

humanistiske data

Meldingsblad for
NAVF's EDB-senter
for humanistisk forskning

Norges Almenvitenskapelige Forskningsråd



**NR
2
1974**

Innhold:

3. REDAKTØRENS SPALTE.
4. NYTT STYRE FOR NAVF'S EDB-SENTER FOR HUMANISTISK FORSKNING.
NORSK LITTERATURKRITIKK 1951 - 1965.
5. TEKSTSØKING.
Tekstene som styrer oss.
Av Jon Bing.
6. KOMITE FOR DB-TERMINOLOGI.
7. EDB SOM FAKTOR I KUNNSKAPSPRODUKSJONEN.
Kommentar til Arne B. Johansen
av Ivar Fønnes.
9. DET TREDJE SYMPOSIUM OM BRUK AV DATAMASKIN I LITTERATUR- OG SPRÅKFORSKNING — CARDIFF, 1. - 5. APRIL 1974.
Av Geir Kjetsaa.
10. EDB-KONSULENT I TRONDHEIM.
11. SPRÅKFORSKARENS KRAV PÅ IN- OCH UTMATNING.
Typussetningens roll vid språklig databehandling.
Av Rolf Gavare.
14. SAMKATALOG FOR HANDESKRIFTMATERIALE.
15. 3RD INTERNATIONAL SUMMER SCHOOL FOR COMPUTATIONAL AND MATHEMATICAL LINGUISTICS, PISA 1974.
By G. Michael Gillow.
17. DATAMASKINELL ELLER TRADISJONELL KORREKTURLESNING?
Av Roald Skarsten.
21. EDB OG PRIMAÆRKILDER.
21. DE NORDISKE ARKIVDAGER — UPPSALA 3. - 4. AUGUST 1974.
Av Ivar Fønnes.
22. PRØVEPROSJEKT OM REGISTRERING AV FOTOGRAFISK MATERIALE.
24. KURS I DATAMASKINELL SPRÅKBEHANDLING, KØBENHAVN JULI/AUGUST 1974.
Av Eirik Lien.
25. EDB-ANVENDELSER I ARKEOLOGI.
Konferanse ved University of Birmingham 4. og 5. januar 1975.
25. NORDISKA SAMARBETSGRUPPEN FOR SPRÅKLIG DATABEHANDLING.
Syfte och sammansättning.
26. ASSOCIATION FOR LITERARY AND LINGUISTIC COMPUTING.
av Kolbjørn Heggstad.
27. MEDARBEIDERE I DETTE NUMMERET
28. MELDINGER OM SENTERETS VIRKSOMHET VÅREN OG HØSTEN 1974.
30. EDB-TJENESTEN VED DET HISTORISK-FILOSOFISKE FAKULTET, UNIVERSITETET I OSLO.
31. DET NORDISKE ØDEGARDSPROSJEKTET — NORSK AVDELING.

Redaktørens spalte

For 10 år siden var det svært få av de humanistiske forskere i vårt land som hadde gjort seg kjent med og tatt i bruk databehandling som ledd i sitt vitenskapelige arbeid. Når en derfor i dag forsøker å ta et overblikk over de humanistiske EDB-aktiviteter, er utviklingen i den siste 10-års perioden egnet til å forbause.

Fra å være et nytt felt preget av eksperimenter og begynnervanskeligheter er humanistisk databehandling i dag blitt et innslag som en regner med ved universitetene. Situasjonen kan for det første illustreres ved det faktum at det nå er en organisert EDB-tjeneste for humanister ved 3 av våre 4 universiteter. Det er også karakteristisk at humanister i flere tilfelle har tatt plass i de styrende organer for universitetenes EDB-sentra. I dag er f.eks. styreformannen for EDB-senteret ved Universitetet i Oslo rekruttert fra det nye arbeidsfeltet — humanistisk databehandling.

I denne sammenheng er det også grunn til å understreke at representantene for det nye anvendelsesområdet i det store og hele er blitt meget vel mottatt både ved universitetenes driftsavdelinger og blant de allerede etablerte brukergrupper. I våre dager inngår humanistene i det praktiske samarbeid ved universitetene på viktige områder som gjelder universitetenes EDB-virksomhet.

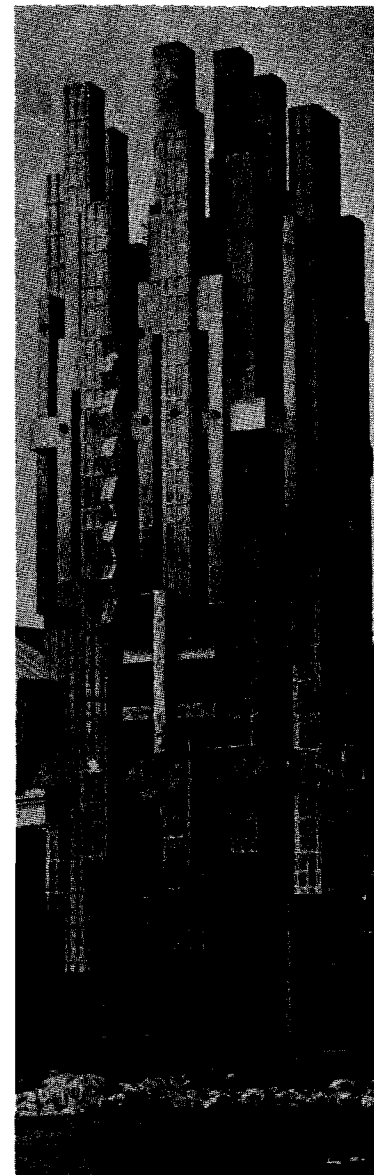
På den annen side er det fremdeles et sterkt behov for å markere de sider ved databehandlingsarbeid som er sentrale for anvendelsen i humanistisk forskning og klargjøre både fellesskap og forskjell i interesser overfor andre brukergrupper. At det finnes fellesskap mellom humaniora og andre arbeidsfelt er åpenbart. Ikke minst gjelder dette arbeidsområder som tekstbehandling og tekstsøking. Dette emnet må absolutt kunne karakteriseres som tverrfaglig, samtidig som det også er et sentralt innsatsområde i utviklingsorganer for offentlig forvaltning og næringslivet. I gjestespalten i dette nummeret drøfter Jon Bing dette temaet i sammenheng med arbeidet innenfor rettsinformatikk.

Den raske utbygging av virksomheten innenfor humanistisk databehandling er også tydelig når en vurderer den mot en nordisk og internasjonal horisont. Ikke minst viktig vil det være i fremtiden å styrke og utbygge videre det nordiske samarbeidet for å sikre at resultatene av utviklingsarbeid kan få en effektiv spredning på nordisk basis.

Derfor vil dette meldingsbladet også ha et åpent vindu mot de andre nordiske land — og utlandet for øvrig. I dette ligger det også en invitasjon til forskere og EDB-medarbeidere utenfor Norge om å bruke vårt meldingsblad som et redskap til spredning av informasjon.

Dette er skorsteinen hos vår nabo, Studentsenteret, Universitetet i Bergen. Fra redaktørens vindu nærliggende å tolke som et symbol på de humanistiske vitenskaper — mangfold og enhet.

*Skulpturen er laget av Arne Vinje Gunnerud.
(foto: AV-avd., Univ. i Bergen).*



Nytt styre for NAVF's EDB-senter for humanistisk forskning



*Universitetslektor Leiv Mjeldheim,
formann i styret
for NAVF's EDB-senter.*

Fagråd A oppnevnte på møtet 17.6. nytt styre for senteret. Styret fungerer fra 1.7. 74 og inntil nytt styre er oppnevnt av det nye fagrådet i 1976.

På møtet ble det også besluttet å oppløse NAVF's EDB-komite for humanistisk forskning som hadde fungert som styre for senteret fra opprettelsen av det i 1972.

Det nye styret har følgende medlemmer:

Universitetslektor Leiv Mjeldheim,
Historisk institutt, UiB.

Avdelingsleder Carl Erik Ellingsen,
Avd. for elektronisk databehandling, UiB.

Dosent Arne Hannevik,
Institutt for litteraturvitenskap, UiO.

Professor Eva Sivertsen,
Engelsk institutt, UNIT.

Professor Egil Pettersen, Nordisk institutt,
UiB er varamann og adm.leder Jostein H.
Hauge, NAVF's EDB-senter sekretær for det
nye styret.

Medlemmene Mjeldheim, Hannevik,
Sivertsen og Pettersen sitter i Rådet for
humanistisk forskning. Ellingsen er medlem
i NAVF's EDB-utvalg.

I forbindelse med reorganisasjonen av
senterets styre har Fagråd A i møte 9.9. 74
vedtatt nytt mandat for det:

Mandat:

1. Styret skal holde seg løpende orientert om driften av sentret. I samråd med sentrets daglige leder skal det sørge for at dette tjener sine formål.
2. Styret har til oppgave årlig å utarbeide forslag til langtidsbudsjett, årsbudsjett og arbeidsprogram og legge dette fram for fagrådet til godkjenning.
3. Styret forbereder og gir tilråding til fagrådet om saker som angår sentrets drift.
4. Styret skal både holde fagrådet løpende orientert om sin virksomhet og utarbeide årsrapport.
5. Styret skal søke nødvendig kontakt med andre organer som arbeider for bruk av EDB i forskningsvirksomhet.

NORSK LITTERATURKRITIKK 1951-1965

Ved Nordisk institutt, Avdeling for nordisk litteratur, Univ. i Bergen har det ei tid vore i gang registrering av aviskritikk om norsk skjønnlitteratur. Registreringa byggjer for det meste på forlagsarkiva, og gjeld eit utval norske forfattarar for perioden 1951-1965.

Stoffet er blitt puncha ved Prosjekt for datamaskinell språkbehandling, Nordisk institutt; korrekturarbeid er i gang, og ein reknar med at materialet kan bli presentert med det fyrste i form av ein katalog som er ordna etter forfattarar og verk.

Idar Stegane.

TEKSTSØKING

Tekstene som styrer oss

Det gamle munnhullet om «ordets makt» burde vært formulert av en jurist — for ingen er som han daglig konfrontert med den viktige rolle ord og tekster spiller i vårt samfunn. Som medlemmer av en rettsstat er vi alle fortrolig med tanken om at grensen mellom rett og urett trekkes av lover, og at lover sammen med forskrifter og andre regelverk regulerer stadig større del av det moderne samfunn. Men det er vel egentlig sjelden at vi retter vår oppmerksomhet mot det selvfølgelig forhold at disse normer for vår adferd er formulert i ord og foreligger som tekster.

Og det er vel også først og fremst i lys av EDB at dette blir viktig. For *tekster* kan behandles maskinelt — i motsetning til regler og normer.

Et av de grunnleggende prinsipper for det juridiske spill er at når man argumenterer for en rettslig norms eksistens eller innhold, må man ta én eller flere rettskildefaktorer som utgangspunkt for sin argumentasjon. «Rettskildefaktorer» er en samlebetegnelse for de kilder som det er tillatt å bygge denne argumentasjonen på, og typiske eksempler er en lovtekst eller en dom. Det finnes nok eksempler på rettskildefaktorer som *ikke* er skrevne (f.eks. sedvanerett) — men de fleste og praktisk viktigste foreligger som tekster. Fra et EDB-synspunkt er denne utgangspo-

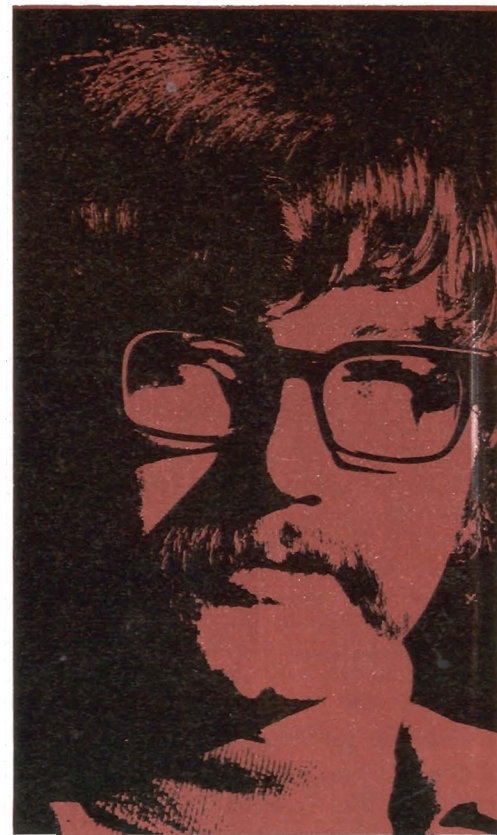
sisjonen nesten utrolig gunstig: Selve det *grunnlag*, de primærkilder, som jurister (og hele rettsvitenskapen) bygger på, er for det meste tekster. Det har derfor vært nærliggende å undersøke hvilke fordeler datamaskinell tekstbehandling kan gi jurister.

Juristen er derfor først og fremst interessert i tekstenes *innhold*, (dette i motsetning til f.eks. en språkforsker som kan være interessert i selve det språk som er brukt i lover). Dernest er det viktig for juristen å få adgang til mulig relevante teksters *nøyaktige* ordlyd — ellers vil han kunne gå glipp av holdepunkter i teksten som indikerer nyanser som er relevante nettopp for det konkrete og helt spesifikke problem han arbeider med.

Tekstsøkesystemer.

Dette har ført til at man for å hjelpe jurister, har laget datamaskinbaserte tekstsøkesystemer. Disse systemene dukket opp i slutten av 50-årene, og allerede den gang valgte man den typen systemer som med et unøyaktig stikkord kalles «fulltekstssystemer». Dette vil si at man leser inn i systemet bare og hele det originale dokumentets fulle tekst (f.eks. hvert ord i en lovtekst). Teksten blir altså *ikke* bearbejdet ved f.eks. tildeling av karakteriserende stikkord, en systematisk klassifikasjon e.lign.

