SLIKETID		22	ÅR	
20	SLIKETID		23	ÅR	
21	SLIKETID	T Vu BD Tor i sirkel, 7602E, L&I/6	24	ÅR	
22	SLIKETID	DIAM: 16.2, H: 1.8	25	ÅR	
23	SLIKETID		26	ÅR	
24	SLIKETID	God	27	ÅR	
25	SLIKETID		28	ÅR	
26	SLIKETID		29	ÅR	
27	SLIKETID		30	ÅR	
28	SLIKETID		31	ÅR	
29	SLIKETID		32	ÅR	
30	SLIKETID		33	ÅR	
31	SLIKETID		34	ÅR	
32	SLIKETID		35	ÅR	
33	SLIKETID		36	ÅR	
34	SLIKETID		37	ÅR	
35	SLIKETID		38	ÅR	
36	SLIKETID		39	ÅR	
37	SLIKETID		40	ÅR	
38	SLIKETID		41	ÅR	
39	SLIKETID		42	ÅR	
40	SLIKETID		43	ÅR	
41	SLIKETID		44	ÅR	
42	SLIKETID		45	ÅR	
43	SLIKETID		46	ÅR	
44	SLIKETID		47	ÅR	
45	SLIKETID		48	ÅR	
46	SLIKETID		49	ÅR	
47	SLIKETID		50	ÅR	
48	SLIKETID		51	ÅR	
49	SLIKETID		52	ÅR	
50	SLIKETID		53	ÅR	
51	SLIKETID		54	ÅR	
52	SLIKETID		55	ÅR	
53	SLIKETID		56	ÅR	
54	SLIKETID		57	ÅR	
55	SLIKETID		58	ÅR	
56	SLIKETID		59	ÅR	
57	SLIKETID		60	ÅR	
58	SLIKETID		61	ÅR	
59	SLIKETID		62	ÅR	
60	SLIKETID		63	ÅR	
61	SLIKETID		64	ÅR	
62	SLIKETID		65	ÅR	
63	SLIKETID		66	ÅR	
64	SLIKETID		67	ÅR	
65	SLIKETID		68	ÅR	
66	SLIKETID		69	ÅR	
67	SLIKETID		70	ÅR	
68	SLIKETID		71	ÅR	
69	SLIKETID		72	ÅR	
70	SLIKETID		73	ÅR	
71	SLIKETID		74	ÅR	
72	SLIKETID		75	ÅR	
73	SLIKETID		76	ÅR	
74	SLIKETID		77	ÅR	
75	SLIKETID		78	ÅR	
76	SLIKETID		79	ÅR	
77	SLIKETID		80	ÅR	
78	SLIKETID		81	ÅR	
79	SLIKETID		82	ÅR	
80	SLIKETID		83	ÅR	
81	SLIKETID		84	ÅR	
82	SLIKETID		85	ÅR	
83	SLIKETID		86	ÅR	
84	SLIKETID		87	ÅR	
85	SLIKETID		88	ÅR	
86	SLIKETID		89	ÅR	
87	SLIKETID		90	ÅR	
88	SLIKETID		91	ÅR	
89	SLIKETID		92	ÅR	
90	SLIKETID		93	ÅR	
91	SLIKETID		94	ÅR	
92	SLIKETID		95	ÅR	
93	SLIKETID		96	ÅR	
94	SLIKETID		97	ÅR	
95	SLIKETID		98	ÅR	
96	SLIKETID		99	ÅR	
97	SLIKETID		100	ÅR	



NKKMs katalogiseringskjema for keramikk.

utsendt et spørreskjema til alle museer som er medlem av NKKM.
2. Utarbeide standardopplegg for registrering, søking og utlistering av data på mikromaskiner. Dette opplegget vil baseres på NKKMs katalogkort og på databasesystemet Dataflex. Med bakgrunn i opplegget for gjenstandsregistrering er det planlagt å utarbeide opplegg for fotografier og billedkunst.

3. Komme med forslag til organisering og plassering av en sentral eller flere regionale databaser for museene. For å få erfaringer på dette feltet skal det gjøres forsøk med en felles database for flere museer. Det vil bli dannet en arbeidsgruppe på ca. 10 museer som skal levere data til denne basen. Forsøket skal gi erfaringer både på den tekniske siden med overføring og innlesing av data og for hvordan det vil være å arbeide med en «felleskatalog» for flere museer.

4. Presentere opplegg for informasjon og opplæring i bruk av edb for museene. I prosjektperioden skal det avholdes kurs i registrering av data, sortering og søking på mikromaskiner.

5. Vurdere nytt datautstyr som kan være til nytte for museene. Her står utstyr for lagring og gjenfinning av bilder som et sentralt felt. Gjenstandsforskning er et fagområde der det visuelle inntrykk spiller en viktig rolle. Registreringsteksten kan ikke erstatte fotografiet, og det er derfor en stor fordel med et system der både tekst og bilde kan fremkalles på skjerm.

6. Vurdere bruk av edb i andre sider av museumsarbeidet (bibliotek, arkiv, tekstbehandling, regnskap, budsjett etc.) Dette er felt som ligger noe på siden av selve prosjektet, men for museene er det viktig å foreta en samlet vurdering av hvilken nytte de kan ha av edb på de forskjellige sektorer. En slik vurdering vil gi bedre utnyttning av de relativt store investeringer som mange museer nå planlegger på edb-sektoren. For å kunne dra nytte av edb til søking eller sortering av gjenstandsmateriale kreves først et omfattende arbeid med å legge inn data. Men på andre områder, som f.eks. tekstbehandling, vil det være mulig å få en umiddelbar rasjonaliseringsgevinst.

Prosjektgruppe

For å lede prosjektet er det oppnevnt et utvalg bestående av konservator Karl Ragnar Gjertsen (Nordmøre Museum), formann, førstekonservator Magne Velure (Hardanger Folkemuseum), fylkeskonservator Nils Georg Brekke (Fylkeskonservatoren i Hordaland), konservator Liv Hilde Boe (Forbundsmuseet i Akershus) og direktør Jostein H. Hauge (NAVFs EDB-senter).

Jeg begynte i forskerstipendet 1. april i år og har arbeidsplass ved Norsk Folkemuseum. Men både planlegging og gjennomføring av prosjektet utføres i nært samarbeid med EDB-senteret, der førstekonsulent Per Vestbøstad er fast kontakt. EDB-senteret har hatt en helt sentral rolle i arbeidet med innføring av edb ved museene, og det er

svært viktig å bygge videre på denne kontakten slik at vi også i fremtiden kan dra nytte av Senterets kompetanse.



Jon Birger Østby har mange års erfaring fra vitenskapelig museumsarbeid, bl.a. som leder for Nomenklaturprosjektet for etnologisk gjenstandsforskning. Dette arbeidet resulterte i nomenklaturer for spise- og serveringsbestikk, drikkestell og melkestell.

Forskerstipend i automatisert arkivinformasjon

Hege Brit Randsborg

Rådet for humanistisk forskning, NAVF, har lenge vært opptatt av den sentrale betydning arkivinstusjoner har for store grupper av humanistiske forskere. På grunn av manglende/mangelfull registrering og katalogisering av arkivmassen, har disse ofte blitt hemmet i sitt arbeid for å utnytte viktig kildetilfang i forskningssammenheng. Både innstillingen til Primærkildekomiteen (1972) og til komiteen for «Edb og primærkilder» (1977) pekte på behovet for å kopiere og distribuere kataloger over primærkildematerialet til forskningsinstusjoner. Ved å ta i bruk edb som teknisk hjelpemiddel kan kopiering/distribuering skje mye raskere enn i dag. På denne bakgrunn har Rådet for humanistisk forskning opprettet et forskerstipend i «Automatisert arkivinformasjon» (varighet: 1 1/2 år) der målsettingen er å

«... utvikle edb-baserte metoder for å bedre oversikten over og tilgangen til Arkivverkets kataloger og arkivmasse, med vekt på den forskningsmessige bruk av kildematerialet».

Forskerprogrammet er knyttet til NAVFs EDB-senter for humanistisk forskning.

Med E. Nysæters «Norske arkivkataloger» (1983) har vi fått en oversikt over hva som fins av kataloger og registre i de forskjellige arkivinstusjoner. Vi vet nå f.eks. at det fins en katalog for Statens

Krigskaskoforsikring 1940-1970, men ikke noe nærmere om hvilke arkivsaker som fins. En naturlig fortsettelse av Nysæters arbeid vil derfor være å overføre katalogene i sin helhet til maskinleselig form.

Ved å ta i bruk datamaskinen i katalogarbeidet, oppnår vi en rekke fordeler. Rettelser og tilføyelser er enklere, kopiering blir langt lettere og gjenfinning av opplysninger skjer fortere. Man kan foreta innskrivingsarbeidet på prinsipielt 2 måter

- 1) Innskriving ved hjelp av tekstbehandlingsprogrammer.
- 2) Innskriving ved hjelp av et registrerings skjema.

Tanken bak stipendiet er å opprette en database der informasjon om de enkelte arkiv i de forskjellige arkivinstitusjoner er lagret. Hovedtyngden i en slik database vil i første omgang utgjøres av katalogene. Man skal kunne søke i databasen etter bestemte kildetyper, emner o.l. Siden dette er siktepunktet med å ta i bruk edb, må vi vurdere hvilken av de to metodene ovenfor som er best egnet.

Metode 1 krever lite av programutrustning på datamaskinen og skiller seg ikke vesentlig fra den innskriving som i dag foretas på skrivemaskin. Metode 2 krever spesialprogrammering og mer forarbeid av arkivaren når han/hun skriver katalogkonseptet. Men denne metoden har den fordel at det innskrevne materialet kan utnyttes mye bedre. Derfor har jeg valgt metode 2 og arbeidet med å utvikle et registrerings-skjema som kan brukes til innskriving av katalogopplysningene.

Hvordan skal et slikt registrerings skjema se ut? Skal det være ett universelt skjema (dvs. et skjema som brukes på alle typer av kataloger) eller skal hver katalog ha sitt skjema? Det første er den enkleste løsningen da man ikke trenger mer enn ett utskriftsprogram. Men da får vi et annet problem. Hvordan skal et slikt skjema struktureres? Det er flere hensyn som må taes her. Skjemaet må være

- 1) enkelt å fylle ut.
- 2) inneholde *alle* opplysninger, både de som trengs for å skrive ut katalogen og de som trengs for å søke i databasen.
- 3) gjøre det enkelt å søke i databasen.

Disse tre hensyn står i en viss grad i motsetning til hverandre. Enkelhet i innskrivingsfasen må avveies mot enkelhet i søkefasen. Hvis skjemaet blir komplisert å bruke, vil ingen være interessert i å benytte det, og drømmen om en felles database faller i grus. Det må også være fleksibelt slik at brukeren til en viss grad kan bestemme utseendet på den katalogen som skrives ut.

Utgangspunktet for et registrerings skjema må være de kataloger som fins i dag. De er organisert etter et visst mønster. De inneholder en innholdsfortegnelse, en innledning, selve fortegnelsen over arkivdokumentene og av og til et stikkordsregister. Innledningen er foreløpig holdt utenom (de er såpass ulike at standardisering er vanskelig). Derfor skal registrerings skjemaets opplysninger kun brukes til å lage fortegnelsen over arkivet (den løpende katalogen), innholdsfortegnelsen og

stikkordsregisteret.

Mitt registreringsskjema (se figur 1) er bygget opp av ulike typer felt. Noen felt inneholder de opplysningene som skal skrives ut i den løpende katalogen (f.eks. hvilket nr. pakken eller protokollen har, hvilken hylle i magasinet den står i), andre felt skal lette søkingen i databasen (f.eks. stikkord til arkivdokumentet), atter andre skal inneholde opplysninger om hvordan katalogsiden skal se ut (f.eks. at innholdet i et felt skal skrives ut med 10 tegn i bredden. Da vil et skjema med 14 tegn i dette feltet bli skrevet ut over 2 linjer). Siden opplysningene skal legges i en database sammen med kataloger fra andre arkivskapere, må vi også ha med felt som forteller hvilken arkivskaper katalogen gjelder for.

Arbeidet med registreringsskjemaet bygger på bruk av ferdigprogrammet DATASTAR for mikromaskiner under CP/M. Ulempen ved å

KATALOG FORM LISTING AND FIELD ATTRIBUTE DEFINITIONS

FIELD NUMBERS

1.1 Løpenummer: _____ 5
1.2 Kode : 5 1.3 Format: _____ 7

2.1 Enke/protokoll nr: _____ 8
2.2 Serie: _____ 9

Arkivstykkets art - innhold:

3.1: _____ 10
 _____ 11
 _____ 12
 _____ 13
3.2: _____ 14
 _____ 15
3.3: _____ 16 3.4: _____ 17

4.1 Tid : 18- 19

5.1 Hyllenr: _____ 20

6.1 Merknad: _____ 21
 : _____ 22

6.2 Klausul: #

Stikkord.

7.1 Geogr. omr: _____ 24
7.2 Personnavn: _____ 25
7.3 Emne : _____ 26

Ekstrafelt

8.1 Fra felt nr: 27
8.2 Tekst: _____ 28
 _____ 29

Arkivskaper:

9.1 Institusjonsnr: _____ 1
9.2 Inst. navn : _____ 2
9.3 Avdeling : _____ 3
9.4 Kontor : _____ 4

Figur 1.

bruke et ferdigprogram er at man ikke alltid klarer å få data skrevet ut igjen på den formen man ønsker. Derfor har jeg også arbeidet med et eget utskriftsprogram skrevet i PASCAL. Problemet med egne programmer skrevet for mikromaskiner er at maskinens hukommelse kan bli for liten. Jeg har tatt utgangspunkt i en ALTOS mikromaskin (8-bits maskin), og den var ikke stor nok til å få inn i programmet en fornuftig måte å dele ord på. Dette er en begrensning som er irriterende. Jeg løste dette problemet ved å lese ord for ord av teksten og teste om det var plass igjen på linja før ordet ble skrevet ut. Hvis ikke, ble hele ordet overført til neste linje. Senere har jeg fått anledning til å prøve en IBM PC (16-bits maskin), og denne har stor nok hukommelse til å kunne foreta fornuftige orddelinger. Figur 2 viser utskrift av en katalogside laget av utskriftsprogrammet på IBM PC.

NUMEDAL OG SANDSVÆR SORENSKRIVERI		RETTERGANG	
Nummer	Arkivstykkeets art - innhold	Tid	Hyllnummer
B. FORHØRSPROTOKOLLER			
1	Autorisert for sorenskriveriet.	13.02 - 24.05	1810 - 1823 4C 004 45
2		24.07 - 15.06	1823 - 1827
3		21.07 - 26.01	1827 - 1831
4	Autorisert for Flesberg og Rollag.	14.04 - 13.08	1831 - 1835
5		13.08 - 08.02	1835 - 1839
6		13.08 - 07.06	1839 - 1847
7		08.06 - 14.01	1847 - 1853
8		26.01 - 24.12	1853 - 1860
SANDSVÆR			
9	Autorisert for sorenskriveriet.	03.03 - 02.10	1810-1827
10		03.10 10.04	1827 - 1834
11	Autorisert for sorenskriveriet. Sandsvør på bindet	06.04 - 08.10	1835 - 1839
12		08.10 - 03.09	1839 - 1847
13		03.09 - 25.11	1847 - 1856
SORENSKRIVERIET.			
14		26.02 - 14.11	1890 - 1896
15		24.11 - 10.06	1896 - 1897
<< Prk. nr. 5070/5080 7961/7964. >>			

Figur 2.

Både registreringsskjemaet og utskriftsprogrammet er nå ferdig fra min side, det som gjenstår er å prøve det i bruk ved de forskjellige arkivinstitusjonene. Neste fase i stipendarbeidet er å legge de dataene som registreres i skjemaet inn i en database og prøve ut ulike edb-metoder for gjenfinning. Dette arbeidet er så smått påbegynt. Jeg har her tenkt å bruke et fritekstsøkesystem (NOVA*STATUS i påvente av at SIFT skal bli implementert. Disse to systemene er i slekt med hverandre). Slike system krever stor plass i hukommelsen og jeg må derfor bruke en større maskin (som f.eks. et universitetsanlegg).



Cand.philol. Hege Brit Randsborg har tidligere arbeidet med historisk demografi. Sammen med Eli Fure og Sølvi Sogner er hun forfatter av Fra tobarnskull til stua full, Universitetsforlaget, 1984.

Frå informasjon til kulturarv NOU 1984, nr. 3

Jostein H. Hauge

«Oppgåva består i å syta for at kjelder til kunnskap om vår tid blir tekne vare på både for vår samtids studiar av eiga tid, og for dei som kjem etter oss. Dei kjeldene kan vera gjenstandar som er leivningar frå vår tid – og dokument i vidaste meining, som ber tankar og sanseinntrykk vidare som skrift, lyd og bilete i ulike media.

Slikt kulturvern er altså eit slag samtidsteneste for ettertida. Problem-et er at vi i vår samtid ikkje veit kva slag teneste ettertida ønskjer av oss. Vi kan gjetta på kva ettertida vil meina at vi burde ha teke vare på og levert vidare. Men når tida kjem og framtida spør, så vil det vera for seint – dersom vi ikkje har gjetta rett.» (Innstillingen s. 21)

*

Innstillingen «Frå informasjon til kulturarv» er et viktig dokument i arbeidet med å trygge informasjon fra ulike medier for skrift, lyd og bilde til bruk i samtid og fremtid. Det regjeringsoppnevnte utvalget med dosent *Berge Furre* som formann har levert en grundig, handlingsrettet og ikke minst godt skrevet innstilling.

