

humanistiske data

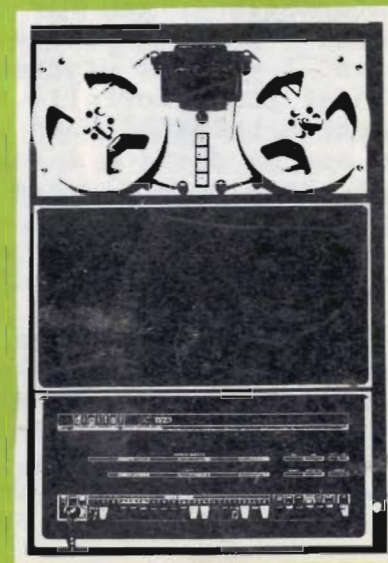
Meldingsblad for
NAVFs EDB-senter
for humanistisk forskning

Norges Almenvitenskapelige Forskningsråd

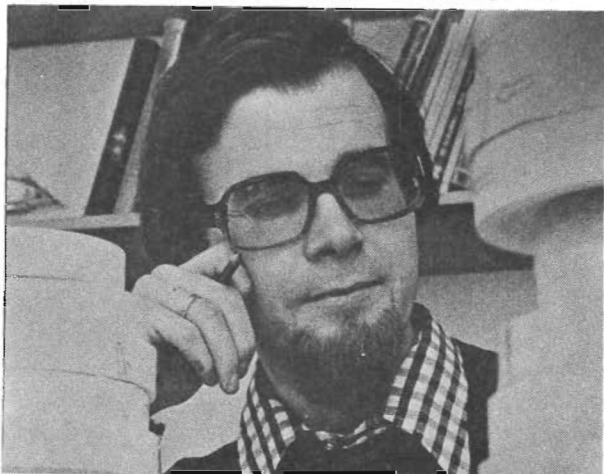


AV INNHOLDET NEVNES:

- *EDB-PROSJEKTER
- *MASKINPROGRAM
- *DATA
- *RAPPORTER
- *MELDINGER



Redaktørens spalte



HUMANISTISKE DATA kommer nå ut med sitt første regulære nummer. En må ha lov til å si at mottakelsen i fagmiljøene av det første nummeret var positiv, og dette gir oss tro på at bladet kan ha en oppgave som informasjonsorgan for EDB-interesserte i de humanistiske fag i vårt land. I en viss utstrekning vil også bladet kunne fortelle fagfolk fra andre arbeidsfelt ved universitetene om noe av det som i dag pågår av EDB-aktiviteter i våre fag.

Innholdet i dette nummeret kan kanskje gjøre det nødvendig å knytte noen bemerkninger til betegnelsen 'meldingsblad'. Som en vil se av stoffet, er bladet ikke bare tenkt som et blad som gir meldinger om arbeidet i NAVF's EDB-senter for humanistisk forskning. Meldingsbladet vil også gi videre ulike typer tilsendt stoff som kan ha

interesse for dem som bruker data-maskiner. For at en gjennom bladet skal kunne gi videre informasjon, må senteret på sin side stå i et nært forhold til de aktive fagfolkene ved universitetene og få innlegg og artikler fra dem til publisering i HUMANISTISKE DATA.

Det er planen at hvert nummer skal ha en 'gjeste-spalte', hvor en presenterer en artikkel som retter søkelyset på et vel avgrenset emne innen det store feltet som humanistisk databehandling er, og som lanserer synspunkter til refleksjon og debatt. I dette nummeret innledes denne serien med et bidrag av Arne B. Johansen: 'EDB som faktor i kunnskapsproduksjonen'. De øvrige bidragsyterne denne gang har også tilknytning til Universitetet i Bergen, men i neste nummer vil en sikre seg en større geografisk spredning også på dette området.

Det har dessverre ikke vært mulig å holde den oppsatte fristen for utgivelsen av det første nummeret av HUMANISTISKE DATA i 1974, men på den annen side er bladet denne gang blitt større enn opprinnelig planlagt. For ikke å ta munnen for full en gang til, vil jeg om neste nummer bare si at det vil komme ut i 2. halvdel av 1974, og så tidlig i høstsemesteret som mulig.

INNHold:

side:

	Redaktørens spalte
1	Gjestespalte
4	EDB-virksomheten i de humanistiske fag ved våre universiteter
14	Oversikt over maskinprogram i Norge til bruk i humanistisk forskning
18	Data i maskinleselig form
25	Nordiske forskarar på kurs i datamaskinell språkbehandling i Bergen
28	Datamaskinen i historieforskninga
29	Datamaskinen i arkeologisk forskning
30	Optisk lesing og datamaskinstyrt fotosetting
33	Arbeidet i senteret høsten 73 og vinteren 74
	Meldinger

EDB som faktor i kunnskapsproduksjonen

Arne B. Johansen

HVA ER DATA?

Det er vanskelig å gi en almengyldig definisjon av hva data er. Grunnen til dette er i første rekke at databegrepet skifter innhold i takt med de ulike oppfatninger (modeller) forskeren har av dataenes opphavssituasjon.

Sett at en arkeolog mener en gitt mengde mennesker produserte en fast mengde fysiske kulturspor av én bestemt type pr. tidsenhet de oppholdt seg i et område. Da blir det viktig å få et mål på mengden av disse kultursporene, enten han nå vil bruke dem til å vurdere folkemengden eller lengden på bosetningsperioden. En annen forsker som mener at produksjon av kulturspor i stor utstrekning varierte uavhengig av folkemengden, vil ikke bruke mengden av kulturspor som data i en slik analyse.

Mens én filolog mener at den skrevne tekst inneholder alle nødvendige data om språk, krever en annen at deler av språkbrukerens sosiale situasjon også må 'følge med' teksten i form av en sosiolingvistisk analyse. En tredje vil i tillegg ha med tonefall og talehastighet.

sin tur kan igjen den 'sosiale situasjon' defineres på et utall måter, dvs. at utallige aspekter av den kan fremheves og tillegges betydning for utformingen av språket.

I praksis betyr dette at det er ufruktbart å betrakte data som selvstendige byggesteiner i kunnskapsproduksjonen. De skal heller oppfattes som resultater av at visse modeller har vært anvendt. Oppbygging av nye modeller er nødvendig dersom en skal bli i stand til å finne nye datatyper. Dette medfører at ny kunnskap nødvendigvis må ha sitt opphav i ideer som utformes før dataene defineres og samles inn.

Grunnen til at enkelte mål og snittfigurer på flintøkser fra skandinavisk steinalder har vært data gjennom noen tiår nå, er ikke at de er udiskutable og selvgitte, men at forskerne i dette tidsrommet har forutsatt én bestemt opphavssituasjon for øksene. Dersom en f.eks. hele tiden antar at øksene har vært brukt til hogging, kan det være viktig å måle variasjoner i eggvinkel for alle økser innen et tidsrom. Var øksene derimot rene statussym-

boler, blir det kanskje fargen eller det umiddelbare helhetsinntrykket som oppfattes som data.

ALMENE KONSEKVENSER AV DATAENES USELVSTENDIGHET.

En viktig følge av dette synspunktet er at hverken innsamling eller håndtering av data må isoleres fra utviklingen av modeller (teorier). Det vil si at 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 en bare kan begrunne det med at 'de muligens kan bli av betydning en gang i fremtiden'.

Dersom en bruker den siste begrunnelsen, betyr det i realiteten at en samler stadig flere data som belyser gamle teorier. Dermed konsentreres det stadig større oppmerksomhet om den etablerte kunnskap, slik at det blir gradvis vanskeligere å anlegge synspunkter som avviker for sterkt. Det blir naturlig at hovedfagsstudenter og andre forskere velger seg emner innen den etablerte ramme, noe som gir denne rammen økt berettigelse. Resultatet er at det virker stadig

mer naturlig å konsentrere både de materielle og de menneskelige ressursene i å utforske og utvide datamassen, uten at en spør om årsakene til at den eksisterer.

Denne selvbekreftelses- og stabiliseringsmekanisme gjør at kunnskapsproduksjonen har lett for å stagnere. Den kan nok suppleres og bygges ut med stadig nye og utmerkede avhandlinger, men de er gjerne detaljutbygging og intern omflytting innen den gamle struktur.

Er det så noe f. eks. i norsk vitenskapshistorie som tyder på at denne konserveringsfaren er noe mer enn en papirtiger?

I arkeologi er det løynefallende at de fleste av de store linjene i vår forhistorie ble trukket opp av pionerene i faget - den gang både menneskelige og økonomiske ressurser var en brøkdel av det vi har i dag. Da fikk vi den teknologiske oppdelingen i stein-, bronse- og jernalder, vi fikk den typologiske arbeidsmåte og vårt syn på innovasjonssentra og kulturpåvirkning i Europa.

Siden den gang har vi filt og flikket og funnet mer av samme slag. Vi føler det ekstra berettiget å bruke mye av våre ressurser til å berge et datasett dersom det er 'en fin bronsealdergrav', dersom det er velegnet for typologiske studier

eller det viser 'klar påvirkning fra provinsialromersk område'.

Dette skyldes selvsagt ikke at tidligere arkeologer hadde en ekstra sikker evne til å treffe på akkurat den rette teorien. Det har heller sin årsak i at vi senere (bevisst eller ubevisst) har nyttet deres teorier til å peke ut de data vi skal samle inn.

Denne 'gamle' datamassen omgir oss på alle kanter som en stadig høyere mur, slik at vi mister evne og dristighet til å se på vår omverden gjennom nye teorier. Naturligvis er det ikke uakseptabelt å anvende nye teorier i moderne arkeologi. Men fordi en først må ta stilling til de gamle teoriene, og fordi arbeidet med dem krever så store ressurser, blir nyskaping en tung prosess.

Hvordan står dette problemet f.eks. i filologi? Har det kommet radikalt nye teorier om språk og språkutvikling innen det norske filologiske miljø etter Sophus Bugge, Magnus Olsen og de andre pionerene? Eller har en også der flikket og bygget videre på deres tankebygninger? Har en i stadig økende grad forbrukt ressursene til å samle flere data om stedsnavn, malføre og skrevne tekster? Er det en stadig økende andel av forskerne som velger seg problemer som kan belyses ved hjelp av disse datamassene? Begrunner en (som arkeologene)

ofte sitt arbeid ved ikke-faglige argumenter, som f.eks. at skolen, interessegrupper, lovverk, naturinngrep osv. krever at en fortsetter i det gamle spor? I så fall har en glemt at de oppgaver som pålegges oss fra de 'ytre instanser', er de som våre forgjengere i faget overbeviste dem om nødvendigheten av.

KONSEKVENSER FOR SYNET PÅ EDB.

Etter mitt skjønn bringer EDB-hjelpemidlet dette problemet enda sterkere inn i bildet.

EDB er særlig karakterisert ved at det gjør dataoversikt og datamanipulering mye lettere, forutsatt at datamengden kan struktureres logisk. Den må også være tilstrekkelig stor til at kostnadene ved input, programmering, kjøring og output er mindre enn det ville koste å utføre det samme med konvensjonelle midler. Det finnes selvsagt også en rekke oppgaver som vanskelig eller overhodet ikke kan løses uten EDB-hjelp.

Det at datamasser lettere kan manipuleres, øker fristelsen til å isolere dataene fra de teoriene som fremskaffet dem. Isolasjonen kan bli så fullstendig og teorien så undertrykket at en i det daglige forskningsarbeid betrakter teorier utelukkende som et etterdata fenomen og som et mer eller

mindre nødvendig resultat av utstrakt datamanipulering.

Resultatet kan bli et kostbart arbeid som gir utmerket manipulerings- evne og oversikt over dataarkiver som bare inneholder detaljer og lite viktige utfyllinger til en gammel struktur. Har en tilstrekkelig god isolasjon mellom teori og data, kan dette arbeidet imidlertid meget vel virke meningsfylt og høyst prisverdig.

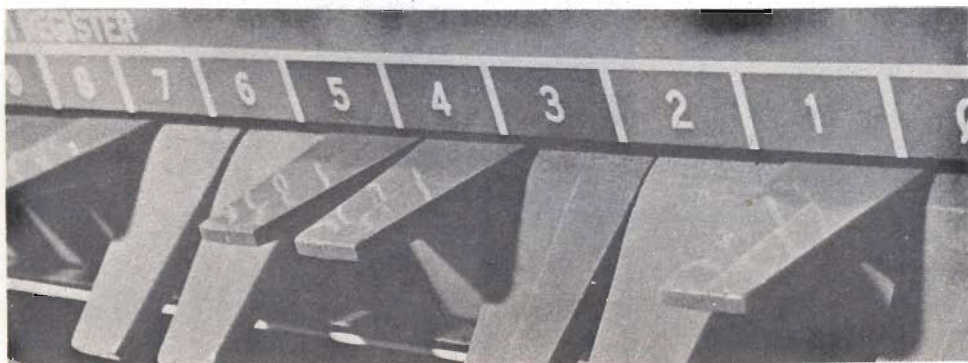
En forsker må ikke glemme at manipulerings- av datamassen ikke kan bringe for dagen informasjon som ikke var 'tenkt inn i' dataene før manipulerings- begynste, dvs. at informasjonen allerede lå gjemt i dataene da de ble samlet inn.

KONKLUSJON:

Bruk av EDB må knyttes nært til behovene hos kunnskapsprodusentene. Dersom EDB isoleres og utvikler arbeidsprosedyrer uten slik kontakt,

betyr det at en lager verktøy i håp om at det skal passe til et eller annet problem. Når brukeren kommer, konstaterer han sannsynligvis at verktøyet er ubrukbart fordi det er beregnet på et problem som er forlatt eller lite aktuelt. Eller han finner seg et problem som passer til verktøyet.

Det må derfor ikke bygges opp noe arkivhold eller noe system for manipulerings- av data, uten at det er svar på et veldefinert forsknings- behov. Dersom vi ikke går varsomt på dette området, risikerer vi å sitte igjen med en forskning der det er hjelpemidlet som bestemmer hvilke problemer vi kan ta opp, i stedet for at problemet alltid burde bestemme hvilke hjelpemidler vi hadde behov for. Faren er kanskje særlig stor innen EDB og annen elektronikk der det er så sterke økonomiske grupper som er interesserte i at deres produkter tas i bruk på flest mulig områder.



NORDISKE FORSKERKURSER 1974

DATAMASKINELL SPRÅKBEHANDLING

I tiden 29. juli - 10. august 1974 vil det ved Københavns Universitet bli arrangert et kurs i datamaskinell språkbehandling.

Kurset vil omfatte såvel teoretisk undervisning og oppgaveløsning som praktiske øvelser på datamaskin. Den teoretiske undervisning vil omfatte grammatikker, setningsanalyse, betydningsanalyse og stilistikk på statistisk grunnlag.