Gjestespalten



Jon Bing

Når juristen skal bruke tekstsøkesystemet, er han ute etter tekster med et bestemt *innhold*. Datamaskinen må imidlertid få en formell beskrivelse av dokumentet. I et fulltekstsystem får denne beskrivelsen i utgangspunktet form av krav til hvilke *ord* dokumentet skal romme. Juristen oversetter altså sitt problem til ord (1) beskriver problemet, og (2) han tror kan være brukt i rettskildefaktorene. Leter juristen f.eks. etter dommer som kan fortelle noe om retten til å kreve inntekstfradrag for utgifter til studiereiser, kan han be om å få dommer som inneholder ordene «skatt», «studier» og «reiser». (I praksis kan teknikken være adskillig mer raffinert.)

Rettslige tekstsøkesystemer av denne typen har nå vært i bruk i flere år, og er i ferd med å bli introdusert også i Norge. De representerer en noe annen løsning enn konvensjonelle dokumentasjons- eller bibliotekssystemer. Der vil f.eks. en vitenskapelig artikkel bli beskrevet ved hjelp av enkelte faste opplysninger (forfatter, trykningsår osv.) og et kort innholdsresymé (ofte bygget opp av emneord). Slike systemer vil altså forutsette en bearbeidelse og tolking av artikkelen som ikke er nødvendig i et fulltekstsystem.

Kobling til språkforskning.

Fulltekstsystemer konfronterer jurister med en lang rekke spørsmål som han ikke er forberedt til å hankses med.

Det har bl.a. ført til at juristene får en oppfordring til å se selve språket i lover,

dommer og andre rettskildefaktorer fra en ny synsvinkel. (I dag foreligger ca. 20 000 ord med lover, og 1,2 millioner ord med rettsavgjørelser i maskinlesbar form.) Konstruksjonen av tekstsøkesystemer forutsetter f.eks. kunnskap om ordfrekvenser og ordbruk — en kunnskap som man tidligere ikke hadde særlig oppfordring til å søke.

I arbeidet med å klarlegge slike forhold, må juristen bygge på det arbeidet som språkforskere allerede har nedlagt, og ta deres resultater og metoder som utgangspunkt. Språkforskningen kommer inn på en helt annen måte enn før som direkte relevante for juristens arbeid — noe som forhåpentligvis kan være fruktbart for begge grupper.

Kobling til andre fagområder.

Det er selvfølgelig heller ikke overraskende at juristen — når han først kommer opp av sin egen konfliktorienterte skyttergrav — oppdager at folk som arbeider innen andre rettsområder har problemer som ligner mye på de jurister forsøker å løse ved hjelp av tekstsøkesystemer. En historiker kan ha en dokumentsamling han konsulterer for å se hvordan de belyser historiske forhold — et tekstsøkesystem ville her kunne gi ham den nødvendige hjelp. En arkeolog har en samling beskrivelser av funnsteder — et tekstsøkesystem kunne tillate ham å lete rundt i denne samlingen uten at beskrivelsene ble formalisert eller kodet. Og vil teologer utnytte den mulighet som er åpnet ved at den norske Bibelen nå finnes i maskinlesbar form?

Det ville sikkert være fruktbart om jurister i

arbeidet med tekstsøkesystemer ble gjort kjent med de problemstillinger som møter forskerne på andre fagområder i deres kamp med tekster som kildemateriale. Et tekstsøkesystem er et enkelt hjelpemiddel — det forutsetter ikke at tekstene blir forhåndsanalysert eller katalogisert. Det vil likevel åpne muligheten for å undersøke om det blant de registrerte tekster finnes momenter som er relevante i forhold til en teori eller problemstilling.

Jus regnes visstnok ikke tradisjonelt som et «humanistisk» fag. Men i alle fall når det gjelder det datamaskinelle aspektet, synes det klart at det er sterke felles interesser mellom rettslig og humanistisk forskning. Det er viktig at ikke tradisjonelle skiller mellom fag skal stenge veien for et naturlig samarbeid.

KOMITE FOR DB-TERMINOLOGI

Norsk språkråd har sett ned ein komite som skal utarbeide ei norsk db-ordbok. (db = databehandling)

Fagspråket innan elektronisk databehandling er (som kjent) svært mykje farga av engelske termar, og målet for komiteen er å finne fram til dei mest brukte ord og uttrykk for å gjere framlegg om norsk fagtermar. Ordboka vil innehalde ca. 1000 termar med korte forklaringar.

Komiteen samarbeider med Prosjekt for datamaskinell språkbehandling, Nordisk institutt, Universitetet i Bergen, om framstilling av boka.

Formann i komiteen er Per Scott, Rasjonaliseringsdirektoratet.

Kolbjørn Heggstad.

EDB som faktor i kunnskapsproduksjonen

Kommentar til Arne B. Johansen av Ivar Fønnes

I forrige nummer av *Humanistiske Data* (HD 1/74) har Arne B. Johansen en artikkel hvor han drøfter forholdet mellom data og problemstilling/teori, spesielt i relasjon til bruk av EDB. Artikkelen inneholder en del synspunkter som kan være verd en kommentar.

Først litt flisespikkeri. Johansen innleder artikkelen med en definisjon av data som neppe kan være særlig fruktbar. Uten at det er eksplisitt formulert, synes hans definisjon å være omtrent slik: Data er materiale som er egnet til å belyse en gitt teori, m.a.o. at data ikke eksisterer uavhengig av en gitt teori. Nå kan jeg imidlertid ikke se at Johansen gjennomfører bruken av ordet «data» i denne betydning i resten av artikkelen. Tvert imot tyder hans formuleringer på at han bruker «data» på vanlig måte, nemlig i ordets egentlige betydning: det som er gitt.

Johansens definisjon av data berører imidlertid det som er hovedpoenget, og et vesentlig sådant, i hans artikkel: Innsamling og håndtering av data må ikke «isolerers fra utviklingen av modeller (teorier)». Data skal tas hånd om «dersom de tjener til å belyse (bekrefte eller avkrefte) en eller flere teorier». Derimot skal de ikke tas hånd om dersom begrunnelsen er at de «muligens kan bli av betydning en gang i fremtiden».

Dette er synspunkter som ikke er nye, men det kan likevel være på sin plass å understreke dem, særlig i forbindelse med bruk av EDB. Oppbygging av store dataarkiver i maskinleselig form, med det formål å stille data til disposisjon for eventuelle interesserte forskere, kan i stor grad bidra til å styre forskningen. I og med at registrering av data for maskinell behandling er en både tidkrevende og kostbar oppgave, er det naturlig at forskere ønsker å ta i bruk data som allerede er maskinleselige. Man kan altså komme i den situasjon at dataarkivene styrer forskernes valg av tema og problemstillinger. Om man ville sette saken på spissen, kunne man hevde at den oversikt over data i maskinleselig form som befinner seg noen sider bak Johansens artikkel i HD 1/74, er svært skadelig fordi den kan påvirke norske forskeres valg av forskningsobjekt.

Rent prinsipielt kan jeg i hovedsaken si meg enig i Johansens synspunkter. Som *ideelt krav* er det viktig å understreke at valg av forskningstema og problemfelt er det primære, og at data samles inn for å belyse de problemer man arbeider med — ikke at problemfelt velges på grunnlag av hvilke data som er best tilgjengelige.

Men *i praksis* tror jeg det er nødvendig og hensiktsmessig å moderere synspunktet noe.

Dette skyldes for det første at faren neppe er så stor som det Johansen hevder, for det andre at vi svært ofte *er nødt til* å la valg av forskningsobjekt være påvirket av datasituasjonen, og for det tredje er det et spørsmål om økonomi (særlig i forbindelse med EDB). Jeg skal forsøke å utdype dette nærmere.

Johansen hevder at dersom man samler inn data fordi «de muligens kan bli av betydning en gang i fremtiden», så betyr det at man «samler stadig flere data som belyser gamle teorier». Dette behøver, etter min mening, ikke nødvendigvis være tilfelle. Det kan meget vel tenkes — og forekommer vel ofte — at man bruker «gamle» data for å belyse nye teorier som går på tvers av tidligere teorier. (Dette forutsetter da at data er bevart i en form som er uavhengig av de teorier tidligere forskere har anvendt på dem). Valg av forskningstema vil naturligvis være begrenset om man forutsetter bruk av en bestemt datamasse, men det er neppe grunnlag for å hevde at man derved bare konsentrerer oppmerksomheten om «den etablerte kunnskap».

I forskningssammenheng vil situasjonen svært ofte være at man har nokså begrensede datamengder og datatyper til rådighet. Særlig gjelder dette ved studier av fortiden. De kilder som historien har overlevert, er

det vi har tilgjengelig. Vi kan ikke etablere nye data som ikke bygger på de overleverte kilder.

Det sier seg selv at forskerne må ta hensyn til dette i sitt valg av forskningstema og problemstillinger. Det er bortkastet tid å velge et tema som ikke kan belyses ved hjelp av de overleverte kilder. Derfor er det hensiktsmessig — og vanlig — at forskere ved valg av forskningsoppgaver også tar hensyn til datasituasjonen. Dette gjelder også ved valg av metode. Vi kan tenke oss at en forsker ønsker å studere norske gårder i eldre tid. For å kunne gi generelle utsagn for hele landet, velger han å undersøke et statistisk representativt utvalg av gårder fra hele landet.

Når det så viser seg at det bare finnes data for en del av gårdene, er det viktig at denne oppdagelse blir gjort før det legges ned et stort arbeid i det metodiske opplegg.

Nødvendigheten av å la datasituasjonen påvirke valg av tema og metode er altså til stede selv om man ikke bruker EDB. I tilknytning til bruk av EDB blir dessuten det økonomiske aspekt et vesentlig argument for å moderere Johannsens prinsipp. Data i maskinleselig form er i mange sammenhenger langt lettere å behandle for analyseformål enn andre data. Til gjengjeld koster det mye både tid og penger å gjøre data maskinleselige. Dette forhold leder naturlig fram til det syn at når først et materiale er gjort maskinleselig, bør det stilles til disposisjon for de forskere som har bruk for det. Det er viktig å unngå at forskere bruker store ressurser på innsam-

ling av data som allerede foreligger på maskinlesbart medium.

Det er selvsagt at en slik politikk medfører at forskeres valg av tema og metoder påvirkes av hvilke data som er tilgjengelige for maskinell behandling. Dette kan i en del tilfeller ha uheldige virkninger, men det fører — som nevnt — ikke nødvendigvis til at man bare konsentrerer sin oppmerksomhet om den etablerte kunnskap, slik Johansen hevder. Dessuten må det kunne anføres at jo mer data som registreres for maskinell behandling, desto mer data får forskerne å velge mellom, slik at valget i stadig mindre grad påvirkes av datasituasjonen.

Når det gjelder bruk av EDB, peker Johansen spesielt på faren for at lett håndterlige datamasser frister til ren datamanipulasjon isolert fra enhver form for teori. Dette er det naturligvis verdt å være oppmerksom på. Selv om ren datamanipulering forekommer uten datamaskiner, er det klart at mulighetene for denne type anvendelse av data (som for andre typer anvendelse) blir langt «bedre» når man kan behandle store datamasser maskinelt.

Dette er imidlertid et problem som gjelder den enkelte forskers holdning til bruk av data, og det har etter min mening lite å gjøre med spørsmålet om hvilke prinsipper som skal legges til grunn for datainnsamling. Det er ingen grunn til å regne med at dataarkiver som er bygget opp av andre, skulle friste mer til ren datamanipulering enn maskinleselige data som forskeren selv har registrert for sitt prosjekt. Muligheten

for ren datamanipulering vil alltid være til stede. Bruk av EDB i forskningen vil nødvendigvis utvide muligheten, og dermed kanskje også fristelsen til denslag bruk av data. Dette er altså en følge av EDB-bruk generelt, uavhengig av prinsippene for datainnsamling.

Om jeg til slutt på grunnlag av ovenstående betraktninger skulle framheve noen prinsipper for tilrettelegging av data i maskinleselig form, måtte det bli følgende:

- Oppbygging av dataarkiver i maskinleselig form (med sikte på at andre skal bruke dem) bør bare skje ut fra et bevisst ønske om å stimulere til bruk av disse data. Man må regne med at eksistensen av slike dataarkiver vil bidra til å styre forskeres valg av forskningsobjekt, slik at f. eks. interessen for lignende datasett som ikke er maskintilgjengelige, blir mindre.
- Ved overføring av data til maskinleselig form bør man ha som mål at den maskinleselige versjon skal være en tro kopi av originalen, dette for å unngå at informasjon går tapt og at senere brukere må basere seg på teorier som innsamleren har lagt til grunn.

Disse prinsipper for innsamling av data tar altså sikte på å stille data til disposisjon for andre forskere uten at skadevirkningene overstiger et akseptabelt nivå.

Hvordan de enkelte forskere håndterer tilgjengelige data i relasjon til sin forskningsoppgave — det må bli forskernes eget ansvar.

Det tredje symposium om bruk av datamaskin i litteratur- og språkforskning - Cardiff, 1. - 5. april 1974.

Det er vanskelig å tenke seg moderne forskning uten internasjonal utveksling av erfaringer. Særlig viktig er slik kontakt når man søker nye veier i forskningen, for eksempel ved anvendelse av EDB.

I slutten av 1960-årene tok derfor en gruppe engelske forskere initiativ til avholdelse av regelmessige konferanser om datamaskinen i språklig og litterær forskning. Den første ble holdt i Cambridge i 1970, den andre i Edinburgh i 1972, og den tredje i Cardiff i begynnelsen av april dette år.

På konferansen deltok mer enn hundre forskere fra et dusin forskjellige land. I løpet av fire dager ble det holdt et førtital foredrag. De beste av disse vil snart bli trykt i bokform. (Jfr. de tidligere utkomne bøkene *The Computer in Literary and Linguistic Research*, Cambridge 1971 og *The Computer and Literary Studies*, Edinburgh 1973).

Et inntrykk man satt igjen med etter forrige konferanse var at forskere som anvender EDB i studiet av språk og litteratur kan deles i to grupper: de

«prossorienterte» og de «problemorienterte». Mens forskerne i første gruppe er mest opptatt av *hvordan* resultatene oppnås ved hjelp av datamaskinen, er forskerne i annen gruppe mer interessert i *hva* som

oppnås: er resultatene faglig betydningsfulle?