Bakgrunnen for opprettelsen av utvalget er at den Lov om avleveringsplikt som ble vedtatt i 1939 ikke på noen måte lenger kan dekke de dokumentasjonsbehov som finnes i dag. Siden loven ble innført for omlag førti år siden, har den tekniske utviklingen ført til at sentralt dokumentasjonsmateriale blir lagret på, og spredd ved hjelp av, metoder som ikke fantes da loven ble til. Ny teknikk har bl.a. ført til at en stadig stigende del av kultur- og vitenskapsformidlingen nå skjer ved hjelp av dokumenter på andre medier enn papir.

Utvalget som ble oppnevnt i 1980, fikk i mandat å gjøre greie for og vurdere situasjonen i Norge i dag med hensyn til arkivering, registrering og oppbevaring av trykksaker, edb-baserte dokument og lyd- og bildemateriale. Utvalget skulle også gjøre greie for hvordan materialet blir gjort tilgjengelig for dokumentasjonsformål.

Videre skulle utvalget vurdere oppbevaringsverdien for de ulike medietyper, i hvilken grad og hvordan de bør avleveres, registreres, oppbevares og gjøres tilgjengelige, og hvem som bør ta seg av disse oppgavene. Løsningene skulle ta hensyn til nordisk og internasjonal standardisering og samarbeid. Utvalget skulle dessuten kunne foreslå særskilte tiltak for presserende oppgaver og ble bedt om å kodifisere sine forslag i form av et lovforslag for området. For å konkretisere forslagene skulle utvalget dessuten også vurdere de økonomiske og administrative konsekvenser av sine forslag, og sette opp en prioriteringsplan for de fremlagte tiltak.

I innstillingen blir de ulike medietyper gjennomgått systematisk under overskriftene Kombinerte dokument, Mikroformer, Edb-dokument (maskinlesbar informasjon), Lydfestinger, Fotografi, Film, Videogram og Kringkasta materiale.

I denne forbindelse vil jeg konsentrere meg om utvalgets forslag m.h.t. dokumentasjon av materiale på maskinlesbare medier.

Edb-dokument

I kap. 2, pkt. 4 gjennomgår utvalget vanskene med å lage og ta vare på det utvalget kaller edb-dokument, dvs. «informasjon som er tilgjengelig på eit medium som kan behandla informasjon ved å bruka elektronisk databehandling» (s. 52).

I kapitlet blir det gitt et bilde av situasjonen i dag når det gjelder omfang, typer og produsenter av slikt materiale. Utvalget ønsker prinsippielt å betrakte edb-dokument på linje med annen dokumentasjon og vil som for andre materialtyper sondre mellom a) allment tilgjengelige edb-dokument og b) ikke allment tilgjengelige edb-

dokument (arkivalia).

Type a er «edb-dokument som alle kan få tilgjenge til gjennom ei abonnementsordning, ved kjøp eller leige, som gåve eller lån.» Typisk for slike er at de er produsert for å nå en stor gruppe mennesker eller dekke informasjonsbehov for spesielle brukergrupper. Denne type edb-dokumenter kan distribueres ved hjelp av ulike tekniske medier, som f.eks. magnetbånd, disketter o.a., men også via dataterminaler slik situasjonen er for informasjon fra databaser – i fremtiden med énveis og toveis kommunikasjonsmuligheter.

Type b er «edb-dokument som berre er tilgjengelege innanfor eit privat område». Karakteristiske eksempler er her edb-materiale produsert i næringsliv, institusjoner og organisasjoner (intern post, regnskap m.v.) eller edb-materiale som enkeltpersoner utvikler for egne formål. For utvalget er det særleg edb-dokumenter av type a som er av interesse.

Utviklingstrekk

Utvalget slår fast at det alt i dag er slik at en stadig større del av dokumentasjonsbehovene i samfunnet blir løst ved hjelp av edb-teknikk. Både vår komplekse samfunnsstruktur og de nye vitenskapelige fremsteg har i dag edb-teknikk som avgjørende forutsetning. Overgang fra numerisk databehandling til bruk av edb ved ord- og tekstbehandling blir gjennomgått i innstillingen.

Den stadig økende og distribuerte bruk av edb fører til at viktige kildetyper mer og mer bare finnes på en edb-tilgjengelig form. Dette er også typisk for informasjon i form av databaser som ofte utelukkende legges til rette for interaktiv bruk via skjermterminal. Den raske tekniske utvikling viser seg også i utviklingen av stadig nye lagringsmedier. Med rette får den nye optiske videoplate stor oppmerksomhet ettersom dette lagringsmediet betyr et gjennombrudd for ensartet lagring av flere medietyper. På en slik plate kan både tall, tekst, foto, film og musikk lagres og utnyttes i kombinasjon.

Arkivering uten mulighet for gjenbruk av materialet er formålsløst. Derfor gjennomgår utvalget også hvilke ulike datatyper som finnes med særlig vekt på forutsetningene for senere utnyttelse. Utvalget skiller her mellom tre typer:

- programpakker (operativsystem), læreprogram eller spill. Her forutsettes ofte en bestemt type maskin for å kunne bruke materialet.
- enkle dataregistre som lagres på samme måte som de presenteres. Disse vil kunne overføres til de fleste datamaskiner og krever ikke spesialprogrammer for å kunne brukes.
- komplekse dataregistre som er laget på en annen måte enn de presenteres på. Her trengs spesiell programvare for å kunne utnytte data, og ofte programmer knyttet til spesielle maskiner.



NOU
NOREGS OFFENTLEGE UTGREIINGAR

Frå informasjon til kulturarv

NOU 1984: 3

*Korleis tryggja informasjon frå ulike media
for skrift, lyd og bilete til bruk i samtid og framtid?
Framlegg til ny pliktavleveringsordning*

Utvalget peker med rette på at det er viktig at en kan skape en arkiveringssituasjon som også gjør det mulig å utnytte både enkle og komplekse dataregistre i ettertid.

Salg og formidling av informasjon i edb-form er relativt nytt i Norge. Det er imidlertid en raskt stigende interesse for å starte utadrettede edb-baserte dokumentasjonstjenester både i offentlige organer, institusjoner og bedrifter. Derfor er det viktig nå å utforme klare regler for en fremtidig oppbevaringsordning.

Alt i dag finnes det delløsninger på arkiveringssektoren: Riksarkivet utarbeidet i 1979 «Regler for avlevering av edb-materiale fra Statsforvaltningen til Arkivverket». Her er det innført prosedyrer for bevaring/kassasjon og utformet krav til teknisk standard og system- og dokumentasjonsstandard. Norges almenvitenskapelige forskningsråd (NAVF) har fått rammekonsesjon fra Datatilsynet til å lagre og utnytte for forskningsformål registre med personopplysninger. Det arbeides også videre med regler for gjenbruk av denne type materiale (se rapporten i dette nr. av Humanistiske Data).

Tilrådinge

Tilrådingene tar som utgangspunkt at edb-dokument og edb-aktiviserte dokument kan bli en publiseringsform som kan få større omfang og betydning enn det trykt skrift og bilde har i dag.

Utvalget tilrår at edb-dokument skal kreves pliktavlevert etter retningslinjer som så langt det er mulig samsvarer med dem som gjelder for andre medietyper. Det heter her at «Edb-dokument som er gjorde alment tilgjengelige i Noreg [skal] avleverast dersom dei er produserte her, dersom dei er produserte i utlandet for norsk utgjevar eller dersom dei er produserte i utlandet og bearbeidde spesielt for den norske marknaden» (s. 94). Gjennom dette hovedprinsipp mener utvalget å fange opp både (fysisk) distribuerte edb-dokument og datakilder som utnyttes i form av databaser.

Siden elektroniske dokumenter ofte kun har kortvarig levetid, setter utvalget en grense på 30 dager. Det vil si at et slikt dokument må pliktavleveres dersom det har vært allment tilgjengelig i et tidsrom på minst 30 dager.

Hva faller utenfor ordningen?

Unntatt fra reglene om pliktavlevering er imidlertid edb-dokument som er produsert i utlandet av en utenlandsk utgiver, og som ikke spesielt er bearbeidet for det norske markedet. Arkivalia skal som før gå til Arkivverket. Unntak gjøres også for ren programvare, selv om utvalget legger sterk vekt på ønskeligheten av å kunne fange opp viktig programutstyr som trengs for å utnytte det datagrunnlag som samles inn.

Situasjonen i dag

Utvalget slår fast at norskproduserte eller norsktilpassede edb-dokument i dag har et lite omfang, også sammenliknet med det materiale av utenlandsk opprinnelse som brukes. Men denne situasjonen vil endre seg: På undervisningssektoren arbeides det på alle nivåer med produksjon av edb-baserte læremidler. Kommersielt informasjons-salg er kommet i gang. Bruk av ulike typer databehandling i forvaltning og forskning vil øke, samtidig som det også innarbeides nye informasjonstjenester der både informasjon om forskning og selve forsknings-dokumentene presenteres ved hjelp av elektroniske medier.

Arkiveringsforhold

Mens distribuerte edb-dokument foreligger i ferdige utgaver, vil innholdet i databaser stadig endres. Alle versjoner av databasene kan over tid derfor ikke tas vare på. Utvalget foreslår her at dokumentasjon av denne type medier bør avtales gjennom separate avtaler mellom databaseoperatør og avleveringsinstitusjon. Utvalget sier likevel at minimum første og siste utgave av en database avleveres. I en del tilfeller kan det likevel være behov for å få alle versjoner av innholdet i en database overlevert. Det gjelder f.eks. børsnoteringer, nyhetsmeldinger etc. Utvalget ser det som ideelt at både data, program og maskinutstyr kan tas vare på, men finner dette klart urealistisk.

Siden gjenskaping av data på nytt datautstyr vil bli et internasjonalt problem i fremtiden, håper utvalget at det vil bli utviklet stadig bedre teknikker for gjenskaping av edb-basert informasjon, men det vil likevel være nødvendig at det ved avleveringen blir gitt så fyldestgjørende opplysninger som mulig om oppbygging av den enkelte datakilde og de programmer som har vært brukt ved behandlingen av data.

Hvem skal ta imot edb-dokument?

Siden edb-dokument funksjonelt tilsvarer dokumentasjonsmateriale på papir, er det naturlig at Nasjonalbiblioteket får i oppgave å skjøtte avleveringsplikten. For å kunne ta seg av denne oppgave må det opprettes en avdeling for mikroformer og edb-dokument med nødvendig edb-personale, datakraft og økonomi til slikt arbeid.

Siden dokumentasjon og lagring av edb-basert informasjon er et nytt felt i Norge, ber utvalget om at det blir satt i gang et prøveprosjekt som kan klargjøre nærmere de ulike arbeidsoppgaver som en generell avleveringsplikt for edb-materiale vil medføre.

Prosjektet «Norsk litteratur i Ungarn»

Edb-database og bibliografi

Anikó N. Balogh



Dr.phil. Anikó N. Balogh er førsteamanuensis i skandinavisstikk (norsk filologi) ved Institutt for germanske og romanske språk, Universitetet ELTE, Budapest. Hun har utgitt bl.a. en norsk-ungarsk parlør og en antologi over 1900-tallets norske prosa. Sammen med András Masát redigerer hun tidsskriftet Papers on Scandinavian Studies.

Ved Institutt for germansk og romansk språk og litteratur ved Universitetet ELTE i Budapest har det pågått et prosjekt om hvordan skandinavisk litteratur er blitt mottatt i Ungarn fra 1870-tallet til våre dager. Prosjektet kom i gang etter initiativ fra prof. *C.J. Hutterer* og *Anikó N. Balogh* og førte frem til bibliografiske samlinger som siden ble publisert i form av hovedoppgaver og doktoravhandlinger.

Siden 1982 er dette prosjektet knyttet sammen med et stort tverrfaglig forskningsprosjekt, som er støttet av MTA og MM (Det Ungarske Vitenskapsakademiet og Kultur- og Undervisningsministeriet i Budapest) og står under ledelse av prof. *Antal Madl*. Dette prosjektet har som mål å utgi en serie bibliografier om hvordan germansk og romansk litteratur er blitt mottatt i Ungarn etter 1945. Utgavene er ment for internasjonalt bruk, med innledning og forklaringer i de respektive språk.

Dette tverrfaglige forskningsprosjektet muliggjorde – for første gang i moderne humanistisk forsknings historie i Ungarn – at vi kunne tenke på å lage en serie edb-bibliografier og en database over disse bibliografiske samlingene.

De to første årene (1981-82) gikk med til å definere databasens sammensetning og opplysningenes typologi. Hver innførsel følger et skjema som er inndelt etter følgende opplysninger:

1. En alfanumerisk streng som betegner hvilken skandinavisk litteratur som ligger i feltet: N står for norsk litteratur, S – svensk

litteratur, D – dansk litteratur, I – islandsk litteratur. P står for primærlitteratur, S for sekundærlitteratur. Til slutt angis nummeret som innførselen har i registeret.

Alle bibliografiske opplysninger i databasen er redundante, dvs. samme informasjon forekommer på forskjellige steder, for å muliggjøre en indre kontroll i systemet.

2. Datafilens navn: P – primærlitteratur, S – sekundærlitteratur.
3. Genre, som betegner genre-poetiske kategorier (bl.a.: R – roman, D – drama, skuespill, T – teaterpremiere, N – fortelling, novelle, P – dikt, L – brev, B – biografi) og litteraturkritikkens kategorier (bl.a. STUD – fagartikkel, essay, studie, forord, etterord, ART – avisartikkel, kritikk, REV – anmeldelse, M – diverse) og utgivelsens form (bl.a.: ANT – antologi).
4. Forfatterens etternavn og fornavn
5. Primærforfatterens navn
6. Oversettelsens ungarske tittel
7. Originaltittelen på norsk
8. Fagverkets tittel i norsk oversettelse
9. Feltets medarbeidere (TRANSL – oversatt av, SEL – utvalgt av, NOTES – anmerkninger, RED – redigert av, ILL – illustrert av, STUD – forord, etterord, innledning, studie).
10. Tidsskriftets data: (= tittel, årgang, nummer, sidetall)
11. Bokens data: (IN – forfatter, tittel, sidetall)
12. Utgivelsens sted, tid, forlag, sidetall.

De bibliografiske dataene er skrevet inn på en Videoton skjerm og opparbeidet med datamaskinen R 40 i KFKI i Budapest (Sentral Forskningsinstitutt for Fysikk, Budapest). Dataene ble tilrettelagt for videre behandling ved hjelp av KFKIs bibliografi-programpakke, som er utarbeidet av *Magda Zimanyi* (forskningsrådgiver i datafag) og Anikó N. Balogh. Det norske registerets analyser ble avsluttet i 1984.

Fra den skandinaviske edb-databasen er det for første gang blitt mulig å utgi en bibliografi om hvordan norsk litteratur er blitt mottatt i Ungarn. Edb-registeret kan brukes til å gi et kvantitativt sett pålitelig bilde av resepsjonshistorien og til å støtte moderne litteratursosiologisk og kulturteoretisk forskning.

De norske bibliografiske samlingene ble analysert fra databasen etter følgende prinsipper:

1. Norske skjønnlitterære verk i ungarsk oversettelse (a) antologier, samlede verk, b) de enkelte forfattere ordnet kronologisk etter den ungarske utgivelsen, c) de enkelte forfattere ordnet etter hovedgenre, innenfor de enkelte genre alfabetisk etter originalens tittel og kronologisk etter utgivelsesår, d) liste ordnet alfabetisk etter oversetterens navn og kronologisk etter oversettelsen).
2. Faglitteratur, litteraturkritikk om norsk skjønnlitteratur i Ungarn

(a) liste ordnet alfabetisk etter norsk forfatter og innenfor forfatteren ordnet kronologisk etter årstall for den ungarske bok eller artikkel, b) liste ordnet alfabetisk etter ungarske forfattere, c) statistiske edb-data, d) edb-navneregister: forfattere, litteraturkritikere, oversettere, redaktører osv.).

Fra den norske databasen utgis ved ELTE som første publikasjon følgende nevnte registre: 1.a, 1.c, 2.a i løpet av 1984.

Tutorial on Machine Translation

Lugano, 2.-6. april 1984

Jostein H. Hauge

1. Innledning

Konferansen samlet 110 forskere i datalingvistikk, dataspesialister og oversettere, i alt vesentlig fra Europa, USA og Canada.

Leder av konferansen var dr. *Margaret King* fra Dalle Molle Institutt for Semantiske og Kognitive studier (ISSCO) i Genève. Dette instituttet koordinerer for tiden EFs oversettelsesprosjekt, EUROTRA, under dr. Kings ledelse.

Samlingen i april var den tredje i en rekke av datalingvistiske konferanser arrangert av ISSCO, hvor tidligere emner har vært datamaskinell semantikk (1975) og automatisk analyse av naturlig språk (1981).

Programmet besto av tre hovedelementer. Det ble gitt oversiktspresentasjoner av utviklingen innenfor maskinoversettelse fra før den 2. verdenskrig til i dag, og gjennomgåelse av lingvistiske og datatekniske metoder av særlig betydning for feltet. Den tredje hoveddelen av konferansen var sentrert rundt presentasjoner av pågående utviklingsprosjekter eller oversettelsessystemer som er i praktisk bruk i dag.