På kurset er det plass til 30 språkforskere som i forveien har et visst kjennskap til bruk av datamaskinelle metoder på lingvistiske arbeidsfelt.

Det felles programmeringsspråket vil være ALGOL (i versjonen UNIVAC NU-algol).

Hovedforelesere blir Martin Kay, USA, Richard Rubinstein UCLA, Henning Spang-Hanssen, Københavns Universitet.

Deltakere vil kunne få dekket sine reise- og oppholdsutgifter.

Søknad om opptagelse på kurset må sendes innen 1. mai til kurslederen: Bente Maegaard, Institutt for anvendt og matematisk lingvistik, Københavns Universitet.

Ved henvendelse til kurslederen vil en kunne få mer detaljerte opplysninger om kurset.

EDB-virksomheten i de humanistiske fag ved våre universiteter

Hensikten med dette oversynet er å gi en første, samlet fremstilling av det EDB-arbeidet som i dag pågår innenfor de forskjellige humanistiske fag ved våre universiteter.

Det som slår en når en samler materiale om denne virksomheten, er at det er i løpet av de siste 10 år at datamaskinen serløst er tatt i bruk i humanistisk forskning hos oss. Etter en forsiktig begynnelse i siste halvdel av 60-årene har anvendelsesområdet ekspandert kraftig i de siste 4 - 5 år slik at EDB i dag er tatt i bruk innen de fleste hovedområder av humanistisk forskning.

Som en kunne vente, finner en at den anvendelse en gjør av databehandling varierer sterkt fra felt til felt og at de ulike prosjekter viser forskjeller med hensyn til i hvilken grad datamaskinens tekniske muligheter utnyttes.

I en rekke av de tiltak som omtales nedenfor, er datamaskinen foreløpig tatt i bruk som et hjelpemiddel til å utføre konvensjonelle arbeidsoppgaver raskere og enklere enn tidligere. På andre felt er det imidlertid utvikling i gang som også setter en i stand til å nytte me-

toder og bruke forskningsstrategier som i det hele tatt ikke kunne tenkes gjennomført uten data-maskinelle redskaper.

En typologisk beskrivelse av det EDB-arbeid som i dag utføres, er et viktig emne, men må likevel utstå til et senere nummer.

De opplysningene om EDB-tiltakene som finnes nedenfor, er hovedsakelig fremkommet som svar på henvendelser til medarbeidere i de enkelte prosjekter, men i tillegg er det ved beskrivelsen nyttet annet tilgjengelig skriftlig materiale. Det har ikke vært ansett som påkrevet at beskrivelsen av de ulike tiltak følger samme mønster eller at alle tiltakene får like mye spalteplass. Spalteplassen skal altså ikke oppfattes som et skjult mål på verdien av det arbeidet som utføres.

De medarbeidere som likevel føler seg stemoderlig behandlet, kan regne med velvilje fra redaktøren dersom de ønsker å gi et mer dekkende bilde av sin virksomhet (jfr. Humanistiske Data nr. 1, s. 3). Skulle det være slik at noen EDB-prosjekter er uteglemt, mottar



redaktøren likeledes gjerne opplysninger om disse til en senere omtale i bladet.

Det kan også opplyses at NAVF's EDB-senter selv vil ta initiativ til å bringe mer omfattende presentasjoner av EDB-arbeidet innen utvalgte prosjekter.

UNIVERSITETET I OSLO

Historisk institutt.

Ved dette instituttet har det i de siste år vært i gang flere ulike typer databehandlingsoppgaver i forbindelse med pågående forskningsarbeider.

Professor Ottar DAHL har siden 1968 ledet forskningsprosjektet 'Innsamling og analyse av voteringsdata i Stortinget'. Som et resultat av arbeidet i prosjektet er data fra Stortingsforhandlinger om voteringer med navneopprop innsamlet og lagt til rette for elektronisk databehandling. Materialet dekker tidsrommet 1814 - 1940. Det er utført analyser på deler av materialet med sikte på å belyse gruppeadferd i Stortinget ved hjelp av kvantitative metoder. En del av de analyser som er foretatt, er utført av hovedfagsstudenter som har arbeidet i tilknytning til prosjektet. Analysearbeidet vil bli ført videre i den kommende tid. Det finnes stensilerte registre over de vote-

ringer som er lagt til grunn for data-innsamlingen.

Professor DAHL har også tatt initiativ til et tiltak kalt 'Innholdsanalyse av Stortingsdebatter fra etterkrigstiden ved hjelp av EDB'. Dette prosjektet er ennå i en tidlig og eksperimenterende fase. Siktepunktet for arbeidet er å prøve ut i hvilken grad EDB kan brukes som hjelpemiddel ved en analyse av innholdsmessige trekk ved debattene (begrepsbruk, temaer, argumenter etc.). Noen utvalgte trontaledebatter er punchet og lagt til rette for maskinell behandling. Analysearbeidet vil bli ført videre i 1974.

Førsteamanuensis Sivert LANGHOLM har i flere år arbeidet med valgundersøkelser fra Christiania i det 19. århundre. Arbeidet foregår som to delundersøkelser. Den ene er en dybdeundersøkelse av Christiania-velgerne i 1868. Ved å sammenstille opplysninger om de ca. 2800 stemmerettskvalifiserte dette året fra en rekke kilder (politiske manntall, skattemanntall, folketellingslister etc.) blir det mulig å analysere data-maskinelt valg materialet med hensyn til en rekke ulike variabler (f.eks. stand, yrke, inntekt, valgdeltakelse i 1868).

Det er også innsamlet data til en langtidsundersøkelse av valgdeltakelsen blant de stemmeberettigede

i tiden 1829 - 1879. De aktuelle personer er identifisert i en rekke forskjellige kilder gjennom et lenge tidsrom, og derved har en skapt grunnlaget for studiet av gjentatt valgdeltakelse hos en og samme velger (eller kategori av velgere) over en rekke valg. Til bruk ved tilretteleggingen av materialet er det bl.a. utviklet et flerdimensjonalt og fleksibelt system for koding av yrkesvariabler.

Et annet EDB-prosjekt ved instituttet er 'Ullensakerprosjektet'. Prosjektet ledes av førsteamanuensis Sivert LANGHOLM og professor Ingrid SEMMINGSEN og har som mål 'å utforske sosiale tilstander og endringsprosesser under tidsrommet for det industrielle gjennombrudd i Norge, samt å legge forholdene til rette for slik forskning. En sentral angrepsmåte er statistisk utnyttelse av massemateriale, på individplanet og ved hjelp av EDB'.

Som datagrunnlag for undersøkelsene har en tilrettelagt folketellinger for Ullensaker fra 1865 og 1875, oversikter over utflyttede fra Ullensaker i disse årene og likeledes emigrasjonslister.

Utvidelse av datagrunnlaget og derved tidsrommet for undersøkelsen foregår kontinuerlig. Overføring av Christiania-tellingen fra 1875 til EDB-tilgjengelig format pågår og vil representere et viktig data-

tilskudd til studiet av norsk samfunnsutvikling i tidsrommet 1860 - 1900.

Til nå har utnyttelsen av det tilrettelagte materialet i hovedsak skjedd i form av studentoppgaver, men det kan nevnes at utvandringsdata også blir utnyttet innenfor rammen av en inter-nordisk undersøkelse.

I sammenheng med disse prosjekter er det utført et omfattende programmeringsarbeid i første rekke av nåværende EDB-konsulent Ivar Fonnes (tidligere vit.ass. ved Historisk institutt) og vit.ass. Elin Mehlum (tidligere programmeringsstipendiat tilknyttet NAVF's EDB-komite for humanistisk forskning).

Norsk leksikografisk institutt.

Ved Norsk leksikografisk institutt har EDB i de siste årene vært tatt i bruk som et hjelpemiddel i ordboksarbeidet. Dosent Dag GUNDERSEN leder instituttets arbeid med registrering av ny språkbruk i norsk og samarbeider med Norsk språkråd. Lignende arbeid er satt i gang i de øvrige nordiske land. I denne registreringen blir alt nytt ordtilfang og nye betydninger og bruksmåter, særlig hentet fra aviser og tidsskrifter, punchet og bearbeidet datamaskinelt.

Ettersom ordtilfanget før punching blir systematisert slik ordboksmateriale vanligvis blir det, ved oppslagsord, tekstblokk, kildehenvisninger etc., og dessuten påført en del grammatiske koder, får en ved hjelp av datamaskinen mulighet til å systematisere og skrive ut materialet på en rekke ulike måter til den etterfølgende leksikografiske eller lingvistiske bruk.

Til i dag er databehandlingen av materialet blitt utført ved Prosjekt for datamaskinell språkbehandling, Nordisk institutt, Universitetet i Bergen (se nedenfor).

Da et slikt materiale ikke lett kan publiseres i trykt form, er den mest umiddelbare nytten av databehandlingen trolig at det kan kopieres som datalister etter hver brukers behov, og raskt stilles til rådighet. Det er å vente at denne formen for publisering etter hvert vil komme til å avløse trykte bøker i mange tilfeller. Norsk leksikografisk institutts nærmeste plan for å utnytte sitt punchede materiale gjelder en håndordbok for bokmål som instituttet skal utarbeide i de nærmeste år.

Norsk stadnamarkiv.

Innen norsk stedsnavnsforskning har en i flere år diskutert hvilken nytte en kunne ha av EDB på dette

feltet. I 1973 ble det etter initiativ fra Norsk stadnamarkiv etablert et samarbeid om et prøveprosjekt med Prosjekt for datamaskinell språkbehandling (PDS), Nordisk institutt, Universitetet i Bergen. PDS tok på seg å utvikle et prøveopplegg for et datamaskinelt navnearkiv på ca. 1000 ord. Arbeidet ble foreløpig avsluttet samme år, og resultatene ble lagt frem på et nordisk møte om stedsnavnsforskning i København sommeren 1973.

Formålet med prøveprosjektet var å få klarhet i de fordeler et datamaskinelt arkiv ville gi. Det materialet som ble tilrettelagt for punching, ble derfor hentet fra eksisterende arkivtilfang. Et stedsnavnsmateriale inneholder opplysninger om f.eks. oppslagsform, stedsangivelse og kartreferanse, lydskrift, andre språklige merknader og kommentarer i fri språklig form.

PDS punchet materialet og utarbeidet programmer som kunne gi systematiserte utskrifter av det (systematisering etter de ulike (morfemindelte) enheter sammensatte ord, ordklassetilhørighet, navnetype, tonelagstype o.fl.)

I et av de første nummer av Forskningsnytt vil det bli gitt en reddegjørelse for tiltaket og de erfa-

ringer en gjorde ved databehandling av denne type ordmateriale.

Talemålsundersøkelsen i Oslo.

Under prosjektleder Eskil HANSSENS ledelse arbeider Talemålsundersøkelsen i Oslo med studier av talemålet i Oslo med henblikk på sosial og regional variasjon i lydsystem, bøyningssystem og syntaks, med hovedvekt på det siste aspektet. Talemålsundersøkelsen i Oslo skiller seg fra andre tiltak innen norsk språkforskning ved at det bygger på utstrakt bruk av gruppearbeid.

Som datagrunnlag bruker prosjektet intervjuer med utvalgte informanter i Oslo. Intervjuene blir transkribert, primæranalysert og punchet for videre EDB-behandling. På det materialet som på denne måten blir maskintilgjengelig, blir det foretatt undersøkelser av ordforråd og frekvenser. Ved hjelp av EDB blir det også foretatt spesialsorteringer for søremner som prosjektets medarbeidere tar opp.

Da den syntaktiske analyse av materialet står sentralt i prosjektets arbeid, blir intervjumaterialet segmentert og klassifisert av de faglige medarbeidere i bestemte syntaktiske enheter og konstruksjonstyper. Resultatet av den manuelle syntaktiske analyse blir deretter punchet og databehandlet for

beregning av frekvenser og fordeling av enheter i materialet.

I likhet med flere EDB-prosjekter i Oslo bruker Talemålsundersøkelsen de standardprogrammer for tekstbehandling (TEXT) som konsulent Ivar Fonnes har utviklet. For den statistiske behandling av de syntaktiske data har konsulent Fonnes laget en tilknytning til den statistiske programpakken DDPP som er tilgjengelig ved dataanlegget ved Universitetet i Oslo.

Innenfor rammen av prosjektet planlegges det også en undersøkelse av Osloboeres språkvekslinger hvor også databehandling vil bli nyttet.

Slavisk-baltisk institutt.

Ved Slavisk-baltisk institutt har det i de siste årene vokset frem et aktivt EDB-miljø hvor arbeidet til nå i hovedsak har vært å nytte datamaskinen i studiet av russiske lyrikeres språk og stil. Det er ved Instituttet således nedlagt et betydelig arbeid med å overføre til maskinleselig form sentrale tekster til forståelse av russisk lyrikk i det 19. og 20. århundre. Av praktiske grunner har en ved Instituttet til nå arbeidet med transkriberte versjoner av diktene.

Av arbeid som har vært utført, kan nevnes at professor Geir KJETSAA har gitt en kvantitativ beskriv-

else av ordforrådet i den russiske dikter Lermontovs dikt. Professor Kjetsaa har også utarbeidet en norm for Pusjkin-tidens dikterspråk med grunnlag i stikkprøver av produksjonen til de diktere i perioden som ansees som litteraturhistorisk mest betydningsfulle og mest produktive. Stikkprøvene er hentet fra verkene til 21 diktere og danner grunnlaget for beskrivelsen av Lermontovs ordforråd. I tillegg arbeides det med fremstilling av en konkordans til Tjutčevs dikterspråk hvor det vil bli foretatt påfølgende litterære undersøkelser. Arbeidet utføres av et team bestående av E. Egeberg, I. Fonnes, S. Gil, G. Kjetsaa og A. D. Perminow.

Vit.ass. Steinar GIL har utarbeidet en sammenlignende stilanalyse av språket hos russiske lyrikere i slutten av forrige århundre og begynnelsen av dette med særlig vekt på studiet av lyrikeren Achmatovas dikterspråk. Hans avhandling 'Forsøk på en kvantitativ stilundersøkelse av substantivene i Anna Achmatovas ordforråd, del 1' er i 1974 godkjent for lisensiatgradsprøven ved Universitetet i Oslo.

Univ.stip. Erik EGEBERG har tilrettelagt dikt av lyrikeren A. A. Fet med tanke på å bruke EDB i semantiske og stilistiske undersøkelser.

Stud.philol. Ernst HANSEN har utarbeidet ordlister for 1700-tallsdikteren Lermontov og er i ferd med å sammenligne disse med 16 stikkprøver tatt fra verk av andre diktere i samme periode.