Denne «interessemotsetning» var atskillig mindre tydelig i Cardiff. Prossorientert forskning med vidløftige utredninger om software og hardware synes å være på vei ut fra disse konferansene. Mens man forrige gang kunne oppleve lidenskapelige konfrontasjoner mellom tilhengere av FORT-RAN og tilhengere av ALGOL, hadde «programsektoren» denne gang fått en langt mindre sentral plass.

Tiden synes mer og mer å bli et presserende spørsmål også for dem som arbeider med EDB. Begrepet «EDB-død» var ofte fremme i diskusjonene, i alle fall på det mer private plan. Med dette menes faren for simpelthen å drukne i de store mengder av informasjon som datamaskinene frem-skaffer. Å kunne bringe de enkelte prosjekter til en heldig avslutning innenfor et noenlunde rimelig tidsrom blir stadig viktigere i en bevilgningsfattig tid. Det er ikke lenger i seg selv meriterende å anvende EDB i sin forskning.

Også denne konferansen gav enkelte eksempler på EDB-dødens herjinger, ikke minst hos gjengangerne. Når man fra konferanse til konferanse taler om titlene på franske 1700-tallsromaner, når man på annet decennium oppholder seg ved en bestemt dikters innflytelse på en annen

dikter — da er det ikke særlig fremdrift i prosjektene.

Men heldigvis var det også forskere som kunne vise til gjennomførte prosjekter av stor forskningsmessig interesse. Her må særlig nevnes Pater Fr. R. Busa fra Venezia. I mer enn tyve år har han arbeidet med sin konkordans til Thomas fra Aquinos veldige produksjon. Nå er konkordansen begynt å utkomme, og den vil omfatte over tyve gedigne bind på tilsammen nærmere ti millioner ord. Selv om Busa har arbeidet med Paven i ryggen, er resultatet imponerende. Sikkert vil konkordansen få den største betydning for den videre utforskning av denne viktige middelalderfilosofen.

Flere av problemstillingene var i og for seg tradisjonelle: sammenligning av diktere ved studium av deres ord- og setningslengde, deres anvendelse av ord og ordklasser. Men det ble gang på gang demonstrert at man selv ved disse «grove» parametre kunne komme frem til overraskende slutninger om de forskjellige dikteres egenart. Forutsetningen er imidlertid et nært samarbeide med statistisk ekspertise.

På enkelte områder har man gjort betydelige fremskritt i løpet av de siste par år. Jeg tenker på homografseparering og lemmatisering, og ikke minst på maskinell oversettelse. Å benytte datamaskinen til

oversettelsesvirksomhet var særlig aktuelt i begynnelsen av 1960-årene. Senere har man kunnet notere økende skepsis og derav synkende bevilgninger. Likevel har engelskmannen T.D. Crawford hatt fin fremgang ved oversettelse av ikke-kunstneriske tekster fra russisk til engelsk. Mens man tidligere bare kunne lage en ord-for-ord oversettelse for å avgjøre om teksten fortjente å bli oversatt for hånd, er man nå i ferd med å skape fullt ut leselige tekster ved hjelp av datamaskinen. Paradoksalt nok er problemet nå at oversettelsene ikke må bli for gode, da det i så fall kan bli vanskelig å oppdage feil.

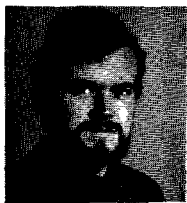
Fremdeles er det stor uenighet om humanistene selv bør lære seg å

programmere. Noen mente at dette ville være altfor tidkrevende, andre betraktet det som en fordel å ha et nærmest mulig forhold til sitt hjelpemiddel. Sistnevnte gruppe gjør unektelig et profesjonelt inntrykk, om enn ikke alle benytter sin datamaskin like flittig som professor Packard fra Los Angeles. Han *puncher* sine greske avhandlinger og får dem ut av datamaskinen i form av de skjønneste «preprints», helt ferdige til offset-fremstilling!

Et resultat av disse konferansene er opprettelsen av Association for Literary and Linguistic Computing (ALLC). Organisasjonen har i dag nærmere 300 medlemmer og utgir *ALLC Bulletin* med den energiske Mrs. Joan M. Smith som redaktør. Det er

grunn til å tro at denne organisasjonen i de kommende år vil bli et sentralt samarbeidsforum for EDB-interesserte humanister (jfr. omtalen av ALLC's arbeid annetsteds i dette nummeret). Hvis datamaskinen virkelig skal bli et verdifullt hjelpemiddel i vår forskning, er det viktig å benytte de muligheter til diskusjon som dette organet frembyr.

Neste konferanse vil bli avholdt i Oxford i 1976. Forhåpentligvis vil Norge da kunne stille med en større delegasjon.



NAVFs EDB-konsulent i Trondheim, cand. philol. Eirik Lien.

EDB-KONSULENT I TRONDHEIM

NAVFs EDB-senter for humanistisk forskning har høsten 1974 opprettet en konsulentstilling i Trondheim, som en videreføring av ordningen med EDB-konsulenter i Bergen og Oslo.

Konsulentstillingen er finansiert av NAVF og Universitetet i Trondheim i fellesskap. Konsulenten er administrativt underlagt NAVFs EDB-senter i Bergen.

Gjennom nær kontakt med senteret skal konsulenten være et bindeledd mellom dette og de humanistiske fagmiljøer i Trondheim, samtidig som han skal yte assistanse i EDB-spørsmål til de humanistiske forskerne på sitt arbeidssted.

Av sentrale arbeidsoppgaver kan nevnes generell opplysningsvirksomhet og konsulenttjenester i forbindelse med forskningsarbeid. Konsulenten skal dessuten ta del i kursvirksomhet om bruk av EDB i de humanistiske fagdisipliner, og han vil også ta på seg programmeringsoppgaver av mer generell interesse.

I konsulentstillingen er ansatt cand. philol. Eirik Lien. Konsulent Lien har universitetsutdanning i både humanistiske og matematiske fag, og har i ca. ett års tid vært

ansatt ved Regnesentret ved Universitetet i Trondheim (RUNIT).

Konsulenten har fast kontor i universitetets lokaler i Lade II, men har også adgang til arbeidsplass ved RUNIT. På NLHT vil konsulenten fra 1975 få disponere terminalutstyr som NAVF har bevilget til stillingen.

Konsulent Lien er interessert i samarbeid med humanistiske forskere både utenfor og innenfor universitetsmiljøet.

Vi vil derfor oppfordre alle humanister som er interessert i EDB-spørsmål å ta kontakt med konsulent Lien. Ved dette vil en ikke minst kunne sikre seg at de konsulenttjenester som en tilbyr humanistene, er best mulig tilpasset de EDB-behov som for tiden står mest sentralt i Trondheim.

Rolf Gavare

Språkforskarens krav på in- och utmatning

Typuppsättningens roll vid språklig databehandling

Vid språklig databehandling av skriftligt textmaterial finns ofta speciella krav på typuppsättningen. In- och utmatning av texter med en större, eller avvikande, teckenrepertoar innebär en stor komplikation. Det finns emellertid nu kraftfulla hjälpmedel för att komma till rätta med dessa problem. I den följande korta översikten behandlas de vanligaste problemen, och i anslutning därtill ges några exempel på lösningar som den moderna tekniken kan erbjuda. (Även om denna artikel främst tar upp textbehandlingens språkvetenskapliga aspekter, så finns dock naturliga beröringspunkter med litteraturvetenskap, biblioteksvetenskap, filologi med flera närliggande discipliner. Problemställningar av likartat slag finns också inom musikkvetenskapen.)

Önskemål beträffande teckenrepertoaren.

Nästan all information i en text kan vara lingvistiskt relevant. Emellertid har den datautrustning som hittills förekommit ofta varit synnerligen otillfredsställande från typografisk synpunkt, vilket avsevärt försvårat en enkel och säker inkodning och en tydlig utskrift. Datatekniken har nu kommit så långt att man i princip kan tillgodose samma typografiska krav som man sedan länge haft i fråga om tryckt text. Vilka önskemål kan då språkforskaren ha beträffande teckenrepertoaren? Ett primärt

krav är ofta att man skall kunna representera såväl versaler som gemena tecken. Redan detta till synes enkla önskemål har ofta inte kunnat uppfyllas på grund av att maskinfabrikanterna inte ansett det nödvändigt med gemena tecken på utrustning som ju normalt varit avsedd för administrativa eller numeriska tillämpningar.

Minst lika viktigt är kravet att kunna återge alla de olika tecken som förekommer i det eller de språk som studeras. Det engelska alfabetet brukar inte vålla några problem eftersom datautrustningen ofta tillverkas i Amerika eller är avsedd för den amerikanska marknaden. Kostnaderna för att på utländsk utrustning få med våra «national-karakterer» (å, ä, ö etc.) är tyvärr i många fall så stora att en del datacentraler låter sig nöjas med de engelska tecknen.

Om de två nu nämnda kraven kan uppfyllas, vilket får anses normalt vid en akademisk datacentral, kan bearbetningar av typografiskt enkla texter, t.ex. inhemska och engelskspråkiga skönlitteratur, presenteras på ett acceptabelt sätt. De verkliga svårigheterna uppkommer emellertid då man arbetar med främmande språk (andra än engelskan). Redan de skandinaviska språken vållar då problem i fråga om återgivningen av å, æ, ö och ø bland annat. Att alfabetiseringsprinciperna är olika i våra språk vållar också

komplikationer som i viss mån är förknippade med tecknens maskinrepresentation.

Också inom ett språk stöter man på dessa problem. Det kan gälla utländskt stoff som citat, titlar på böcker och filmer, namn etc, det kan gälla inlånade ord som behållit sin ursprungliga form, t.ex. i fråga om accentuering, och det kan gälla vetenskapligt språk eller fackspråk där speciella tecken och symboler används. Studerar man äldre texter behövs också ofta tecken som inte brukas i modern skrift.

Rikedomen på olika accenttecken, eller andra diakritiska tecken, vållar särskilda problem för många språk. Genom den mycket begränsade teckenuppsättning som brukar finnas på utskriftsenheterna kan man i allmänhet bara ha tillgång till några få accentuerade bokstäver. Man kan lösa detta genom att diakritikon behandlas som särskilda symboler, vilka trycks ut över, respektive under, bokstaven. Även om det är svårt att få en tydlig utskrift av alla diakritiska sammanställningar, oberoende av bokstavens form, kan detta ändå ofta ge ett tillfredsställande resultat. Problemen blir mest uttalade då inte enbart gemena autografem (vokaltecken) accentueras, exempelvis vid markering av tryckaccent eller prosodi.

da:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo
da:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo
:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo: enivəsi'feit *sprojk,da:ta jøtə'bo

Junkturen, logogram och andra symboler förorsakar också ofta bekymmer. För att ta några tecken som exempel: — (tankstreck), ° (grad), § (paragraf) × (multiplikationstecken), ‰ (promille), de synnerligen varierande konventionerna för bruket av olika citationstecken («, », „, “, ”, „, ‚, ‘ och’) samt användningen av ð och í i spanskan. Listan kan göras mycket lång.

En del typografiska stil- och gradväxlingar har också uppenbar lingvistisk relevans och bör därför kunna representeras även i datorutskrifter. Bokstäver och siffror i över- eller underkant har likaså särskild betydelse (M^{le}, 3rd. in-8° . etc.)

De hittills behandlade problemtyperna kan man stöta på i ordinär text med vanliga, latinska skrivtecken. En hjälplig lösning kan man då ofta komma fram till även med normal datautrustning. Av naturliga skäl accentueras svårigheterna då det undersökta språket har tecken ur icke-latinska alfabet, exempelvis grekiska, kyrilliska, herbreiska eller arabiska eller ett fonetiskt alfabet. I flerspråkiga texter, lexikon mm, hopar sig dessa svårigheter. Lämna man så den alfabetiska skriften och ser på en syllabografisk skrift som den japanska, en logografisk skrift som den kinesiska eller den piktografiska hieroglyfskriften så har svårigheterna hittills varit nästan oöverstigliga.

Hur skall vi då komma till rätta med dessa problem? Låt oss se vad det finns för utrustning som är lämplig för vårt ändamål.

Hur inmatningsproblemen kan lösas.

Texter som skall matas in har ofta en tryckt förlaga, och man frågar sig kanske i första hand om det finns något slag av optisk klartextläsare som kan identifiera normal tryckt text. Svaret på frågan blir i praktiken oftast nej. Anledningen ligger i de många begränsningarna hos dagens optiska läsare. För det första är tillförlitligheten ofta dålig: svårighet att skilja mellan vissa tecken, t.ex. punkt och komma, h och b eller å och ä, fläckar i papperet och ojämnt tryck liksom genomtryck från papperets baksida vållar störningar. Likaså får man ha restriktioner på papperskvalitet, sidformat, textytans placering på sidan etc. För det andra är konstant typvidd ett normalt krav. Detta innebär att man brukar vara begränsad till att använda utskrifter från skrivmaskin eller radskrivare. Flertalet tryckstilar har individuell typvidd. Slutligen är de flesta läsare starkt begränsade till igenkänning av vissa specialkonstruerade typsnitt avsedda just för optisk läsning — OCR — eller till vissa andra enkla snitt som finns hos radskrivare och skrivmaskiner. Typerna är hos dessa i allmänhet groteska, dvs har bokstavslinjer utan schatteringar (hårstreck). Numera finns några läsare som kan programmeras att känna igen olika typsnitt, t.ex. COMPUSCAN och REI-Reader, men många av de tidigare begränsningarna kvarstår, bland annat vållar stilblandningar problem. Det torde dock dröja flera år innan vi kan använda en flexibel och tillförlitlig optisk läsare till en rimlig kostnad. Om den text man önskar bearbeta

maskinellt är relativt nyligen tryckt, varmed här menas efter omkring 1960, bör man kontrollera möjligheten att sätteriet kan ha sättremsor eller -band kvar och om man i så fall kan få utnyttja dessa. Detta kan ge en unik möjlighet att få hela textmaterialet direkt i maskinläsbar form och med all erforderlig typografisk information inkodad. Vissa typer av sättremsor och -band kräver emellertid också att datacentralen har speciella läsare för dessa.

