humanistiske data

Meldingsblad for
NAVF's EDB-senter
for humanistisk forskning

Norges Almenvitenskapelige Forskningsråd

NR
1-2
1975

Innhold:

Redaktørens spalte

Dette nummer av Humanistiske Data kommer ut sterkt forsinket, og fører til at det i 1975 ikke blir mulig å gi ut mer enn ett nummer. Nå er kanskje ikke dette noen stor tragedie. Erfaringene til nå har vist at vi kanskje startet med litt for optimistiske forhåpninger om den skrivetrang som bladet ville utløse. NAVF's EDB-senter driver for sin del også informasjonstjeneste gjennom andre kanaler enn meldingsbladet — i første rekke gjennom sine konsulenttjenester i Oslo, Trondheim og Bergen.

På den annen side mener vi å ha fått bekreftet at det finnes et reelt informasjonsbehov som dette meldingsbladet kan dekke, og at det på sin måte kan bidra til øket kontakt mellom EDB-interesserte humanistiske forskere.

Når vi har kalt bladet et meldingsblad er det også for å kunne stå fritt til å servere stoffet som «blandet drops» — med skiftende blandingsforhold. Det vises i dette nummeret ved at vi både har med stoff av EDB-teknisk karakter, orientering om faglige emner innenfor humanistiske fag og diverse typer meldinger av interesse for våre EDB-miljøer.

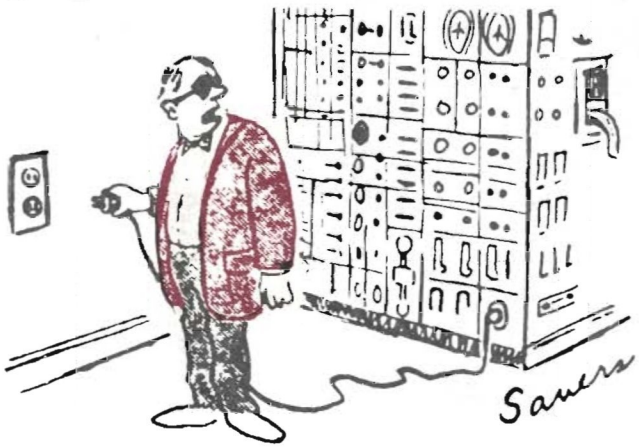
Når vi diskuterer hvordan en best kan legge forholdene til rette for en formålstjenlig bruk av datamaskin i humanistisk forskning, kommer en snart inn på spørsmål som gjelder opplæring og utdanning. Generelt formulert er ofte problemene stilt slik: Er det en forutsetning at en humanistisk forsker som i sitt arbeid ønsker å bruke EDB, må lære seg å programmere selv? Hvis han ikke selv skal lære å programmere, men bare skaffe seg en allmenn orientering om databehandling som bakgrunn for sin omgang med EDB-personale, hvordan skal han da best erverve den nødvendige innsikt? Hvilke undervisningstilbud finnes i universitetsmiljøene for den humanist som ønsker å sette seg inn i og utnytte datamaskinelle metoder?

I dette nummeret er det to bidrag som særlig tar opp slike emner: Gjestespalten denne gangen er et intervju med professor Svein Nordbotten, Universitetet i Bergen, om fagstudiet informasjonsvitenskap, som i lærerplaner og undervisningsemner også direkte er rettet mot å gi en bakgrunn for databehandling innenfor humanistiske fag. I en oversiktsartikkel gjennomgår konsulent Roald Skarsten en del litteratur som enten kan gi en introduksjon til databehandling eller gi et innblikk i EDB-arbeid innen humanistiske fag.

Andre steder i bladet finnes enkeltbidrag med orientering om pågående prosjektarbeid i Norge.

Spørsmål om utdanning og opplæring både i organisert form og i form av egenopplæring, er et sentralt emne som vi vil forsøke å behandle noe videre i neste nummer av Humanistiske Data.

Ikke minst er det aktuelt å spørre dem som selv har tatt EDB i bruk om deres syn for eksempel på spørsmålet om humanisten selv må lære seg å programmere eller om en med tilfredsstillende resultat kan støtte seg på dyktige EDB-konsulenter.



"Without me, you're a **nobody**, see!"

Gjestespalten

INFORMASJONSVITENSKAP

Professor Svein Nordbotten svarer på spørsmål fra Humanistiske Data.

Professor Svein Nordbotten, du er professor i informasjonsvitenskap ved Universitetet i Bergen. Kan du gi en kort beskrivelse av ditt fagområde?