Britisk institutt.

Forsk.stip. Einar BJORVAND har tilrettelagt Edmund Spensers 'Fowre Hymnes' for datamaskinell språkbehandling og har fått utarbeidet en konkordans til dette verket. Konkordansen er nå publisert (Univ. forlaget 1973).

Hovedfagsstudent Sturla KOLSTAD har overført til EDB en samling nyhetsreportasjer og lederartikler fra 'The Times' og 'Daily Mirror' og vil nytte materialet til undersøkelser av ordforrådet i tekstene.

Også innenfor studiet av gammelengelsk er databehandling tatt i bruk idet dosent Arthur O. SANDVED har gjort verket 'Lives of Saints' av den gammelengelske forfatter Ælfric tilgjengelig for databehandling. Til materialet er det utarbeidet en konkordans. Dosent SANDVED vil bl.a. nytte de gammelengelske tekstene til studiet av kasusbruken etter preposisjoner.

Institutt for musikkvitenskap.

Institutt for musikkvitenskap er i ferd med å ta i bruk EDB i flere

forskningsprosjekter. Under ledelse av univ.lektor Jon-Roar BJØRKVOLD arbeides det med en komparativ analyse i 1800-tallets romanse-genre, med utgangspunkt i musikk-verk av komponistene Schubert, Schumann, Kjerulf og Grieg. Det planlegges videre en musikk-sosiologisk undersøkelse av førskolebarns sangmiljø i Oslo, der EDB ikke bare er tenkt brukt på rent sosiologisk materiale, men også ved analyse av selve sangmelodiene. Det er under utarbeidelse et forslag til maskinleselig kode som vil dekke hele det semantiske innhold i et standard notebilde.

Univ.lektor Arvid O. VOLLSNES arbeider med et musikkvitenskapelig prosjekt hvor EDB er tatt i bruk ved en undersøkelse av polyfoni og samklang i Farteln Valens musikk, i første omgang i hans motetter.

UNIVERSITETET I BERGEN

Nordisk institutt.

Prosjekt for datamaskinell språkbehandling (PDS) startet virksomheten i 1967 da det ble opprettet en amanuensis-stilling i språklig databehandling ved instituttet. Ved PDS arbeider der i dag foruten lederen, førsteamanuensis Kolbjørn Heggstad, en vitenskapelig assistent, to programmerere og to kontorassistenter.

I de årene PDS har vært i virksomhet har en hatt som hovedoppgaver

- 1) å skape et eget programmerings-system på universitetets dataanlegg til bruk ved datamaskinell språkbehandling,
- 2) å bygge opp et arkiv med maskintilgjengelige tekster i norsk språkforskning og
- 3) å bygge opp et maskinbasert ordarkiv over et stort norsk ordtilfang.

I løpet av perioden har en utviklet diverse programsystemer til oppgaver innen Nordisk institutt og ved andre institusjoner. Det er bl.a. utviklet et programsystem for tekstbehandling kalt STRIL i samarbeid med George M. Gillow.

En har i dag et tekstarkiv som omfatter ca. 6 mill. løpende ord fra norsk skjønnlitteratur, faglitteratur og fra aviser.

Videre er det utarbeidet en norsk ordregistrant som i dag omfatter ca. 55.000 oppslagsord i bokmål og ca. 40.000 oppslagsord i nynorsk. Vokabularet, som er grammatisk klassifisert, er bygd opp av ordmaterialet i en rekke ordbøker og ordsamlinger. Ut fra de grammatiske kodene kan en oppslagsforms bøyingsformer automatisk genereres. Prosjektet samarbeider med Norsk språkråd om denne oppgaven, og registranten vil bl.a. bli brukt som et hjelpemiddel i normeringen av norsk språk.

I 1969 tok PDS initiativet til et opplegg for databehandling av tilveksten i norsk ordforråd etter 1945 og samarbeidet inntil 1971 med Norsk språknemnd om dette tiltaket. Da ble samarbeidet utvidet til også å omfatte Norsk leksikografisk institutt. Norsk språk-nemnds del av virksomheten blir ført videre av Norsk språkråd. Institusjonene ekserperer og overfører til hullbånd materiale som så blir lest inn og lagt til arkivet over nyordstilfang ved PDS. Ulike sorteringer og lister blir deretter produsert fra materialet til bruk ved institusjonene i Oslo.

Etter oppdrag fra Universitetsforlaget har en ved PDS i gang et norsk-tysk ordboksprosjekt hvor en bygger opp et ordboksmateriale i maskintilgjengelig form. I forbindelse med dette prosjektet er det blitt utviklet et programopplegg for fotosats.

I årenes løp har PDS ytt verdifull assistanse til språkforskere og andre som har ønsket å sette seg inn i de muligheter EDB gir i forskningsarbeidet. For fremtiden vil PDS konsentrere innsatsen om å videreføre arbeidet med et datamaskinelt arkiv for norsk språk. De muligheter for fleksibel og rask søking, kopiering av materiale og tillegg av data som EDB gir, vil gjøre et slikt arkiv til et verdifullt hjelpemiddel både i

norsk språkforskning og i andre vitenskaper hvor en arbeider med data om norsk språk. Viktige deler av arbeidet vil bli konsentrert om forsøk med å utvikle automatiske ekserperingsprogram slik at en med grunnlag i en ordregistrant automatisk kan få lagt til registranten de nye ord som de ulike deler av tekstarkivet inneholder.

Prosjekt for datamaskinell språkbehandling gir ut serien Norske Språkdata hvor forskningsrapporter og data fra prosjektet blir presentert.

Innenfor rammen av prosjektets virksomhet arbeider vit.ass. Gulbrand ALHAUG med en forskningsoppgave om prinsipper for sammensetninger av ord i moderne norsk, hvor både maskintilgjengelig avismateriale og bokmålsdelen av ordregistranten blir nyttet. Arbeidet tar sikte på å analysere sammensetninger med substantivisk forledd fra såvel fonematisk som grammatisk og semantisk synsvinkel.

Ved Nordisk Institutt, Avd. for norrøn filologi, arbeider univ. lektor Magnus RINDAL med datamaskinell behandling av Barlaams og Josephats saga. Tekstene som benyttes, er et gammelnorsk håndskrift fra annen halvdel av 13. århundre, som inneholder ca. 80.000 ord. Opplysninger om håndskriftvarianter er innarbeidet i tekst-

grunnlaget. Teksten er punchet og korrekturlest. Den tilrettelegges nå for en automatisk, fullstendig ordeksperering, en analyse av språket, og en ny utgave.

Førsteamanuensis Marina MUNDT ved samme avdeling brukte for noen år siden EDB ved en stilistisk undersøkelse av Laxdoela saga. For å sammenligne språket i denne sagaen med ordforrådet i andre sagaer ble Håkonar saga, Sturlas Islendinga saga og Knytlinga tilrettelagt for databehandling. Til sagatekstene ble det utarbeidet ord-indeksér.

Forsk.stip. Per-Bjørn PEDERSEN ved Avd. for nordisk språkvitenskap forbereder en analyse av stilistiske forhold i nynorske litterære tekster fra omkring århundreskiftet i samarbeid med NAVF's EDB-senter. Det blir for tiden punchet et tekstmateriale hvor det senere blir innarbeidet kodete opplysninger til de tekstdeler som studiet omfatter. Materialet vil bli behandlet både kvantitativt og kvalitativt.

Klassisk institutt.

Ved Klassisk institutt har professor Knut KLEVE med forsk.stip. Jan SONGSTAD som nærmeste medarbeider ledet et forskningsprosjekt kalt 'Papyrene fra Herculaneum' siden 1971. Prosjektet samarbeider med NAVF's EDB-senter. Formålet er ved hjelp av EDB å fremskaffe materiale

for studiet av epikureismen og for å tolke papyrustekster som er overlevert i fragmenter.

For dette formål er det tilrettelagt for maskinell behandling Epikurs egne skrifter og andre nærstående filosofiske verk, og alle skrifter av filosofen Philodem som finnes blant papyrustekstene fra Herculaneum. Materialet omfatter i dag 90 tekster på tilsammen ca. 5000 moderne boksider. Størst mulig troskap mot foreleggenes greske språkform er tilstrebet.

Arbeidet ved prosjektet vil i første omgang ta sikte på å utarbeide konkordanser til tekstene. I forbindelse med den nyttgivelsen av Philodems verk 'Om Gudene' 1. bok som professor Kleve vil foreta, venter en at materialet også vil være til nytte i arbeidet med lakuneproblemene.

Religionsvitenskapelig institutt.

Professor Peder BORGES og univ.stip. Roald SKARSTEN har i de siste år drevet et EDB-prosjekt med det mål å fremskaffe leksikografiske hjelpemidler (KWIC-konkordans, ordlister, frekvenslister etc.) til studiet av skriftene til den jødiske teolog Philo av Alexandria (som levde ved vår tidsregnings begynnelse).

Ut fra en vitenskapelig standardutgave på gresk og med tillegg av

senere publiserte tekstfragmenter er det tilrettelagt den mest fullstendige tekstsamling som finnes av Philos verker. Materialet består av ca. 1/2 mill. løpende ord og diakritiske tegn er tatt med.

Arbeid pågår nå med å lage KWIC-konkordans til de 38 bøkene som skriftsamlingen er inndelt i, og en konkordans til en av bøkene er blitt publisert på instituttet (1972). Konkordansen vil innen prosjektet bli brukt til studiet av eksegetisk terminologi og stilmønstre i kommentarverker til Det gamle testamente.

Romansk institutt.

I samarbeid med NAVF's EDB-senter for humanistisk forskning har professor Lars Otto GRUNDT høsten 1973 satt i gang en systematisk undersøkelse av naturvitenskapenes ordforråd og syntaks i mellomfransk. Første etappe er en analyse av Traite de Cosmographie skrevet i 1432 av astronomen Jean Fusoris. Verket er blitt punchet, og programmer er under utarbeidelse med henblikk på en beskrivelse av bokstavenes fordeling, redundans og informasjonsverdi. Denne beskrivelsen vil tjene som grunnlag for neste etappe hvor en vil utarbeide en naturvitenskapelig terminologi, samt en oversikt over elementære setningsstrukturer i mellomfransk.

Russisk institutt.

Hovedfagsstudent Malvin IVERSEN forbereder et materiale for databehandling i tilknytning til sitt arbeid 'Predikativ bruk av adjektiv ved former av kopula i moderne russisk', og samarbeider med NAVF's EDB-senter.

Arbeidet bygger på et eksempelmateriale hentet fra skjønnlitterære tekster og bruksprosa som er skrevet etter 1950. Med tanke på den senere lingvistiske behandling blir materialet inngående klassifisert etter en rekke språklige trekk før overføring til maskinlesbar form starter. Ved EDB-analysen vil en ved søking etter ulike kriterier prøve å avdekke forhold som regulerer bruken av denne leddkategorien i moderne russisk.

Tysk institutt.

EDB vil her bli tatt i bruk av en hovedfagsstudent, Knut SCHRØDER, som arbeider innen feltet feilanalyse. NAVF's EDB-senter yter bistand til tiltaket. Som datagrunnlag brukes 90 frie stiler på tysk skrevet av gymnasiaster med utvidet tyskpensum (ca. 18.000 ord). Stilmaterialet blir tilrettelagt for datamaskinell behandling på en slik måte at både elevenes versjon og 'rettingene' (som klassifiseres) blir tatt vare

på. I løpet av våren vil data-maskin bli brukt til å sortere og statistisk behandle materialet som grunnlag for det videre arbeid med feilanalysen.

Etno-folklorisk institutt.

Ved dette instituttet er det satt i gang et prosjekt innen studiet av folkedans. Under ledelse av magister Egil BAKKA, Trondheim er det utarbeidet et system for å gi en maskin-lesbar representasjon av opplysninger knyttet til eller hentet fra 89 filmopptak av bygdedans fra Telemark og Aust-Agder (f.eks. informant, bygd, dansetype og dansvariasjon). Det er til nå tilrettelagt et prøvemateriale på 2100 hullkort og innledende programmer for behandling av materialet er utviklet. Ansvarlig for programarbeidet er Bruce Taylor, EDB-avdelingen, Universitetet i Bergen.

NAVF's EDB-senter for humanistisk forskning.

Ved NAVF's EDB-senter har adm.leder Jostein H. Hauge tatt initiativ til en overføring av norsk avis-materiale fra det 20. århundre for datamaskinell analyse. Til i dag er det tilrettelagt et materiale som består av en ukes avisstoff fra hver av avisene Dagbladet, Morgenbladet, Bergens Tidende og Morgenavisen hentet fra 1900 og

1925 (til sammen ca. 600.000 ord). Utvidelse med materiale fra 1950 planlegges.

Alt avisstoffet unntatt annonser og tabeller er punchet slik det foreligger i avisnumrene. Det er foretatt en inndeling av materialet i stofftyper ved hjelp av et kodesystem. Materialet er primært tilrettelagt for studiet av moderne norsk verbalsyntaks, men det representerer også et verdifullt grunnlagsmateriale for studiet av utviklingstendenser i norsk avisspråk i vårt århundre.

Historisk institutt.

Ved Historisk institutt har det fra 1969 pågått arbeid med en ny statistisk bearbeidelse av Folketellingen av 1801. Ansvarlig for tiltaket er Historisk institutt, Universitetet i Bergen, Statistisk sentralbyrå og Riksarkivet. Folketellingen av 1801 er en hovedkilde for studiet av norsk samfunnsutvikling på 1700-tallet og den er et naturlig utgangspunkt for studiet av det norske samfunn på 1800-tallet. Materialet er derfor et meget benyttet referansepunkt i historisk forskning. Den tilrettelegging av det statistiske materialet for data-maskinelle analyser som nå foregår i Bergen, utføres slik at alle tilgjengelige data fra folketellingsprotokollene blir punchet uten noen form for koding.

Protokollen omfatter opplysninger om ca. 880.000 personer. En viktig del av virksomheten har vært å utvikle programmer til å oversette den tekstuelle informasjonen til tallkoder som er velegnede for statistisk bearbeidning. Dette problemet er løst ved et sett av automatiske analysesystem der et av hovedsiktepunktene har vært å gjøre det enkelt å foreta manuelle forandringer til den koding maskinen foretar. Ansvarlig for programmeringsarbeidet er vit.ass. Jan Oldervoll, Historisk institutt. Primærpunchingen er nå så godt som fullført, og det er ventet at folketellingen i sin helhet vil være analyseklar sommeren 1975.