Om nu texten måste inkodas manuellt så kan det vid användning av standardutrustning, såsom hålräms- eller hålkortsstansar, magnetbandsinkodare och textskärmsterminaler, uppstå svårigheter om typuppsättningen är stor eller om icke-latinska tecken skall inkodas. Det finns vissa hålrämsstansar där skrivaren har utbytbart typhuvud och det finns avancerade magnetbandsinkodare med stor teckenuppsättning avsedda för den grafiska industrin. Dessa alternativ kan möjliggöra inkodning av t.ex. grekiska eller ryska texter utan användning av kryptiska kodifieringar som försvårar arbetet. Hålrämsalternativet gäller av naturliga skäl främst enspråkiga texter. Ställs större krav på teckenrepertoaren vid inkodning är man hänvisad till någon typ av grafisk bildskärmsterminal. Det finns för närvarande flera typer av skärmar med snabb teckengenerator som är direkt programmerbar. Man kan då alltefter behov omdefiniera varje knapp på tangentbordet och direkt vid nedtryckningen av en tangent se det önskade tecknet korrekt återgivet på skärmen. I stället för att använda tangentbordet kan det

ibland vara lämpligare att peka på den symbol man önskar med en s.k. ljuspenna bland en uppsättning som visas på skärmen. Sådana snabba grafiska skärmar med inbyggda minidatorer för teckengenereringen är bland annat IMLAC PDS och DEC GT.

I fråga om vissa tillämpningar kan det bli tal om rent grafisk inmatning t.ex. av kinesiska skrivtecken eller handskrift. Detta kan ske exempelvis med ljuspenna på en grafisk bildskärm, skrivning på någon typ av rasterplatta, t.ex. RAND-Tablet, mekanisk kurvföljning eller optisk avsökning. Val av metod är så beroende av forskningsinriktningen att några allmänna lösningar knappast kan anges.

Hur det språkliga materialet lagras internt i datorn efter inläsningen är sedan i princip likgiltigt, men man bör om möjligt här ta hänsyn till de speciella alfabetiseringsregler som finns i respektive språk för att underlätta senare sorteringar och sökningar på texten. Alfabetiseringsprincipernas inverkan på tecknens interna lagringsform är mycket väsentlig, men detta kan inte närmare tas upp här.

Utmatning av text med stor typuppsättning. Även om man i fråga om inkodningen av en text är hjälpt av en enkel och säker metod att överföra informationen, så är kraven på överskådlighet och lättlästhet i utdata än större. Bland annat vänder ju sig resultatutskrifterna normalt till en större publik än de personer som är direkt knutna till forskningsprojektet. Vi skall därför till

slut se på våra möjligheter till en korrekt återgivning av «svår» text vid olika typer av utmatning.

Dagens typografiskt mest avancerade metod är den datorstyrda fotosättningen. Resultatet av en sådan sättning kan antingen bli till en tryckt bok eller till mikrofilm. Man har tillgång till hela den typografiska arsenalen. Typuppsättningen är mycket stor — man kan vid en och samma körning ha tillgång till ett tusental olika tecken som dessutom kan varieras till storlek och lutning. Helt nya tecken kan på tämligen enkelt sätt infogas i systemet. Sättmaskiner av detta slag är bland andra DIGISET, LINOTRON, PHOTON och SEACO. (Jfr Geir Berges rapport i Nr. 1. 1974 av Humanistiske Data.) Detta tillvägagångssätt kräver en hel del programmeringsarbete och är, bland annat med hänsyn till kostnaderna, mest aktuellt för en slutlig resultatredovisning, knappast för arbetsmaterial.

COM — Computer Output Microfilm — är en teknik som på allvar har börjat konkurrera med det vanliga utskriftsförfarandet vad gäller stora datamängder. Av denna utrustning finns två principiellt skilda typer. Den ena arbetar med ungefär samma teckenrepertoar som en vanlig radskrivare, eventuellt med tillägg av kursiv och fet stil, och saknar därför intresse om en stor uppsättning speciella tecken fordras. Den andra typen registrerar på film den information som genereras på en inbyggd snabb grafisk bildskärm. Teckengenereringen tillgår då på samma sätt som vid den nyss nämnda inkodningen via grafiska skärmar.

dvs genom uppritning av små, raka vektorer. Exempelvis BENSON Cathographe arbetar enligt denna princip. Från COM-utrustningen kan man få både mikrofilm i rullar och s.k. microfiche. Läsapparater finns nu till förhållandevis lågt pris — också sådana med möjlighet till kopiering direkt på papper.

De grafiska bildskärmarna kan naturligtvis även användas för interaktiv inspektion av texter. En förutsättning är då, som ovan, att det finns en snabb vektor/teckengenerator som styrs av en intern minidator om man önskar tecken utanför standarduppsättningen. Särskilda kopieringsenheter kan också anslutas till flertalet bildskärmar.

En mycket intressant utveckling har ägt rum i fråga om snabba elektrostatiske radskrivare/kurvritare (printer/plotter). Arbetsprincipen är att elektriskt ladda upp papperet i ett tätt raster av punkter och därefter låta det passera genom en färglösning där laddade partier blir svärtade. Upplösningen kan vara så god som 8 punkter per mm, vilket ger ett punktraster som knappast är urskiljbart med blotta ögat. En bokstavstyp av normal skrivmaskinsstorlek kan då ha en vidd av 16-20 punkter. Dessa slag av skrivare finns i en mängd olika utföranden, dels konstruerade som radskrivare, då med inbyggd teckengenerator på 64 - 128 olika tecken, dels som kurvritare (plotter) där man är helt fri att mata ut vilka grafiska mönster som helst och dels som kombinerad radskrivare och kurvritare. På den senare varianten kan alltså t.ex. diakritiska tecken ritas på skrivtecken utan att användaren själv behöver svara för

genereringen av varje punkt i varje skrivtecken på raden. Är det däremot fråga om utskrift av icke-latinska tecken eller någon speciell stil eller grad av vårt normala alfabet behövs en kurvritande skrivare. Denna typ av utrustning är inte dyrare än konventionella radskrivare. VERSATEC Matrix, VARIAN Statos och GOULD är välkända namn i detta sammanhang. Liknande skrivare finns för registrering på

värmekänsligt papper. Bläckstråleskrivare med programmerbar teckengenerering torde också bli ett intressant alternativ inom de närmaste åren. Naturligtvis gäller det här liksom tidigare att en del standardprogram behöver skapas för att utrustningen skall kunna utnyttjas effektivt för de speciella lingvistiska tillämpningarna. Det är emellertid ett engångsarbete.

Det torde av denna korta översikt ha framgått att det finns mycket goda möjligheter att underlätta det språkvetenskapliga arbetet med skriftligt textmaterial. Man kan slippa tungarbetad kodning av en mängd tecken i texten och i stället arbeta med den i typografiskt korrekt form. Detta är redan nu en realitet.

Samkatalog for håndskriftmateriale

I juni 1974 ble det avgitt en Innstilling om samkatalog for håndskriftmateriale fra det utvalget som Riksbibliotekstjenesten oppnevnte i juni 1972.

Bakgrunnen for opprettelsen av utvalget var at «Riksbibliotekstjenesten ønsker å få vurdert muligheten for en standardisert katalogisering av arkivalia av privat proveniens i norske samlinger og om MARC-formatet som et internasjonalt standardformat for bibliografisk beskrivelse kunne være et egnet grunnlag for en slik standardisering.»

Følgende mandat ble lagt til grunn for utvalgets arbeid:

1. vurdere muligheten for og ønskeligheten av en samkatalog for manuskripter og brev som provenienstmessig er privatarkivalier i norske samlinger,
2. foreslå retningslinjer for en eventuell koordinert katalogisering og indeksering av dette materiale,

3. vurdere om en slik samkatalog kan baseres på EDB-program utviklet i NORMARC-prosjektet eller andre tekniske løsninger

og

4. foreslå organisering og finansiering av en slik samkatalog for håndskriftmateriale.

Utvalget har funnet det ønskelig å få gjennomført et prøveprosjekt med maskinleselig katalog før det tas stilling til et endelig forslag om utformningen av en norsk samkatalog for privatarkivalia. Målet for et slikt prøveprosjekt må være å overføre til maskinleselig form katalogeksempler utvalget har mottatt gjennom sin spørreundersøkelse, og et representativt utvalg av katalogopplysninger fra en større samling. Følgende sider ved et eventuelt fremtidig driftsprosjekt vil avklares: økonomi, data-innsamling og -registrering, katalogproduksjon og spredning av informasjon til brukere av systemet.

Utvalget foreslår at prøveprosjektet utføres av NAVF's EDB-senter for humanistisk forskning, og med representanter for utvalget i en styringsgruppe for prosjektet.

Riksbiblioteket har i overensstemmelse med utvalgets innstilling henvendt seg til NAVF med ønske om samarbeid med NAVF's EDB-senter. Saken er nå til behandling i Rådet for humanistisk forskning og i NAVF's EDB-senter. En venter at NAVF kan gi sin uttalelse om det foreslåtte prøveprosjektet i løpet av høsten 1974. Innstillingen om samkatalog for håndskriftmateriale foreligger nå trykt i serien Skrifter fra Riksbibliotekstjenesten (nr. 3 1974). Interesserte vises til denne for ytterligere informasjon.

3rd international summer school for computational and mathematical linguistics, Pisa 1974.

I participated in the Summer School from the 19th August to the 6th September, that is to say in everything except the first week's courses which were of an introductory nature and thus optional.

I give in the following an overview over those courses I attended.

1. B. HALLPARTEE — INTRODUCTION

1. *B. Hall Partee - Introduction to Logic II.*

Among the topics covered: Predicate calculus — formal systems and models — syntax and semantics for such models — Tarskian truth conditions.

The course was a useful introduction to some of those aspects of formal logic more relevant to the study of linguistics. Of particular value to me was an insight into the problems of quantification. The course was well prepared and the lecturer thoroughly competent.

2. *M. Kay — Automatic morphological, syntactic and semantic analysis.*

This course covered for the most part the first two aspects of its title. Its declared aim was to impart the philosophy of approach to these problems exemplified by the General Syntactic Processor of R. Kaplan. The key to

this approach lies in the recognition of the non-deterministic nature of the systems required to perform such analyses. The parallel to other fields with which I am well acquainted (multitasking within an operation system) was for me particularly interesting.

The seminar gave one a chance to discuss some more practical details of implementation on computer systems.

The presentation was, as always with professor Kay, clear and deceptively simple.

3. *S. Petrick — Introduction to LISP.*

This course, while not without value, fell to some extent between two stools; rather too concise for linguists with no programming experience, somewhat superficial and slow moving for those of us with previous computational experience.

4. *Ch. Fillmore — Semantics.*

Among topics covered: Basic concepts — syntax — semantics and pragmatics — semantic representation contrasted with a theory of comprehension — semantic fields — the projection problem — text semantics — scope — a classification of semantic notions — valence — lexical semantics — interpretive and gene-

rative semantics contrasted — case grammar.

Professor Fillmore is a very well known and respected linguist who has done much in the field of semantics, so any criticism of his work or lectures from me seems a little pretentious. However, the presentation of systematic ideas by continued exemplification seems to have certain weaknesses. It tends to take a long time to present fairly simple ideas and in so doing it obscures to some extent just the systematisation it is trying to present. It also requires an extremely high standard of presentation if it is to succeed wholly.

5. *B. Hall Partee — Logic and semantics.*

Among the topics were: Methods of linguist and logician compared and contrasted — the work of Richard Montague — his approach to connection between syntax and semantics — relationship between Montague and transformational grammars.

As with her first course, professor Partee's presentation of this rather specialized topic was extremely workmanlike. The formalism of Montague grammars is well worth a closer study.

6. *E. Bach — Semantics in generative grammar.*

Topics covered included: Background — standard theory («Aspects», Katz) — generative semantics — extended standard theory («interpretivist») — critique and comparison between generative school and extended standard theory — A number of new mathematical results about the relation between TG's and recursively enumerable sets - the «almost» equivalence of the two approaches (generative and interpretive) and the consequent need for «weaker» systems for the description of natural languages.

This course promised well and indeed contained a number of very interesting ideas and results. As with professor Fillmore, however, I feel that the rather discursive style of presentation tends to obscure the clarity of the underlying ideas. One is again left with the feeling that the same notions could with advantage have been presented more concisely.

7. *T. Winograd — Current topics in computational semantics.*

Topics covered: A perspective — history — the system SHRDLU (the «blocks world») — representation of knowledge — new directions.

Professor Winograd gave an altogether excellent course. The presentation of the

«blocks world» with its attempt to treat a very simple universe of discourse in depth was extremely interesting as was Winograd's self-criticism and comments about the lessons learned from the project. It illustrated clearly how closely the fields of linguistics and artificial intelligence are related.

8. *J. Lyons — Spatio — Temporal expressions, causality, mood and modality.*

The title covers the topics of this course fairly adequately. Professor Lyons interesting classification system here presented showed the same tendencies as Fillmore and Bach, here in a rather more extreme form.

9. *W. Woods — Advanced problems in syntax and semantics for intelligent machines.*

Topics covered: Basic and Augmented Transition Network grammar formalism — their relation to finite state and Turing machines and thus to transformational grammar — Wood's ATN parser program — special additional mechanisms — semantic interpretation in the LUNAR system — description of the speech recognition project at Bolt, Baranek and Newmann in which professor Woods is currently participating.

This course was undoubtedly the best course of the school from my point of view. Of both practical and theoretical interest and expertly presented, professor Woods managed to convey a great deal of information in a very short time. I can see a real pos-

sibility of the practical application of the techniques presented to our local problems of text processing.

There were also a number of single lectures. The standard here was not very high. I would like, however, to mention David Hays two lectures (plus three discussion sessions) on cognitive structures. The ideas presented here, bordering on psycholinguistics, are as yet not very well developed but are undoubtedly interesting.

I attended altogether 108 hours of lectures and seminars during a period of three weeks.

From my point of view the summer school has been very valuable. It has enabled me to consolidate my rather haphazard knowledge of linguistics somewhat and given me a better understanding of the methods involved.