Informasjonsvitenskap er studiet av informasjonens generelle kjennetegn, hvordan den oppstår, representeres, oppbevares, formidles og utnyttes, hvordan informasjonsstrukturer og informasjonsprosesser formaliseres og systematiseres, og hvordan moderne metodologi og teknikk, herunder elektronisk databehandling og telekommunikasjon, kan utnyttes for effektiv samling og utnytting av kunnskap.

Er dette et nytt fagfelt ved universitetene? Finnes det parallelle fag ved de andre universitetene i vårt land eller i utlandet?

Det kan svares både ja og nei på dette spørsmålet. I den form vi har gitt faget her i Bergen, er det relativt nytt og finnes ikke ved mange universiteter. Men som de fleste nye fag bygger også informasjonsvitenskap på eldre komponenter som har vært tatt opp ved mange universiteter. Jeg kan her nevne emner som systemteori, datamaskinorienterte emner og anvendelse av datamaskinen som et verktøy i numeriske beregninger.

Hvilken plass har databehandlingsemner i faget?

Elektroniske datamaskiner får en stadig større betydning når det gjelder å samle og utnytte kunnskap. Deres egenskaper og muligheter er derfor et sentralt studieobjekt innen vårt fag. I vårt studieopplegg for grunnfagstudenter legger vi stor vekt på at studentene skal lære å bruke datamaskinen på en fornuftig måte ved løsning av ulike typer av problemer. Senere i studiet får studentene velge hvilken plass de vil gi de tekniske hjelpemidler.

I Bergen er informasjonsvitenskap et fag felles for Det samfunnsvitenskapelige fakultet og Det historisk-filosofiske fakultet. Hva er grunnen til det?

Historisk ble maskinell databehandling som fag, først tatt opp ved de matematiske institutter i tilknytning til numeriske beregninger. Etterhvert oppdaget også andre at datamaskinen kunne nyttes til mer enn å addere og subtrahere numeriske data. Forskere fra mange disipliner oppdaget at datamaskinen generelt kunne nyttes for å samle, systematisere, lagre og behandle symbolske data etter programmerte regler. Ved Universitetet i Bergen så humanister og samfunnsvitenskapelige forskere på et tidlig tidspunkt datamaskinenes muligheter som verktøy for å ta vare på og behandle store tekster og masser av samfunnsvitenskapelige data. På det tidspunkt var humanister og samfunnsvitere i Bergen samlet i ett fakultet og de gikk derfor sammen om å få etablert faget informasjonsvitenskap. Selv om samfunnsvitere senere har dannet sitt eget fakultet, er faget informasjonsvitenskap i Bergen blitt et felles fag for de to fakulteter.

Men finnes det områder hvor de to fakultetene har ulike interesser i relasjon til faget informasjonsvitenskap?

Jeg vil tro at de to fakulteters interesse med hensyn til faget informasjonsvitenskap slett ikke er så forskjellig som man kunne tro. La meg få illustrere hva jeg mener med noen eksempler. Begge fakulteter er karakterisert ved fag som arbeider med store datamasser, og de bør derfor være sterkt interessert i den utvikling som for tiden pågår med sikte på databasemetodikk for å kunne lagre slike masser slik at de står til disposisjon for utnyttelse av mange brukere for ulike formål. Samfunnsvitere har gjennom mange år vært opptatt av datamaskinen som redskap for statistisk behandling av kvantitative data. Men nå viser det seg at også historikere, literaturforskere og språkforskere har behov for hyppighetstabeller og statistiske tester i sin forskning. Humanistene har på sin side sett på behandlingen av tekstlige data som sitt spesielle interessefelt, men så viser det seg at samfunnsvitere og andre nå peker på at også de har behov for metoder som automatisk kan analysere og trekke innholdet ut av erfaringsmateriale beskrevet i tekstlig form. Min oppfatning er at selv om problemene faglig sett er vesensforskjellige vil de informasjonsvitenskapelige metoder i stor utstrekning være de samme både for humanister og samfunnsvitere.

Kan du gi en kort redegjørelse for hva det legges vekt på i undervisningen i HF-delen av faget.