Styret for Kjeldeskrift-fondet og Kommissjon for Diplomatarium Norvegicum har tatt opp arbeidet med en revidert og videreført utgave av Regesta Norvegica. Regesta Norvegica skal gi en kronologisk ordnet fortegnelse over alle kjente dokumenter til Norges middelalderhistorie, trykte og utrykte, med innholdsgjengivelse og alle nødvendige opplysninger.

Arbeidet med tilretteleggingen av kildematerialet blir for den største del utført ved de historiske institutter i Oslo, Bergen og Trondheim.

Som ledd i publiseringsarbeidet tilrettelegges nå ved Historisk

Institutt i Bergen de første bindene av Regesta Norvegica for optisk lesing. Det er planen å utarbeide ulike registre ved hjelp av datamaskin og tilrettelegge en magnetbåndversjon av trykkmanuskriptet som kan brukes direkte ved fotosetting.

Historisk museum.

Førstekonservator Asbjørn E. HERTEIG er leder for prosjektet 'Den vitenskapelige bearbeidelse av Bryggematerialet' hvor også EDB blir tatt i bruk som et hjelpemiddel i samarbeid med NAVF's EDB-senter.

De arkeologiske undersøkelsene på Bryggen i Bergen i tidsrommet 1955 - 1968 brakte for dagen et allsidig materiale som omfatter flere hundre tusen enheter. For å kunne utnytte dette materialet vitenskapelig og musealt har det vært nødvendig å overføre samtlige lokaliseringssdata til EDB-lesbar form. Ca. 1/5-del av disse data er tilrettelagt for punching, og overføring av data til maskinlesbar form pågår kontinuerlig. Arbeidet med tilrettelegging av materialet på Universitetets dataanlegg pågår og likeledes programutvikling for den vitenskapelige systematisering av det.

I de senere faser av analysearbeidet vil det være ønskelig å anvende EDB

ved rekonstruksjon av det keramiske materialet og som utgangspunkt for vitenskapelig analyse av funnstoffet og statistiske undersøkelser.

Ved Historisk museum, Avd. A, pågår det i samarbeid med NAVF's EDB-senter arbeid med å tilrettelegge steinaldermateriale for datamaskinell behandling. Det er utarbeidet en egen type avmerkingskort ('mark sense card') for museets data og de praktiske prøver vil foregå i løpet av vinteren 1974.

Innenfor rammen av Historisk museums forskningsprogram 'Hus og husbunad på Vestlandet' har Fylkeskonservatoren i Hordaland, og Historisk museum, Avd. B, utviklet et samarbeid om registrering og systematisering av kulturhistorisk grunnlagsmateriale som kommer inn til arkivet gjennom årlige kulturgeografiske registreringer og katalogiseringsprosjekt ved lokalmuseene. Materialet punches fortløpende, og en tar sikte på å bygge opp et 'museumsnettverk', basert på et sentralt kulturhistorisk oversiktsarkiv.

Med utgangspunkt i dette materialet har fylkeskonservator Nils G. BREKKE tatt opp et forskningsprosjekt kalt 'Vestnorsk folkekunst', hvor målsettingen er å gjennomføre en faktoranalyse av alle motiv som inneholder ornamenter, symbol- og figurframstillinger. Analysen tar

sikte på å klarlegge materialets morfologiske og ikonologiske strukturer, og som et ledd i dette arbeidet bygges det opp et generelt ikonografisk arkiv og motivregister over norsk folkekunst.

Når det gjelder bruk av EDB, er det til i dag i samråd med NAVF's EDB-senter utarbeidet katalogkort og opplegg for å representere den analytiske beskrivelse av arkivmaterialet i datamaskinell form. Arbeidet har for øvrig vært konsentrert om den terminologiske grunnlagssystematikk som den maskinelle behandling forutsetter.

UNIVERSITETET I TRONDHEIM

Musikkvitenskapelig institutt.

Ved instituttet har det i flere år vært arbeidet for å ta datamaskinen i bruk i forskningsarbeidet i musikkvitenskap. Professor Hampus HULDT-NYSTRØM og vit.ass. Tore SIMONSEN leder arbeidet med prosjektet NORSTIL som har til måsetting å utvikle et EDB-verktøy til bruk i stilistiske analyser av musikk. Det legges stor vekt på å gjøre dette verktøyet så lett tilgjengelig som mulig for forskere som ikke har tidligere erfaring med EDB.

For dette formål er det utviklet en innmatingskode for transkribering av primærinformasjon i note-

skrift. Det foreligger videre et programsystem som bl.a. gir muligheter for overføring og lagring av data i databasen, syntaktisk korrekturlesing av data, 'korrekturlytting' av data (i samarbeid med Institutt for Akustikk, UNIT-NTH og med Sveriges Radios Elektronmusikkstudio, Stockholm), søking på databasen og for analyser av data. Analysedelen av programsystemet er skrevet slik at en forsker med forholdsvis elementær kjennskap til programmering vil være i stand til å implementere evt. egne analyseprogram i NORSTILs programsystem.

Det er til analysebruk kodet og lest inn ca. 300 skandinaviske polsdanser fra ulike trykte kilder. Disse vil bli brukt til en sammenlignende analyse av 'dialekt'-trekk i folkemusikk.

Univ.lektor Arne HOLEN har nylig avsluttet et arbeid om komponisten Heinrich Schutz' harmonikk, hvor EDB har vært brukt til akkord- og akkordparsøking og til statistisk behandling.

Historisk institutt.

I den siste tiden har det vært kontakt mellom medarbeidere i Det nordiske Ødegårdsprosjektet ved instituttet og Regnesentret ved Universitetet i Trondheim. En har drøftet muligheten for et prøveprosjekt for noe av det materiale

en samler inn i Trondheim. Målet er å tilrettelegge og analysere data fra et delområde av Ødegårdsprosjektet for å kunne vurdere hvorvidt det er ønskelig å ta i bruk datamaskin i instituttets arbeid innenfor dette nordiske samarbeidsprosjektet.

BEHOV FOR KURS OG INFORMASJON

NAVF's EDB-senter for humanistisk forskning forsøker gjennom sitt personale ved universitetene og gjennom kontakt med fagområdene å danne seg et bilde av de informasjons- og kursbehov som i dag eksisterer i tilknytning til EDB-anvendelser i de humanistiske fag.

For å få et sannest mulig inntrykk av situasjonen ønsker senteret også kontakt med de enkelte aktive eller potensielle EDB-brukere som har synspunkter på disse spørsmålene. Vi ser gjerne også at det kan komme i gang en meningsutveksling om emnet i dette bladet. Kontakten kan opprettes så formelt eller uformelt som den enkelte ønsker.

THIRD SYMPOSIUM ON THE USE OF COMPUTERS IN LITERARY AND LINGUISTIC RESEARCH - CARDIFF 1. - 5. APRIL 1974

Symposiet i Cardiff er det tredje i denne serien hvor en tar opp sentrale emner i tilknytning til databehandling av tekster for språkvitenskapelig eller litterær analyse (Cambridge 1970, Edinburgh 1972).

På symposiet i april vil det bli gitt 40 presentasjoner av EDB-arbeidet innenfor disse arbeidsfeltene:

Klassiske studier.
Programutrustning for litterær databehandling.
Statistisk analyse.
Tekst- og vokabularstudier.
Stilistikk.
Tekstkritikk.
Sosiolingvistik.
Leksikografi.
Tematisk analyse.
Konkordanser.
Problemer i forbindelse med inn- og utdata.
Moderne og orientalske språk.
Syntaktisk analyse.

I neste nummer av HUMANISTISKE DATA vil det bli gitt en nærmere redagjøreelse for en del av de emner som ble behandlet på symposiet i Cardiff.

Oversikt over maskinprogram i Norge til bruk i humanistisk forskning

For å gi en første oversikt over de programmer som til nå er utarbeidet i vårt land til bruk i humanistiske databehandlingsoppgaver, har en ved NAVF's EDB-senter foretatt en rundspørring i de forskjellige miljøene om den programutvikling som har foregått. I de følgende spalter gir vi videre de opplysningene som er kommet inn og viser til de personer som bør kontaktes for nærmere opplysninger om programmene. For de fleste EDB-miljøene gjelder det at det i tillegg til de mer generelle programprodukter foreligger programmer for spesialiserte oppgaver i tilknytning til de enkelte faglige prosjekter. Slike programmer blir bare i liten utstrekning omtalt separat.

Den oversikten over EDB-orientert forskningsarbeid som finnes i dette nummeret, vil i en del tilfeller gi supplerende bakgrunnsmateriale.

NAVF'S EDB-SENTER FOR HUMANISTISK FORSKNING

Ved NAVF's EDB-senter har en del av arbeidet med programutvikling til nå hatt som mål å utvikle basal

programutrustning som grunnlag for den applikasjonsrettede programmering. Av slikt programarbeid kan nevnes utvikling av kommunikasjonsprogram for tilknytning til UNIVAC 1110 ved Universitetet i Bergen og kommunikasjonssystem for parallell bruk av flere terminaler til senterets maskin PDP 11/20.

Det er dessuten utviklet en rekke programmer av mer almen interesse. Av slike kan nevnes:

- * Program for å lese inn tekst fra papirbånd. Programmet kan ta hånd om de fleste papirbåndkoder.
- * Program for korrigering av tekst innlest fra papirbånd.
- * Ordseleksjonsprogrammer (maske-system) for enkelt å kunne gjenfinne spesifiserte formelle strukturer i en tekst. Programmet er tenkt å skulle inngå som en del av et generelt tekstbehandlingssystem.
- * Ordlisteprogrammer for konstruksjon og lagring av ordlister i en form egnet for terminalbruk. I tilknytning til disse er en i ferd med å utvikle et programsystem for terminalbruk som gjør

det mulig fra terminal å slå opp i en ordliste, sette forskningsrelevant informasjon til språklige enheter i en tekst ('tags' og siden viderebehandle teksten med deres tilleggskoder.

- * Mindre programmer for kvantitativ analyse av tekster, blant annet ord- og setningsstatistik

Alle de foran nevnte programmene er skrevet i Assembler på PDP 11/20.

- * For bruk på UNIVAC 1110 ved Universitetet i Bergen er det i COBOL skrevet et program for dannelsen av KWIC-indekser.

(Dette er en representasjon av et tekstmateriale hvor alle ord skrives ut på egen linje i sin kontekst (Key Word In Context) med referanse til hvor i teksteordet forekommer.)

Dessuten er det utviklet programmer i tilknytning til behandling av greske tekster og arkeologisk materiale på UNIVAC 1110.

Ved henvendelse til senteret vil en kunne få nærmere opplysninger.

UNIVERSITETET I BERGEN

Prosjekt for datamaskinell språkbehandling, Nordisk institutt.

Ved Prosjekt for datamaskinell språkbehandling (PDS) er det i løpet av de 6 - 7 siste årene lagt ned et stort arbeid i å utvikle programmer for tekstbehandling. Av slike kan nevnes:

- * Program for Innlesing av papirbånd (også TTS-bånd).
- * Program for lagring av tekst.
- * Korrigeringsprogram.
- * Program for utarbeidelse av en rekke ulike typer frekvensordlister.
- * Program for utarbeidelse av KWIC-Indekser og konkordanser. Med konkordansprogram menes her program som gir en konvensjonell konkordans hvor et bestemt ord, foruten å bli skrevet ut i sin kontekst, også blir uthevet som oppslagsord på en egen linje. En slik representasjon karakteriseres undertiden med termen KWOC (Key Word Out of Context).
- * Tekstbehandlingssystemet STRIL, skrevet i PL/1 og Assembly, utviklet i samarbeid med sjefskonsulent George M. Gillow. STRIL-systemet dekker en rekke

av de mest brukte rutiner innenfor datamaskinell språkbehandling og er bl.a. karakterisert ved fleksible maske-program for mønstergjenkjenning. Dette tekstbehandlingssystemet er skrevet for IBM 360 og kan i dag ikke brukes ved det nye UNIVAC-anlegget ved Universitetet i Bergen, bortsett fra enkelte komponenter som er skrevet om i ALGOL.

Ved PDS foreligger det dessuten en serie programmer som nyttes i Prosjektets arbeid med de ulike typer ord- og tekstarkiv som er bygd opp ved Institusjonen.

Nærmere opplysninger fås ved henvendelse til lederen, førsteamanuensis Kolbjørn Heggstad, Nordisk Institutt.

Historisk institutt.

I forbindelse med forskjellige EDB-prosjekter innen historisk forskning er det skrevet en rekke programmer av mer generell natur.

- * Program for dannelse av KWIC-Indekser hvor brukeren har mulighet for å spesifisere ord som ikke skal være med.
- * Program for å lage f.eks. sted- og navneregister fra en tekst.

- * Identifiseringsprogram som ut fra ett eller flere datasett kopler sammen alle opplysninger som gjelder samme person.
- * Typesettingsprogram til produksjon av magnetbånd som kan styre settemaskin av typen Linotron 505.
- * I forbindelse med en ny statistisk bearbeidelse av folketellingen av 1801 er det bl.a. laget et klassifiseringsprogram for å sette tallkoder på tekstuell informasjon, f.eks. yrkestermer.

Alle programmene er skrevet i PL/1 og kan foreløpig bare kjøres på IBM-anlegg. For nærmere opplysninger: Vit.ass. Jan Oldervoll, Historisk Institutt.

Religionsvitenskapelig institutt.

Ved Religionsvitenskapelig institutt er det til databehandling av greske tekster utarbeidet et KWIC-Indeksprogram som er spesielt tilpasset gresk språkmateriale.

- * Indeksprogrammet er lagt opp slik at det kan behandle tekster hvor alle diakritiske tegn i gresk er tatt med. Programmet fjerner alle diakritiske tegn før sortering og setter dem på plass igjen når sorteringsoperasjonene er fullført.

Nårmere opplysninger fås ved henvendelse til univ.stip. Roald Skarsten, Religiøsvitenskapelig Institutt.

UNIVERSITETET I OSLO

Det meste av den programutvikling som refereres nedenfor, er utført i løpet av de siste årene av Ivar Fønnes, som nå er konsulent for NAVF's EDB-senter i Oslo.

En vesentlig del av utviklingsoppgavene har hatt tilknytning til forskningsarbeid ved Historisk institutt.

Alle programmer er skrevet i FORTRAN for bruk på CDC 3300 ved Universitetet i Oslo.

* Programsystemet TEXT.

TEXT er et generelt programsystem for tekstanalyse og består av 27 enkeltprogrammer. Systemet omfatter programmer for

- a) Innlesing og redigering av tekstinformasjon.
- b) Informasjonssøking.
- c) Frekvensordlister av diverse typer.
- d) KWIC-Indekser.
- e) Diverse beregninger - ordforråd, leselighetsindeks (LIX), bokstavfrekvenser osv.
- f) Tilrettelegging (koding) av

data for statistisk behandling med programpakken DDPP.