The school was more oriented towards linguistics as such than the title would suggest and the computational background required was minimal. Courses of a similar nature can thus be recommended to a wide range of humanists with potential rather than actual computational interests.

The courses of Woods and Kay will be of direct practical value to me in the programming of morphological and syntactic analysis programs.

The entire proceedings of the school were recorded on video tape and are to be published in the course of 1975. The reader is referred to this publication for further details.

Datamaskinell eller tradisjonell korrekturlesning?

| | | |
|--------------------------------------------------------------------------------------------------------------|--------------|---------------|
| 112 αὐτοῦσε ἐπιούσο καὶ ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 112 αὐτοῦσε ἐπιούσο καὶ ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 113 καὶσε ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 113 καὶσε ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 114 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 114 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 115 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 115 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 117 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 117 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2C 2E11 | IIFFFFFANSE 5 |
| 119 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 119 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 120 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 120 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 127 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 127 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2C 2E11 | |
| 128 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 128 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 133 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 133 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 133 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 133 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2C 2E11 | |
| 134 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 134 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2K 2E11 | |
| 136 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 136 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2C 2E11 | |
| 137 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I 137 ἐπιούσε ἑρῶσε ἐπιούσε ἑρῶσε τε καὶ ἑρῶσε I | FABE 2C 2E11 | IIFFFFFANSE 7 |

Bildet viser del av en side av en maskinutlisting etter datamaskinell sammenligning av to utgaver av boken *De mundo*. Understrekingene viser hvor linjene avviker fra hverandre og ytterst til høyre er det opplysninger om avvikene.

Når man lager en datamaskinleselig tekstutgave, representerer korrektoren en stor og viktig del av arbeidet. At det er nødvendig med korrekturlesning er innlysende for alle, men at omfanget av korrekturarbeidet overgår selve punchearbeidet er ikke umiddelbart like innlysende. Det er imidlertid min erfaring, etter, i samarbeid med prof. P. Borgen, å ha laget en

datamaskinleselig utgave av Philo av Alexandrias samlede verker. Tilsvarende erfaringer uttrykker A. Q. Morton, som har laget flere datamaskinleselige tekstutgaver, «The real work in producing a perfect text of the Homeric corpus is not in the typing out of the text, as beginners are inclined to think, but in proofreading the text.» 1) Ut fra erfaringene med den tradisjonelle

form for korrekturlesning falt det naturlig å prøve andre former for korrektur. Jeg har derfor i forbindelse med punching av noen bøker fra hellenistisk tid forsøkt datamaskinell korrekturlesning. Denne korrekturmetoden er ikke ny, men den har ingen tradisjon i det norske humanistiske EDB-miljø og metoden har fått overveiende kritisk omtale fra personer med erfaring i datamaskinell tekstbehandling. 2) Metoden går i korthet ut på at dataene punches to ganger og at man lar en datamaskin foreta en sammenligning med angivelse av alle ulikheter mellom dem. Ut fra originalen kan man så konstatere hvilken skrivemåte som er den rette.

Den umiddelbare reaksjon på en slik framgangsmåte er at puncheutgiftene, som er store nok ved én gangs punching, vil bli fordoblet og dermed altfor store til at framgangsmåten kan ha praktisk verdi. B. Munk Olsen, som har laget en datamaskinleselig utgave til bøker av J. Saresberiensis, argumenterer på denne måten. 3) Ut fra mine erfaringer sto det klart for meg at denne argumentasjonen ikke kunne være avgjørende for en sammenligning av tradisjonell korrekturlesning kontra datamaskinell korrekturlesning. Jeg har derfor foretatt en konkret sammenligning av de to korreksjonsmetodene for å finne fram til den som er mest effektiv og mest økonomisk i praksis. 4)

For å sammenligne de to utgaver av en tekst som er punchet to ganger har jeg skrevet et program som fungerer på følgende

måte. (kfr. bildet). Første linje i hver bok (som svarer til en rekord) sammenlignes med hverandre i datamaskinen. Hvis de er helt like, skrives det ingen meldinger og sammenligningen fortsetter parvis med de påfølgende linjer. Når to linjer ikke er like, blir begge linjer med de behørlige referanser skrevet ut, den ene over den andre, og de ulikheter som finnes, blir markert i utskriften ved hjelp av understrekning. Hvis det kun er en eller to feil på linjen, skrives det ut en melding om dette, slik at man ikke behøver å lete på linjen etter flere feil enn dem som er angitt. Hvis den ene linjen mangler ett ord i forhold til den andre, eller det er hoppet over en linje, skrives det ut en melding om dette. Med disse opplysningene er det meget raskt gjort å finne fram til ulikhetene og feilene.

Den datamaskinelle korrekturlesningsmetoden bygger på den forutsetning at det er små sjanser for at to som puncher gjør samme feil på samme sted i teksten. Hvor ofte dette vil skje kan beregnes statistisk. Hvis de to som puncher hver har 0,5% feilanslag, og disse er tilfeldig fordelt, kan antallet av felles feil beregnes til $(\frac{0,5}{100})^2 = 0,0025\%$ eller 0,025 promille. Det vil si at i en bok med ca. 100000 anslag (ca. 50-60 sider) kan det være 2 til 3 feil som ikke blir funnet fordi de ikke framtrer som tekstmessige ulikheter.

En tilsvarende bok med 0,5% feilanslag inneholder 500 feil som det etter den tradisjonelle metode blir korrekturleserens oppgave å finne. Ved en sammenligning av de to metodene er det naturlig å stille følgende spørsmål: 1) Hvor mange av disse

henimot 500 feil klarer en korrekturleser å finne innen bestemte tidsrammer. 2) Er det rimelig å anta at feilanslagene fordeler seg tilfeldig?

La oss ta det siste spørsmålet først. Det viste seg, som man kunne vente, at feilanslagene ikke var helt tilfeldig fordelt. De viste tvertimot en tydelig tendens til å samle seg i de lengre ordene, og dermed kan antallet felles feil bli flere. Dette forhold er oversatt av B. Munk Olsen. 5) De lengre ordene (mer enn fem anslag, aksenter o.l. inkludert) utgjør imidlertid størsteparten av det totale antall anslag. Hvis vi på grunnlag av stikkprøver, antar at de utgjør 75% og hvis vi for enkelthets skyld regner med at alle feilene finnes innen de lengre ordene, så blir feilprosenten 0,66 og regnestykket blir følgende $(\frac{0,66}{100})^2 = 0,0044\%$ eller 0,044 promille. Det vil si at fire-fem feilanslag ikke blir oppdaget av totalt 100000 anslag. Det er en begrensning for anvendelsen av denne metode at den baserer seg på en forholdsvis tilfeldig fordeling av feilanslagene. Hvis puncheforskriftene blir svært kompliserte og punchingen følgelig ikke blir direkte avskrift, er det muligheter for en spesielt skjev fordeling. Ved punchingen av Tertullians verker gjaldt 70% av punchefeilene endringer som skulle foretas i forhold til teksten. 6) Uansett hvilken metode man skal anvende ved korrigeringen svarer det seg å ha enkle puncheforskrifter hvis man vil unngå mange feil, og eventuelle endringer i teksten er det ofte lettere å la datamaskinen forta.

R. Wisbey er avvisende overfor metoden fordi de som puncher er tilbøyelige til å gjøre

«certain natural errors» som ikke blir oppdaget. Den feiltype det her er tale om er sannsynligvis forvekslinger av bestemte (nabo)taster for han sier videre at med to punchemaskiner med forskjellig arrangerte tastaturer vil faren for slike feil bli sterkt redusert. 7) En del stikkprøver i de feil som er begått har imidlertid ikke gitt grunnlag for å anta at forvekslinger av bestemte (nabo)taster er en spesielt hyppig feilårsak.

Det første spørsmålet om hvor mange feil en korrekturleser klarer å finne innen bestemte tidsrammer er det litt komplisert å svare på, fordi det er avhengig av korrekturleserens generelle dyktighet og dernest av variasjoner i hans prestasjoner. Både fysiske og psykiske faktorer influerer på en korrekturleser. I dette tilfelle var det særlig grunn til å anta at kjennskap til at man «konkurrerte» med en datamaskin ville føre til en skjerpet konsentrasjon og dermed influere på prestasjonen. I forbindelse med retting av en av bøkene ble korrekturleseren informert om dette eksperimentet. Resultatet ble at han fant hele 95% av de feilene som ble funnet ved den maskinelle sammenligning allerede i første korrektur-omgang. Det er et ekstremt godt resultat og den sannsynligste forklaring synes å være kjennskapet til at man konkurrerte med og ble kontrollert av datamaskinen. For at denne sammenligningen skulle være av verdi var det derfor nødvendig å anvende ordinære korrekturresultater.

To bøker som var korrekturlest på ordinær måte viste at henholdsvis 70% og 77% av de feilene som ble funnet ved den datamaskinelle sammenligning ble funnet

etter én korrekturromgang. I dette eksperimentet utgjorde korrekturtiden ca. 40% av punchetiden. Det vil si at man kunne lese 21/2 gang korrektur på den tid det tok å punche teksten for annen gang.

En sammenligning av resultatene fra den ene av bøkene som var korrekturlest tre ganger med resultatene fra den datamaskinelle korrekturlesing viste at korrekturleseren ikke hadde funnet så mange feil som datamaskinen. Sammenligningen med den andre boken som var korrekturlest fire ganger ga samme resultat. Resultatet var faktisk bedre for den boken som var korrekturlest 3 ganger, idet bare 3% av de feilene som var blitt oppdaget av maskinen, ikke var blitt oppdaget av korrekturleseren. Dette illustrerer at vanlig korrekturlesing kan ha en varierende effekt og det viser at selv ved tre til fire korrekturlesninger er det vanskelig å finne de «siste» feilene. Ikke i noen av disse to tilfellene hadde korrekturleseren funnet feil som skyldtes felles feilpunching og som derfor ikke kunne oppdages av maskinen. En slik feil ble funnet i den førstnevnte boken. Den teoretiske beregning av antallet mulige feil som «unnslipper» ved denne metoden motsies altså ikke av disse resultatene.

Samme person gjentok stundom samme feil i samme ord. Muligheten for at den som puncher har en tendens til å gjøre feil som er karakteristisk for vedkommende gjør det nødvendig med to forskjellige personer til å punche.

På bakgrunn av disse resultater mener jeg at den beste metoden m.h.t. effektivitet er den datamaskinelle korrekturlesing. Forut-

satt at man ønsker å få en tilnærmet 100% riktig datamaskinleselig tekstutgave, og det er vel det vanligste, er den også økonomisk vurdert den beste metoden. Det kan kanskje være tilfeller hvor man ved punchingen av data har et formål som ikke nødvendigvis er en nesten feilfri tekst, og i slike tilfeller svarer det seg selvsagt best med én gangs punching og ingen korrektur, eventuelt én gangs korrektur, ettersom et flertall av de feil som finnes ved vanlig korrekturlesing blir funnet første gang. Den beregning som vanligvis foretas og som resulterer i at man velger vanlig korrekturlesing, forutsetter at det er tilstrekkelig med to gangers korrekturlesing for å få en henimot feilfri tekst. Det er midlertid på dette punktet man forregner seg, som dette eksperimentet klart viser.

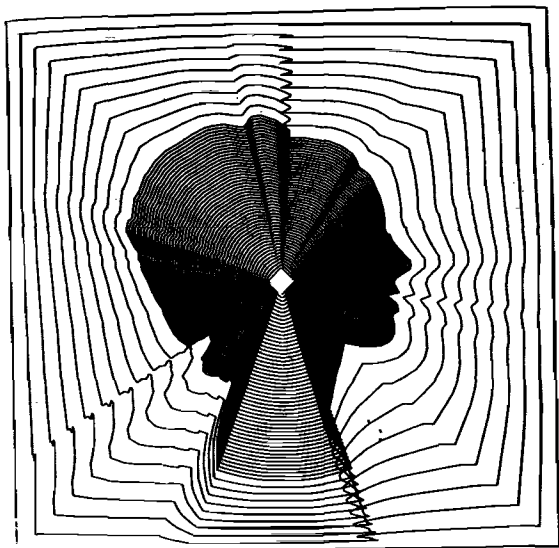
Hvis man ikke kan akseptere de feilene som blir oppdaget ved den beskrevne datamaskinelle sammenligning, så kan man, på samme måte som man ellers ville øke antallet korrekturromganger, bare øke antallet puncheomganger. Metoden byr også på andre klare fordeler. En fordel er at selve oppdateringen av teksten delvis kan skje på en måte som ikke medfører nye punchefeil under rettingspunchingen. Hvis den ene linjen, eller deler av den, er riktig skrevet i motsetning til den andre, så kan man slippe rettingspunching ved at man bruker den riktig punchete linje (rekord) i den nye (rettede) utgave av teksten.

En stor fordel er den innsparing som finner sted ved at korrekturoperasjonene kun skjer én gang mot ellers f.eks. tre til fire ganger. Man sparer her først og fremst

administrasjon og mye tid for den som utfører disse rettingoperasjonene, enten man nå har en programmerer til å gjøre det eller man må gjøre det selv.

Datamaskinell korrekturlesing kan derfor anbefales til dem som vil ha tilrettelagt data i maskinleselig form med færrest mulig feil i.

- 1) A.Q. Morton and A.D. Winspear, «It's greek to the computer», 1970, s. 29.
- 2) B. Munk Olsen, *Anvendelsen af elektronisk databehandling ved løsningen af filologiske oppgaver: konkordanser, indices verborum*. Københavns Universitet, Romansk institut 1968, s. 74-75 og Roy Wisbey, *Concordance Making by Electronic Computer: Some Experiences with the «Wiener Genesis»*. *Modern Language Review*. Vol. LVII nr. 2 (1962) s. 165 og s. 170.
- 3) Kfr. B. Munk Olsen. Op.cit. s. 75.
- 4) NAVF's EDB-senter v/adm. leder J.H. Hauge har støttet dette forsøket med bl.a. tilsagn om dekning av ekstra puncheutgifter.
- 5) B. Munk Olsen, Op.cit. s. 75.
- 6) Joseph K. Siberz og Joseph G. Devine. *Literary Data Processing Conference Proceedings*, Yorktown Heights. Utg. Jess B. Bessinger, Jr., Stephen M. Parrish, Harry Arader. IBM Corporation, Armonk, New York 1964, s. 139.
- 7) R. Wisbey, Op.cit. s. 16.