2. Utviklingen fra starten fram til 1966

I sitt foredrag «Early History of Machine Translation» tok *B. Buchmann*, ISSCO, Genève, for seg utviklingen fra 1933 til 1966. Han tok utgangspunkt i de forsøk som ble gjort i 1933 av franskmannen G. Artsruni og russeren P. P. Smirnov-Trojanskij.

Trojanskis oppfinnelse inkorporerte mange trekk ved maskinoversettelse som også finnes i dag. Han klargjorde også stadier i oversettelsesprosessen som har stått sentralt i teoretiske studier frem til i dag, bl.a. skillet mellom *analyse*, *overføring* (transfer) og *produksjon* (generation), og begrepet *mellomspråk* (intermediary language).

Tanken bak betegnelsen mellomspråk var at på tvers av ulike overflatestrukturer har alle språk et felles logisk innhold som kan brukes ved oversettelse ved at det utvikles et univerelt logisk fellesspråk.

2.1 De første forsøkene med datamaskin

Den egentlige starten av maskinoversettelse kom i 1946-47, da engelskmannen A. D. Booth og amerikaneren D. H. V. Britten utarbeidet et program for detaljert to-språklig koding av ordene i en ordbok og implementerte det i en datamaskin.

For å løse problemene med maskinoversettelse, trodde man at det var tilstrekkelig å oversette de konstante semantiske enhetene i et språk med de tilsvarende i målspråket. Forsøk viste likevel at det var nødvendig med grammatisk informasjon. Problemene med alternative ordbetydninger ble forsøkt løst med *mikroglossarer* tilpasset de teksttyper som skulle oversettes. Denne type strategi for å unngå betydningsproblemer har vært mye brukt i senere systemer.

2.2 Konferanse om maskinoversettelse

Den første konferansen om maskinoversettelse ble holdt ved Massachusetts Institute of Technology (MIT) i 1952. Der ble det lansert en rekke nye metoder for maskinoversettelse. Det ble fra mange hold påpekt at en form for syntaktisk analyse burde inngå i oversettelsesarbeid.

I de forsøkene som fulgte, ble stadig nye strategier forsøkt. Bl.a. ble det gjort prøver med å lette maskinoversettelsen ved å skrive tekstene inn i en enkel og regelmessig syntaks, som også til dels var tilpasset strukturen i det språk som det skulle oversettes til. Slike metoder er senere blitt forsøkt med hell.

Det ble også lansert ideer om en én-til-mange oversettelse i motsetning til de rådende språkparoversettelsene, ved at det ble introdusert et vippespråk (pivot language) som kunne brukes som mellomstasjon i oversettelsen til mange ulike målspråk.

2.3 Georgetown-forsøket

I 1954 ble det gjennomført et eksperiment ved Georgetown University som skulle vise verden at maskinell oversettelse fra russisk til engelsk var mulig. Oversettelsen ble en suksess: den økte interessen for maskinoversettelse enda mer og åpnet nye finansieringskilder. I løpet av de neste ti år ble det investert over 20 millioner dollar av ulike amerikanske

statsinstitusjoner.

En tilsvarende utvikling fant sted i Sovjet, der det fra 1954 av ble etablert en rekke forskningsgrupper som arbeidet med maskinoversettelse. Også i England ble slikt arbeid tatt opp, bl.a. ved Cambridge University.

2.4 Første generasjons systemer

Typisk for de oversettelsesprosjekter som fulgte Georgetown-eksperimentet, var såkalt «brute force empirical approach». Reglene baserte seg i høy grad på et leksikon, og systemene virket slik at *program og data var vevet inn i hverandre*, noe som senere skulle vise seg å være katastrofalt ved utbyggingen av systemene. System av denne type kalles ofte *første generasjons oversettelsessystemer*.

2.5 Større realisme etter 1957

Etter noen års prøving og feiling med blandet suksess, og parallelt med Noam Chomskys publisering av nye metoder for formell språkbeskrivelse i 1957, inntraff det en reorientering i flere oversettelsesmiljøer. Flere som tidligere hadde vært optimistiske, mente nå at man burde oppgi den empiriske og heuristiske tilnæringsmåten og basere seg på mer grunnleggende lingvistisk viten som basis for oversettelsessystemene.



Dr. Margaret King, som ledet konferansen i Lugano, er koordinator for EFs oversettelsesprosjekt.

Det realistiske mål ble nå av stadig flere sett på som enten å skaffe frem grovoversettelse for skumlesing (f.eks. av teknisk litteratur) eller oversettelser som inngår som ledd i et mer omfattende, konvensjonelt drevet oversettelsesarbeid.

2.6 ALPAC-komiteen

I Amerika ble det i 1964 bestemt av National Science Foundation å foreta en evaluering av resultatene av de store investeringer som var foretatt i den foregående ti-årsperioden. Det ble oppnevnt en komité (Automatic Language Processing Advisory Committee), senere kjent som ALPAC-komiteen. Komiteen ønsket å vurdere om maskinell oversettelse kunne konkurrere med tradisjonell oversettelse i en praktisk arbeidssammenheng. Komiteen konkluderer bl.a. med at «... We do not have useful machine translation. Further there is no immediate or predictable prospect of useful machine translation».

Resultatene var øyeblikkelige og katastrofale for forskningsmiljøene. Pengestrømmen tørket inn og tilgjengelige prosjektmidler ble kanalisert inn i mer teoretisk orientert lingvistisk arbeid.

3. Utviklingen fra 1966 til i dag

I sitt foredrag om utviklingen fra 1966 til i dag, la *Effi Amaniadou* og *Susan Warwick* vekt på at effekten av ALPAC-komiteens arbeid også hadde positive virkninger. Arbeidet med maskinoversettelse ved amerikanske universiteter fikk riktignok dramatiske tilbakeslag, men samtidig kom datamaskinell lingvistikk til å få en oppblomstringstid. I det ti-året som fulgte ble det utarbeidet et solid kunnskapsfundament innenfor analyse av naturlig språk, som senere arbeid med maskinoversettelse har nytt godt av.

3.1 Nye forutsetninger

Den enorme kapasitetsøkning som datamaskinene etter hvert fikk og de langt mer sofistikerte edb-metoder førte til et helt nytt fundament for maskinoversettelse. Samtidig førte teknologien til at databehandling av tekstmateriale dramatisk økte mengden av dokumenter som ble publisert.

I Japan er språkoversettelse et sentralt element i kravspesifikasjonen for det nye konseptet for 5. generasjons datamaskiner. I dag har Japan en rekke fremtredende forskningsmiljøer i datalingvistikk og flere avanserte oversettelsesprosjekter. Mange mener at Japan i fremtiden vil få en ennå mer fremtredende posisjon på feltet.

3.2 Ulike typer maskinoversettelse

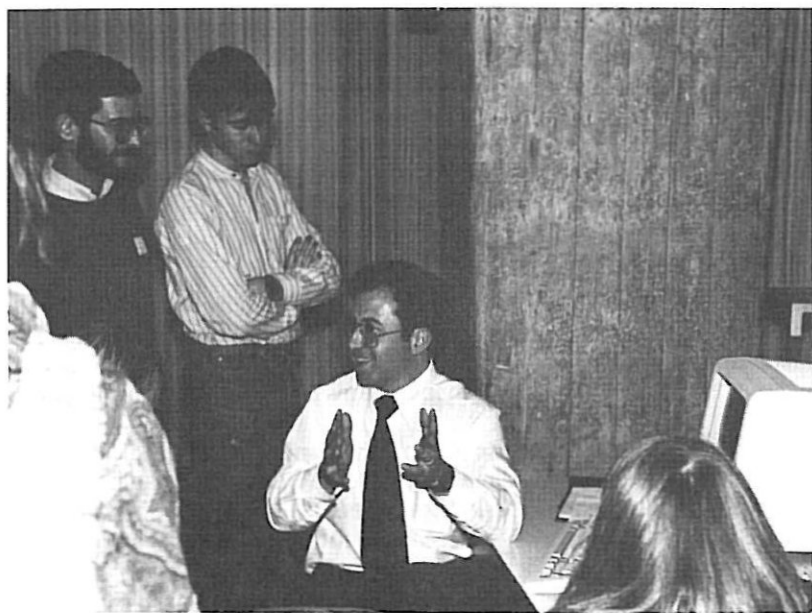
Innenfor maskinoversettelse er det de senere år foretatt en sontring i

ulike former for oversettelse. Mens en tidligere hadde som mål en «ren» maskinoversettelse, er ambisjonene etter hvert senket til *maskinassistert oversettelse* idet en erkjenner at den faglærte, menneskelige komponent er avgjørende. Med «menneskeassistert maskinoversettelse» ønsker en å karakterisere systemer der en spesiell tilretteleggelse av tekstene har funnet sted før oversettelsen (pre-editering), eller systemer som gir oversetteren mulighet til å assistere datasystemet (dvs. løse problemer) under selve utførelsen av oversettelsen.

«Maskinassistert menneskeoversettelse» får en når en datamaskin foretar en oversettelse som siden blir korrigert og rettet (post-editert) av en oversetter. I de fleste tilfelle vil dette være nødvendig, men jf. beskrivelsen av TAUM-METEO-prosjektet nedenfor.

3.3 Andre generasjons systemer

Mens systemene tidligere var empirisk basert og gjerne konsentrert om det enkelte ord, blir nå oversettelsen indirekte. Det betyr bl.a. at det som basis for oversettelsen blir gitt en for formålet fullstendig strukturell beskrivelse av utgangsspråket før oversettelsen finner sted. Oversettelsen foregår nå via en mellomform. Denne mellomformen kan være av typen Interlingua eller foregå etter overføringsprinsippet (transfer, jf. innledningen ovenfor).



Professor Alan K. Melby, Brigham Young University, viser konferansedeltakere sin arbeidsstasjon for oversettere.

3.4 Interlingua

Målet er her å finne en språkuavhengig representasjonsform som kan være møtested for de to språkene. Forsøk på å skape en slik plattform har foregått i mange miljøer (bl.a. Grenoble), men blir nå helst oppfattet som et fjernt utviklingsmål.

3.5 Overføring

En rekke prosjekter arbeider i dag etter denne modellen. Fasene består i 1) utarbeidelse av en strukturell beskrivelse av utgangsspråket som til dels også inkorporerer semantisk informasjon, 2) overføring av strukturelle elementer fra ett språk til et annet og 3) produksjon av tekst i målspråket.

Et sentralt punkt i arbeidet med maskinassistert oversettelse er hvordan en kan øke systemenes slagkraft ved å inkorporere formalisert semantisk kunnskap, inklusiv tekstinherent kunnskap og såkalt kunnskap om verden i den morfologiske og syntaktiske beskrivelse.

3.6 Målområdet

Maskinoversettelse er i dag primært innrettet mot å gi et effektivt bidrag til internasjonal informasjonsoverføring. Utgangspunktet er nøkternt sett overføring av fakta uten stilistisk fullkommenhet. Derfor er arbeidet mest rettet mot den industrielle sektor og internasjonale organisasjoner. Det skilles ofte mellom tre typer oversettelse:

- 1) høykvalitet oversettelse for internasjonale klienter
- 2) hurtigoversettelse av konsentrater, møtereferater osv.
- 3) grovoversettelse for senere redigering av oversettere.

4. Eksempler på systemer for maskinoversettelse

4.1 Systran

Systran-systemet ble utviklet av Peter Toma, California fra 1960 av, og er kanskje det mest kjente systemet. Trolig kan det karakteriseres som et første generasjons system som kontinuerlig er bearbeidet med tanke på å inkorporere nye konsepter fra maskinoversettelse. Det ble primært utviklet for oversettelse fra russisk til engelsk og innkjøpt av American Air Force. Senere ble oversettelse engelsk-fransk utviklet, og EF kjøpte systemet i 1975 etter å ha fått utført en omfattende inventering av tilgjengelige systemer.

4.2 TAUM-gruppen

Universitetet i Montreal begynte i 1965 å utvikle ekspertise i analyse av naturlig språk. Mellom 1968 og -71 utarbeidet en prosjektgruppe, TAUM, et basissystem for oversettelse etter et andre generasjons

konsept. Systemet ble i 1977 utviklet videre under en stor kontrakt med den kanadiske stat til et system for automatisk oversettelse av værvarslere fra engelsk til fransk (TAUM-METEO).

Det rapporteres at systemet gir 80 prosent korrekt oversettelse av værvarslene. Disse rapportene er imidlertid skrevet i en form for telegrafisk stil som bl.a. unngår bøyde verb.

4.3 METAL

Linguistics Research Center ved University of Texas har arbeidet med maskinoversettelse siden 1961, først med et teoretisk siktemål. Siden 1971 har Siemens i München vært den viktigste finansielle bidragsyter. *Jonathan Slocum* opplyste på konferansen i Lugano at METAL-systemet til nå har vært prøvd på ca. 1000 sider autentisk tekst med oversettelse fra tysk til engelsk. Det blir rapportert fra 45 til ca. 80 prosent korrekt oversettelse, dvs. oversettelse som ikke senere blir forandret av profesjonelle oversettere.

4.4 SUSY

Ved Universitetet i Saar ble det første analysesystemet for moderne skrevet tysk ferdig i 1960. Det ble foretatt en utvikling av en prototype for oversettelse fra russisk til tysk i 1971. I 1972 ble det av flere institutter i fellesskap opprettet en institusjon for elektronisk språkforskning (SFD), støttet av det tyske forskningsråd. Hovedvekten ved SFD har hele tiden vært på forskning i maskinoversettelse. I løpet av det siste ti-året er det utviklet et generelt, flerspråklig oversettelsessystem SUSY (Saarbrücker Übersetzungssystem).

4.5 TITUS

TITUS-systemet er utviklet til bruk for den internasjonale tekstilorganisasjonen ITF. TITUS ble startet i 1969 og viser bl.a. hvordan maskinassisterte systemer for oversettelse kan inngå som et hjelpemiddel i et omfattende edb-basert informasjonsarbeid. Filosofien bak TITUS er å oversette sammendrag av dokumenter med tekstilteknisk innhold til en standard representasjonsform. Med basis i denne standardiserte oversettelse kan man automatisk produsere oversettelser i alle de offisielle språkene i organisasjonen. Sammendrag av tekster blir også automatisk indeksert og lagret i en felles-språklig database, slik at brukerne av informasjonssystemet kan søke i databasen i alle språk og få utskrevet resultatet i den ønskede språkvariant.

4.6 CULT

Chinese University Language Translator (CULT) ble påbegynt i Hong Kong i 1969. Systemet er blitt brukt til å oversette det kinesiske

matematiske tidsskriftet «Mathematica Sinica» og senere «Acta Sinica Physica» til engelsk. I CULT er det lagt stor vekt på pre-editering av dokumentene bl.a. med syntaktiske og semantiske indikatorer.

4.7 LOGOS

LOGOS Corporation ble grunnlagt i 1969 og hadde i 1971 bl.a. med støtte fra den amerikanske regjering utviklet et engelsk-vietnamesisk oversettelsessystem. Av andre språk det har vært arbeidet med, er engelsk, spansk, russisk, fransk og farsi. For tiden står språkparet tysk-engelsk i fokus. LOGOS er et andre generasjons system med en oversettelsesmodul som hevdes å være språkuavhengig.

4.8 GETA

Groupe d'Etude pour la Traduction Automatique (GETA) ved Universitetet i Grenoble har arbeidet med oversettelsesproblemer siden 1972, og bygger på tidligere forsøksvirksomhet fra 1969 av. For tiden arbeider gruppen bl.a. med oversettelse av russisk-engelsk basert på et testmateriale bestående av vitenskapelige og tekniske sammendrag. GETA-gruppen samarbeider også med EFs oversettelsesprosjekt.



Lugano i Sveits er en populær konferanse- og turistby.

4.9 Brigham Young University

Ved Brigham Young University har professor Alan K. Melby deltatt i utvikling av en fremtidig arbeidsstasjon for oversettere. Det nye ved Melbys arbeidsstasjon (nå i sin 3. versjon) er at systemet ikke forutsetter store maskinelle ressurser. Alle komponentene utenom selve maskinoversettelsesprogrammet kan få plass på en mikrodatamaskin med platelager. Den maskinoversatte teksten kommer derimot fra et større dataanlegg. Den grunnleggende forutsetning er at oversetteren med letthet skal kunne veksle mellom tekstbehandling, ordboksoppslag, innførsler i ordboken, post-editering av maskintekst osv.

4.10 EUROTRA

Planleggingen av EFs felles oversettelsesprosjekt (EUROTRA) startet i 1979 (se også Humanistiske Data 2-82).

I november 1982 ga ministerrådet i EF klarsignalet til å etablere «a machine translation system of advanced design capable of dealing with all the official languages of the Community».

EUROTRA vil bygge på overføringsprinsippet (transfer) for derved å skape et møtepunkt for 42 språkpar. Målet er å utvikle slike former for språkbeskrivelse at en overføringskomponent vil være i stand til å relatere beskrivelsen av utgangsspråket til en (metodisk sett) identisk representasjon av målspråket som oversettelsen så tar sitt utgangspunkt i.

Systemet vil bli svært modulært oppbygd både for å bli robust og for lett å kunne endres og utbedres. Det er f.eks. et ønske at systemet *alltid* skal være i stand til å gi en eller annen form for oversettelse.

Både i den strukturelle analysen og ved design av ordbøkene har EUROTRA forsøkt å bygge inn den viten som er nådd i arbeidet med tidligere systemer, og har bl.a. lagt stor vekt på å inkorporere semantisk informasjon.

Dette referatet er et sammendrag av en mer fullstendig fremstilling som vil komme ut i Senterets rapportserie i løpet av høsten 1984.