Av årsaker jeg nettopp har redegjort for tror jeg ikke det er mulig å skille ut en humanistisk og en ikke-humanistisk del av informasjonsvitenskapen. Vårt grunnfagsstudium utgjør for eksempel hva jeg mener er det minimum av generelle kunnskaper om informasjonsvitenskapelige metoder det er behov for enten studentene skal arbeide med humanistiske eller samfunnsvitenskapelige anvendelser. Blant de kursemner vi tilbyr undervisning i, og som bør være av spesiell interesse for studenter som vil arbeide med humanistiske anvendelser, kan jeg nevne ikke-numerisk informasjonsbehandling, filbehandling, automatisk tekstanalyse, søkesystemer, databaseteori og -systemer, formelle maskiner og formelle språk.

Hvilke mål sikter undervisningen mot og hvilke sektorer i yrkeslivet kan utdanningen rekruttere til?

Vårt siktemål er å utdanne kandidater som med støtte i sin faglige bakgrunn innen humanistiske, samfunnsvitenskapelige eller andre fag skal kunne gå inn i forskningsmiljøer, offentlige institusjoner og organisasjoner og i private bedrifter og bidra til konstruksjon av hensiktsmessige informasjonssystemer til støtte for sine arbeidsgiveres primære arbeidsoppgaver.

Jeg antar at du etter at du kom til Bergen har fått en nærmere kontakt med humanistiske fag enn tidligere. Hvordan er det naturlig for deg å karakterisere disse fagene i forhold til dem du ellers kjenner?

Det er riktig at jeg har fått både større kontakt med humanister og innsikt i deres fag i løpet av de snart fire år jeg har vært i Bergen. Likevel føler jeg meg på ingen måte kompetent til å vurdere humanistiske fag i forhold til andre. Selv forbauses jeg imidlertid stadig av hvor mange problemstillinger innen humaniora som kan spesifiseres og angripes ved hjelp av metoder som jeg trodde bare hadde anvendelse i natur- og samfunnsvitenskapene. Jeg kan igjen nevne matematisk-statistiske metoder. Det som vel kanskje karakteriserer de humanistiske vitenskaper sett ut fra min synsvinkel er det enorme og enda lite utnyttede kildemateriale disse vitenskapene har til disposisjon.

Er det noen områder innenfor de humanistiske fag som du synes er særlig interessante fra et informasjonsvitenskapelig synspunkt?

Ja, jeg er meget opptatt av formell språkteori og dens betydning sett fra en informasjonsvitenskapelig synsvinkel. Vi vet alle at datateknologien allerede har revolusjonert arbeidet på mange felter, og at samspillet mellom menneske og datamaskinen er usedvanlig viktig i denne sammenheng. Vi

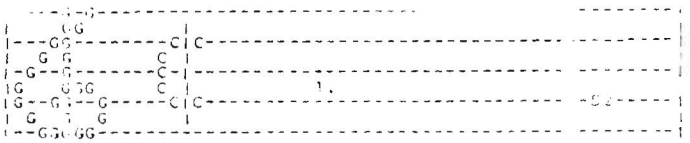
vet også at idag er primærbrukeren avhengig av selv å være — eller å kunne benytte seg av en dataekspert for å få utnyttet datamaskinens muligheter. Dette kan gi dataekspertene en urimelig sterk og farlig posisjon. Jeg ser det som en viktig oppgave at forholdene legges til rette for at også den alminnelige mann skal kunne nyte godt av det datamaskinen kan gi. En forutsetning for at dette skal kunne skje er imidlertid at brukerne kan kommunisere med datamaskinen uten forutgående kurser og studier. Med andre ord, brukerne må i størst mulig utstrekning kunne uttrykke sine problemer i en fri form som for dem virker naturlig og få datamaskinens reaksjon på samme form. Dette betyr at vi må løse problemet med å la maskinen oversette det naturlige språk til en form den selv kan arbeide videre med, og når den har funnet svaret, eller eventuelt krever flere opplysninger, må den kunne uttrykke seg i et språk som er avpasset etter det brukeren har anvendt. Dette er i grunnen ikke annet enn det innholdsanalytiske problem språkfolk har arbeidet med gjennom mange år.

Som tidligere omtalt er datamaskinen et viktig verktøy for en rasjonell informasjonsbehandling også i forbindelse med humanistisk forskningsarbeid. Mener du at de humanistiske forskere som ønsker å bruke informasjonsvitenskapelige metoder selv bør erverve seg kompetanse på feltet eller er det tilstrekkelig å samarbeide med andre med slik kunnskap i forbindelse med planlegging og gjennomføring av EDB-tiltak?