IPT – Improved Programming Technology

**80 karakterer x 24 linjer, overførings-
hastighet opp til 9600 baud, karakter-
mode**

**som ovenfor, men med blokk-mode,
printer, interface, videoc utgang, cursor
kontroll etc.**

*Dataspråk: Den anglo-norske språkkode man
braker når ingen andre enn dem som allerede
forstår det, skal forstå det.*

Software som inkluderer: RTX, Disc, Magnetic Tape and Cassette Operating System. BASIC. FORTRAN. Utility and Library programs. File Manager. 360 Cross Assembler.

Alle moduler leveres med det samme interfacesnitt – Facit SPI for parallell dataoverføring som gjør dem enkle å tilpasse enhver minicomputer. Interface finnes til flere minicomputere på markedet.

Datasystemet for det rasjonelle firma. Markedets mest moderne minicomputer-baserte multijobboperativsystem for administrative oppgaver. Arbeider on-line (fra dataskjermer) og i sann-tid, d.v.s. man arbeider i takt med firmaet. Leveres som komplett "turn-key"-system, komplett med hardware og software tilpasset Deres firma, installert og igangsatt med fullstendig vedlikeholdsavtale.

Systemet er ingen "stålboks", men kan vokse i takt med firmaet på såvel hardware som software siden.

LPS-11 periferalsystemet for laboratoriebruk. Inkluderer:

en 12-bits A/D omformer med 8 kanals multiplexer og lysdiodtablå (light emitting diode) for 6 sifre.

en forforsterker for fire innganger.

en programmerbar realtidsklokke og dobbel «Schmitt trigger».

og en 16-bits «buffered» digital input/output.

En DECwriter for inntasting/utskrift med 30 tegn i sekundet.

Laborativeversjonen av RT-11 Realtidsoperativsystemet som betjener BASIC og FORTRAN.

Samt fire ukers opplæring.

EDB OG PRIMÆRKILDER

I overensstemmelse med vedtak i NAVF's styre 12.9. 1973 har Rådet for humanistisk forskning og Rådet for samfunnsvitenskapelig forskning nedsatt et utvalg for å utrede bruken av datamaskinleselige media som kildespredningsform.

Opprettelsen av dette utvalget har sammenheng med de vurderinger som kommer til uttrykk i Innstilling fra NAVF's Primærkildekomite (avgitt november 1972).

Primærkildekomiteens innstilling er for tiden til uttalelse ved universitetene og oppbevaringsinstitusjonene.

Utvalgets mandat:

- I. Beskrive og vurdere hvilke tekniske løsninger som kan komme på tale når det gjelder avbildning av forskjellige kategorier forskningsmateriale.
- II. Utrede muligheter for systemløsninger som vil gi størst mengde samkjørbare data på tverrfaglig basis innenfor de humanistiske og samfunnsvitenskapelige fag, idet en også tar hensyn til behovet for samkjøring av nytt og gammelt forskningsmateriale.
- III. Vurdere kostnader ved valg av ulike tekniske løsninger på kort og lang sikt. På kort sikt ved å komme fram til kostnadene ved å punche materialet. På lengre sikt ved å vurdere kostnadene ved bruk av maskinleselige overføringsformer sammenlignet med de fotografiske teknikker.

IV. Tjene som sakkyndig organ for dem som skal utarbeide detaljerte planer for kildeoverføring.

Utvalgets sammensetning:

Adm.leder Jostein Helland Hauge,
NAVF's EDB-senter for humanistisk forskning, *formann*.

Overbibliotekar John Brandrud,
UB, UiB.

Amanuensis Jon Eivind Kolberg,
Institutt for samfunnsfag, Univ. i Tromsø.

Vit.ass. Jan Oldervoll,
Historisk institutt, UiB.

Forskningsleder Per Torsvik,
NAVF's sekretariat for mediaforskning, UiB

Universitetslektor Joan Veim,
Institutt for informasjonsvitenskap, UiB.

Førstesekretær Dag Tveit Bråthen,
NAVF, er administrativ sekretær.

Utvalget hadde sitt konstituerende møte 23. september i Bergen.

DE NORDISKE ARKIVDAGER — UPPSALA 3. - 4. AUGUST 1974.

Av Ivar Fønnes.

De nordiske arkivdager — som holdes hvert 3. år — var denne gang lagt til Uppsala i tiden 3.-4. august. Hovedoverskriften var: *Arkivene i dataalderen*.

Den største del av tiden var avsatt til temaet *Databaser*. Her ble det redegjort for arkivprosjekter som anvender EDB i de forskjellige nordiske land. Opplegget var beregnet på arkivarer som ikke hadde egen EDB-erfaring, og diskusjonen ble derfor ikke rettet mot konkrete EDB-problemer. Det kan likevel være verdt å trekke fram noen viktige synspunkter som det syntes være bred enighet om:

- a) Arkivdata som skal inn på maskinleselige media, bør i størst mulig utstrekning overføres i *originalform*.
- b) Arkivarer som skal delta i EDB-baserte prosjekter, bør skaffe seg kunnskaper i EDB.
- c) Arkivarene må komme med på et tidlig stadium ved planlegging og opplegg av EDB-systemer for arkiver.

Et annet tema var *Datamakulering*, hvor det ble redegjort for regler som er utarbeidet i Sverige.

Det tredje tema gjaldt *Varigheten av databærere i EDB-systemer*. Her ble det redegjort for en systematisk svensk undersøkelse av magnetbånd som data-lagringsmedium. Det framgikk av undersøkelsen at man ikke kan lagre magnetbånd så svært lenge før antall feil i data blir et problem.

Arkivdagene ble avsluttet med omvisning i det nye Riksarkivet i Stockholm.

Prøveprosjekt om registrering

Gamle bilder er Kulturhistorisk kildemateriale

Norsk Kulturråd henvendte seg i begynnelsen av 1974 til NAVF med søknad om bistand til å gjennomføre et prøveprosjekt med EDB-registrering av fotografier. Bakgrunnen for henvendelsen var at en lenge har vært klar over at gamle fotografier ofte representerer et verdifullt kulturhistorisk materiale og at innsatsen når det gjelder innsamling og registrering av dette materialet er et forsømt område i vårt land.

Til å utrede aktuelle spørsmål i forbindelse med registrering og bevaring av gamle fotografier nedsatte Norsk Kulturråd i 1972 et ad-hoc utvalg (Fotoutvalget). Dette utvalget har under sitt arbeid med registreringsspørsmålene funnet det ønskelig å få klarlagt om en i det fremtidige innsamlings- og registreringsarbeid bør legge opp virksomheten på en slik måte at databehandling kan tas i bruk ved utnyttelsen av bildematerialet.

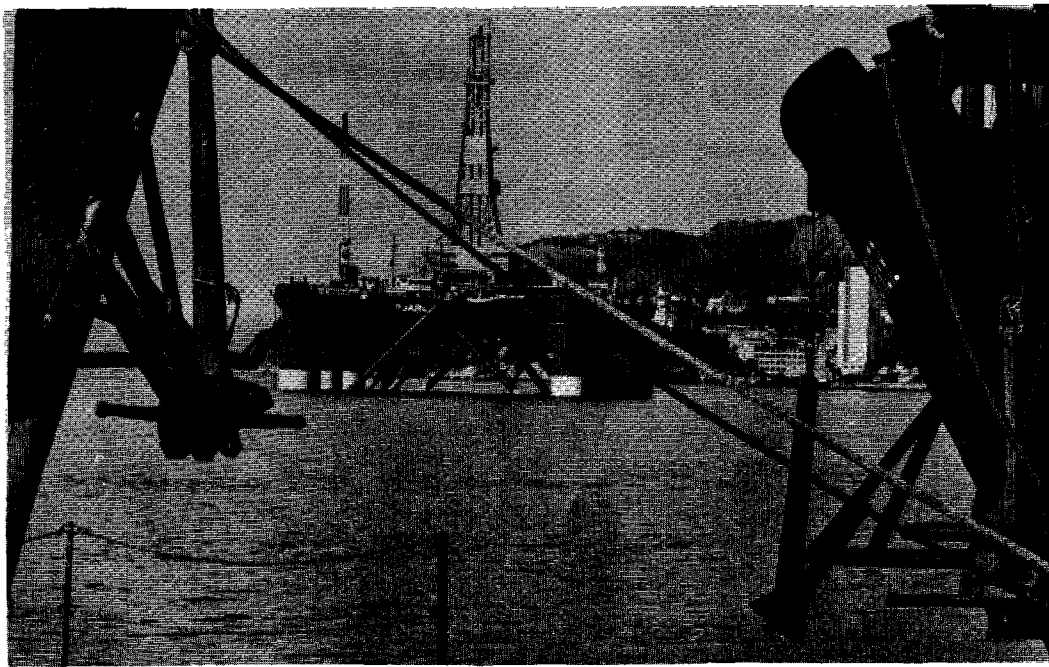
Henvendelsen fra Norsk Kulturråd gikk derfor ut på å søke samarbeid med NAVF



Kollektivtransport (hesteomnibus) i Bergen i 1890-årene. (fra Billedsamlingen, Univ.bibl., Univ. i Bergen).

av fotografisk materiale

— nye bilder blir det



«Oljeboring» på Bergen havn.
(foto: Birkhaug & Omdal).

gjennom EDB-senteret for humanistisk forskning for å kartlegge mulighetene og metodikken for EDB-registrering av fotografier. For å kunne utprøve mulighetene i praksis ble det foreslått å organisere et prøvedataarkiv for et representativt utvalg fotografier.

Etter tilråding fra styret for EDB-senteret godkjente NAVF i vår et samarbeid mellom Fotoutvalget og EDB-senteret i Bergen. Samarbeidet er kommet i gang i høst, og det er nedsatt en styringsgruppe som foruten representanter fra senteret består av arkivar Sofie Rogstad, Privatarkivkommisjonen og vit.ass. Jan Oldervoll, Historisk institutt, Univ. i Bergen.

Det viktigste arbeidet i første fase vil være å komme frem til et hensiktsmessig registreringsskjema som både kan tillate en fleksibel søking etter de «konvensjonelle» arkivopplysninger og som dessuten gir adgang til seleksjon av bilder med utgangspunkt i motivopplysninger om bildene.

Kurs i datamaskinell språkbehandling, København juli/august 1974

Kolbjørn Heggstad avsluttet sin artikkel i forrige nr. av Humanistiske Data med å opplyse at årets kurs (det tredje i rekka bevilget av Nordiske forskerkurser) holdes i København.

Kurset ble avviklet i tida 29. juli til 10. august, og kurssted var Schæffergården.

I alt var vi 30 deltakere: 14 fra Danmark, én fra Finland, to fra Island, 6 fra Norge og 7 fra Sverige. Forelesere var prof. Martin Kay og dr. Richard Rubinstein, begge fra University of California, og prof. Henning Spang-Hanssen fra Københavns Universitet. Universitetslektor Bente Mægaard, også fra Københavns Universitet, var kursleder.

Kurset vekslet mellom forelesninger, gruppearbeid og bruk av datamaskinen på RECKU (Det regionale EDB-center ved Københavns Universitet), en UNIVAC 1110. Dagen var hovedsakelig delt i to økter, regulert av måltidene, og disse øktene var igjen delt i to, regulert av kaffepauser. Perioden før lunch var viet teoriforelesninger, mellom lunch og middag var det gruppearbeid, eventuelt bruk av datamaskinen på RECKU. De som av en eller annen grunn ikke ønsket å bevege seg dit, kunne benytte en terminal installert på Schæffergården som var tilknyttet datamaskinen på RECKU. Det ble av og til kødannelser ved den terminalen.

Allerede første dagen ble opplegget for kurset diskutert ved at hver av deltakerne

fortalte om hvilke prosjekter de var i gang med, og ut fra dette ble 6 arbeidsgrupper dannet:

- morfologisk analyse
- syntaktisk analyse
- grammatikktesting
- semantisk analyse
- ortografi
- stilistikk

I og med at omtrent halvparten av tida var avsatt til gruppearbeid, gav kurset rik anledning til eksperimentering og prøving av arbeidsmetoder innenfor disse feltene. Men to uker er likevel kort tid til å utarbeide og utteste ferdige programmer, slik at de fleste gruppene baserte seg på å arbeide med strategier for hvordan problemene burde angripes og å grovskissere hvordan prosedyren kan se ut. Så kan hver og en i sitt lønnkammer arbeide videre og prøve å få det hele til å fungere. Kurset skulle i alle fall ha gitt et godt grunnlag å arbeide videre på.

Den teoretiske delen av kurset var viet tre forskjellige emner:

- morfologisk og syntaktisk analyse (som i bruk av metode er praktisk talt like)
- programmeringsteknikk
- statistikk

Den morfologiske og syntaktiske analysen var basert på et kompendium utarbeidet av Martin Kay. Kompendiet er en metodologisk

innføring i mulige framgangsmåter for å la datamaskinen gjøre en slik analyse. Framstillingen, både i kompendiet og i forelesningen, bærer preg av at den er et *forslag* til løsning. Å gi et fyldestgjørende resymé av den foreslåtte metodikken her, ville bli altfor plasskrevende.

I den delen som tok for seg programmeringsteknikk, delte forsamlingen seg i to grupper: én for de som har skrevet minst ett program — og fått det til å virke («advanced group»), og én for alle de andre. Programmeringsspråket som ble brukt, var SIMULA, et språk utviklet ved Norsk regnesentral i Oslo. Undervisningen i nybegynnergruppa ble supplert med en kort oversikt over NU-ALGOL (en versjon av ALGOL utviklet ved Regnesentret ved Universitetet i Trondheim) som SIMULA bygger på og er en videreutvikling av. Etterhvert som kurset skred fram, økte nybegynnergruppas antall på bekostning av den avanserte.