Alle programmer i TEXT er lagt til rette slik at brukerne selv kan utføre alle deler av databehandlingen. Kurs i bruk av systemet holdes regelmessig, og programbeskrivelser er utarbeidet. En nærmere presentasjon av systemet er gitt i en stensil fra 1971.

* HISO-programmene.

HISO er et programsystem for behandling av rubriserte protokoll-data (f.eks. folketellingsdata) i tekstform. Systemet består av 18 enkeltprogrammer som omfatter

- a) Innlesing, konsistenssjekking og redigering av protokoll-data.
- b) Opplegg av dataarkiv, samsortering av flere datasett.
- c) Sortering av rubrikkilister, ordlister samt navnelister med alle opplysninger om de enkelte enheter.
- d) Maskinell koding og tilrettelegging for statistisk behandling med programpakken DDPP.

Henvendelse: Konsulent ved NAVF's EDB-senter for humanistisk forskning, Ivar Fønnes, Postboks 1102, Blindern, Oslo 3.

* Voteringsanalyseprogrammer.

Et sett analyseprogrammer er utarbeidet ved Historisk institutt for bruk i voteringsanalyser. Programmene forutsetter at data er innlest og tilrettelagt med program-

pakken DDPP, og de er utviklet for å kunne foreta spesielle analyser som ikke dekkes av DDPP. Pr. 1.1. 1974 foreligger 12 programmer som omfatter

- a) Sortering av representanter etter stemmegivning.
- b) Guttman-skalaer.
- c) Klyngeanalyse på grunnlag av samsvar i stemmegivning.
- d) Samhold innen grupper og polarisering mellom grupper.
- e) Korrelasjon mellom par av voteringer.

Alle voteringsanalyseprogrammene og HISO-programmene er lagt til rette slik at brukerne kan anvende dem på egen hånd. Programbeskrivelser er utarbeidet.

Henvendelse: Vit.ass. Elin Mehlum, Historisk Institutt eller konsulent Ivar Fønnes.

UNIVERSITETET I TRONDHEIM

Musikkvitenskapelig Institutt.

Ved dette instituttet foreligger programsystemet NORSTIL for analyse av oppnotert musikk. Det kan benyttes til analyse av flerstemt musikk, men egner seg i sin nåværende form best til enstemt musikk

* NORSTIL består av fire deler:

1. Innmatningskode for å transkribere primærinformasjon i noteskrift.

2. En Intern maskinkode (kalt analysekode) for å lagre primærinformasjon i et datamaskinlager.
3. Et program for korrekturlesing og innlesing av data.
4. Et utvidbart sett analyseprogrammer.

En benytter i dag SPSS (Statistical Package for Social Sciences) for analysen, men systemet er lagt opp med tanke på at egne analyseprogrammer skal kunne implementeres.

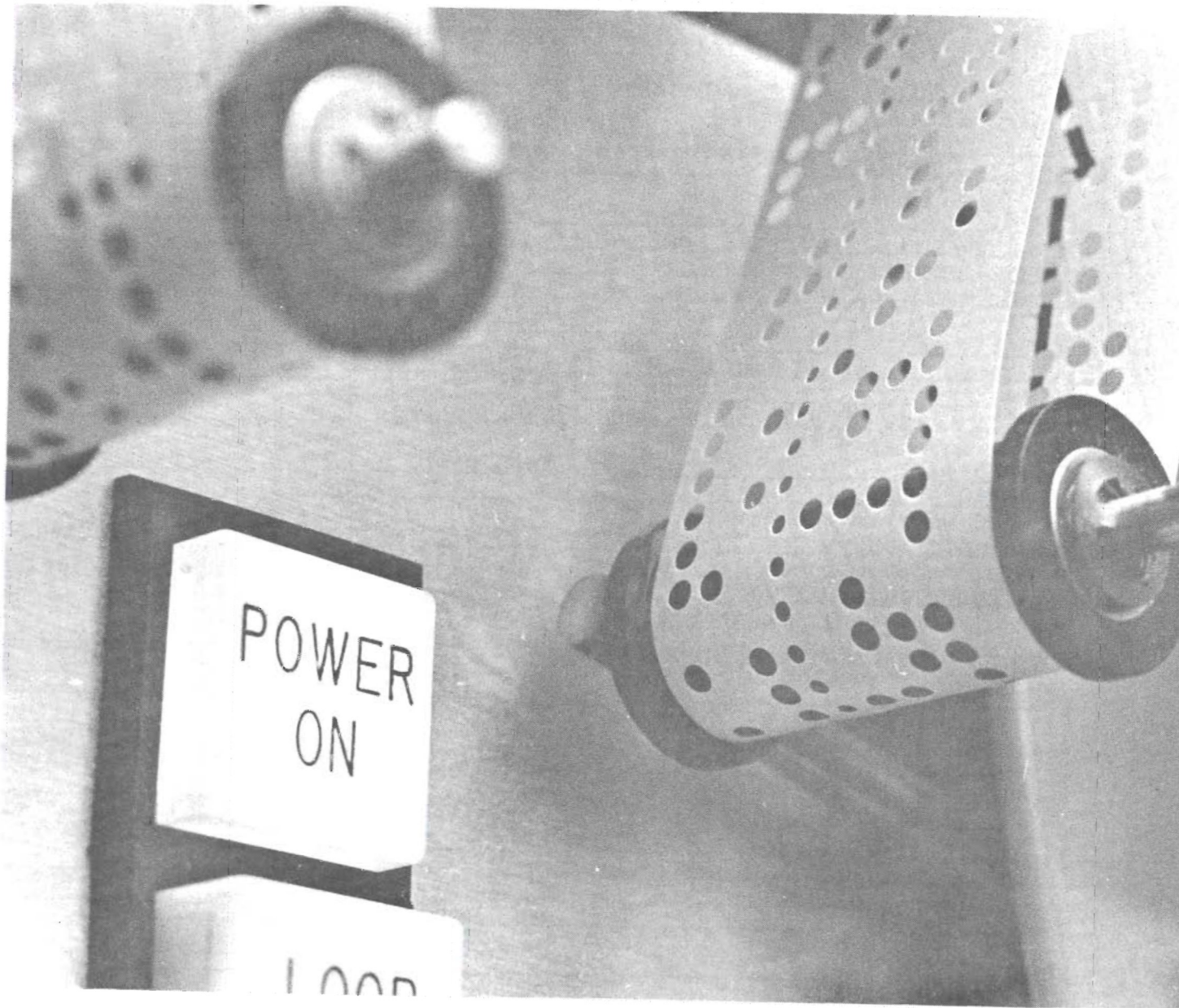
Systemet er utviklet på regnearklegget ved Universitetet i Trondheim (UNIVAC 1108), og er skrevet i ALGOL. Med små endringer kan det brukes på andre anlegg utstyrt med ALGOL kompilator.

Arbeidet, som har vært ledet av vit. ass. Tore Simonsen, er nærmere beskrevet i hans publikasjon 'NORSTIL 70. Et datamaskinsystem for stilanalyse av musikk', Trondheim 1973.

Nærmere opplysninger om programsystemet gis av forfatteren.

BEKLAGER

Den relativt lange tilblivelsestiden for dette nummeret har ført til at opplysningene på et par punkter ikke gir den dagsaktuelle situasjon. Disse forholdene vil vi gjøre rede for i neste nummer. Konklusjon: Den humanistiske forskning er i stadig utvikling!



D A T A I M A S K I N L E S E L I G F O R M

Forkortinger:

F = Datasettet er fullt tilgjengelig.
 A = Datasettet er tilgjengelig etter avtale.
 I = Datasettet er ikke tilgjengelig.

UiB = Universitetet i Bergen.
 UiO = Universitetet i Oslo.
 UNIT = Universitetet i Trondheim.

MB = magnetbånd, PB = papirbånd, HK = hullkort, PL = plattelager.
 OPT = optisk lesbart.

INNHold				HENVENDELSE
<u>AMERIKANSK</u>				
PROSA:				
Brown Corpus of American English	1 mill. ord	MB	F	Bertil Sundby, Engelsk inst., UiB.
Brown Corpus of American English	1 mill. ord	MB	F	Eirik Lien, RUNIT, Trondheim.
Brown Corpus of American English	1 mill. ord	MB	F	NAVF's EDB-senter, Bergen.
<u>ENGELSK</u>				
POESI:				
E. Spenser: Fowre Hymnes	9.200 ord	MB	A	Einar Bjorvand, Britisk inst., UiO
PROSA:				
Ælfric: Lives of Saints	108.000 ord	MB	A	Arthur O. Sandved, Britisk inst., UiO.
'The Times/Daily Mirror' 1/5-6/5-1972	72.000 ord	MB+ PB	A	Sturla Kolstad, Britisk inst., UiO.
ORDBØKER:				
The New Merriam Webster Pocket Dictionary	42.000 oppslagsord	MB	A	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UiB
Webster's Seventh New Collegiate Dictionary (ukomplett)	130.000 oppslagsord	MB	A	
Haugen: Norsk-engelsk ordbok	60.000 oppslagsord	MB	F	
<u>GRESK</u>				
PROSA:				
Filosofiske tekster av Philodem (ca. 40%), Epikur og andre epikureere, Plutarch og Sextus Empiricus	1 mill. ord	PB	F	Knut Kleve, Jan Songstad, Klassisk inst., UiB.
Philos samlede verker (38 bok- titler) + fragmenter	1/2 mill. ord	MB+ PB	A	Roald Skarsten, Peder Borgen, Rel.vit.inst., UiB.
POESI:				
Hesiod: Theogoni	5.100 ord	MB	A	NAVF's EDB-senter, Bergen.
Homer: Odysseen	66.000 ord	MB	A	
<u>RUSSISK</u>				
POESI:				
A.A. Fet: Osnovnoe sobranie stichotvorenij. Leningrad 1959.		MB	A	Erik Egeberg, Slavisk-baltisk inst., UiO.

' Anna Achmatova: Sočinenija I-II, New York 1967-68 (samlede dikt).			A	Steinar Gil, Slavisk-baltisk inst., UiO
Konstantin Bal'monts diktsamling 'Budem kak solnce' (1969)			A	
Innokentij Annenskijs samlede dikt med unntak av tragediene på vers.			A	
Aleksander Bloks diktsamling 'Gorod' (1960)			A	
F.I. Tjutčevs dikt (1966)		MB	A	Geir Kjetsaa, Slavisk-baltisk inst., UiO
E.A. Baratynskijs samlede dikt (1957)		MB	A	
Stikkprøver på versdikting fra 21 utvalgte diktere (skrevet i tiden 1828-1841)		MB	A	
M.J. Lermontovs samlede dikt (1961)		MB	A	
<u>NORSK</u>				
SAKPROSA:				
' Trontaledebatten i Stortinget 1953	ca. 100.000 ord	MB+	A	Ottar Dahl, Historisk inst., UiO.
Trontaledebatten januar 1949	76.500 ord	PB	A	
Trontaledebatten desember 1945	52.500 ord	MB+	A	
Lærebøker i historie for 4. klasse i grunnskolen (5 verk)	95.000 ord	MB	A	Biørn Berg, Pedagogisk inst., UiO
Ulike avis-kronikker om kvinnesak (1972)		MB+	A	Eskil Hansen, Tove Berg, Inst. for nordisk språk og litteratur, UiO.
' Arnholm: Avtalerett		MB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UIB (Det er ikke lest korrektur på største delen av dette materialet.)
Berulfsen, Djupedal, Lundebj: Ordet		PB	F	
Willy Dahl: Perspektiver		MB	F	
Tove Ellefsen (red.): Makt og monopol på Marienlyst (nyn.)		PB	F	
Gerhardt Gran: Essays i utvalg		PB	F	
Hageberg (red.): Omkring 'Peer Gynt' (nyn. og bokm.)		PB	F	
Longum: Frontlinjer		PB	F	
Torild Skard: Norsk etterkrigsradikalisme		MB	F	
Sigurd Skirbekk: Den samfunnsvitenskapelige tenkemåte		MB	F	

Lærebok i musikk og musikkhistorie for gymnas og lærerskole		MB	F	
Lærebok om forming i grunnskolen og ungdomsskolen		MB	F	
Trontaledebatt I - 1949		MB	A	
Trontaledebatt II - 1956		MB	A	
Mykland: Omkring 1814		PB	F	
Nærbøvik: Portretter i norsk historie		PB	F	
Ramsøy (red.): Det norske samfunn		PB	F	
Paasche: Kong Sverre		PB	F	
Jens Arup Seip: Tanke og handling i norsk historie		PB	F	
Skartveit: Revolusjon i Guds navn (nyn.)		PB	F	
Steen: Fra norsk historie		PB	F	
Stigen: Gud i moderne filosofi		PB	F	
Wellek & Warren: Litteraturteori		PB	F	
SKJØNNLITTERATUR:				
Borgen: Barndommens rike		MB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UIB.
Camus: Den fremmede		PB	F	
Elster: Farlige folk		MB	F	
Elster: En korsgang		MB	F	
Falkberget: Den fjerde nattevakt		MB	F	
Falkberget: Nattens brød (4 bind)		PB	F	
Falkberget: Eli Siursdotter		PB	F	
Falkberget: Bør Børson jr.		PB	F	(Det er ikke lest korrektur på størstedelen av dette materialet)
Garborg: Trøtte mænd (riksmål)		PB	F	
Garborg: Den burtkomne faderen (nyn.)		PB	F	
Trygve Gulbrandsen: Det blåser fra Dauingefiell		PB	F	
Trygve Gulbrandsen: Og bakom synger skogene		PB	F	
Hoel: Essays og epistler		PB	F	
Hoel: Trollringen		MB	F	
Kielland: Fortuna		PB	F	
Kielland; Garman & Worse		MB	F	
Kielland: Novellletter		PB	F	