Jeg tror det er meget viktig å presisere de roller informasjonsvitenskapen kan og bør spille i humanistisk forskningsarbeid i de nærmeste årene. Informasjonsvitenskapelige metoder er idag blitt et omfattende felt som vokser raskt. Det er derfor ikke rimelig å anta at en enkelt forsker både vil kunne beherske et humanistisk fag og informasjonsvitenskap fullt ut. I humanistisk forskningsarbeid tror jeg derfor det vil bli behov for følgende personellkategorier: For det første vil det være behov for humanistiske forskere med kunnskap om og erfaring i bruk av informatiske metoder, forskere med hva jeg vil kalle sekundærkompetanse i informasjonsvitenskap/EDB oppnådd for eksempel ved et grunn- eller mellomfagsstudium i informasjonsvitenskap. Dernest vil det være behov for konsulenter med primærkompetanse i informasjonsvitenskapelige metoder, men med innsikt i de humanistiske anvendelsesfelter det arbeides på, oppnådd for eksempel gjennom hovedfagseksamen i informasjonsvitenskap og grunn- og mellomfagseksamener i historisk-filosofiske fag. Til slutt vil det kreves personale for forskning, utvikling og undervisning i informatiske metoder. Av dette personalet vil det kreves høy primærkompetanse i informasjonsvitenskap, dvs. hovedfagseksamen og forskerkompentanse i dette fag. Hvilken sekundærkompetanse denne gruppen av personale har, vil jeg tro er mindre avgjørende. I tillegg til dette akademiske personale, vil det selvsagt også være behov for teknisk datapersonale som programmerere, operatører og personale til dataregistrering etc.

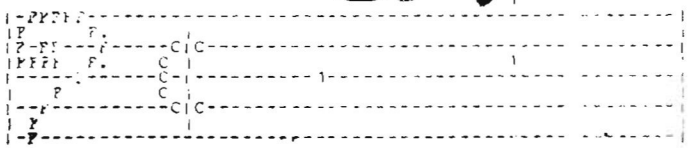
Er du villig til å spå noe om den fremtidige utvikling innenfor humanistisk databehandling?

Humanistisk databehandling er selvsagt ikke noe mål i seg selv. Jeg er imidlertid overbevist om at omfanget av viktige humanistiske forskningsoppgaver som kan dra fordel av informasjonsvitenskapelige metoder og datamaskiner er meget stort. I dag er humanistisk databehandling uten tvil ennå i sin begynnelse, og det skyldes vel at det enda er få forskere som har den nødvendige kompetanse i informasjonsvitenskap. Ved Institutt for Informasjonsvitenskap har vi nå den store glede av å se at tallet på studenter, med tidligere eksamener fra det Historisk-filosofiske fakultet øker. Jeg vil tro at disse studentene, etterhvert som de blir ferdige, vil bidra til å gjøre datamaskinen til et langt hyppigere anvendt hjelpemiddel i humanistisk forskning.



Jon-Roar Bjørkvold:

EDB og Musikk



I norsk musikkvitenskapelig miljø har den enkelte student og forsker til nå stort sett foretatt sine stilanalyser for hånd, det være seg bestemmelse av akkorder og deres sammenhenger, analyse av motivbruk, undersøkelse av formspørsmål e.l.

Analyser der kvantifiseringer av enkeltfenomener og musikalske sammenhenger utgjør en viktig del, vil med bruk av EDB kunne gjennomføres på et vesentlig større materiale og med langt høyere presisjonsnivå enn hva tilfellet har vært til nå.

Det var for 4-5 år siden etter en undersøkelse jeg foretok av melodiske særtrekk i Kjerulfs Welhaven-romanser at tanken om utvikling av EDB-analyser på musikk først meldte seg.

Ved Institutt for musikkvitenskap, Universitetet i Oslo, er vi så heldige også å ha tidligere realister med betydelig bakgrunn i databehandling blant våre hovedfagstuderende.
Sammen med to av disse studentene, Petter Henriksen og Tor Sverre Lande, dannet jeg en arbeidsgruppe med det mål å utvikle EDB-analyser på musikk. Selv har jeg ingen databehandlingsbakgrunn og fungerer i arbeidsgruppen bare som musikolog og koordinator. Arbeidet, som for alvor kom igang først i 1972-73, ble støttet finansielt av NAVF's EDB-komité for humanistisk forskning.

Første siktemål var å lage en kompakt datamaskinleselig kode som bevarer hele innholdet i standard musikknotasjon, såvel horisontale/melodiske som vertikale/harmoniske sammenhenger. Vi så det samtidig som vesentlig at koden ikke ble for kryptisk, men at den for leselighetens skyld lå mest mulig opp til et vanlig notebilde. Det foreløpige resultat av dette arbeidet forelå våren 1974 da innlesningskoden «Musikode» forelå ferdig til bruk.