Nytt for årets kurs var en kort innføring i kvantitativ lingvistikk/språkstatistikk gitt av prof. Henning Spang-Hanssen. Det er begrenset hva som kan sies om dette emnet på to forelesninger, men i og med at han knyttet metodikken til konkrete eksempler, ble det likevel utbytterikt. Vi fikk i det minste et begrep om hva en ønsker å måle med denne typen statistikk.

Disse to forelesningene i språkstatistikk må ha gitt mersmak fordi flere av kursdeltaker-

ne ymtet frampå om at de godt kunne tenke seg et sommerkurs som bare tok for seg den kvantitative sida av datamaskinell språkbehandling. Idéen er i alle fall med dette publisert.

I og med at vi både bodde, spiste og hadde undervisning på Schæffergården, fikk vi god anledning til uformelle samvær, både faglig og sosialt. Fordi gruppearbeidet tok en god del av tida på kurset, var det viktig at en kunne få kontakt med de andre gjennom disse samværene. I et fag som befinner seg på utviklingsstadiet, er det av stor betydning at diskusjoner og samtaler kan foregå i mindre, uformelle grupper. En har, spesielt i dette emnet, behov for å høre hva likesinnede holder på med.

Den siste dagen var avsatt til oppsummering av gruppenes arbeid og til drøfting av den datamaskinelle lingvistikks framtid i Norden. Det er nemlig klart at årets kurs var det siste som Nordiske forskerkurs vil gi støtte til, i alle fall blir det ikke aktuelt med støtte derfra de første åra. De har gjort sitt til å sørge for at et nordisk samarbeid innen dette fagområdet kan komme i gang. Og nettopp fordi disse forskerkursene har vist seg å dekke et behov, er det synd om det videre samarbeidet skal strande på manglende økonomiske midler. Den kontakten og det miljøet som er etablert gjennom disse tre åra, må ikke gå tapt. Det må fortsatt kunne arrangeres mindre kurs og konferanser på nordisk basis innen dette emnet.

For at det skal kunne skje, ble det dannet en

nordisk samarbeidskomité for datamaskinell språkbehandling. Hovedoppgaven blir å sørge for kontinuiteten i dette arbeidet. (Se omtale av samarbeidsgruppa annetsteds i dette nummeret).

Hvis fruktene av komitéens arbeid blir slik det tegnet til på avslutningsmøtet, burde datamaskinell språkbehandling på nordisk basis ha en framtid. Og da vil heller ikke denne kursserien ha vært forgjeves.

EDB-ANVENDELSER I ARKEOLOGI

Konferanse ved University of Birmingham 4. og 5. januar 1975.

Konferansen er den tredje i serien «Computer Applications in Archaeology» som University of Birmingham arrangerer. Hensikten med konferansen er å bringe sammen arkeologiske forskere og dataspesialister for å drøfte den siste utvikling innen EDB-orientert arkeologi og diskutere fremtidig virksomhet.

Av emnene på konferansen i januar 1975 kan nevnes:

Statistiske metoder og modeller (inkl. «Cluster Analysis»), informasjonsbehandling, grafiske teknikker og registrering av funndata.

Interesserte kan få nærmere opplysninger ved å skrive til:

Miss S. Laflin, Computer Centre, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, U.K.

NORDISKA SAMARBETSGRUPPEN FÖR SPRÄKLIG DATABEHANDLING.

Syfte och sammansättning.

Vid den allmänna slutdiskussionen den 9 augusti 1974 under sommarskolan i språklig databehandling i Köpenhamn tillsattes en kommitté med namnet Nordiska samlarbetsgruppen för språklig databehandling (Nordic Committee for Computational Linguistics). Dess syfte skulle vara att verka för fortsatt aktivitet på området, i första hand genom anordnande av sammankomster av olika slag (t.ex. symposier och sommarskolor) och främjande av informationsflödet. Till ledamöter utsågs de tre som lett sommarskolorna 1972-1974, nämligen Sture Allén, Göteborg, Kolbjörn Heggstad, Bergen, och Bente Maegaard, Köpenhamn. Gruppen fick fria händer att till sig ansluta ytterligare ledamöter. Särskilt nämndes vikten av att Finland och Island blev representerade.

På samarbetsgruppens vägnar
Sture Allén

Redaksjonell merknad:

Det er senere meldt at gruppen har supplert seg med Baldur Jónsson fra Island.

Association for literary and linguistic computing

Association for
Literary and
Linguistic
Computing

BULLETIN

Summer Term 1974
Vol. 2 No. 2

Association for Literary and Linguistic Computing, oftast omtala som ALLC, blei oppretta i 1973. Ideen blei først lagt fram på eit symposium i Cambridge i 1970, der ein del engelske språk- og litteraturforskarar hadde komme saman for å drøfte spørsmål av felles interesse i samband med datamaskinell behandling av tekster. Nokre få utanlandske forskarar hadde blitt invitert, m.a. eit par frå Noreg. I 1972 blei eit tilsvarende symposium halde, denne gongen i Edinburgh og med noko større internasjonal deltaking.

Karakteristisk for desse samlingane var at mange av dei reint praktiske vanskane som humanistane hadde når dei skulle gå i gang med maskinell tekstbehandling, blei drøfta. Korleis skulle ein t.d. punche tekstene, eller var dei kanskje puncha tidlegare i samband med andre prosjekt? Burde ein lære seg å programmere sjølv, og i så fall kva for eit programmeringsspråk var det mest velegna? Slike og mange tilsvarende spørsmål er svært aktuelle for humanistar som kanskje er blitt overtyste om at datamaskinelle metodar har mykje å gi, men som enno er usikre på korleis dei skal komme i gang. Den tilslutning desse første arrangementa fekk, viste at eit organisert framhald av tiltaket var ønskjeleg, og på dette grunnlaget blei ALLC oppretta som internasjonal organisasjon.

I første punktet i vedtektene heiter det: «Its

principal purpose is to provide a means of communication for all those concerned with literary and linguistic research by computer». Ein såg det slik at i datamaskinell tekstbehandling var det svært mange problem som var felles både for litteraturforskarar og lingvistar, derfor skulle det vere plass for begge i organisasjonen. Vidare er det grunn til å understreke at ALLC også vil gi eit tilbod mellom dei store internasjonale arrangementa. Kongressar er det mange av i vår tid og dei har utvilsamt sin verdi, men dei dekker berre ein del av det kommunikasjons-behovet som moderne forskning har.

ALLC satsar på å bygge ut si verksemd i åra som kjem. Førebils har dei ei kongressrekke som har arrangement annakvart år. (Sjå melding i dette nummeret om årets arrangement i Cardiff.) Ein del seminar om spesielle emne er tillyste. Det er oppretta arbeidsgrupper innan fleire emne, m.a. om opplegg av datamaskinelle tekstarkiv, metodar i arbeidet med dramatiske tekster, leksikografi og velegna programmeringsspråk. ALLC gir 3 gonger i året ut sitt eige meldingsblad Bulletin som inneheld rapportar, organisasjonsnytt, lister over aktuelle arrangement, bokmeldingar o.s.v.

ALLC har vakse fram av eit engelsk initiativ og det har derfor til no vore naturleg med ein viss engelsk dominans i sekretariat og styre. For å skape ei meir internasjonal jamvekt har ein no bygd ut eit system med

nasjonale tillitsmenn som skal vere ansvarlege for organisasjonen i vedkommande land. Dette både for at ALLC's arbeid skal bli kjend, og for at aktivitetar i landet skal bli rapporterte internasjonalt. Tillitsmennene er samstundes «korresponderande redaktørar» i Bulletin, og skal koordinere innsendt materiale.

Ut frå si målsetjing valde ALLC å basere arbeidet på eit aktivt sekretariat og at interesserte skulle teikne medlemskap i organisasjonen.

I punkt 2 i vedtektene heiter det: «Graduates of any University, and persons of the standing of graduates, shall be eligible for Membership on application to the Honorary Secretary, subject to the approval of the Committee». Det er også mogeleg for andre spesielt interesserte å søkje om medlemskap. Bibliotek og andre institusjonar kan teikne kollektivt medlemskap. (For tida kostar individuelt medlemskap £ 3 og kollektivt £ 7.50.)

Medlemskap gir fritt tilsendt Bulletin, rett til reduserte prisar ved kjøp av publikasjonar og deltaking på arrangement av ALLC.

Individuelle medlemmer har stemmerett og vel pr. post styre for organisasjonen.

Underskrivne har tatt på seg å representere ALLC i Noreg, og vil gjerne ha kontakt med fagfolk som vil vite meir om ALLC. Eg vonar med denne presentasjonen å ha fått fram litt av det som skil denne organisasjonen frå andre som arbeider delvis innan same område. Det er naturleg å

nemne International Association for Applied Linguistics (AILA) som i denne samanhengen er opptatt med datamaskinelle metodar ved språkinnlæring og framstilling av leksikografiske hjelpemiddel. Ved The International Conferences on Computational Linguistics er ein for det meste opptatt av det desse metodane har tilført generell språkteori og grensar i sitt arbeidsområde nær opp til simulert intelligens. Etter mi meining utfyller desse organisasjonane kvarandre, og ALLC har sitt viktige ansvarsområde. ALLC er først og fremst organisasjonen for alle dei språk- og litteraturforskarar som arbeider med tekster og som ønskjer å ta datamaskinene i bruk i sitt arbeid, men som på grunn av den store arbeidsinnsats og dei store kostnader det ofte fører med seg, vil ha ei planlagt verksemd der ein byggjer på tilgjengeleg røynsle.

OBS!

INTERNASJONAL EDB-KONFERANSE

Idet bladet går i trykken, kommer det melding om at "The Second International Conference on Computers and the Humanities" vil finne sted ved University of Southern California, Los Angeles, i tiden 3. - 6. april 1975. Det er nå anledning til å sende inn forslag om foredrag innen språk, litteratur, musikk, historie og arkeologi. Nærmere opplysninger om arrangementet fås ved henvendelse til NAVF's EDB-senter for humanistisk forskning.

MEDARBEIDERE I DETTE NUMMERET:

STURE ALLÉN, professor i språklig databehandling, Avdelningen för språklig databehandling, Göteborgs Universitet.

JON BING, amanuensis ved Institutt for privatrett, Avd. for EDB-spørsmål, Universitetet i Oslo.

IVAR FONNES, konsulent i Oslo for NAVF's EDB-senter for humanistisk forskning.

ROLF GAVARE, driftsleder ved Avdelningen för språklig databehandling, Göteborgs Universitet.

G. MICHAEL GILLOW, sjefskonsulent ved NAVF's EDB-senter for humanistisk forskning.

KOLBJØRN HEGGSTAD, førsteamanuensis ved Prosjekt for datamaskinell språkbehandling, Nordisk institutt, Universitetet i Bergen.

GEIR KJETSAA, professor ved Slavisk-baltisk institutt, Universitetet i Oslo.

EIRIK LIEN, konsulent i Trondheim for NAVF's EDB-senter for humanistisk forskning.

ROALD SKARSTEN, universitetsstipendiat ved Religionsvitenskapelig institutt, Universitetet i Bergen.

IDAR STEGANE, vitenskapelig assistent ved Avd. for nordisk litteratur, Nordisk institutt, Universitetet i Bergen.

Meldinger om senterets virksomhet våren og høsten 1974

1. Organisasjonsforhold.

Som omtalt annetsteds i dette nummeret, ble det i juni oppnevnt nytt styre for senteret. Gjennom det nye styret er senteret blitt nærmere knyttet til NAVF's Fagråd A.

I løpet av meldingsperioden har senteret fått to nye medarbeidere. Cand.mag. Bjørn Eide er engasjert i 1/2-stilling som operatør. Eide vil også ta del i senterets prosjektassistanse. Fra 1. juli har senteret hatt i arbeid en konsulent i Trondheim. I stillingen ble cand.philol. Eirik Lien engasjert (se egen omtale av konsulent-tjenesten i Trondheim).

2. Planleggingsarbeid.

Da senteret startet sitt arbeid i 1972, var det naturlig å forsøke å få et overblikk over arbeidsfeltet de kommende år gjennom konsulentbistand og prosjektsamarbeid på en rekke ulike fagområder.

Det ble imidlertid snart klart at det området som senteret skulle virke innenfor — de humanistiske fag — fremviser meget store og differensierte brukerbehov. Det ville f.eks. være arbeidsoppgaver nok innenfor fagene språkvitenskap eller arkeologi alene.

For å skaffe seg et oversyn over EDB-behovene besluttet derfor NAVF's EDB-komité for humanistisk forskning vinteren 1974 at senteret skulle innhente og bearbeide opplysninger om EDB-situasjonen i de humanistiske fag i dag og gi en oversikt over de mest sentrale

EDB-behov i fagmiljøene. Dette arbeidet ble avsluttet i mars 1974.

Behovsoversynet bekreftet at det ligger store EDB-oppgaver og venter i så godt som alle humanistiske fag. Sentrale EDB-behov synes å være knyttet til veiledning, instruksjon og kursvirksomhet. Et stort antall av de humanistiske forskere i vårt land tilkjenner behov for å få en faglig orientert innføring i hvordan EDB kan komme til anvendelse på bestemte fagområder. Behovet for en prosjektrettet konsulenthjelp blir sterkt understreket. Av anvendelsesområder synes bruk av EDB i arkivarbeid å være ett av de sentrale. Likeledes er synet på verdien av datamaskinell behandling av tekster utvetydig i uttalelsene fra de humanistiske forskerne.

Det behovsoversynet som ble lagt frem, ble av senterets styre våren 1974 brukt som grunnlagsmateriale ved utarbeidelsen av en langtidsplan for senteret for perioden 1974-1977. Forslaget til langtidsplan ble oversendt NAVF til behandling i mai og vedtatt av Fagråd A i juni.