ALLC Louvain-la-Neuve, 2.-6. april

Ole Lauvskar

Den ellefte internasjonale konferansen til Association for Literary and Linguistic Computing (ALLC) fann stad i Louvain-la-Neuve (Belgia) 2.-6. april. I alt var det nesten 200 deltakarar. Dei fleste var ifrå Europa og Nord-Amerika, men jamvel Asia, Latin-Amerika og Afrika var

representerte. Konferansen var delt i to parallelle sesjonar med om lag 45 foredrag til saman. Dessutan vart det arrangert to arbeidsgrupper, ei for litteraturstatistikk og ei for databaser.

Det var svært mange ulike emne på konferansen. Den einaste fellesnemnaren såg ut til å vera edb i språk- og litteraturfaga. Emna var studium av forfattarskap, stilistikk, litteraturstatistikk, tekstkritikk, syntaktisk analyse, morfologi, fonologi, terminologi, leksikografi, ordbøker ved hjelp av edb, leksikalske databaser, naturleg språk, programvarer og databehandling av tekstar. Som ein skjønar var det noko for einkvar smak. Men sidan konferansen var så omfemnande, var det uråd å handsama kvart emne utførleg.

Konferansen var tospråkleg ved at foredraga og spørsmåla/svara veksla mellom fransk og engelsk. Sidan arrangørane var fransktalande, vart også mykje av den generelle informasjonen gjeven på fransk. Det såg ut til at for svært mange av deltakarane var ikkje det noko problem.

Det var to innlegg ifrå Noreg. Dei vart presenterte av *Patrick Chaffey* og *Kari Anne Rand Schmidt*, begge ifrå Britisk institutt ved Universitetet i Oslo.

Patrick Chaffey la fram det einaste innlegget som hadde fått merkelappen terminologi. Han fortalte om prosjektet ADNOM, som han leiar. ADNOM er ei forkorting for Administrative Nomenclature. Målet for prosjektet er å standardisera bruken av engelske namn på norske institusjonar.

Kari Anne Rand Schmidt presenterte prosjektet sitt som går ut på å avgjera om «The Equatorie of the Planetis» er skriven av Chaucer eller andre. Ho valde i innlegget å sjå berre på ein variabel, antal ulike ord i tekstdelar med ei viss lengd. Ho diskuterte vanskanane med å velja kor lange slike tekstdelar skulle vera. Ho hadde jamført teksten nemnt over med ein sikker tekst av Chaucer, «A Treatise of the Astrolabes». Då hadde ho brukt ein type/token test (eigentleg: t-test). Ho hadde delt dei to verka opp i like store tekstdelar. Det hadde ho gjort fleire gongar, men ho hadde variert lengda på tekstdelane. Det synta seg at utfallet på testen ifrå gong til gong var ganske avhengig av kor lange desse tekstdelane var. Ein kunne spørja seg kva den «optimale» lengda på ein tekstdel skulle vera.

Etienne Evrard ifrå Belgia konsentrerte seg også om å sjå på rikdommen av ordtilfanget. Han presenterte ein metode som nytta kvadratisk regresjon. Målet var å tilpassa ei krumma (eigentleg parabolisk) line til ein funksjon som uttrykkjer fordelinga av orda for ulike tekstlengder. Ved hjelp av ein slik metode hevda han å få eit mål på kor mykje ordtilfanget varierer og eit anna mål på kor vidt ordtilfanget stagnerer med aukande tekstlengde (det vil seia at forfattaren nyttar etter kvart dei same orda om att).

Ifrå Belgia vart det dessutan presentert to andre foredrag innan litteraturstatistikk. Det eine presenterte arbeidet med å undersøkje om Pierre Lacos (1741-1803) hadde lukkast med å skjula sin eigen stil når

han skreiv brevromanen «Les Liaisons dangereuses». Det andre foredraget såg på tre satiriske stykke skrivne på latin og verseform av ein jesuittisk prest og hans elevar ifrå det 17. århundre. Desse vart jamførte med sju stykke på verseform skrivne av andre relevante forfattarar. Ved hjelp av faktoranalyse, klarte ein å skilja dei tre førstnemnte ifrå dei andre.

Gregor Büchel og *Nico Weber* ifrå universitetet i Bonn i Vest-Tyskland presenterte Kant-prosjektet. I Bonn har dei skrive inn utvalde delar av Kants samla verk på edb. Målet for prosjektet er så langt som råd er å få ein Kant-indeks, det vil seia eit oversyn over filosofiske termar – med forklaringar. Det har også vorte brukt ein del kvantitative metodar i prosjektet. Ved sidan av frekvenslister, har dei arbeidd ein del med å finna den «semantiske avstanden» mellom orda. I tillegg har dei studert kollokasjonar.

Yaacov Choueka ifrå Israel snakka om uttrykk som inneheld fleire ord (han konsentrerte seg mest om to ord). Uttrykket skal ha ei meining ut over det at to ord er attmed kvarandre. Til dømes er «etter måten» eit fast uttrykk, medan «etter båten» ikkje er det. Ved å sjå på dei mest frekvente para (to ord attmed kvarandre), vil det vera mange av dei som ikkje er faste uttrykk. Choueka presenterte då ein metode som filtrerte bort dei para som ikkje var faste uttrykk. Eg fekk ikkje heilt tak i metoden/formelen, men han tok bl.a. omsyn til kor *relativt* ofte det



Louvain-la-Neuve er ein heilt ny by – noko som arkitekturen ber eit sterkt preg av.

andre ordet i paret følgte etter det første. Han hadde testa denne metoden på forsøkspersonar, og det synte seg å vera eit godt samsvar med metoden og det som forsøkspersonane svara (dei fekk vita det første ordet i uttrykket og skulle gissa det andre).

Det vart også presentert ein del programvare på konferansen. Ved ZUMA (Zentrum für Umfragen, Methoden und Analysen) i Mannheim, Vest-Tyskland har dei utvikla programpakken TEXTPACK V for tekstanalyse. Pakken skulle vera høveleg både for kvantitativ innhaldsanalyse og for meir kvalitativ tekstanalyse. Blant anna gjev programmet lemma-forma av dei ulike orda. Det gjev også multiple ordkombinasjonar. Ved Ohio State University i Columbus, USA har dei laga pakken LEXED. Pakken som er berekna på mikromaskiner, har fleire program for lingvistisk analyse.

Fleire foredrag tok føre seg databaser. Det kom bl.a. fram at ein arbeider med å laga databaser både for tysk (i Bonn) og for hollandsk (i Leiden, Nederland). Elles var dei fleste foredraga innafor dette feltet ifrå Italia. I Pisa har dei blant anna arbeidd med å laga eit internasjonalt oversyn over det som er gjort av databehandling innafor dei klassiske faga.

Joan M. Smith ifrå England presenterte arbeidet som ISO (International Standards Organisation) har gjort med å standardisera teikna som vert brukte til å representera språk og annan informasjon.

Det er klart at med så mange ulike emne, vil ikkje alle foredraga samla same interessa. Dei ulike foredragshaldarane hadde dessutan berre 20 minutt til rådvelde for kvart foredrag – pluss 10 minutt til spørsmål. Det er ikkje så enkelt å presentera fleire årsverk på så kort tid.

Konferansen var godt arrangert og vertskapet klarte å skapa ei hyggeleg sosial ramme rundt han. Konferansen var også fin som bakgrunn for fagleg kontakt med kollegaer ifrå andre land.

Conference on Computers in English Language Research

Windemere, England, 20.-23. mai 1984

Knut Hofland

Denne konferansen var den femte i rekken og samlet 60 deltakere fra 15 land. Konferansen var også et ledd i feiringen av British Council's 50-års jubileum, og British Council hadde bidratt med både finansiell

støtte og lokal arrangørhjelp. Hovedarrangør var professor *Geoffrey Leech* ved University of Lancaster.

Utgangspunktet for de fleste av foredragene var tilrettelegging eller bruk av tekstkorpora i forskning.

Henry Kucera, Brown University, åpnet konferansen med foredraget «An electronic thesaurus as a computational linguistic problem». Han startet med å diskutere hvordan forholdet mellom antall ordformer og lemma er påvirket av fordelingen av ordklasser og størrelsen av korpus/ordbok. Videre vil forholdet mellom substantiv og verb være avhengig av størrelsen på ordboken. I Brown korpus er det en grammatisk tvetydighet i ordformer på 12 %. Men ofte er en av formene i klart flertall. Kucera gikk deretter over til elektronisk publisering av ordbøker og tesauri. Det en ønsker av et program for ordbokhjelp ved tekstbehandling, er f.eks. synonym for ord som skrivemerket peker på. Ordet kan være *left* og systemet bør da gi *departed*. Dette vil kreve både analyse og produksjon av ordformer fra



Nelson W. Francis (t.v.) og Stig Johansson, som har ledet arbeidet med henholdsvis Brown Corpus og Lancaster-Oslo/Bergen Corpus.

systemet. Et korpus har for lite data i denne sammenheng og ved Brown bygger en opp en database som i dag har 160.000 ordformer. Den skal utvides til å omfatte 200.000 ordformer og ca. 60.000 lemma.

Jan Svartvik og Mats Eeg-Olofson, Lund universitet, orienterte om det nye prosjektet TESS, «Text segmentation in speech», som har fått ekstern støtte fra flere kilder, bl.a. Ericsson Information Systems. Prosjektet består av to deler: en videre analyse av det eksisterende korpus og en reversering av reglene fra denne analysen for produksjon av tale fra skrevet tekst (talesyntese). Korpuset er merket i tone-enheter (intonasjonsmessige sammenhengende stavelser), og analysen ved hjelp av forskjellige parserprogrammer skal både gjøres for individuelle tone-enheter og for sekvenser. Det vil bli brukt mikromaskiner i det videre arbeid, både fordi en da er uavhengig av universitetets datasentral og fordi talesynteseutstyret lettere kan tilkobles en mikromaskin.

To av foredragene handlet om nye korpora som var bygget opp eller planlagt bygget opp. Mahavir P. Jain, India, planlegger et korpus parallelt til Brown og LOB over engelsk brukt i India. India har 22 millioner innbyggere som bruker engelsk som hovedspråk og er dermed det tredje største engelskspråklige land i verden (foran Canada med 20 millioner). Formålet med prosjektet er å kunne sammenligne korpuset med LOB og Brown, derfor er tekstene hentet fra tekstmateriale produsert i 1961. Målet er bl.a. å identifisere «indianismer» og å finne hvilke kjennetegn som særtegn India-engelsk. Korpuset vil bli merket med opplysninger for dette formål. Det arbeides også i India med et korpus fra 1981.

Yang Hui-zhong, Kina, beskrev et engelsk korpus med tekster fra tekniske emneområder for bruk i undervisning. Det er økende interesse for engelsk i Kina og frekvenslister for de hyppigste ordene vil være grunnlag for undervisningen. Korpuset består av 2000 prøver på 500 ord innen 10 emneområder.

Det var flere foredrag fra universitetet i Lancaster. *Susan Blackwell* gikk gjennom de enkelte trinn i merkingen av LOB korpuset. *Geoffrey Sampson* refererte fra arbeid med manuell analyse av eksempler fra LOB korpuset. Disse ble maskinelt lagret i en tre-bank og var grunnlag for statistikk til den syntaktiske analysen. *Roger Garside* redegjorde for de enkelte trinn i denne analysen: Først tilordning av mulige konstituent-tagger til en ordklasse-tag, deretter søking etter visse syntaktiske mønstre for korreksjon av tagger og til slutt en entydiggjøring av tagger basert på statistiske metoder. *Fanny Leech* gikk videre inn på den probabilistiske parseren hvor det bl.a. var brukt Markovkjeder. *Eric Atwell* bruker en modifisert utgave av taggingprogrammene for å finne feil i tekster. Dette gjelder både stavefeil der det feilstavete ordet fremdeles er et gyldig ord, f.eks. *from* i stedet for *form*, og de tilfeller der galt ord er valgt. Han utvikler nå et stort leksikon basert på LOB korpus, Longman og Collins ordbøker.

Stig Johansson, Oslo, gikk gjennom en rekke eksempler fra LOB

korpuset og pekte på en del problemer med å velge rett tag i forbindelse med post-editering.

Gert van der Steen, Amsterdam, startet med å beskrive QUERY, et program for å søke i et tekstkorpus, umerket, tagget og/eller lemmatisert. Det tar tre minutter å gå igjennom det taggete Brown korpus. Ønsker om utvidelser har initiert et nytt program som fra starten er blitt skrevet med portabilitet for øye. I Nederland har forskjellige miljøer innen humanistisk databehandling blitt enige om standard Pascal som utgangspunkt for programmering med et felles subrutinebibliotek. Dette åpner mulighet for lettere utveksling av program. Den nye syntaktiske mønstergjenkjenneren gjør bruk av en korpusgrammatikk, en mønster (pattern) grammatikk, en parser generator, en transducer og tilslutt en rapportgenerator for presentasjon av resultater.

Jan Aarts, Nijmegen, gikk gjennom erfaringer fra seks års korpusrelatert forskning og arbeidet med å lage «den lingvistiske database». Denne ble presentert av *Theo van der Heuvel*. *Nelleke Oostdijk* demonstrerte ved hjelp av en overbevisende videosekvens, den såkalte «Linguistic Work Bench». Der kan en ved manuell analyse eller med mer eller mindre maskinell assistanse legge inn morfologisk og syntaktisk analyse og lagre denne i en database. Senere kan deler av data kalles frem, presenteres på ulike måter og arbeides videre med.

Det var ellers presentasjoner fra miljøer i Liege, Brüssel, Leuven, Venezia, Camerino, Göteborg, Uppsala, Helsinki, Belfast og Birmingham. Det var parallelle foredrag i flere av sesjonene. Dette forhold og det høye deltakerantallet understreket at arrangementet har forandret karakter fra et seminar til en konferanse. Neste konferanse blir i Lund i 1985 (mai/juni) med Jan Svartvik som arrangør.

Studiereise til England 21.-25. mai 1984

Per Vestbostad

Målet med reisa var dels å studera edb-bruken ved engelske museum, dels å sjå på bruken av interaktiv video ved nokre undervisnings-institusjonar. Eg reiste i lag med Jon Birger Østby, forsker-NAVF i edb-metodar for kulturhistoriske museum.

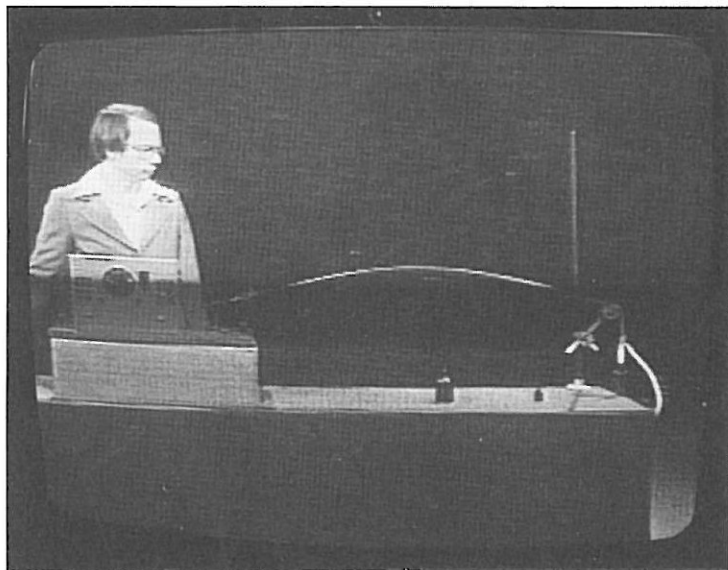
Bruk av interaktiv video i høgare undervising

Onsdag 23/5 vitja vi **The Open University** ved Milton Keynes. Det er eit friundervisings- og breviskule-universitet, som m.a. har drive pedagogiske forsøk med interaktiv bruk av video-plater. Vi møtte her *Diana*

Laurillard frå avdelinga for undervisingsteknologi. Ho la fram resultatane av eige forskingsarbeid, som m.a. synte at studentane vert passiviserte ved *lange* video-sekvensar, slik at dei nyttar tilhøyrande datamaskinlagra spørsmål dårlegare enn når læraren legg inn hyppige avbrot. Ho meinte elles at Massachusetts Institute of Technology har eit av dei fremste forskingsmiljøa når det gjeld datalagring på videoplater.

Den videoplatea som Open University hadde produsert sjølv, var diverre utlånt til eit seminar den veka vi var der, så det oppsettet fekk vi ikkje sjå. Før læreprogramma vert festa til plate, er dei grundig utprøvde på video-bandspelar. Her nyttar dei ein Apple II kopla til Panasonic bandspelar og Luxor fjernsyn. Vi fekk demonstrert eit «klassisk» læreprogram frå fysikk på dette utstyret. Når det galdt videoplatespelar, hadde Mrs. Laurillard best tru på Philips sin nyaste modell.

Ved **Imperial College, Department for Mechanical Engineering** møtte vi *Nick Rusby*, som leiar ein programvaresentral for interaktive læreprogram (CEDAR). Dei har utvikla læreprogram for industrien og offentlege etatar (særleg postverket), medan skuleverket ikkje hadde råd til å kosta utvikling av slike program. Men London University sin AV-avdeling hadde laga ei forsøks-plate med demonstrasjonsmateriale frå ulike fagfelt. Plata er laga for Philips-spelarar og kan kjøpast for



Eit modellforsøk i eit interaktivt læreprogram på videoplate. Forsøket syner korleis endringar i bølgelengda verkar på ein brubane.

kring £20.