Kjelland: Nye novelletter		PB	F	
Kjelland: Sne		PB	F	
Kjerkegaard: Begrepet Angst		PB	F	
Kjerkegaard: Frykt og Bæven		PB	F	
Kristiansen: Jomfru Lide		PB	F	
Krokann: I Dovresno (nyn.)		MB	F	
Lagerløf: Kjørekaeren (norsk oms.)		PB	F	
Lie: Familien på Gilje		MB	F	
Lie: Troll		MB	F	
Lo-Johanson: Lykken (norsk oms.)		PB	F	
Omre: Utvalgte noveller		PB	F	
Sandemose: En sjømand går i land		PB	F	
Sandemose: Det svundne er en drøm		PB	F	
Sandemose: En flyktning krysser sitt spor		MB	F	
Scott: Det gyldne evangelium		MB	F	
Scott: Kilden		PB	F	
Undset: Fattige skjebner		PB	F	
Undset: De kloge jomfruer		PB	F	
Undset: Jenny		PB	F	
Undset: Madame Dorthea (ukompl.)		PB	F	
Undset: Kristin Lavransdatter (ikke Kransen) Husfrue		PB	F	
Korset		PB	F	
Vesaas: Fuglane		MB	F	
Vinie: Ferdaminne		MB	F	
NYHETSSTOFF:				
Aftenposten 16/4 - 29/4-1968 (utdrag)	ca. 282.000 ord	MB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UiB
Morgenbladet 2/5-16/5-1968	ca. 225.000 ord	MB	F	
Dagbladet 2/5-16/5-1968	ca. 213.000 ord	MB	F	
NTB 1/2-30/4-1971 (utdrag)	ca. 1 1/2 mill. ord	MB	F	
Løpende tekst hentet fra Bergens Tidende 5-10/3 1900 og 2 - 7/3-1925. Dagbladet 1-7/7 1900 og 6-11/7-1925. Morgenavisen 8-13/9 1902 og 7-12/9 1925. Morgenbladet 2-6/1 1900 og 5-10/1 1925.	Ca. 1/2 mill. ord	PB+ MB	A	Jostein H. Hauge, NAVF's EDB-senter, Bergen

TIDSSKRIFT OG UKEBLAD:				
Farmand 1969 nr. 47,48,49,51,52		PB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UiB.
Illustrert 1968 nr 47-52 og 1969 nr 1 - 5		PB	F	
Magasinet for alle 1968 nr 47-52 og 1969 nr 1-5		PB	F	
Det Nye 1969 nr 49 og 50 1970 nr 1 - 3		PB	F	
Mitt Liv 1970 nr 2-4		PB	F	
ORDBOKMATERIALE:				
Haugen: Norsk-engelsk ordbok	60.000 oppslagsord	MB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UiB.
Avispråk: Ekserperte oppslagsord normert til en grunnform og klassifisert etter ordklasse, en tekstblokk der ordet forekommer m/kildeangivelse og stofftype	ca. 25.000 oppslagsord	MB+ PB	A	Dag Gundersen, Norsk leksikografisk inst., UiO.
ANDRE TEKSTER:				
Fortellinger fra Olav den helliges saga		PB	F	Nordisk inst., Prosjekt for datamaskinell språkbehandling, UiB.
Sverresoga (nyn.)		PB	F	
I manns minne (nyn. og bokm.)		MB	F	
Prinsessene som dansa i åkeren - eventyr frå Rogaland (nyn.)		MB	F	
Når trommene talar (nyn.)		PB	F	
Barlaams saga ok Josaphats (gno)		MB	A	Magnus Rindal, Nordisk inst., avd. B, UiB.
Håkonar Saga (gl.isl)	100.000 ord	MB	A	Marina Mundt, Nordisk inst., avd. B, UiB.
Knytlinga Saga (gl.isl)	50.000 ord	HK	A	
Islendinga Saga (gl.isl)	100.000 ord	HK	A	
150 eksamensbesvarelser i norsk stil for 9. og 10. klasse i ungdomsskolen		PB	A	Jostein H. Hauge, NAVF's EDB-senter, Berger
Aaskow: Anvisning til den rette og bestemte brug af de lægemidler hvormed de kongelige krigsskibe på deres søetog forsynes (Kbh 1778)	44.000 ord	MB+ PB	A	Øivind Larsen, Seksjon for medisinsk historie, UiO.
Bang: Medicinische Praxis (Kbh. 1791)	100.000 ord	PB	A	
Hedin: Handbok för pracktika läkare-vetenskapen (Sth. 1796)	100.000 ord	PB+ MB	A	

TALE:				
Transkripsjoner av intervju med afasi-pasienter	675 ord	MB	A	Inger Moen, Britisk inst., UiO.
Transkriberte intervjuer	50.000 ord	MB	A	Eskil Hansen, Inst. for nordisk språk og litteratur, UiO.
ARKIVMATERIALE				
HISTORIE:				
Folketelling fra Ullensaker 1865 og 1875	ca. 13.700 personer	MB+ PB/ HK	A	Ingrid Semmingsen, Sivert Langholm, Historisk inst., UiO.
Personer født i Ullensaker og bosatt i Kristiania 1865 og 1875	ca. 3.900 personer	MB+ PB	A	
Personer født i Ullensaker og bosatt i øvrige Akershusbygder 1865 og 1875	ca. 5.000 personer	MB+ PB	A	
Personer emigrert fra Ullensaker 1866 - 1899	1670 personer	MB+ HK	A	
Voteringer ved navneoppøp i Stortinget 1814-1940		MB+ HK	A	Ottar Dahl, Historisk inst., UiO
Kyrkjebok for Os (S.Bergenhuss) 1796-1800, Ordrett registrert	12.000 personoppl.	MB	I	Jan Oldervoll, Historisk inst., UiB. (Registreringen blir avsluttet i 1974) (Skal utgis):
Folketellingen av 1801 (ordrett registrert)	900.000 personer	MB	A	
Manntall fra 1663 og 1666 for Os prestegield (S.Bergenhuss)	2000 personer	MB	F	
Regesta Norvegica, 2 bind	3 mill. tegn	MB+ OPT	I	
Emigrasjonsprotokoll for Bergen havn 1895-1905	3 mill. tegn	MB	A	
Priskuranten for Bergen Børs 1815-1880	200.000 tegn	MB	A	
Biografiske data om Stortingets medlemmer 1814-1973	2600 repr. + 8000 repr. perioder	HK+ PL	A	Kjell Sælen, S. Rokkan, H. Valen, Sosiologisk inst., UiB.
Biografiske data om norske riksråder 1814-1973	ca. 370 riksråder	HK+ PL	A	
KULTURHISTORIE:				
Arkitektur og husbunad på Vestlandet (bare Hordaland og Sogn og Fiordane foreligger nå)	ca. 10.000 funn	PB	A	P. Gjærder, Historisk museum, avd. B, UiB.
ARKEOLOGI:				
Lokaliseringsdata vedr. arkeologisk materiale fra Bryggen i Bergen	ca. 10.000 funn-nr. er reg.	PB	F	A. Herteig, Middelaaldersamlingen, Historisk Museum, UiB.

MEDISINSK HISTORIE:				
Pasientprotokollen (1863-1874), Smaalenes Amtssygehus		MB	A	Øyvind Larsen, Seksjon for medisinsk historie, UiO.
<u>MUSIKK</u>				
Fartein Valens korverk (ukompl.)	ca. 1200 kort	HK	F	Arvid Vollsnes, Inst. for musikkvitenskap, UiO
64 barnesanger	4.600 kort	HK	A	Petter Henriksen, Inst. for musikkviten- skap, UiO.
3 Vinje-romanser av Edv. Grieg + 3 romanser fra F. Schuberts 'Die schöne Müllerin'	2.400 kort	HK	A	
Polsdanser fra Norge, Sverige og Finland	20.000 noter	HK+ MB	A	Tore Simonsen, Musikkvitenskapelig inst., UNIT.

Nordiske forskarar på kurs i datamaskinell språkbehandling i Bergen

av Kolbjørn Heggstad

Innom ramma av den nordiske kulturavtalen er det oppretta samarbeidsorgan som aktivt skal støtte forskarutdanninga i Norden. Frå 1960-talet har det vore arrangert kurs der forskarar frå dei nordiske landa har fått høve til å møtast og delta i vidareutdanningskurs med internasjonalt kjende forskarar som forelesarar.

Innan datamaskinell språkvitenskap har det vore relativt liten aktivitet i Norden, berre nokre få stader kan ein tale om har hatt eit arbeidsmiljø. Ein av dei første stadene der ein tok opp datamaskinelle metodar innan språkvitenskap var ved Gøteborg Universitet med professor Sture Allén som den leiande. Det var derfor naturleg at professor Allén, Avdeling for språkleg databehandling, tok initiativet til det første forskarkurset. Kurset blei halde i august 1972 i Marstrand, og var på alle måtar eit vellukka kurs som var lagt opp med forelesningar om programmeringsspråk, datastrukturar og språkvitenskaplege forskningsoppgåver. Lærarar var professor Martin Kay, University of Irvine, California,



Arrangøren av Nordisk sommerskole i datamaskinell språkbehandling, førsteamanuensis Kolbjørn Heggstad t.v. sammen med foreleserne på kurset: prof. Martin Kay og prof. John C. Brown t.h.

og instructor Ronald Kaplan, Harvard University, og professor Sture Allén. Professor Allén foreleste særleg om leksikologiske forskningsoppgåver.

I ein rundbordsdebatt på slutten av Marstrand-kurset kom det fram sterke ønskje om eit framhald av denne form for nordiske forskarkurs i datamaskinell språkvitskap. Under skrivne tok derfor initiativet og søkte Styret for Nordisk sommarskole for forskarutdanning om å få arrangementet av eit nytt kurs i same emne lagt til Universitetet i Bergen 1973. Resultatet av søknaden var positivt og kurset kunne haldast i Bergen 13. - 25. august. Lærarar ved kurset var professorane Martin Kay og John Seely Brown, begge frå University of Irvine, California. Professor Martin Kay hadde ansvaret for den generelle delen av kurset og gav ei rekkje oversynsforelesningar. Professor John Seely Brown foreleste over datamaskinell semantikk.

På grunn av at kurset for ein stor del skulle leggjast opp for gruppearbeid, blei øvre grense for deltakarantal sett til 25. Det var om lag 60 som søkte om opptak. I innbydinga var det gjort merksom på at kurset primært var for språkforskarar som tidlegare hadde fått ei elementær innføring i DB. Dette kurset var lagt opp med praktiske øvingar der ein skulle ta opp relevante

lingvistiske problemstillingar og ein kunne såleis ikkje avsetja tid til elementærkurs i DB.

Det var under planlegginga av kurset innhenta løyve frå Norges allmenvitenskapelige forskningsråd til å få etablert eit samarbeid med NAVF's EDB-senter for humanistisk forskning om gjennomføringa av den delen av kurset som ville kreve datamaskinkapasitet. Ved førebuinga og under kurset hadde ein tilgang til senterets dataanlegg, PDP 11/20, og ein disponerte to terminalar herifrå. I tillegg hadde deltakarane også tilgang til Universitetets UNIVAC 1110 og tre terminalar. Størst bruk gjorde ein av NAVF's dataanlegg, der ein hadde utarbeidd og lagt til rette fleire spesielle programsystem til bruk ved kurset og kopla tre av terminalane til.

Dette kurset var lagt opp slik at ved sida av å gje teoretisk undervisning, ville ein prøve å aktivisere deltakarane med terminaløvingar og andre arbeidsoppgåver. Både Kay og Brown la derfor opp til praktiske demonstrasjonar og øvingar gjennom sine forelesningar. Deltakarane var inndelte i 5 arbeidsgrupper som kvar hadde sine hjelpeinstruktørar ved terminalane.

Etter ei veke med forelesningar og øvingar blei det halde eit

helge-seminar, der lærarar og deltakarar saman diskuterte seg fram til kva prosjektopplegg som ein i den siste delen av kurset skulle arbeide med. Siste veka blei såleis ei veksling mellom forelesningar, framdriftsmøte for prosjekta og diskusjonar om DB-arbeidet i arbeidsgruppene. Arbeidsdagen var frå kl. 0900 til kl. 1700. I grupperomma og NAVF's dataanlegg med 3 terminalar var tilgjengelege til ca. kl. 2400, og mange av deltakarane nytta seg av dett

Kurset i Bergen 1973 må delvis vurderast saman med kurset i Marstrand i 1972, og ein hadde også bygd på dei røynslene ein fekk i Marstrand. Det særleg ved Bergens-arrangementet var den store vekt ein hadde lagt på det å skape ein situasjon der lingvist kunne få direkte kontakt med programmering og anna DB-arbeid. Skal humaniora få praktisk nytte av DB-teknikken, trengst det at ein del aktive forskarar innan dei ulike områda er i stand til sjølvstendig å ta stilling til kva for hjelp dei kan få. Berge kurset 1973 kunne berre ha ein initierande verknad. Mange av deltakarane fekk der for første gong høve til å prøve DB i fagleg samanheng.

Eit slikt 'workshop'-opplegg som ein denne gongen hadde, krev svært nøye planlegging. Med hovudlærar

frå USA og dermed heller tungvint kommunikasjon, kunne det lett ha blitt vanskeleg å gjennomføre etter planen. Når ein likevel fekk eit tilfredsstillande resultat, må ein tilskrive det den store røynsla desse amerikanske universitetslærarane hadde med denne arbeidsforma og den uvanleg store innsatsvilje og evne til samarbeid dei lokale programmerarane under leiding av sjefskonsulent G. M. Gillow ved NAVF's EDB-senter, synte.

I den teoretiske delen av kurset konsentrerte ein seg særleg om semantikk denne gongen. Dette er ein del av lingvistikken som saman med 'artifisiell intelligens' meir og meir opptar forskarar, og ikkje minst forskarar innan datamaskinell lingvistik. I Norden har det vore liten aktivitet innan datamaskinell semantikk, og på kurset i Bergen kunne ein ikkje gje meir enn eit grovt oversyn.

Kurset i Marstrand og Bergen utfylte i stor monn kvarandre, og tilsaman har desse etter underskrivne si meining vore med å leggje eit godt allment grunnlag i datamaskinell lingvistik for ei gruppe aktive forskarar i Norden. Det var god geografisk spreiding på deltakarane og dei representerte ei rad ulike deldisiplinar innan lingvistikken.

Datamaskinell språkvitskap må enno ei stund karakteriserast som eit nytt arbeidsområde som er i stadig utvikling. I ein slik etableringsfase er det særleg viktig at dei som alt arbeider innan området eller står i ferd med å setje nye prosjekt igang, har god kontakt seg i mellom og held seg nøye orientert om det som hender andre stader.

Det er derfor svært positivt at Styret for Nordiske sommarskolen for forskarutdanning no har gjort kjent at det også i 1974 vil bli arrangert eit tilsvarande kurs. Arrangøren denne gongen vil bli universitetslektor Bente Maegaard, Institutt for anvendt og matematisk lingvistik, Københavns universitet. Informasjon om dette kurset vil bli sendt til universiteta med det første.

?

TRENGER DE ASSISTANSE AV
EDB-KONSULENT?

NAVF's EDB-senter for humanistisk forskning har blant sitt personale medarbeidere som har til oppgave å yte konsulentassistanse til forskere i de humanistiske fag.

Konsulentene har erfaring fra EDB-arbeid i ulike humanistiske disipliner og vil kunne drøfte EDB-opplegg med humanistiske EDB-brukere og yte konsulentassistanse i de ulike faser av arbeidet.