I langtidsplanen blir det lagt vekt på at senterets mest sentrale innsatsområder skal være kurs- og informasjonsvirksomhet, prosjektassistanse og generell programutvikling. Planen gir også en rettesnor for den interne fordeling av innsatsen på de ulike målområdene. Innenfor programutvikling blir senteret i første omgang gitt i oppgave å legge til rette et brukerorientert tekstbehandlingssystem ved

Universitetet i Bergen og i Trondheim. Når det gjelder prosjektassistanse, peker planen særlig på bistand til fagmiljøer som ønsker å få vurdert verdien av å ta i bruk datamaskin ved behandling av systematisert forskningsmateriale. At vitenskapelig bearbeidet arkivmateriale har en sentral plass i mange humanistiske fag viste også det tidligere omtalte behovsoversynet med all tydelighet.

3. Prosjektsamarbeid.

En omtale av senterets viktigste samarbeidstiltak vinteren 1974 er gitt i forrige nummer. Enkelte av disse er ført videre i denne perioden, mens andre er avsluttet.

Av samarbeidstiltak som er avsluttet, kan nevnes et prøveprosjekt som senteret har utført med databehandling av et steinaldermateriale for Historisk museum, Universitetet i Bergen. En videreføring av dette arbeidet vurderes nå av museet. Senteret har i denne perioden avsluttet de databehandlingsoppgaver for hovedfagsstudenter ved Tysk og Russisk institutt, Universitetet i Bergen, som ble omtalt i forrige nummer.

Ytterligere bistand er i perioden gitt bl.a. til forskere ved Romansk institutt, som arbeider med mellomfranske tekster og til et forskningsprosjekt innen stilistikk ved Nordisk institutt. Samarbeidet med Klassisk institutt om databehandling av papyrusmateriale er ført videre med sikte på å avslutte utarbeidelsen av konkordanser til materialet ved slutten

av året. Middelaldersamlingen, Historisk museum, Universitetet i Bergen (Bryggenundersøkelsene i Bergen) har også i denne perioden fått adgang til senterets utstyr og har mottatt konsulenthjelp i forbindelse med etablering av en database for de arkeologiske funndata. De av museets sivilarbeidere som står for EDB-driften, har fått arbeidsplass i senteret.

Senterets konsulent i Oslo har gitt konsulentassistanse til enkeltforskere og gruppeprosjekter. Som oversikten over igangværende arbeid i forrige nummer viste, er det ved flere institutter ved Universitetet i Oslo aktive EDB-miljøer. I perioden er støtte gitt bl.a. til Historisk institutt, Slavisk-baltisk institutt, Britisk institutt og Talemålsundersøkelsen i Oslo. En nærmere omtale av konsulentens deltakelse i EDB-arbeidet ved instituttene er denne gang innarbeidet i meldingen fra EDB-tjenesten ved Det historisk-filosofiske fakultet ved Universitetet i Oslo.

Sommeren 1974 ble NAVF's konsulenttjeneste i Trondheim etablert. Den tilsatte konsulenten, Eirik Lien, har i løpet av høstsemesteret orientert seg i de humanistiske fagmiljøene i Trondheim. Gjennom kontakt med representanter for instituttene og enkeltforskere har han derved lagt et grunnlag for prosjektassistanse og generelt utviklingsarbeid i fremtiden.

Av større samarbeidstiltak som er etablert siden siste rapport, kan nevnes et prøveprosjekt i Bergen om databehandling av opplysninger om eldre fotografisk materiale. Samarbeidspartner er Norsk Kulturråd (konferer

separat melding om tiltaket). Resultatene av dette prøveprosjektet, som planlegges avsluttet i løpet av våren 1975, vil bli rapportert i et senere nummer av meldingsbladet.

I løpet av sommeren og høsten deltok senteret i NAVF's behandling av trykningssøknad for Norsk Landbruksordbok. Fagråd A innvilget i høst søknad om tilrettelegging for trykking via filmsats av landbruksordboken. I dette arbeidet, som vil foregå i Oslo, får NAVF's EDB-senter ansvaret for det EDB-tekniske opplegget. Senterets konsulent i Oslo vil få faste konsulentoppdrag i tilknytning til prosjektet, hvor det for øvrig etter planen blir engasjert en programmerer i 1/2-stilling over en 2-års periode. Utarbeidelsen av et magnetbånd for filmsetting vil også fremskaffe for forskningsformål et verdifullt landbruksvitenskapelig ordmateriale — i første rekke til bruk ved Norsk leksikografisk institutt.

4. Kurs- og seminarvirksomhet.

I april inviterte senteret til et seminar i Bergen om databehandling av klassiske tekster. Foreleser var dr. Stephen Waite, Dartmouth College, Hannover, U.S.A. Dr. Waite fikk også tid til separate drøftinger med medarbeidere innen EDB-prosjekter ved Universitetet i Bergen.

Seminarvirksomheten i Oslo har i perioden vært drevet i samarbeid med Forskningsgruppen for databehandling i humaniora. I vårsemesteret ledet lektor Tor Hultman, Malmø et seminar om databehandling av språket hos svenske gymnasiaster. Professor Svein Nordbotten, Universitetet i Bergen ga i mai en presentasjon av faget informasjonsvitenskap.

I høst arrangerer senteret et kurs i NU-ALGOL i Bergen i samarbeid med Avdeling for elektronisk databehandling. Som et supplement til innføringskurset — som blir gitt i forelesningsform — gir senteret gruppeøvinger i programmering med bruk av senterets terminalutstyr. De fleste deltakerne på kurset (ca. 15) arbeider i vitenskapelige stillinger i ulike humanistiske fag ved Universitet i Bergen.

For å få ny kunnskap og nye impulser å bringe videre har senterets personale også benyttet seg av de kurs og seminar som tilbys på feltet i utlandet. En henviser her til sjefskonsulent Gillows rapport om sommerskolen i data-maskinell og matematisk lingvistik i Pisa og konsulent Liens orientering om det som foregikk på den 3. nordiske sommerskolen i data-maskinell språkbehandling i København. En kan dessuten nevne at konsulent Berge i juni deltok på det nordiske forskerkurset «Data-logi med inrikning på små databaser» i Uppsala og at tekstbehandlingssystemet TEXT ble presentert i et foredrag av konsulent Fønnes på den internasjonale konferanse om innholdsanalyse i Pisa i september.

5. Programmeringsarbeid.

En stor del av aktivitetene innenfor programmering har naturlig nok direkte tilknytning til det aktuelle prosjektsamarbeidet som senterets personale deltar i. I enkelte tilfeller kan dette samarbeidet gjøre nødvendig at en inkorporerer nye funksjoner i allerede tilgjengelig programutrustning. Ikke sjelden betyr imidlertid samarbeidet at en må utvikle

nye programmer for det foreliggende problemområdet eller nye varianter av programmer. Det er på den annen side også klart at det undertiden bygges opp mer generelle programmer av slike som i første omgang har sammenheng med spesifikke prosjekter.

I tillegg til det programmeringsarbeidet som er motivert av bestemte forskningsoppgaver, foregår det også i senteret utviklings- og evalueringsarbeid innenfor programmering av mer grunnleggende karakter. Senterets personale i Bergen har i perioden videreutviklet et kommunikasjonssystem mellom senterets maskin PDP 11/20 og UNIVAC 1110, Universitetet i Bergen. Systemet simulerer en terminal-konsentrator slik at man over en enkelt telefonlinje kan knytte en rekke terminaler til UNIVAC 1110 via forbindelse med PDP 11/20. Systemet gjør bruk av PDP's operativsystem, noe som gir en

meget fleksibel terminaldrift. Flere terminaler kan utføre separate oppgaver parallelt, samtidig som terminalbrukeren kan nytte lokalmaskinens funksjoner (lagring, utskrift etc.).

Av produksjonsorientert utviklingsarbeid kan nevnes at en har videreutviklet programmene for innlesing av ulike typer papirbånd på PDP 11/20.

Med basis i et program for KWIC-indeksjer av gammel-engelsk tekstmateriale har en gått videre for å gi programmet større fleksibilitet, med bl.a. det formål at programmet senere kan inngå i et standard tekstbehandlingssystem.

Som ledd i arbeidet med et tekstsistem for UNIVAC-anlegg er det også drevet forsøksarbeid for å finne frem til effektive metoder for lagring av større mengder tekstdata. I

forbindelse med den fremtidige utvikling av tekst-programmer er også flere programmeringsspråk (APL, SIMULA o.fl.) blitt evaluert.

Senterets konsulent i Oslo har i denne perioden særlig videreutviklet programmer i tilknytning til programsystemet HISO og TEXT (se Humanistiske Data nr. 1/74) for å inkorporere funksjoner som pågående analysearbeid gjør nødvendig. Av slikt arbeid kan nevnes videreutvikling av sorteringsprogrammer i HISO og planlegging av et program for analyse av data for familierestruksjon. I tilknytning til tekstbehandlingssystemet TEXT er det bl.a. under utarbeidelse programmer for koding av tekstmateriale på grunnlag av kriterier i teksten.

Det er planen å gi en ny og ajourført oversikt over programsituasjonen for de humanistiske fag i et av de første numrene av Humanistiske Data.

EDB-tjenesten ved det historisk-filosofiske fakultet, Universitetet i Oslo

Oversikt over virksomheten 1973-74.

EDB-tjenesten ble opprettet ved vedtak i fakultetsrådet av 2. mai 1973. Tjenesten ble bygget opp rundt den konsulentstilling som NAVF hadde opprettet fra 1. mars s.å., og et EDB-utvalg ble oppnevnt bl.a. for å fungere som styre for EDB-tjenesten.

EDB-tjenesten har sett det som sin oppgave å videreføre og videreutvikle den virksomhet som allerede var i gang på feltet

EDB ved fakultetet. Virksomheten foregår i nær forbindelse med NAVF's EDB-senter for humanistisk forskning, i og med at konsulenten er administrativt underlagt dette senteret.

Følgende oppgaver har stått sentralt i EDB-tjenestens første arbeidsår:

1. Vedlikehold og videreutvikling av generelle programsystemer.

Programsystemet TEXT (for tekstana-

lyser) og HISO (for rubriserte protokolldata) samt en samling analyseprogrammer for voteringsdata er i stadig bruk blant forskere og studenter ved fakultetet, TEXT ved en rekke institutter, de to siste ved Historisk institutt.

Det er EDB-tjenestens oppgave å vedlikeholde programsystemene, gi konsulentbistand og veiledning slik at de kan anvendes av brukerne, og videreutvikle

programmer og systemer i samsvar med brukernes behov. Denne virksomheten er av avgjørende betydning for bruk av EDB i forskning ved fakultetet.

2. Undervisning og seminarer.

Kurs i bruk av eksisterende programmer og systemer er nødvendig for en tilfredstillende løsning av oppgavene i pkt. 1 ovenfor. I perioden har det vært holdt brukerkurs både i TEXT og HISO.

Videre har EDB-tjenesten arrangert kurs i programmeringsspråket LISP og et kurs i anvendelse av LISP i syntaktisk analyse. Høsten 1974 holdes et kurs i programmeringsspråket COBOL.

Kursene har dels vært holdt av EDB-konsulenten og dels av hjelpelærere som er trukket inn fra andre steder (bl.a. EDB-senteret, Univ. i Oslo).

Konsulenten har også stått sentralt i arrangement av tverrfaglige seminarer som knytter seg til EDB-metoder i humaniora. Seminarene har vært holdt i regi av Forskningsgruppen for databehandling i humaniora og NAVF's EDB-senter for humanistisk forskning.

3. Konsulentarbeid i tilknytning til prosjekter/programmering.

Konsulentarbeid og veiledning har særlig vært gitt i forbindelse med bruk av de nevnte programsystemer (pkt. 1) og programpakken DDPP (statistikk). Men også i andre sammenhenger gis det bistand, først og fremst i forbindelse med metodiske spørsmål og statistikk. Dessuten har det vært utført

spesialprogrammering for flere prosjekter. Av prosjekter som har fått bistand, kan følgende nevnes:

Talemålsundersøkelsen i Oslo (Nordisk inst.), Ullensakerprosjektet, Voteringsanalyser, Innholdsanalyse av stortingsdebatter (Historisk inst.), stilistiske studier av russisk poesi (flere forskere ved Slavisk-baltisk inst.), gammelengelsk språk, moderne engelsk avisspråk (Britisk inst.), tekster av Homer og Hesiod (Kunsthistorie), dataregistrering av musikk (Musikkvitenskap). Flere kunne nevnes, og noen nye er på planleggingsstadiet.

DET NORDISKE ØDEGÅRDSPROSJEKTET NORSK AVDELING

For Det nordiske ødegårdsprosjektet, norsk avdeling er det ved Regnesentret, Universitetet i Trondheim utført en undersøkelse av lønnsomheten av å systematisere data fra dette historiske forskningsprosjektet ved hjelp av EDB. Arbeidet ble utført under ledelse av forsker Eirik Lien, nå konsulent for NAVF's EDB-senter i Trondheim.

Om tiltaket er det utarbeidet en rapport (RUNIT-rapport A74009) som viser hvordan programpakken SPSS kan brukes til å analysere dataene fra et undersøkelsesområde (Stjørdal) og hvordan de kodes. Dataene er hovedsakelig numeriske (årstall, antall brukere på forskjellige tidspunkt, størrelsen

på åker, eng, avling, takst o.l., gårdens avstander til sjø, nabo, kirke, hovedvei), men også ikke-numeriske (gårdens navn og eiere, jordbunnstype o.l.).

Interesserte kan ved henvendelse til Regnesentret, Universitetet i Trondheim få tilsendt den omfalte rapport om prøveprosjektet.

Redaksjonen avsluttet 20. nov.

HUMANISTISKE DATA blir utgitt av NAVF's EDB-senter for humanistisk forskning i Bergen.

Senterets administrative leder, Jostein H. Hauge, har det redaksjonelle ansvar for meldingsbladet.

De som ønsker å få bladet tilsendt, kan bestille det ved henvendelse til senterets adresse:
Villavei 10,
Boks 53,
5014 Bergen-Universitetet.

Innlegg kan sendes til samme adresse.

Ad. emblemet på forsiden.

På omslaget har vi denne gang tatt i bruk NAVF's nye emblem som ble introdusert i forbindelse med Rådets 25-års jubileum.

I emblemet, som er laget av tegneren Alf Næsheim, symboliserer stjernen makrokosmos mens cellen med kromosomstavene symboliserer mikrokosmos.