Etter råd frå Nick Rusby vitja vi også **Council for Educational Technology**, og fekk med oss eit utval av deira publikasjonar (inkl. publikasjonsliste).

Endeleg var vi også i kontakt med *Ian Mercer* ved **Geological Museum**, som held på å laga eit interaktivt video-program til bruk for publikum. Programmet er under utprøving, og dei rekna med at publikum kunne nytta det nokså snart.

Oxford University Computing Service

Tysdag 22/5 hadde eg kontakt med *Dr. Ruth Glynn* ved Oxford University Computing Service (OUCS). Ho hadde m.a. ansvaret for å få maskinlagra tekstar velberga gjennom fotosetjaren deira, som er den einaste fotosetjaren ved engelske universitet, eit anlegg (Miles 33) til omlag 1,5 mill. kr. For tida arbeidde dei med å finna rimelege løysingar som kunne skriva ut det ferdige formatet på vanleg papir, for teksten vart teken ut på dyrt fotosats-papir. Dei gjev ut eit eige meldingsblad til fotosats-brukarane («LASERCOMP NEWS»).

OUCS driv også eit nasjonalt TEXT ARCHIVE med mange britiske og utanlandske maskinlagra tekstar. Mykje av dette er lagra i rå-format, og det skortar på kapasiteten til å arbeida tekstane fram til normalisert format.

Dr. Glynn fortalde også om to større databasar som var i arbeid der:



Det nye universitetsanlegget til The Open University ved Milton Keynes.

Peter Fraser held på å byggja opp eit gresk personnamn-arkiv, medan arkeologane registrerer opplysningar om greske vase-former og -dekorasjonar.

Eg informerte om den statistiske programpakken STAR, som er utvikla ved NAVFs EDB-senter for humanistisk forskning. Dr. Glynn er arkeolog av utdanning, og ho var svært interessert i å sjå nærare på STAR.

Reisa knytta såleis nye kontaktar, og nyttig informasjon gjekk begge vegar. Det var nok mest å henta på video-feltet, men det er også viktig å knyta kontaktar på dei felta der vi sjølve har noko å gje.

Museum Documentation Association

og opplegg for registrering og edb-rutiner ved britiske museer

Jon Birger Østby

Organisasjon

I 1977 startet 10 britiske nasjonalmuseer et arbeidsprogram for å utvikle edb-opplegg for museer, og dette ble opptakten til MDA (Museum Documentation Association).

MDA holder til ved Duxford Airfield, og fra 1982 er MDA blitt et offisielt organ med bevilgninger direkte fra The Office of Arts and Libraries. Samtidig ble det foretatt en omorganisering, og MDA har i dag to typer medlemmer. Den ene gruppen består av nasjonalmuseer, og 19 av disse museene er i dag kommet med. Den andre gruppen består av de 9 Area Museums Councils som vi finner i Storbritannia, og alle museer som er tilsluttet et slikt regionalt museumsråd har dermed fulle rettigheter som medlemmer i MDA.

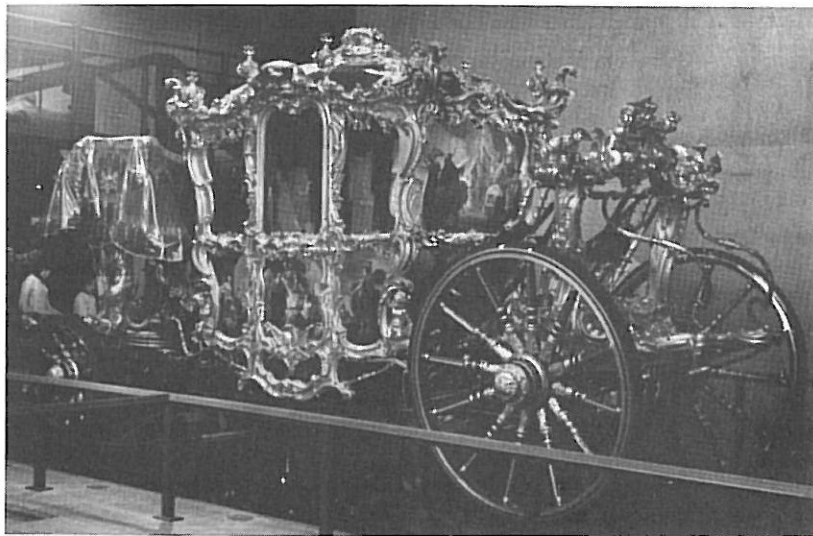
MDA startet med et personale på 3 i 1977 og har i dag 10 ansatte. Budsjetten for 1984 er på ca. 1,6 mill. kroner. Halvparten dekkes over offentlige bevilgninger og den andre halvdel skal dekkes gjennom inntekter på oppdragsvirksomhet, salg av systemer og katalogkort, forskningsbidrag og bidrag fra nasjonalmuseene.

Katalogkort og skjemaer

I samarbeid med forskjellige museer har MDA utviklet 22 forskjellige katalogkort. Dette spekteret dekker kort for en rekke formål fra et generelt kort for gjenstandsregistrering til spesialkort for foto, kunst, drakt, geologi, naturhistorie og en rekke andre gjenstandstyper. Til

hvert kort er det laget en egen instruksjonsbok. MDA selger årlig ca. 400.000 av disse katalogkortene, og de brukes nå av 3-400 museer. Men det er bare en del av disse museene som har registrert materialet på edb.

MDA har bare en rådgivende status, og man har til nå vært forsiktig med å dirigere museenes virksomhet på katalogsektoren. Dette har altså resultert i en stor flora av forskjellige kort, og man ser det positivt at det på denne måten har vært mulig å tilfredsstille en rekke spesielle ønsker fra forskere med forskjellige spesialfelter. Men samtidig har det vist seg å være vanskelig å operere med så mange forskjellige opplegg, og det kommer stadig ønsker om nye. Det pågår derfor nå en diskusjon om det ikke vil være riktig å gå inn for en mer restriktiv linje på dette feltet. Samtidig opplever man konflikten mellom større museer med stor ekspertise og behov for katalogkort som gir muligheter for en detaljert beskrivelse, og mindre museer som ikke har de samme behov. Dette er en konflikt som vi jo kjenner her i landet også. MDA har forsøkt å løse dette problemet med å la museene velge mellom et generelt kort og en rekke spesialkort, men dette er heller ingen uproblematisk vei å gå. Selv vil jeg holde en knapp på den linjen man har valgt med NKKM-kortet, der en har tilrettelagt muligheten for en detaljert beskrivelse og til dels overlatt til museene å velge i hvilken grad de vil utnytte denne muligheten.



En av kroningsvognene på London Museum.

Edb-sektoren

MDA har utviklet sitt eget program GOS. GOS er et program for produksjon av kataloger og registre. Det foreligger ingen muligheter for direkte søking i databasen. Det er sterke ønsker om å få utviklet program for direkte søking, men MDA ser ingen muligheter for å skaffe ressurser til dette arbeidet nå.

GOS kan ikke kjøres på mikromaskiner. Noen store museer som National Maritime Museum, British Museum og Science Museum satser på å bygge ut egne store dataanlegg, men de fleste museene satser på egne mikromaskiner for registrering og så katalogproduksjon ved MDA.

MDA gir råd for innkjøp av datautstyr, men det er museene selv som foretar den endelige avgjørelsen. Ofte blir det derfor lokale kontakter med spesielle leverandører som blir utslagsgivende, og dette har resultert i en rekke forskjellige anlegg. Dette har skapt problemer når det gjelder avlevering av data, men MDA skal nå få utstyr til å spille av forskjellige typer disketter.

Ca. 40 museer registrerer data på egne mikromaskiner, og 15 museer sender data til registrering ved MDA. Men utbredelsen av mikromaskiner øker sterkt, og i løpet av 1985 regner man med at rundt 100 museer vil ha egne mikromaskiner for registrering. Museene sender så inn sine data på disketter eller tape. De bestemmer selv hvilke rubrikker som skal databehandles og hvordan de vil ha ordnet kataloger og registre. Utskriftene leveres på papir eller mikrofiche.

Det foreligger ingen planer om å opprette noen felles database for kulturhistoriske museer i Storbritannia. MDA har kopier av det registrerte materialet, men det er de enkelte museene som har eiendomsretten. MDA har ikke lagerkapasitet på dataanlegget for en så stor database. Straks et oppdrag er avsluttet spilles dataene over på tape for å frigjøre plass for nye data, og MDA har heller ikke programvarer for direkte søking. Men i og med at det eksisterer et stort standardisert katalogmateriale i maskinlesbar form, vil produksjon av en felles database først og fremst være et juridisk og teknisk problem. Det finnes derimot en fellesbase for Natural History. Denne databasen ligger i Manchester, og den skal dekke hele Storbritannia.

Ved MDA er det reist spørsmål om museumspersonell burde bli flinkere til å analysere egne behov og ressurser i forbindelse med innføring av edb, og om det kan være ren vanetenkning som får dem til å begynne med å legge inn gjenstandskatalogene. Nykatalogisering og edb-registrering av eldre gjenstandssamlinger er et kostbart og et tidkrevende arbeid. Ved innføring av edb har MDA derfor ofte anbefalt større museer å begynne med oppstillingskataloger. Ved en av de store nasjonalmuseene har man regnet ut at en omlegging av gjenstandskatalogene vil kreve 50 stillinger i 20 år. For å komme i gang med et prosjekt som både vil være nyttig og overkommelig er det derfor foreslått å

begynne med oppstillings- og utlånskataloger. Samtidig er det forslag om å registrere alle gjenstander som det er spørsmål om, slik at det bygges opp en database for den mest brukte delen av samlingene.

Publikasjoner

MDA utgir en rekke rapporter og andre publikasjoner om gjenstandsregistrering og om bruk av edb. For dem som er interessert i løpende informasjon om det som skjer i Storbritannia på dette feltet, vil jeg anbefale et abonnement på MDA Information. Dette er et meldingsblad som også har korte artikler. MDA Information kommer ca. 4 ganger i året, og et abonnement til Norge koster ca. kr. 75. Adressen er *MDA, Duxford Airfield, Duxford, Cambridgeshire CB24QR, U.K.*

Konklusjon

MDA er en interessant institusjon med tanke på diskusjon om fremtidig organisering av edb-utvikling og katalogarbeid for norsk museumsvesen. Vi kan ha mye å lære av arbeidsmodellen og finansieringsformen, og vi kan dra nytte av det materialet som er utviklet. Men MDA har ingen fast tilknytning til et større edb-miljø slik vi har det her i landet gjennom NAVFs EDB-senter for humanistisk forskning. Dette er en kontakt jeg tror det er viktig å bygge videre på fordi vi gjennom dette samarbeidet får tilgang til bredere ekspertise og ressurser på edb-sektoren, enn det vi kan regne med at museumsetaten vil ha muligheter for å bygge opp alene.

Norsk Arkivseminar, Fana Folkehøgskole, 24.-26. mai 1984

Anne Hals

Hege Brit Randsborg

Norsk Arkivseminar har deltagere fra alle institusjoner innen Arkivverket (dvs. Riksarkivet og statsarkivene). Denne gangen hadde man konsentrert seg om 2 emner:

1. Innstilling om styringsstruktur og ansvarstilhøve i Arkivverket (Bjørkvikutvalgets innstilling)
2. Edb og Arkivverket.

Avdelingsarkivar *Ivar Fønnes* holdt hovedforedraget under emne 2.

Vi vil her gjengi hovedpunktene i foredraget.

Arkivverkets bruk av edb er nært knyttet til den offentlige forvaltnings bruk av edb. Det er i år omtrent 20 år siden norske statsinstitusjoner begynte å ta i bruk edb og bygge opp dataregistre på magnetbånd. Statistisk sentralbyrå (SSB) var en av de aller første, senere har en rekke andre étater og institusjoner fulgt etter. Det er etablert en helt ny type arkivmateriale som er spesielt verdifullt fordi datateknikken gjør det så lett å bearbeide, ordne, systematisere, finne fram i osv. Samtidig er det spesielt sårbart fordi lagringsmediet er nokså ømfintlig og lite arkivbestandig.

Situasjonen i dag er denne:

I forvaltningen ligger det mange tusen magnetbånd med arkivmateriale som ingen har den fulle oversikt over. Mange av båndene er så gamle at informasjonen kan ha gått tapt eller vil kunne gå tapt før den blir avlevert. Hva som er foretatt/foretas av kassasjon har ingen oversikt over.

I Arkivverket fins hverken utstyr som kan behandle magnetbånd/data av en slik størrelsesorden eller personellressurser/driftsmidler til å ta fatt på problemet for alvor.

Arkivverket er forpliktet til å ta imot edb-materiale og står altså i realiteten ikke fritt til å velge *om* eller *når* man skal ta i bruk edb. Man er nå stilt overfor valget mellom å komme i gang straks eller å gi fra seg muligheten til å få kontroll over situasjonen. Bekymringen knytter seg først og fremst til alderen på de eldste magnetbåndene og mengden av nytt materiale som øker i akselererende tempo.

Aktuelle edb-oppgaver for Arkivverket

Fonnes tok for seg følgende hovedområder for Arkivverkets edb-arbeid:

1. Mottak av arkivmateriale som avleveres på edb-media, oppbevaring og utbygging av tjenestetilbud til Arkivverkets brukere.

Behandlingen av edb-arkivalier reiser en rekke problemer som krever både arkivfaglig innsikt og edb-kompetanse. Det gjelder kassasjon og eventuell bearbeidelse av materialet og det gjelder dataformater, registre og dokumentasjon.

2. Deltagelse i utformingen av forvaltningens edb-systemer, dvs. systemer som produserer materiale som senere skal arkiveres.

I Riksarkivet har man konsentrert seg om edb-baserte journalføringsystemer fordi disse gir gode muligheter for raskere og mer effektivt arkivarbeid.

Arbeidet på dette feltet har omfattet tre oppgaver:

- a) utarbeidelse av retningslinjer for edb-basert journalføring
- b) konsulentbistand til institusjoner som planlegger bruk av edb i journalføringen

c) samarbeid med R-direktoratet om utvikling av edb-systemer for journalføring.

En rekke departementer og andre institusjoner arbeider nå med planer om å ta i bruk edb-basert journalføring, og noen er allerede i gang. Konsulentbistand anses nødvendig for å sikre en forsvarlig arkivdannelse i de institusjoner det gjelder.

I samarbeid med R-direktoratet forsøker man å få fram et programsystem som skal være et tilbud til institusjoner som vil bruke edb i journalføringen - NOARK. Systemet ventes i prøve-drift fra årsskiftet. Riksarkivets rolle er først og fremst å delta i utforming av kravspesifikasjoner, men Riksarkivet vil også være med i selve utviklings- og utprøvningsfasen.

3. Automatisert arkivinformasjon, dvs. bruk av edb for å lette framfinningen i arkivmaterialet (via kataloger og registre).

På dette feltet har NAVF bevilget et forskerstipend i 1 1/2 år fra 1.8.1983 og *Hege Brit Randsborg* er i gang med å lage et standardssystem for katalogisering. Systemet er nå under utprøving. (Se for øvrig egen rapport i dette nr. av Humanistiske Data).

I tillegg til stipendiatens arbeid er det i lengre tid arbeidet med bruk av edb i produksjon av kataloger ved Norsk privatarkivinstitt.

En annen aktuell form for automatisert arkivinformasjon, er å legge eksisterende registre til forskjellige typer dokumenter inn i en database for direkte søking fra terminal. Dette kan f.eks. være aktuelt for tinglysdokumenter. Bl.a. har statsarkivar *Egil Øvrebø* presentert et opplegg for etablering av en slik database, kombinert med lagring av selve dokumentene på videoplate. Øvrebøs forslag vil være et utmerket prøveprosjekt for å prøve ut videoplatens muligheter i arkivsammenheng.

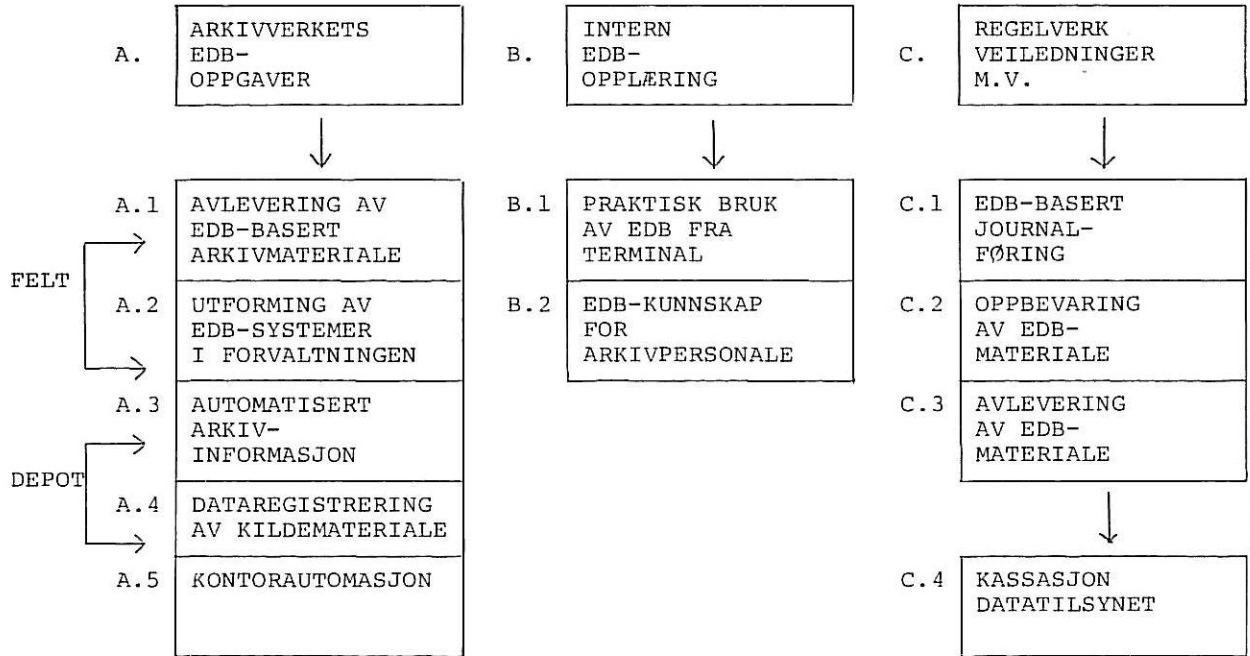
4. Dataregistrering av historiske kilder, f.eks. emigrasjonsprotokoller, kirkebøker m.v. for å lette framfinning m.v. i kildene.