Deres navn og adresse er:

Ivar Fønnes, Historisk institutt,
Universitetet i Oslo,
Postboks 1102-Blindern,
OSLO 2.
Telefon: 46 68 00,
lokal 9919.

✱

Geir Berge, NAVF's EDB-senter for
humanistisk forskning
v/Univ. i Bergen,
Postboks 53,
5014 Bergen-Univ.
Telefon: 21 00 40,
lokal 3240.

I samarbeid med Universitetet i Trondheim har NAVF opprettet en lokal konsulenttjeneste i Trondheim i 1974. Konsulenten vil tiltre ved slutten av vårsemesteret. Nærmere opplysninger om konsulenttjenesten vil bli gitt i neste nummer av HUMANISTISKE DATA og gjennom separat kunngjøring ved Universitetet i Trondheim.

Datamaskinen i historieforskninga

Rapport frå ein konferanse i Uppsala — juni 1973 av Jan Oldervoll

I tidsrommet 24. - 30. juni 1973 vart det skipa til ein konferanse i Uppsala over emnet 'Computers In History'. Konferansen hadde mellom 30 og 40 deltakarar frå ei rekkje land i Europa og frå Nord-Amerika.

Historikarar har etter kvart teke datamaskinen i bruk på ei rekkje forskjellige område. For ikkje å få for stor tematisk sprelling, konsentrerte konferansen seg om bruk av datamaskin og kvantitative metodar i økonomisk og sosial historie. Av konkrete emne som vart tekne opp, kan ein nemna: Analyse av hushaldstruktur, bruk av datamaskin i historisk geografi, metodiske problem i samband med databehandling av folketeljingsmateriale.

Eit gjennomgåande tema i diskusjonen av både desse emna og andre, var dei problema som oppstår på grunn av at dataførebuing er svært kostesam. Prekodning vart for det meste sett på som lite tenleg. Ved forskjellige institusjonar er det dei siste åra utvikla system for

meir eller mindre automatisk koding av skjematisk materiale.

Det som kjennetelknar alle desse systema, er at dei krev punching av helle den teksten analysen skal basera seg på. Hovudprinsippet i kodinga er også den same. Men dei praktiske utforminga er likevel svært forskjellig frå prosjekt til prosjekt, alt etter kor kompleks materiale ein har for seg. Eit system kan åleine knapt dekkja dei ulike behov prosjekta har.

Eit anna problem som stod sentralt, var dei vanskar ein har med å samordna eksisterande forskningsprosjekt med same problemstilling, men med utgangspunkt i forskjellige geografiske område. Her støyer ein på problem både av teknisk og av kjeldemessig art.

På konferansen diskuterte ein også i kor stor grad ein kunne lita på dei ymse statistiske testar historikarar og andre samfunnsvitskapsfolk brukar. Her kom det ikkje til noko avklaring.

Det siste problemet kan på mange måtar seiast å vera typisk for konferansen. Det vart lagt fram for konferansen ei rekkje rapportar om ymse forskningsprosjekt der datamaskinen stod meir eller mindre sentralt. Men på grunn av at dei fleste deltakarane først og fremst var historikarar og ikkje datamaskinfolk, kom dei historisk-metodiske problema til å stå i framgrunnen, ikkje dei som galdt bruk av datamaskinen. Dette vil ein prøvje å retta på til neste gong ein skal skipa til ein slik konferanse. Det kan mellom anna gjerast ved å gjer emnefeltet for konferansen vesentlig trongare.

Elles var det overmåte nyttig å få diskutera gjennom den måten både ein sjøl og andre gjer ting på, både for å stadfesta at den måten ein arbeider på, er fruktbar, og for å få impulsar til brige og til nye måtar å bruka datamaskinen på.

Folkrörelse-prosjektet i Uppsala, med professor Carl Göran Andrä i spissen, var eit framifrå vertskap

Datamaskinen i arkeologisk forskning

Rapport fra en konferanse i Birmingham 5. og 6. januar av Asbjørn E. Herteig

Som et resultat av den stadig økende interesse for bruken av EDB-teknikker i analysen av arkeologiske data ble det i 1972 arrangert en konferanse ved EDB-senteret til Universitetet i Birmingham.

Denne konferansen var så vellykket at man satte seg som mål å arrangere tilsvarende sammenkomster hvert år.

En fullstendig rapport fra konferansen i 1972: *Computer Applications in Archaeology* er trykt i *Science and Archaeology* 9, George Street Press, Stafford, 1973.

Konferansen i Birmingham var interessant på flere måter, kanskje først og fremst ved å understreke den veldige utvikling som har funnet sted de senere år i forholdet mellom arkeologer og EDB-spesialister. Ved de konferanser som ble arrangert ut over 60-årene var det også deltakere fra en rekke forskjellige disipliner; av matematikere, statistikere, EDB-folk og humanister, men sistnevnte gruppe og spesielt arkeologene, var jevnt

over statistiker i en dialog som nok gikk de fleste langt over hodet.

Til manges forundring ble det blant matematikere og statistikere registrert en nesten påfallende sterk interesse for arkeologiske problemer. De fant nemlig i arkeologien en kilde med veldige datamasser som det kunne øses av med sikte på å teste egne metoder og teknikker. Men det var i de aller fleste tilfeller teorier og teknikker som ikke kunne anvendes i praksis. Annerledes nå ved den konferansen som ble holdt 5. og 6. januar i Birmingham. Det var fremdeles et betydelig og nødvendig innslag av EDB-ekspertise og statistikk, men det arkeologiske aspekt slo sterkere igjennom, og det var i hovedsaken tale om anvendbare teknikker og metoder. En rekke av foredragene bygget direkte på analyser av arkeologiske informasjoner, fra registrering av gravningsdata, EDB-baserte fredningsregistre, generelle søkemetoder for gjenfinning av data til statistiske metoder og modeller, herunder forskjellige numeriske metoder for

evaluering av jordundersøkelser som undertiden innleder eller under spesielle forhold helt erstatter arkeologiske gravninger.

Særlig interesse knyttet seg til en redegjørelse for en del spesielt avanserte muligheter som EDB-teknikken nå stiller til disposisjon for så vel 'skrivebord'- som 'feltarkeologen'.

Konferansen syntes ellers å bekrefte at England nå er kommet for fullt på dette felt innen arkeologien. Frankrike som var meget tidlig ute med avanserte tilbud, synes å gå sine egne veier, mens tyske arkeologer foreløpig holder seg mer reservert i bakgrunnen.

Blant de mange prosjekter det arbeides med fra engelsk hold, er en standardisering av klassifikasjonsmetodikk for alle engelske museer. Det kan ventes et forslag til en internasjonal standard fremlagt allerede til sommeren i forbindelse med International Council of Museums (ICOM's) generalkonferanse i København.

Optisk lesing og datamaskinstyrt fotosetting

inntrykk fra en studiereise — av Geir Berge

Undertegnede oppholdt seg noen dager i Stockholm og Uppsala i juli 1973 for å samle informasjon om datastyrt fotosetting og optiske lesere. Jeg vil her forsøke å gi et overblikk over de erfaringer som ble gjort på denne studiereisen.

1. FOTOSETTING

Teknisk oversikt.

Det eksisterer to hovedtyper av slike settere:

- a) Matrisesettere (f.eks. Linotron 500).
- b) Digitalsettere (f.eks. Digiset 40T2).

Begge typer tar som inndata magnetbånd, som foruten teksten som skal settes, inneholder operasjonskoder for setteren, så som opplysninger om store eller små bokstaver, kursiv osv. For matrisesettere finnes karaktersetten på gjennomskinnelige plater (grids), og ved å belyse rett plate på riktig sted og velge korrekt forstørrelse, får en det ønskede tegn på filmen. Det er klart at mengden av tegn som det er mulig å sette med denne teknikken, er begrenset av tegnsettet på de ek-

sterende plater. Skal tegnsettet utvides, må en lage nye plater med de ekstra karakterene som er ønskelig, og dette faller relativt kostbart.

Settere av den digitale typen arbeider etter helt andre prinsipper. Her lagres informasjonen om karaktersetten digitalt, vanligvis på et platelager. Denne digitale informasjonen brukes av en liten datamaskin til å generere en karakter med ønsket utseende ved hjelp av et katodestrålerør. Dette bildet overføres så til filmen. Denne teknikken gir full frihet når det gjelder oppbygningen av det karaktersetten som brukes. En kan f.eks. lage en karakter, som når den hentes frem, gir et komplekst fotografisk bilde.

Ved siden av bruk av denne teknikken i vanlig tekstfremstilling, kan en tenke seg at et anvendelsesområde kan være fremstilling av notebilder. Det synes mulig å generere et karaktersetten som inneholder de nødvendige 'byggesteiner' for dannelse av et notebilde, for så å la setteren produsere et partitur på samme måte som ved setting av vanlig tekst. Jeg kjenner ikke til om dette har vært prøvd, men det synes å være et

felt hvor en relativt stor økonomisk gevinst kunne oppnås.

Fordeler ved bruk av fotosetting. Forenklet kan en si at prosessen ved fotosetting er:

- a) Manuskript + operasjonskoder
---> Magnetbånd 1 + klartekst for korrekturlesing.
- b) Magnetbånd 1 + korreksjoner
---> Magnetbånd 2 (+ klartekst for evt. korrekturlesing).
- c) Magnetbånd 2
---> Fotosatt film (evt. microfiche eller mikrofilm).

Økonomi.

For kunden vil setteprosessen koste omtrent det samme for datastyrt fotosetting som setting med tradisjonelle metoder dersom trykkeriet skal utføre både operasjonene a, b og c. En kan derimot oppnå en betydelig kostnadsreduksjon dersom en overleverer trykkeriet et felfritt magnetbånd hvorfra en direkte får fotosatsen. Med andre ord: Kunden utfører selv arbeidet som faller under a og b, og lar bare trykkeriet ta hånd om c.

Som talleksempel kan nevnes at vit. ass. Jan Oldervoll, Historisk inst., Universitetet i Bergen med denne metoden oppnådde å få satt en bokside for ca. 10 kr. mot ca. 80 kr. ved tradisjonelle metoder under arbeidet med settingen av et register over historiske dokumenter (Regesta Norvegica).

En må riktignok fra brutto besparelse trekke utgifter forbundet med pkt. a og b, som omfatter punching av manus, og nødvendig programmering for å kunne få med operasjonskodene for styring av fotosetteren, samt utgifter til maskintid. Til tross for dette kan en regne med en netto besparelse som vil være større for kvantitativt store arbeider. En annen faktor som må tas med, er at kunden for senere prosjektarbeid allerede har ervervet en viss programmeringskompetanse innen området, og derved kan en slippe med lavere utgifter til programmering.

Oppdatering av et innlest materiale.

Når først et grunnmateriale er punchet og innlest på f.eks. magnetbånd, er det senere relativt enkelt å foreta korreksjoner. Det kan være tale om nye innskudd, sletting av visse sekvenser, forandringer i andre osv. Den nye informasjonsmengden kan da legges på en egen file, som i sin tur flettes sammen med den gamle versjonen for å gi en ny oppdatert utgave.

Representasjonsmuligheter.

Ettersom materialet ligger på maskinleselig medium (f.eks. magnetbånd), kan en bruke standardprogram og evt. egne programrutiner for å danne sorterte lister, indekser osv. Et eksempel på slik bruk er arbeidet med Nusvensk Frekvensordbok utført i Gøteborg av gruppen 'Språkdata' under ledelse av professor Sture Allén.

Diverse.

Det bør også nevnes at de moderne fotosetterne gir mulighet for utdata i form av mikrofilm eller mikrofiche direkte. Sikkerheten mot feil under settingen er også meget god, forutsatt et feilfritt magnetbånd (2), da hele prosessen går automatisk (se ovenfor).

Til slutt vil jeg igjen peke på de mulighetene som ligger i friheten ved valg av karakterset for de digitale setterne. Så vidt vitnes er det nettopp installert en slik setter i Trondheim. I Uppsala har firmaet Almquist & Wicksell i bruk en Digiset 40T2 og var i gang med innkjøring og uttesting av denne under mitt besøk der.

Konklusjon.

Av det foregående synes det da som en kan ha størst fordeler av datastyrt fotosetting dersom ens materiale tilfredsstillende ett eller flere av følgende punkter:

1. Et stort materiale.
2. Materialet skal oppdateres og trykkes flere ganger over et tidsrom.
3. Flere 'representasjoner' (sorterte lister etc.) av materialet ønskes satt.
4. Datamaskinell behandling av materialet er ønskelig. Som eksempel her kan nevnes et arbeid ved Prosjekt for datamaskinell språkbehandling (UIB), hvor en norsk-tysk ordbok på magnetbånd blir 'snudd' for å danne grunnlag for en tilsvarende tysk-norsk ordbok.

2. OPTISK LESING

Under oppholdet i Stockholm ble to firmaer som arbeidet med optisk lesing kontaktet, AB AURIGA og AB FREDR. WAGNER.

AB AURIGA

Dette firmaet disponerer 2 lesere av typen Recognition Equipment og en Almex. Dette er kostbare maskiner med stor kapasitet, ca. 3600 tegn/sekund. For å oppnå minimalt antall feil ble det anbefalt å benytte spesialpapir for optisk lesing (OCR papir), og det ble da oppgitt en feilprosent på 1 pr. 20 - 30.000 tegn.

Leserne gjenkjente alle tegn som inngår i et EBCDIC-kodesett, og som er skrevet i en standard type, fortrinnsvis OCR/B-font. Dog kunne ikke utstyret skille mellom punktum

og komma, og videre ville de norske karakterene Ø og Å by på problemer.

Brukeren er gitt visse utslettingsmuligheter idet en med spesialkarakterer kan få slettet (ikke lest)

- (i) foregående karakter,
- (ii) 2 foregående karakterer,
- (iii) Inneværende linje.

Når det gjelder priser, opererte AB AURIGA med følgende satser i Sv.kr.:

Initialkostnader for startprogrammering fra 500 kr. og oppover.

Linjepris fra 0,005 kr.

Pris pr. produksjonstilfelle ...
150 - 200 kr.

AB FREDR. WAGNER

Dette firmaet drev foruten optisk lesing på servicebasis også salg av optiske lesere av typen Compu-Scan system 170, en leser som i prisen inkluderer en FACIT papirbåndpunch for utdata med kapasitet 150 b/s.

Leseren ble levert i to versjoner:

- (i) Lesehastigheten 100 b/s, pris 400,000 Sv. kr.
- (ii) Lesehastigheten 200 b/s pris 450,000 Sv. kr.

Leserne var ellers identiske så vidt jeg forstod.