Dette feltet er dominert av forskningsmiljøene, men Arkivverket har vært med, først og fremst Statsarkivet i Bergen. Vi har hatt stor nytte av denne registreringen - f.eks. har framfinningen i emigrasjonsmaterialet blitt veldig mye raskere og dermed mindre arbeidskrevende. Edb-avdelingen i Riksarkivet har ikke ressurser til å gjøre noe på dette området på kort sikt. RHD (Registrerings-sentral for historiske data, Universitetet i Tromsø) anses derfor som en viktig støttespiller som avlaster Arkivverket på dette området.

Intern edb-opplæring og kompetanseoppbygging

Edb-opplæring er viktig for at Arkivverket skal kunne gi råd og stille krav vedrørende arkivmateriale på edb-media samt kunne utnytte edb-teknologien i sitt eget arbeid. En stor del av Arkivverkets ansatte

AKTIVITETER KNYTTET TIL EDB



har fått elementær opplæring i edb og tekstbehandling, og de har fått en viss erfaring i praktisk bruk av datamaskin. De resterende grupper vil få tilsvarende opplæring når forholdene ligger til rette for det (f.eks. ved maskinanskaffelser).

Fonnes understreket at det var viktig at arkivpersonalet får innsikt i en del mer generelle edb-spørsmål, og særlig slike som er relevante i arkivsammenheng. Et sentralt område er her kunnskaper om datalagringsmedia – lagringskapasitet, arkivbestandighet, oppbevaringsmåter osv., og naturligvis hva slags muligheter og begrensninger som er forbundet med ulike medier.

Arbeid med regelverk, veiledninger m.v.

Riksarkivaren har laget et utkast til retningslinjer for edb-basert journalføring. I retningslinjene legges det vekt på bl.a. følgende:

- a) Krav til både edb-systemet og til driftsrutinene knyttet til det.
- b) Ulike løsninger tillatt forutsatt at de tilfredsstiller visse minimumskrav til arkivfaglig forsvarlighet.
- c) Krav om fast avgrensede journalperioder.
- d) Krav om bestemte utskrifter, løpende og ved slutten av en journalperiode.
- e) Meldeplikt til Riksarkivaren med rett for ham til å gripe inn dersom han finner det nødvendig.

I 1979 ble det laget avleveringsregler med retningslinjer/veiledning for oppbevaring av edb-materiale. Disse bør justeres på enkelte punkter (bl.a. kravene til magnetbåndformatet), men vi mangler ennå tilstrekkelige kunnskaper bl.a. om varigheten av disketter.

Et vanskelig område for Riksarkivaren er forholdet til Datatilsynet. Personregisterloven har i realiteten fratatt Riksarkivaren all formell kompetanse når det gjelder kassasjon av personregistre. Riksarkivarens kassasjonsregler kan altså uten videre bli satt til side dersom Datatilsynet stiller krav om sletting av arkivverdig materiale. Dette gjelder både edb-basert materiale og arkivalia på papir dersom det inneholder «følsomme» personopplysninger. Personregisterloven er nå under revisjon, og Riksarkivaren vil arbeide for å få kompetanseproblemet mer avklart i den reviderte loven.

Fonnes kom til slutt inn på status og utbyggingsplaner for edb-ressursene i Arkivverket.

Status pr. i dag

Riksarkivet har etablert en midlertidig edb-avdeling fra 1.11.1983 med to stillinger, avdelingsleder og arkivar, hvorav bare den siste er en fast stilling. Avdelingens arbeidsområde omfatter de ovenfornevnte oppgaver og dessuten arbeid med anskaffelse og drift av utstyr, planlegging, utredning og andre spørsmål med tilknytning til edb. Av utstyr

disponerer *Riksarkivet* halvparten av en Altos mikromaskin med adgang til 2-3 terminalarbeidsplasser.

Fra NAVF's instrumenttjeneste er det lånt en énbrukermaskin (IBM PC) med stort lager. Den skal benyttes til utviklingsarbeid, først og fremst av stipendiaten, men også av edb-avdelingen. NAVF's instrumenttjeneste har også stilt til disposisjon en terminal for kjøring mot eksterne anlegg (Blindern).

Ved *SA Trondheim* er det installert et Discovery dataanlegg med 3 arbeidsplasser, ved *SA Hamar* en énbrukermaskin (IBM PC), ved *SA Bergen* en KONTIKI énbrukermaskin og ved *SA Stavanger* en Victor énbrukermaskin i samarbeid med Interkommunalt arkiv (IKA). Alle steder er det utbyggingsmuligheter.

Arkivverket har nå altså edb-utstyr fra 5 forskjellige leverandører og med like mange ulike diskett-formater. Dette gjør overføring av data mellom institusjonene vanskelig, men det finnes konverteringsmuligheter. En bedre samordning av utstyret vil være et mål på litt lengre sikt.

Utbyggingsplaner

I budsjettet for 1985 er det foreslått at edb-avdelingen skal bygges ut med 3 faste stillinger: Avdelingsleder, systemingeniør og edb-konservator (driftstilling). Dette vil gi avdelingen i alt 4 stillinger. For de påfølgende år er det foreslått 1 edb-arkivar, 1 programmerer og 1 edb-konservator.

Utstyrsforslaget for 1985 er på 3 mill. kr og skal dekke Riksarkivet og de 5 statsarkivene utenom Trondheim.

Et hovedmål for Arkivverkets virksomhet er *å komme á jour med avlevering av edb-arkivalier*. Dette er et ambisiøst mål som vil kreve en betydelig innsats av ressurser. Men det er viktig at Arkivverket greier å gjennomføre dette, ellers vil man etter all sannsynlighet komme enda mer på etterskudd i 1990-årene. Dessuten øker risikoen for tap av materiale for hvert år som går. Det er også et mål for de nærmeste år å medvirke aktivt i *utformingen av edb-systemer for journalføring og informasjonssøking i forvaltningen*. Arkivverkets oppgave er her å bidra til systemløsninger og driftsopplegg. Jo mer erfaring og kompetanse etaten skaffer seg innen databehandling, jo mer aktivt og konstruktivt kan den medvirke i dette arbeidet.

I løpet av de nærmeste år bør bruk av edb også bli en integrert del av *Arkivverkets daglige virksomhet* på en rekke områder. Hvor fort man skal gå fram og hvor langt man kan komme på dette feltet, vil i all hovedsak avhenge av mulighetene for å anskaffe egnet utstyr og programvare.

De forskjellige problem som det ble pekt på i Fønnes' foredrag, ble drøftet i mindre grupper og deretter i plenum. Emnene som ble tatt opp, var:

inspeksjon og avlevering av edb-arkivalia
katalogisering på edb
nye lagringsmedia
edb og journalføring
edb og arbeidsmiljø
edb og teknisk avdeling
Arkivverkets edb-kompetanse – opplæring og utvikling
registerarbeid – med og uten edb
Arkivverket og lovverket.

Anne Hals er førstearkivar ved Riksarkivet.

Hege Brit Randsborg er forsker-NAVF i automatisert arkivinformatikk.

COLING 84

Stanford University, California, 2.-6. juli 1984

Knut Hofland

Dette var den tiende COLING konferanse og samtidig det 22. årsmøte i Association for Computational Linguistics (ACL). Konferansen samlet ca. 500 deltakere. Det var 120 foredrag i løpet av de fire konferansedagene med fem parallelle sesjoner slik at det var mulig å få med ca. 25 foredrag.

Nesten halvparten av foredragene ble holdt av amerikanere og ellers var Japan, Storbritannia, Vest-Tyskland og Frankrike godt representert. Fra Norden var det to svenske og to finske foredrag. Foredragene var gruppert med 1-3 foredrag under ca. 15 hovedemner. Det emnet som hadde flest foredrag (ca. 1/4) var parsing i kombinasjon med emnene syntaks, programmeringsspråk og logikk. Mange foredrag gjaldt også maskinoversetting (nesten halvparten fra Japan), generering av naturlig språk, grensesnitt og pragmatikk. Det ble holdt flere paneldiskusjoner, bl.a. om kvantitativ lingvistikk, naturlig språk og databaser og maskinlesbare ordbøker.

Geoffrey K. Pullum, University of California, Santa Cruz, holdt et inspirerende foredrag der han først diskuterte om naturlige språk kan beskrives som regulære sett, med en kontekstfri eller kontekstsensitiv grammatikk. Han henviste til artikler som er under publisering som for første gang viser at noen språk ikke kan beskrives med en kontekstfri grammatikk. Til slutt gav han en karakteristikk av forskjellige grammatikkmodeller.

ATN (Augmented Transition Network) grammatikker har tidligere vært en gjenganger i mange foredrag om parsing og syntaks. I dag er det forskjellige funksjonelle grammatikker som er dominerende. En funksjonell grammatikk er en grammatikk som inneholder lister med en rekke egenskaper som igjen er et attributt/verdi par, f.eks. person: tredje, tall: entall. Eksempler på slike grammatikker er Lexical Functional Grammar (LFG, Kaplan og Bresnan), Functional Unification Grammar (FUG, Martin Kay), Generalized Phrase Structure Grammar (GPSG, Gazdar og Pullum) og PATR-II (Stanford Research Institute, SRI). *Lauri Karttunen*, SRI og Center for the Study of Language and Information (CSLI, Stanford), diskuterte lingvistiske aspekter til et nytt system for analyse ved hjelp av egenskapsgrammatikker. Dette er en utvidelse av PATR-II systemet og inneholder bl.a. muligheter for negasjon og disjunksjon i egenskapslistene, noe som andre systemer mangler. *Martin Kay*, XEROX, presenterte FUG i forbindelse med maskinoversetting. Aspekter i forbindelse med PATR-II systemet ble diskutert av *Fernando Pereira* og *Stuart Shieber* fra SRI. Sistnevnte system er både implementert i LISP (Interlisp og Zetalisp) og PROLOG. PROLOG er ellers på frammarsj som programmeringsspråk innen datamaskinell lingvistikk. I flere foredrag var implementasjonen gjort i PROLOG, bl.a. LFG.

I tillegg var det demonstrasjoner og utstilling av bøker. Xerox demonstrerte 1108 Lisp-maskin og programvaren LFG (Lexical Functional Grammar – program for syntaktisk og semantisk analyse) og LOOPS (programvare for ekspertsystem). Også Symbolics viste sin Lisp-maskin (noe kraftigere og flere ganger dyrere enn Xerox 1108). På



*Hoover Tower,
Stanford University.*

denne maskinen demonstrerte folk fra Stanford Research Institute (SRI) systemet TEAM (Transportable English Access Data Manager), dvs. en engelsk-språklig databasekommunikasjon som kan settes opp for forskjellige databaser. Det ble ellers demonstrert en kanadisk termbank, et system for undervisning i datamaskinell lingvistikk, HAM-ANS - Hamburg Dialogue System og to systemer for datamaskin-assistert oversetting. Det var få demonstrasjoner i forhold til antall foredrag og de gode maskinressurser og nettverksmuligheter som var tilgjengelige.

Martin Kay fra Xerox var arrangør av konferansen og han ble valgt til ny formann i den internasjonale komiteén for datamaskinell lingvistikk. Neste COLING, i 1986, blir arrangert i Bonn, Vest-Tyskland.

Litteratur

Proceedings of Coling 84, 2-6 July 1984, ACL (kan bestilles fra ACL Secretary-Treasurer Dr. Donald E. Walker, Bell Communications Research, 445 South Street, Morristown, NJ 07960, USA, for \$30 + \$18 flypost Europa).

COLING 84 sommerskole

Nytt av året var arrangementet av en sommerskole i tilknytning til COLING. Det var seks kurs som ble holdt uken før konferansen. Vertskap for flere av kursene var «Center for the Study of Language and Information» (CSLI) ved Stanford University. Dette er et samarbeidsprosjekt mellom flere av avdelingene ved Stanford University, Stanford Research Institute (SRI), Xerox Palo Alto Research Center (PARC) og Fairchild. CSLI har en stor maskinpark tilgjengelig, bl.a. 110(!) Xerox Dandelion Interlisp-D maskiner, flere VAX, en DEC-20 og flere laserskrivere som alle er koblet i et nett.

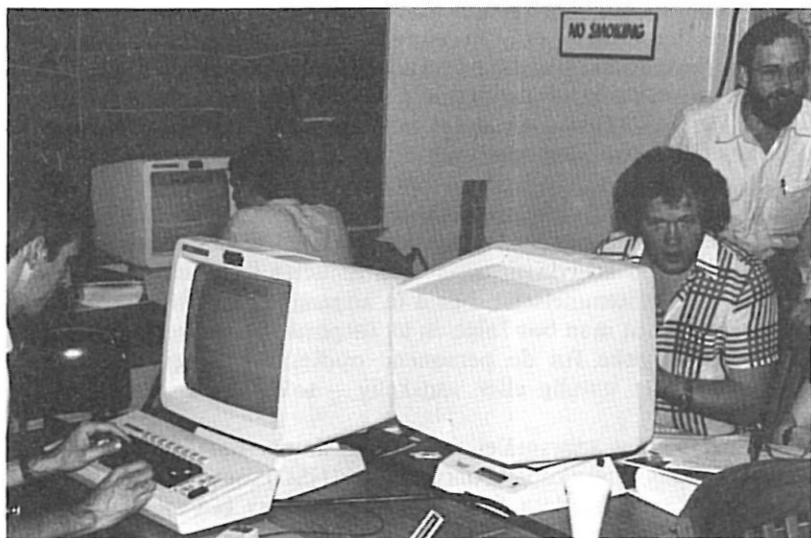
Jeg deltok på kurset «Parser Construction Techniques» med *Henry Thompson*, University of Edinburgh som foreleser. Kurset ga en innføring i en rekke programmer (gjenkjennerne, generatorer og parsere) for automatisk analyse og syntese av naturlig språk. Hver kursdag besto av to økter med forelesning og øvelser i hver. En gjenkjenner kan teste om en setning som er innlest i maskinen, er grammatisk ifølge den maskinlagrede grammatikken. Denne kan ha form av et nettverk (finit transisjonsnettverk) eller et hovednettverk og flere delnettverk (rekursive transisjonsnettverk). Dessuten ble det gjennomgått en parser (som gir setningens struktur hvis den er grammatisk) for kontekstfrie grammatikker som bruker et nettverk (chart) som datastruktur (eksempel på en ikke-deterministisk parser) og en såkalt Marcus-parser som har spesielle strategier for testing av grammatisk struktur (bl.a. look-ahead, eksempel på deterministisk parser).

Foreleseren hadde lagt de enkelte parserne på en felles fil-maskin, og deretter kunne deltakerne hente frem en parser fra denne maskinen og overføre den til sin egen maskin for modifikasjon og utprøving.

Foreleseren hadde lagt kurset opp på en pedagogisk fin måte og deltakerne fikk en svært god praktisk innføring i parserkonstruksjon. Det var 1-2 deltakere på hver Xerox Lisp maskin og vi fikk understreket hvor rasjonell en slik maskin er til denne type arbeid.

De andre kursene som ble holdt var «Lisp as Language» med *Brian Smith* fra Xerox, «Prolog for Natural Language Analysis» med *Fernando Pereira* fra SRI, «Situation Semantics» med *David Israel* fra BBN og *John Perry* fra Stanford, «Machine Translation» med *Brian Harris* fra Ottawa og *Alan Melby* fra Brigham Young University og «The Sound Structure of Language» med *Mark Liberman* fra Bell Laboratories.

Sommerskolen var svært vellykket og dette er et tiltak som framtidige arrangører av COLING bør fortsette med.



Noen av deltakerne ved COLING 84's sommerskole. Foreleseren Henry Thompson fra University of Edinburgh står bak til høyre.

Nytt fra RHF/NAVF

NAVF-innstilling om oppbevaring og gjenbruk av personidentifiserbare forskningsdata

Innstillingen forelå ferdig desember 1983. I løpet av første halvår i år har den vært til behandling i de ulike fagråd i NAVF og NAVFs styre.

Bakgrunnen for at NAVF ønsket en utredning om lagring og gjenbruk av forskningsdata var uklarheter mht. tolkning av forskerens taushetsplikt etter forvaltningsloven. Taushetspliktreglene i forvaltningsloven gjelder alle med forskningsoppgaver i offentlige forskningsinstitusjoner (universitet, høyskoler, etc.) eller som mottar offentlig støtte (fra forskningsråd, departement e.l.).