Som inndata godtok leserne alle karakterer, fortrinnsvis i OCR/B-font, men en kunne selv definere de karakterer leseren skulle godta. Dette betyr at en selv kan bygge opp sitt private karaktersett og også forandre dette senere om dette er ønskelig. Det skulle således ikke være noe til hinder for at maskinen kunne lese f.eks. gresk tekst. Videre kunne utstyret skille mellom punktum, komma og Ø, Å etc.

Kravene til inndata var videre at de var skrevet på OCR-papir med marg større enn 1 tomme på alle kanter. En kunne heller ikke skrive tettere enn 5 linjer pr. tomme, og det ble anbefalt å bruke engangsfargebånd (FILM RIBBON) ved skrivingen.

Karaktersett for utdata var valgfritt, og en kunne som nevnt her benytte papirbånd, eller om ønskelig magnetbånd.

For lesing på servicebasis ble følgende satser i Sv.kr. brukt:

pr. kjøring	kr. 60,00.
pr. lest blankett	kr. 0.40.
pr. lest linje	kr. 0.02.
pr. lest tegn	kr. 0.005.

Med lesere av denne typen, som er relativt rimelige i innkjøp, kan en vurdere om innkjøp av en 'sentral' leser for lesing av flere gruppers skrevne materiale er aktuelt. En slik sentral leser ville da overflødiggjøre nevnte gruppers

hullbåndpuncher, idet registreringen bare trenger en standard elektrisk skrivemaskin.

Denne besparelsen må vurderes mot avskrivningsutgifter, utgifter til drift, service etc., for den optiske leseren. Det er videre vanskelig å si noe om driftsstabiliteten for en leser av nevnte type, og hvor lett eller vanskelig det evt. er å få en tilfredsstillende serviceordning etablert, men det synes likevel å være en dataregistreringsmetode som vil bli stadig mer aktuell.

Et annet alternativ for enkelte prosjekt som i dag registrerer mye data på hullbånd, kan være å gå over til 'vanlig maskinskriving' som over sendes et OCR-firma for lesing på servicebasis. Et meget løst overslag viser at en punchedames årsproduksjon ville koste av størrelsesorden 12,000 N.kr. å få lest hos AB FREDR. WAGNER. Prisen er her, som det fremgår av satsene ovenfor, avhengig av ord pr. linje, linjer pr. side og antall sider som oversendes for lesing pr. gang, et beløp som må vurderes mot de utgifter en papirbåndpunch medfører.

Konklusjon.

Av det over nevnte fremgår det at optiske lesere med en rimelig kapasitet allerede i dag tilbys til konkurransedyktige priser, og det er sannsynlig at utviklingen vil gjøre denne formen for dataregistrering stadig mer attraktiv.

Arbeidet i senteret høsten 73 og vinteren 74

En oversikt over senterets virksomhet frem til høsten 1973 er gitt i HUMANISTISKE DATA nr. 1, 1973. Som omtalen der viser, har senterets personale tatt fatt på en rekke arbeidsoppgaver f.eks. innen programutvikling som er av en slik art at de strekker seg inn i den perioden som denne rapporten gjelder.

a) OPPLYSNINGSVIRKSOMHET OG KONSULENTTJENESTER

Senteret har i meldingsperioden ført videre sitt arbeid med å orientere forskere om databehandling av humanistisk forskningsmateriale og med å gi konsulenthjelp for prosjekter som er i ferd med å utforme EDB-opplegg.

Det er f.eks. holdt kontaktmøter i senteret med representanter for 'Sekretariatet for registrering av faste kulturminner i Norge' og med Norsk Kulturråds fotoutvalg hvor emner i tilknytning til EDB-registrering av tiltakenes materiale har vært drøftet.

Senterets konsulenter har gitt hjelp til redaktørene i Norsk Stadnamleksikon og Norsk Landbruksordbok i forbindelse med de pågående drøftinger om overføring av ordboks-

materialet til maskinleselig form. I dette samarbeidet har særlig spørsmålet om bruk av EDB i trykningsarbeidet stått sentralt.

Senterets leder presenterte senteret på NAVF's konferanse for engelsk, tysk og romansk språkvitenskap i Sarpsborg høsten 1973.

Ellers har senterets personale i Oslo og Bergen deltatt i en rekke samtaler med enkeltforskere og prosjektmedarbeidere om bruk av EDB i deres arbeid.

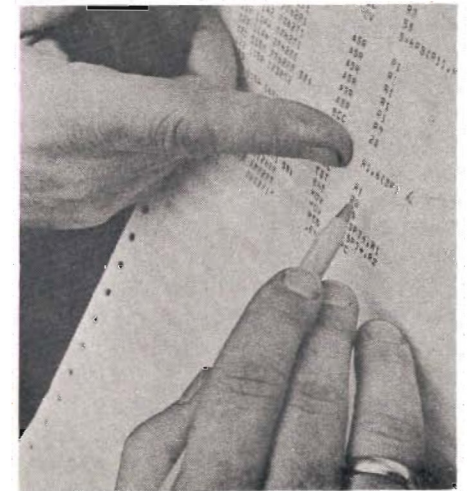
b) KURS- OG SEMINARVIRKSOMHET

Etter initiativ fra senteret ble det i høst arrangert en rekke sammenkomster for EDB-interesserte i Bergen. På møtene drøftet en særlig metoder for å klassifisere språklige enheter og legge kodede opplysninger av faglig art til et språklig forskningsmateriale ('tagging'). De terminalbaserte programmene som senteret har utviklet for slike formål, ble demonstrert.

I november arrangerte senteret et introduksjonsseminar om bruk av EDB Innen humaniora for forskere og studenter ved Universitetet i

Trondheim. For demonstrasjonsformål var det etablert terminalforbindelse til senterets datamaskin i Bergen, og programmer for interaktiv tekstbehandling ble presentert og drøftet.

Etter invitasjon fra senteret besøkte professor Sture Allén fra Gøteborgs Universitet Universitetet i Oslo i november hvor han ledet et seminar vjet virksomheten ved Avdelingen for språklig databehandling i Gøteborg. Foruten å redegjøre for pågående arbeid orienterte professor Allén om hovedtrekkene i den nye forskerutdanningen i språklig



databehandling ved dette universitetet.

Senterets konsulent i Oslo har i vinter ledet en serie øvelser i programmeringsspråket LISP hvor en også har foretatt innledende eksperimenter med automatisk syntaktisk analyse av norske setninger. Det kan ellers nevnes at senteret i perioden har lagt planer for et kortere seminar om bruk av EDB ved litterære analyser. Foreleser på seminaret, som vil bli arrangert slutten av april, blir professor Stephen Waite, Department of Classics, Dartmouth College, Hanover, USA. Flere forskere som arbeider med selvstendige oppgaver og medarbeidere knyttet til prosjekter, har fått opplæring i å nytte senterets datamaskin til redigering av sitt maskintilgjengelige forskningsmateriale, som i de fleste tilfeller består av tekster.

c) PROGRAMUTVIKLING

Den delen av programmeringsarbeidet som er knyttet til senterets datamaskin, har i denne meldingsperioden hovedsakelig vært drevet på to hovedområder - utviklingsarbeid innenfor interaktiv tekstbehandling ('tagging-system') og arbeid med å skape et hensiktsmessig kommunikasjonsopplegg mellom PDP 11/20 og UNIVAC 1110 ved Universitetet i Bergen (se også HUMANISTISKE DATA nr. 1, s. 8 og oversikt over programmer i dette nummer). En metodisk beskrivelse av

programmeringsarbeidet på disse felt vil bli gitt i et senere nummer av HUMANISTISKE DATA.

Senteret har i denne perioden også utviklet programmer i tilknytning til et prøveprosjekt for Historisk museum, Universitetet i Bergen hvor prøvematerialet består av arkivopplysninger for en del av museets steinaldermateriale. Data er registrert på spesielt utarbeidede avmerkingskort ('mark sense card') og overført til magnetbånd med bistand fra Psykologisk Institutt. Senteret har fra magnetbånd tilrettelagt materialet på sin egen datamaskin for de testoppgavene, i første rekke sorteringer, som i løpet av våren vil bli utført i samarbeid med Historisk museum.

Arbeidet på UNIVAC 1110 har vært konsentrert om programutvikling i forbindelse med databehandling av greske tekster og arkeologisk funnmateriale fra Bryggen i Bergen.

Det er videre utarbeidet konkordansprogram for gammelengelske tekster og program for sortering av feiltyper i et tekstmateriale som består av Bergens-elevs frie stiler på tysk.

Senterets konsulent i Oslo har arbeidet videre med å overføre sentrale analyseprogrammer for historisk forskning til bruk på CYBER-anlegget, i første rekke HISO-

systemet (programsystem for behandling av rubriserte protokolldata i tekstform) og voteringsanalyseprogrammer. Dessuten har konsulenten brukt en vesentlig del av sin tid til å vurdere hvilke muligheter CYBER-anleggets standard programutrustning gir for å løse humanistiske databehandlingsoppgaver.

d) PROSJEKTSAMARBEID

En stor del av det utviklingsarbeid som er omtalt i punktet ovenfor, har direkte tilknytning til prosjektsamarbeid som senteret har påtatt seg. Som det går frem av oversikten over EDB-virksomheten annetsteds i dette nummeret, har senteret samarbeid i gang med en rekke enkeltpersoner og brukergrupper. Senterets personale i Bergen har således flere samarbeidstiltak i gang innenfor tekstbehandling hvor både moderne norsk, norrønt, tysk, russisk og gresk materiale blir tilrettelagt eller analysert. I tillegg er et samarbeid i gang om EDB-opplegg for systematisert materiale, i første rekke arkeologisk og kulturhistorisk arkivmateriale.

Senterets konsulent i Oslo samarbeider med forskere fra flere språk-institutter som på egen hånd eller ved bistand fra konsulenten, benytter tekstbehandlingssystemet TEXT. Dessuten har konsulenten et utstrakt samarbeid med de EDB-prosjekter innen historie som blir drevet ved Universitetet i Oslo.

IFIP CONGRESS 74

INTERNATIONAL FEDERATION FOR
INFORMATION PROCESSING

STOCKHOLM 5. - 10. AUGUST 1974

I tiden 5. - 10. august 1974 vil det bli holdt en internasjonal kongress om informasjonsbehandling i Stockholm.

IFIP arrangerer hvert 3. år internasjonale kongresser og har som målsetting:

'to promote information science and technology;
to advance international cooperation in the field of information processing;
to stimulate research, development and application of information processing in science and human activity;
to further the dissemination and exchange of information on information processing;
to encourage education in information processing.'

På kongressen i Stockholm vil det bli holdt 41 foredrag av inviterte forelesere og 200 påmeldte presentasjoner. Det vil bli arrangert 10 paneldiskusjoner.

Nedenfor siteres emneområdene fra kongress-folderen:

Computer Hardware and Architecture Software.
Mathematical Aspects of Information Processing.

Technological and Scientific Applications.

Applications in the Social Sciences and the Humanities.

System for Management and Administration.

Social Implications of Computers.

Som en vil se, er det en egen seksjon for EDB-anvendelse i de humanistiske og samfunnsvitenskapelige fag.

Det kan nevnes at professor Joseph Raben er invitert til å gi en forelesning om 'Computers and the Humanities'. Professor Raben er redaktør av tidsskriftet med samme tittel som foredraget.

De som er interessert i nærmere opplysninger om kongressen, kan få fylldig informasjon ved å henvende seg til:

IFIP Congress 74,
c/o Stockholm Convention Bureau,
Strandvagen 7 c,
S-114 56 STOCKHOLM
Sweden.
Telefon: 08/63 04 45

SOMMERSKOLE I DATABEHANDLING I ITALIA

PISA 12. AUGUST - 5. SEPTEMBER 1974

'La logique, le semantique, et les ordinateurs.'

Nærmere opplysninger hos:

Professor A. Zampolli,
CNUCE,
Via S. Maria 36,
56100 PISA,
Italia.

SPALTEPLASS I HUMANISTISKE DATA

Vi vil her nytte høvet til å fornye det tilbudet om spalteplass som ble fremsatt i det første nummeret av HUMANISTISKE DATA.

Tilbudet har foreløpig ingen avgrensning så sant emnene faller innenfor det arbeidsområdet som NAVF's EDB-senter er en eksponent for.

Av ulike typer bidrag kan nevnes: artikler i forbindelse med seminar og kongresser, erfaringer fra EDB-arbeidet i bestemte prosjekter eller generelle omtaler av slike, orientering om litteratur en har funnet verdifull eller om programprodukter som er utviklet - ute eller i vårt land.

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.

Lay-out og trykk:

Universitetets trykkeri

Bergen

SOMMERSKOLE I LITTERÆR STATISTIKK

CAMBRIDGE, ENGLAND - 13. - 19. JULI
1975

I 1975 vil Association for Literary and Linguistic Computing søke å arrangere en sommerskole i litterær statistikk.

Fra ALLC's Bulletin sakses følgende:

'There will be three types of sessions: main lectures, guest lectures, tutorials. The main lectures will be given by Mr. Norman Thomson (IBM) and will include estimation and confidence intervals, design of experiments, analysis of variance. Professor Joseph Raben (CUNY) has been invited to give the keynote lecture on the Sunday evening immediately preceding the course; other guest lecturers invited are the Rev. A. Q. Morton (Scotland), Professor R. W. Bailey (America), and Professor Y. T. Radday (Israel). In addition to tutorial sessions for exercises based on the lectures, there will be discussions on the statistical validity of the work carried out, and practical demonstrations.'

De som ønsker å bli holdt orientert om planene fremover, bes melde sin interesse til:

Mr. M. H. T. Alford
University of Cambridge
Literary and Linguistic Computing
Centre,
Sidgwick Avenue,
CAMBRIDGE CB3 9DA,
U.K.

MEDARBEIDERE I DETTE NUMMERET:

GEIR BERGE er konsulent ved NAVF's EDB-senter for humanistisk forskning.

KOLBJØRN HEGGSTAD er førsteamanuens ved Nordisk institutt, Universitetet i Bergen og leder 'Prosjekt for datamaskinell språkbehandling'. Medlem av NAVF's EDB-komite for humanistisk forskning i en rekke år.

ASBJØRN E. HERTEIG er førstekonseruator ved Middelaldersamlingen, Historisk museum, Universitetet i Bergen. Han er prosjektleder for de arkeologiske undersøkelser på Brygge i Bergen.

ARNE B. JOHANSEN er prosjektleder for 'Hardangerviddaprojektet for tverrvitenskapelig kulturforskning'. Han har i flere år vært medlem av NAVF's EDB-komite for humanistisk forskning.

JAN OLDERVOLL er vitenskapelig assistent ved Historisk institutt, Universitetet i Bergen, hvor han særlig arbeider med instituttets EDB-prosjekter.