I NAVFs rammekonsesjon er det åpnet adgang for at forskningsdata kan overføres til arkivinstitusjon for å sikre mulighet for fremtidig bruk av slike data for oppfølgingsundersøkelser, panelstudier, testing og kontroll av forskningsresultater, etc. Datatilsynet presiserte imidlertid at dette bare kunne gjennomføres dersom regler om taushetsplikt ikke var til hinder for en slik ordning. Om taushetsplikten etter forvaltningsloven var et reelt hinder for overføring og gjenbruk av forskningsdata, tok ikke Datatilsynet stilling til. Den usikkerhet som dermed var knyttet til anvendelsen av taushetspliktsbestemmelsene i forvaltningsloven, førte til at det oppstod tvil om langtidslagring og gjenbruk av personidentifiserbare forskningsdata overhodet lot seg gjennomføre innenfor rammen av eksisterende lovgivning.

Utvalget fikk således i mandat å vurdere taushetsplikten i forvaltningsloven i forhold til spørsmålet om langtidslagring og gjenbruk. Utvalget skulle videre foreslå et opplegg for arkivordning for forskningsdata som ville være akseptabel innenfor gjeldende lovverk.

Utvalget konkluderte med at forskningsdata kan overføres fra forsker og lagres ved arkivinstitusjon uten at dette i seg selv bryter med taushetspliktsbestemmelsene. For å få adgang til gjenbruk av data var det enighet om at man bør følge én av følgende fremgangsmåter: enten innhente samtykke fra de personene opplysningene gjelder, eller – dersom dette er umulig eller vanskelig – søke om dispensasjon fra departement.

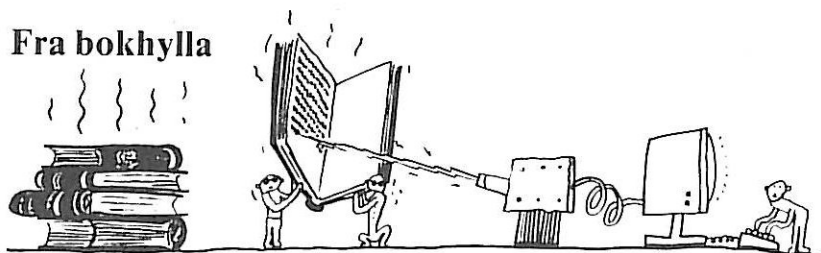
Når det gjelder spørsmålet om arkivordning, har Utvalget vist til Norsk samfunnsvitenskapelig datatjeneste (NSD) som aktuell arkivinstitusjon for forskningsdata fra prosjekt støttet av NAVF.

Som en oppfølging av innstillingen har NAVFs styre vedtatt å oppnevne et råd som skal innstille hvilke forskningsdata det er behov for å overføre til eget forskningsdataarkiv samt hvilke data som skal kunne frigis for gjenbruk. Sekretærfunksjonene skal ivaretas av Datafaglig sekretariat ved NSD. Dette sekretariatet er opprettet for å ivareta en del funksjoner som bl.a. NAVFs rammekonsesjon forutsetter blir oppfylt.

Thore Gaard Olaussen

Datafaglig sekretariat/NSD

Fra bokhylla



Bjørn-Andersen, N., Earl, M., Holst, O. and Mumford, E. (eds.): *Information Society – For Richer, For Poorer.* North-Holland, New York and Amsterdam, 1982. 320 s. ISBN 0 444 86426 1 Pris: \$47/Dfl. 110.

Information Society – for Richer, for Poorer er en samling av utvalgte foredrag holdt på en internasjonal konferanse i London i 1982. Denne konferansen – som hadde tittelen «The Transition to an Information Society: The Distribution of Benefits and Risks Associated with Micro-Electronic Applications» – samlet 120 representanter fra statsadministrasjon, industri, fagforeninger, media, forskningsinstitusjoner, universiteter og forbrukergrupper. Deltakernes varierte bakgrunn reflekteres i hovedtemaene for foredragene: informasjonssamfunnets særtrekk, statlig edb-politikk, samfunnsgrupper som er spesielt utsatt for datateknologiens konsekvenser og edb i arbeidslivet. Rapporter avgitt av arbeidsgrupper om framtidige forskningsområder er også inkludert i boka.

K.N.

Draskau, Jennifer, & Høedt, Jørgen (eds): *'The World of LSP' II* (1984). Nyt Nordisk Forlag, Copenhagen, 1984. 300 pp. ISBN 87-17-03432-9 Price: D.kr. 198.

'The World of LSP' II (1984) is a 50% enlarged, fully updated version of the data register *The World of LSP* (LSP = Language for Specific Purposes) – the first comprehensive overview of LSP pedagogic and research activities. The register covers 200 institutions representing all five continents. Some of the projects listed are based on computational methods.

MELDINGER

Nytt svensk studietilbud i datalingvistik

I høst starter Göteborgs universitet en ny fire-årig utdanning i datalingvistik. Instituttene for språkvitenskapelig databehandling, allmenn språkvitenskap, informasjonsbehandling og, i mindre utstrekning, filosofi skal samarbeide om undervisningen. Noen av emnene som vil bli tatt opp er: programmering, automatisk språkanalyse, formaliserte beskrivelser av naturlig språk og utvikling av informasjonssystemer.

Til de 24 studieplassene har det meldt seg 275 søkere. Utsiktene for å få arbeid etter endt utdanning blir regnet for å være gode.

System for utveksling av data om barnespråk

Ved Carnegie-Mellon University i USA har prof. *Brian MacWhinney* etablert et system for utveksling av data om barnespråk. Systemet vil gjøre det mulig for forskere å utveksle edb-kodete, transkriberte data med eller uten koder.

Forskere inviteres til å delta i dette systemet ved bidrag av egne data. Til å begynne med vil det legges vekt på engelskspråklige data. Støtte til utveksling av data på andre språk vil imidlertid ta til i løpet av ett års tid.

Interesserte forskere kan ta kontakt med:

Associate Professor Brian MacWhinney, Dept. of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213, USA.

News from ICAME

The eighth issue of ICAME NEWS – the Newsletter of the International Computer Archive of Modern English – appeared in May. This issue contains the following articles:

Jan Aarts: The COB: A Linguistic Data Base

Hans van Halteren: User Interface for a Linguistic Data Base

Gert van der Steen: On the Unification of Matching, Parsing and Retrieving in Text Corpora

Pieter de Haan: Relative Clauses Compared

Eric Akkerman: Verb and Particle Combinations: Particle Frequency Ratings and Idiomaticity

Inge van der Hurk et al.: To Strand or not to

A report on the 1983 Conference on the Use of Computers in English Language Research is also given.

ICAME NEWS no. 8 can be ordered from the Norwegian Computing Centre for the Humanities.

*

Papers given at the ICAME-sponsored Fourth Conference on the Use of Computers in English Language Research in Nijmegen, Holland, in 1983 have been published in a volume which also includes other related articles:

Jan Aarts and Wilhelm Mejs (eds.): *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*. Rodopi, Amsterdam, 1984. 229 pp. Price: Hfl. 50.

Wittgensteins etterlatte skrifter til Universitetet i Bergen

Da filosofen Ludwig Wittgenstein døde i 1951 etterlot han seg 18.000 sider notater. 12.000 av disse er skrevet inn i maskinleselig form av en prosjektgruppe i Tübingen i Vest-Tyskland, mens resten er under innskriving i Norge.

Disse tekstene blir nå samlet ved Universitetet i Bergen. Filosofisk institutt ved UiB skal stå for den filosofisk-faglige ledelse av dette prosjektet, som også tar sikte på å utvikle edb-metoder for å bearbeide stoffet. I den forbindelse er det etablert et samarbeid med NAVFs EDB-senter for humanistisk forskning, som vil lede det edb-faglige arbeidet i prosjektet.

Fra Wittgensteinprosjektet i Tübingen har en tysk forsker, *Michael Kulemann*, kommet til Bergen. Han har sin arbeidsplass på NAVFs EDB-senter der han i ca. tre måneder vil arbeide med standardisering av kodene i det materialet som er skrevet inn i Tyskland og med implementering av programmer han har utviklet for prosjektet.

Prosjekt i automatisk indeksering

I januar i år søkte NAVFs EDB-senter for humanistisk forskning og Institutt for rettsinformatikk, Universitetet i Oslo NORDINFO om midler til et prosjekt i automatisk indeksering. Automatisk indeksering omfatter maskinelle metoder for utplukking av stikkord fra en løpende tekst, og har spesiell relevans for tekstsøkesystemer, dvs. systemer for

lagring og gjenfinning av informasjon i mer eller mindre tekstlig form. I prosjektet ønsker man å konsentrere seg om lingvistiske metoder, nærmere bestemt lemmatisering (dvs. reduksjon av ord til grunnform), identifisering av ordklasser og oppsplitting av sammensatte ord.

Prosjektgruppen består av forsker *Tove Fjeldvig* (Inst. for rettsinformatikk), cand. philol. *Anne Golden* (UiO) og *Øystein Reigem* (Senteret). I prosjektet vil man ha et nært samarbeid med docent *Benny Brodda* og docent *Gunnel Källgren* ved Institutionen för lingvistik ved Stockholms universitet, som arbeider med nært beslektede problemer. I tillegg er det knyttet kontakt med professor *Fred Karlsson* ved Institutionen för allmän språkvetenskap, Helsingfors universitet og universitetslektor *Bente Maegaard* ved Institut for anvendt og matematisk lingvistik, Københavns universitet.

NORDINFO har i prinsippet stilt seg velvillig til prosjektet, men har ønsket seg en sterkere orientering mot BDI-området (BDI = bibliotek, dokumentasjon og informasjon), og en bedre kontakt med BDI-miljøene. NORDINFO anser splitting av sammensatte ord som spesielt interessant for BDI-området, og mente at det i første omgang er behov for en utredning om dette emnet.

For å gi prosjektgruppen anledning til å arbeide videre med saken, bevilget NORDINFO en sum tilsvarende tre forskermånedsverk. Som ledd i arbeidet vil det også bli arrangert et møte med norske BDI-medarbeidere i høst.

Ved hjelp av midler fra de nordiske samarbeidsnemndene for humanistisk og samfunnsvitenskapelig forskning (NOS-H og NOS-S) vil det bli arrangert møter med de lingvistiske samarbeidspartnerne i Norden. Det første blir holdt i Stockholm i september. Til dette første møtet vil en også invitere fagfolk fra de nordiske BDI-miljøene.

Prosjektsamarbeid som betalte oppdrag

Rådet for humanistisk forskning opprettet fra 1.1.81 som en prøveordning en konsulentstilling ved NAVFs EDB-senter for humanistisk forskning som skulle finansieres ved betalte oppdrag som Senteret påtar seg. Denne ordningen er nå blitt permanent.

Oppdrag kan omfatte generell konsulentbistand, systemering, programmering og driftsoppgaver i forbindelse med edb-prosjekter. De som er interessert i å få nærmere opplysninger om denne samarbeidsordningen, kan ta kontakt med Senteret:

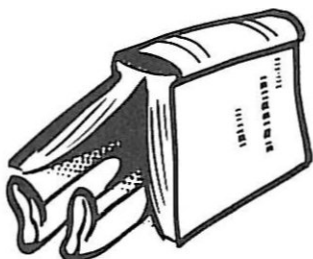
NAVFs EDB-senter for humanistisk forskning

P.b. 53

5014 Bergen-Universitetet

Tel: 05-212954/55/56

Nytt i biblioteket



Konferanserapporter

Mason, R.E.A. (ed.): Information Processing 83. Proceedings from IFIP 9th World Computer Congress, Paris. North Holland, 1983. 964 s.

Online Information. 7th International Meeting. London 6-8 December 1983. Oxford, 1983. 482 s.

Perlmutter, Linda/Keren, Carl (eds.): The application of mini- and micro-computers in information, documentation and libraries. Proceedings of the International Conference. Amsterdam, 1983. 801 s.

Brown, Kenneth Richard (ed.): The Challenge of Information Technology. Proceedings of the forty-first FID Congress, Hong Kong, 13-16 September 1982. Amsterdam, 1983. 356 s.

Arkeologi og kulturhistorie

Bølviken, Erik o.a.: Correspondence analysis: an alternative to principal components. Offprint World Archaeology. London, 1982. 19 s.

Light, Richard B./Roberts, D. Andrew: Microcomputers in museums. Duxford, 1984. 78 s.

Sarasan, Lenore/Neuner, A.M. (eds.): Museum Collections and Computers. Report of an ASC Survey. Lawrence, 1983. 292 s.

Edb-emner

Duke, John: Interactive Video: Implications for education and training. England, 1983. 148 s.

Ennals, John Richard: Beginning micro-PROLOG. New York, 1984. 196 s.

Illingworth, Valerie: Dictionary of computing. New York, 1983. 393 s.

Touretzky, David S.: LISP. A gentle introduction to symbolic computation. New York, 1984. 384 s.

Wilensky, Robert: LISPcraft. New York, 1984. 385 s.

Diverse

Baum, Joan: Computers in the English class with particular attention to The City University of New York. New York, 1983. 78 s.

Demografiska Databasen vid Umeå universitet. En presentation av verksamheten. Umeå, 1980. 125 s.

- Nordlund, Gunnar: Arkiven i dataålder. En studie i hur en ny teknik föder nya problem. Stockholm, 1982. 43 s.
- O'Shea, Tim/Eisenstadt, Marc: Artificial intelligence. Tools, techniques and applications. New York, 1984. 497 s.
- Oppbevaring og gjenbruk av personidentifiserbare forskningsdata underlagt lovbestemt taushetsplikt. Innstilling fra et utvalg nedsatt av NAVFs styre. Oslo, 1984. 43 s.
- Ragsdale, Ronald G.: Datamaskiner i skolen. Retningslinjer for planleggingen. Oslo, 1983. 151 s.
- Sandøy, Helge (red.): Talemål i Bergen. 1/83 - 2/83. Bergen, 1983.
- Winston, Patrick Henry: Artificial Intelligence. Addison-Wesley, 1984. 527 s.

Konferanser

Edb-konferanse for de humanistiske fag

NAVFs EDB-senter for humanistisk forskning arrangerer en nasjonal edb-konferanse i dagene 14.-16.november. Konferansen blir holdt på Vestlia høyfjellshotell på Geilo.

Formålet med konferansen er å samle humanistiske fagmedarbeidere som er interessert i edb til en drøfting av edb i humaniora. Konferansen søker å gi en dekkende presentasjon av den edb-aktivitet som foregår i dag og få fram perspektiver på fremtidig bruk av edb i f.eks. historiske fag, gjenstandsfag, språkfag og litteratur.

Det vil også bli satt opp tverrfaglige sesjoner knyttet til bruk av datautstyr og ulike metoder for kildetilrettelegging, informasjonsgjenfinning og publisering. Gjennom gjesteforedrag vil konferansen i tillegg sette søkelyset på den teknologiske utvikling fremover, hva som skjuler seg bak betegnelsen «informasjonssamfunnet» og vurdere vilkårene for de humanistiske verdier i et teknologisk samfunn.

Flere opplysninger fås ved henvendelse til:
NAVFs EDB-senter for humanistisk forskning
Tlf. (05)212954/55/56

RIAO 85

Det internasjonale symposiet «Computer Assisted Information Retrieval» skal avholdes 18.-20. mars 1985 i Grenoble, Frankrike. Symposiet blir arrangert av Centre des Hautes Études Internationales d'Information Documentaire.

Konferansens hovedtemaer blir forskningsframskritt i informasjons-gjenfinning og hvordan kunstig intelligens, informatikk og brukerne selv kan bidra til denne forskningen.

Flere opplysninger kan fås fra:
Centre des Hautes Études Internationales d'Informatique Documentaire, 36 bis, Rue Ballu, 75009 Paris, France.

Den 12. internasjonale ALLC konferanse

Association for Literary and Linguistic Computing (ALLC) inviterer til konferansen «Computers in Literary and Linguistic Research», 5.-8. juni 1985 ved Université de Nice, Frankrike. Konferansen vil legge vekt på bruk av datamaskinelle metoder i tekststudier (leksikometri og lingvistisk statistikk). Den som ønsker å holde et foredrag må sende et sammendrag innen 30. november 1984 til:

Professeur Étienne Brunet, Institut national de la langue française (URL9), 98, bd E. Herriot, 06007 Nice, France.

Frist for påmelding til konferansen er 31. mars 1985. Flere opplysninger og påmeldingsskjema fås fra ovenstående adresse.

International Conference on Data Bases in the Humanities and Social Sciences

ICDBHSS/85 skal avholdes i Grinnell, Iowa, USA 22.-24. juni 1985. Formålet med konferansen er å samle humanister og samfunnsvitere for å utveksle informasjon om opprettelsen, bruken og vedlikeholdet av fagrelevante databaser. Foredrag vil bli holdt om bl.a. mikromaskiner, videodisk, amerikanske og internasjonale kommunikasjonsnettverk, on-line kataloger og elektroniske meldingssystemer. Konferansen vil også omfatte utstillinger og demonstrasjoner.

Flere opplysninger fås fra:

Thomas F. Moberg, ICDBHSS/85 Coordinator, Grinnell College, P.O. Box 805, Grinnell, Iowa 50012-0810, USA.

ICCH '85

ICCH '85 – International Conference on Computers and the Humanities – skal finne sted ved Brigham Young University, Utah, USA, i dagene 26.-28. juni 1985. Sponsorer for konferansen er Brigham Young University Humanities Research Center, Conferences and Workshops sammen med Association for Computers and the Humanities. Foruten foredrag om forskning og undervisning i edb i de fleste humanistiske fag, vil det bli holdt en utstilling med bøker, tidsskrifter, og program- og maskinvare.

Sammendrag av foredrag må være mottatt innen 15. november 1984, mens påmeldingsfristen er 15. mai 1985. For ytterligere opplysninger skriv til:

Randall Jones/ICCH85, Humanities Research Center, 3060 JKHB, Brigham Young University, Provo, Utah 84602, USA.

SUMMARY

Digitalisering av bilder

Digitalization of pictures

Curator Jan Wiig at the Norwegian Museum of Science and Industry describes the principles of digitalization – a new technique for the transferral of museum and archive pictures to a permanent medium. He also discusses existing storage systems and their drawbacks.

Wiig gives a run-down of the demands museums and archives should make on satisfactory picture storage. On this basis he outlines an ideal system in light of current and future technological and financial possibilities. As the initial costs of a digitalization system are high, Wiig believes that a national centre for this purpose will be necessary.

Automatisk alfabetisering av transkribert tibetansk

Automatic sorting of transcribed Tibetan

Researcher Jens Braarvig is in the process of editing and transcribing a tibetan sutra (a Buddhist canonical scripture), the original of which is in the possession of the University Library, Oslo. Computing Officer Espen Ore at the Centre has designed the software for the project and in this article he describes an algorithm for automatic sorting of transcribed Tibetan.

An earlier method was described by Peter Nancarrow in the ALLC Journal, Vol. 1, No. 1, 1980. Nancarrow's algorithm is based on entering data on subscript/superscript letters and unvoiced initials along with the text and tables of the sorting sequence residing on files. The method described in this article uses standard transcribed Tibetan as input and also allows the user to define his own transcription.

Pedagogisk programvare for datastøttet undervisning

Pedagogical software for CAI

College Lecturer Lars Vavik distinguishes between four main methodological categories of computer-assisted teaching: computer-assisted instruction, simulation, model building and information processing. Analyses of these four categories are given.

Computer-assisted instruction expands the possibilities of textbooks, simulations imitate systems which previously could not be dealt with in the classroom, model building creates a framework for self-generating learning, and information processing increases the accessibility of new data. However, the gap is big between the possibilities of data technology and the quality of many so-called pedagogical programs. A striking number are based on simple principles of stimulation, response and feedback. This situation will improve as computer

capacity increases and better «author tools» are developed. Good software should be interactive, flexible and controlled by its users, thus requiring a high degree of pedagogical insight.

Extensive use of CAI in languages, music and art is hardly relevant. Many «drill programs» are of minor pedagogical importance, generation of pictures, music and speech is still in a stage of experimentation and «composition tools» need more testing. The different forms of CAI are more useful in history, social studies and related subjects. In any case, CAI software must be an asset to instruction and at the same time demonstrate the computer's possibilities and limitations.

PLATO og datamaskinassistert læring PLATO and computer-assisted instruction

Systems Consultant Rune Midtvedt describes Control Data's PLATO, one of the oldest and best-known computer-assisted instruction systems. In the US it is used on all levels of education and in many company training programs.

On the one hand, PLATO is a comprehensive collection of courseware. On the other hand, it is a method of instruction which takes advantage of the computer's possibilities. An extensive use of graphics and animation ensure a clear presentation of the course material. In combination with a touch-sensitive screen this makes PLATO particularly suited to simulation. PLATO also makes it easy to add explanatory texts and illustrations. In addition, several terminals can be linked in a network, thus enabling a teacher to «check in on» students and comment on their progress.

Thirdly, PLATO consists of various «author languages» for the development of programs. These can be used without previous knowledge of programming.

PLATO's video station makes it possible to portray constructions and courses of events in a realistic way by combining video, graphics, sound, text and calculations. The video station can be used for demonstrations, as an ADP-based «reference book», and for computer-assisted instruction.

Teknikk og undervisning Technology and teaching

Director Jostein H. Hauge at the Centre interviews Lars Skjold Wilhelmsen, Manager of the Department of Audiovisual Instruction at the University of Bergen since its establishment in 1972. This department is a service institution for the production of teaching aids, information programs etc. with the aid of video, photos and sound. It also arranges courses in the use of this type of technology.

Wilhelmsen points out that up to now humanists have mediated knowledge mainly through the spoken and written word. However, the Humanities can also benefit from mediation via video technology and data processing.

In Wilhelmsen's opinion, the optical videodisc combined with microcomputers opens up new possibilities for learning. For the mediation of the Humanities the relevance of this form of technology is clear, as the use of pictorial material is fundamental to especially the arts subjects. Using the videodisc in language teaching results in the interaction of visual and auditive impressions. The Department of Audiovisual Instruction is planning on experimenting with

videodiscs in connection with the development of a teaching program in art history.

Wilhelmsen believes that due to new students' proficiency in data processing, technology-based teaching methods must be adopted at the university level. He suggests establishing university fora for computer-assisted instruction in which topic-related insight, pedagogical knowledge and technical expertise are represented.

EDB-prosjekt for kunst- og kulturhistoriske museer

ADP project for museums of art and cultural history

In cooperation with the Centre Norwegian museums of art and cultural history have carried out experiments with data processing since the mid-70's, mainly in connection with data entry of artefact information and the production of catalogues. In recent years many museums have acquired their own micro-computers, a development which has created the need for increased support and cooperation. With this in mind the Council for Research in the Humanities has established a two-year research fellowship connected to the Centre. The fellowship has been awarded to Jon Birger Østby, who started work at the Norwegian Folk Museum on April 1st.

Østby reports that a goal of his project is to coordinate existing ADP systems, thus enabling museums to unite around joint solutions concerning both data technology and machine-readable source materials. The following tasks are planned: 1. Conduct a survey of the museums' use of and need for computer services. 2. Prepare a standard system for data entry, searching and printed output of microcomputer data. 3. Make suggestions for the organization and location of one central or several local data bases for museums. 4. Present a plan for information on and training in ADP for museums. 5. Evaluate new computer equipment which can be of use to museums. 6. Evaluate the use of data processing in other aspects of museum work.

Forskerstipend i automatisert arkivinformasjon

Research fellowship in automated archive information

The Council for Research in the Humanities has established a research fellowship (duration: 1 1/2 years) in automated archive information. The goal of this fellowship is to develop computer-based methods for improving researchers' access to catalogues and source materials held by archives institutions. The fellowship has been awarded to Hege Brit Randsborg, who works at the National Archives in Oslo in close connection with the Centre.

A data base is to be established in which information (mainly in the form of catalogues) on each Norwegian archive will be stored. It will be possible to search in this data base for various source types, subjects etc. Randsborg describes in detail a form she has developed which can be used for the entry of catalogue information. She has also started work on the entry of this information into a data base and the testing of different free text retrieval systems.

Frå informasjon til kulturarv

From information to cultural heritage

This is the title of a proposal made by a committee appointed by the Norwegian government in order to evaluate means of securing information stored on various media for present and future use.

Director Jostein H. Hauge at the Centre reports that the background for the appointment of this committee is that existing legislation does not cover documents stored on the new technological media. Hauge gives an overview of the contents of the proposal, and concentrates on the committee's suggestions concerning documentation of material on machine-readable media.

The committee outlines the problems connected with the production and storage of data processed documents, which are on the increase. In their opinion all documents of this kind, and the first and last versions of data bases, should be handed over to the National Library for storage.

Prosjektet «Norsk litteratur i Ungarn»

The project «Norwegian literature in Hungary»

Dr. Anikó N. Balogh at the ELTE University, Budapest outlines a project at the Institute of Germanic and Romance Languages and Literature on the reception of Scandinavian literature in Hungary from 1870 to today. Since 1982 this project has been connected with a comprehensive, interdisciplinary project with the goal of publishing a series of computer-stored bibliographies and establishing a database on the reception of Germanic and Romance literature in Hungary after 1945.

Balogh describes the composition of the database and the typology of the information it contains, along with the principles for the analyses of the Norwegian bibliographical register which have taken place. These analyses were recently completed, and some of the registers will be published in the course of 1984.

Tutorial on Machine Translation

This conference took place in Lugano, Switzerland, April 2-6. Director Jostein H. Hauge at the Centre reports that it was attended by 110 researchers in computational linguistics, computer specialists and translators from Europe, USA and Canada. The leader of the conference was Dr. Margaret King from the Dalle Molle Institute of Semantic and Cognitive Studies (ISSCO) in Geneva. At present ISSCO is coordinating the EEC's translation project EUROTRA.

The program of the conference was divided into three: 1. the development of machine translation from before the Second World War to today, 2. linguistic and computational models of special importance to the field, 3. on-going developmental projects and translation systems in present use. Hauge gives a summary of the speeches given on these topics.

ALLC Louvain-la-Neuve, 2.-6. april

Computing Officer Ole Lauvskar at the Centre reports on the 11th international conference held by the Association for Literary and Linguistic Computing in Belgium. This conference, which was attended by 200 people, was divided into

two parallel sessions covering a total of 45 papers. Workshops on literary statistics and data bases were also arranged.

The subjects dealt with were wide-spread and included authorship studies, stylistics, literary statistics, textual criticism, syntactical analysis, morphology, phonology, terminology, lexicography, computer-generated dictionaries, lexical data bases, natural languages, software, and data processing of texts.

Lauvskar gives summaries of papers given at the conference on literary statistics, software, and data bases.

Conference on Computers in English Language Research

This conference – the fifth of its kind – took place in Windemere, England, May 20-23. Senior Computing Officer Knut Hofland at the Centre reports that it was arranged by Professor Geoffrey Leech at the University of Lancaster and was attended by 60 people from 15 countries. The speeches given dealt mainly with the preparation and use of text corpora in research.

Henry Kucera, Brown University, USA spoke on electronic thesauri, whereas Jan Svartvik and Mats Eeg-Olofson, Lund University, Sweden presented the new project TESS, «Text Segmentation in Speech.» Two speeches were on corpora of India-English, and a corpus of technical English for the teaching of Chinese students was described. Several presentations were given on work based on the Lancaster-Oslo/Bergen corpus. Speakers from Holland included Gert van der Steen, Amsterdam, who described QUERY, a program for information retrieval from a text corpus, and Jan Aarts, Nijmegen, who spoke on, among other things, work on «the linguistic data base.»

The next conference will take place in Lund, Sweden, in May/June 1985.

Studiereise til England 21.-25. mai 1984

Study visit to England, May 1984

Senior Computing Officer Per Vestbøstad at the Centre reports on a visit he and Research Fellow Jon Birger Østby made to England. The purpose of this tour was to study the use of both data processing at English museums and interactive video at educational institutions. Østby gives a separate report on the former.

The Open University has carried out pedagogical experiments on interactive video, whereas the Department for Mechanical Engineering at Imperial College runs a software centre for interactive video programs (CEDAR). At London University a test disc of demonstration material from various subjects has been made. The Geological Museum is also currently testing an interactive video program, designed for the benefit of visitors.

A visit was also paid to Oxford University Computing Service, where developmental work is being carried out on the only university-owned phototypesetter in England. OUCS runs a national Text Archive which contains numerous British and foreign machine-stored texts. Work on large data bases is in progress at the Service: one of Greek personal names, the other of Greek vase shapes and decorations.

Museum Documentation Association

Research Fellow Jon Birger Østby reports on his visit to Museum Documentation Association in Duxford, England. MDA was started in 1977 by 10 British

national museums as a program aimed at developing ADP systems for museums. MDA is now an official agency of which 19 national museums and Great Britain's 9 Area Museum Councils are members.

In cooperation with various museums MDA has developed 22 different index cards which are used by 300-400 museums. However, these cards do not satisfy all of each museum's needs, a problem which MDA is trying to solve.

MDA has developed the program GOS for the production of registers and catalogues. GOS is designed for mainframes and can therefore only be employed by some large museums.

About 40 smaller museums enter their data on their own microcomputers, whereas 15 send their data to MDA for entry. However, it is expected that in the course of 1985 about 100 museums will have micro's at their disposal. All catalogues are produced at MDA on the basis of data sent by members on diskette or tape.

MDA issues a number of reports and other publications, among them the quarterly newsletter MDA Information. MDA's address: MDA, Duxford Airfield, Duxford, Cambridgeshire, CB2 4QR, U.K.

Norsk Arkivseminar, Fana Folkehøgskole, 24.-26. mai 1984 Norwegian Archive Seminar

Senior Archivist Anne Hals and Research Fellow Hege Brit Randsborg report that one of the two main subjects dealt with at this seminar was data processing and archives institutions. They give a summary of the lecture held by Department Archivist Ivar Fønnes on this subject.

Fønnes gave an overview of relevant data processing tasks for archives institutions, which include: 1. reception of ADP-stored archive material, and the expansion of services to the users of archives, 2. Participation in the shaping of the governmental administration's ADP-systems, 3. Automated archive information, 4. Data entry of historical sources in order to facilitate retrieval etc.

Other subjects Fønnes discussed were internal training in data processing and work on guidelines for journal entry, storage etc. He also spoke on existing data processing resources in archives institutions and plans for expansion.

COLING 84

The 10th COLING conference, arranged by Martin Kay from Xerox, took place at Stanford University, California, 2-6 July. Senior Computing Officer Knut Hofland reports that it was attended by 500 people and that 120 speeches were given in five parallel sessions.

The most common topic at the conference was parsing in combination with syntax, programming languages and logics. Many speeches also dealt with machine translation, generating of natural language, interface and pragmatics. Panel discussions were arranged on e.g. quantitative linguistics, natural language and data bases, and machine-readable dictionaries. Demonstrations were given of software on LISP computers, a Canadian term bank, a system for the teaching of computational linguistics and two systems for machine translation.

A summer school was also arranged this year, which consisted of six courses held the week before the conference. Hofland took part in the course «Parser

Construction Techniques», led by Henry Thompson from the University of Edinburgh. An introduction was given to a number of programs (recognizers, generators and parsers) for automatic analysis and synthesis of natural language.

Meldinger

News

This autumn the University of Gothenburg will start a new four-year course on computational linguistics. Some of the subjects that will be dealt with are: programming, automatic linguistic analysis, formalized descriptions of natural language and the development of information systems.

Associate Professor Brian MacWhinney announces that he has established a System for the Exchange of Child Language Data at Carnegie-Mellon University in the USA. Researchers are invited to join this system by contributing their data. Write to: Associate Professor Brian MacWhinney, Dept. of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania, 15213, USA.

12.000 pages of the philosopher Ludwig Wittgenstein's posthumous works have been transferred to machine-readable form in Tübingen, West Germany - the remaining 6.000 pages are in the process of receiving the same treatment in Norway. The University of Bergen has taken over the whole project as a joint venture between the Department of Philosophy and the Centre. The West German researcher Michael Kulemann is currently staying at the Centre in order to work on the standardization of codes and implementation of programs he has developed for data processing of the material.

The Norwegian Research Centre for Computers and Law together with the Centre have applied for funding for a project on automatic indexing. This project will deal mainly with linguistic methods, i.e. lemmatization, identification of parts of speech and splitting of compound words. So far the project group has received funding for three months of further research work, besides meetings with cooperating linguists and library and documentation personell in Scandinavia.

Forthcoming conferences:

National computing conference for the humanities - Geilo, Norway, 14-16 November 1984. Arranged by the Centre.

RIAO 85 - symposium on computer assisted information retrieval, Grenoble, France, 18-20 March 1985. More information from: Centre des Hautes Études Internationales d'Informatique Documentaire, 36 bis, Rue Ballu, 75009 Paris, France.

12th International ALLC Conference - Nice, France, 5-8 June 1985. More information from: Professeur Étienne Brunet, Institut national de la langue française (URL9), 98, bd E. Herriot, 06007 Nice, France.

ICDBHSS/85 - International Conference on Data Bases in the Humanities and Social Sciences, Grinnell, Iowa, USA, 22-24 June 1985. More information from: Thomas F. Moberg, ICDBHSS/85 Coordinator, Grinnell College, P.O. Box 805, Grinnell, Iowa 500112-0810, USA.

ICCH '85 - International Conference on Computers and the Humanities, Provo, Utah, USA, 26-28 June 1985. More information from: Randall Jones/ICCH85, Humanities Research Center, 3060 JKHB, Brigham Young University, Provo, Utah 84602, USA.

Forts. fra 2. omslagsside.

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

Det er ingen grunn til å venta



administrative computer systemer
353 STRAUME
tel.(05)33 14 40



**Alle ventar på UNIX -
Det nye standardsystemet på datamaskiner**

Du treng ikkje venta lengre. ACOS har brukt UNIX-utstyr frå 1982 og vi er svært nøgde med resultatet.

Vi har vald å satsa på Altos som Unix-partnar, og det har vi ingen grunn til å angra på.

Altos har maskiner frå dei minste arbeidsstasjonar til dei meir vidløftege systema med rom for tilkopling av 16 skjermar. Det som kjenneteiknar Altos-produkta er likevel at alle maskintypane kan nytta Unix operativsystem og 3 forskjellige nettverksystem, m.a. ETHERNETT. På denne måten kan fleire hundre terminalar «snakka» med kvarandre.

Dei fleste maskinene kan dessutan nytta CP/M og MSDOS som alternative operativsystem.

Altos leverer òg maskiner med mange typar prosessorar - Intel 8086 og 80186 og Motorola 68000 (fleire er i kjømda).

For dei som tvilar på at Altos er eit godt Unix-alternativ, viser vi til tidskriftet BYTE frå august 1984. Der var det ingen tvil.

Universitetet i Bergen har no mange installasjonar frå ACOS, og vi er ikkje uvillige til å hjelpa til med fleire.

Du har oppgåva - Vi har løysinga