Samtidig ble det arbeidet for å finne fram til en struktur som koden vår kunne «blåses opp til» ved innlesning i datamaskinen. Det viste seg snart at analyse av musikk krever manipulering av betydelige datamengder og derfor krever lagringsmedia utenfor hurtighukommelsen. Lagringsproblemene er pr. idag langt på vei løst.

Innpunching av musikken i kodet form har etter hvert vist seg å være en svært tidkrevende prosess, ja, så tidkrevende at et større analyseprosjekt, der hele notebildets sammensatte informasjon skal tas vare på, nærmest vil forby seg selv. Oppmuntret av musikkforskere har en derfor ved NAVF's EDB-senter for humanistisk forskning arbeidet med planer om å sette igang forsøksarbeid knyttet til innspilling av den musikalske informasjon direkte over til dataleselig form via et elektronisk orgel. Det har videre ut fra samme målsetting det siste halvår utviklet seg et lovende samarbeid med Matematisk institutt, avdeling D, ved prof. Ole Johan Dahl, og Fysisk institutt, Kybernetisk avdeling, ved dosent Lars Walløe. Det tas sikte på å utvikle et system, som via et elektronisk orgel både skal kunne spille musikken direkte inn, og, hva som også er meget vesentlig, sørge for musikalsk avspilling av koden som et ledd i korrigerings-prosedyren.
Ved gjennomføringen av disse planene vil en samarbeide også med NAVF's EDB-senter. Et musikkfirma i Oslo har tilbudt et egnet orgel til gunstig pris, slik at prosjektet også fra denne siden sett ligger godt an. At såvel professor Dahl som dosent Walløe, i et møte i mai d.å. med NAVF's EDB-senter, hevdet at prosjektet også rent datavitenskapelig og kybernetisk sett vil kunne by på interessante problemstillinger, borger for at denne siden

av arbeidet vil bli tatt godt vare på. Det synes således nå å kunne oppstå et lovende tverrfaglig miljø ved Universitetet i Oslo, der ikke bare ett institutt hjelper et annet, men der faktisk tre ulike miljøer kan dra vitenskapelig nytte av et felles prosjekt.

Innspillings- og avspillingsanlegget er primært tenkt bygget av hovedfagsstudenter og det arbeides i øyeblikket med å finne dugelige studenter som vil påta seg dette som hovedfagsarbeid.

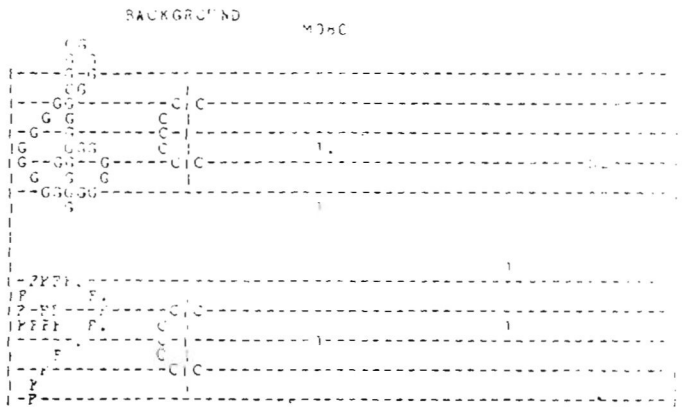
I mellomtiden er flere hovedoppgaver under utarbeidelse der utvikling og konkret musikalsk applisering av analyseprogrammer står i sentrum. En del analyseprogrammer er alt ferdig og prøvet konkret ut på musikk. Selv holder jeg på med en musikksosiologisk undersøkelse der førskolebarns sang søkes kartlagt både rent musikalsk og også sosiologisk (prosjektet har tittelen «Vårt musikalske morsmål» og bygger på et materiale samlet inn på tre daghjem i Oslo i løpet av året 1974-75), og der databehandling er tenkt tatt i bruk i analysen.

Til tross for, eller kanskje heller nettopp fordi EDB i musikkvitenskapelig sammenheng ser ut til å være en meget fruktbar tanke, finner jeg som musikolog imidlertid å ville reservere meg aldri så lite.

Det er grunn til å understreke at de fascinerende muligheter EDB synes å gi musikkforskeren aldri kan, slik jeg ser det, bli noe mer enn et viktig supplement i det musikkanalytiske arbeid.

Musikkanalyse fordrer noe mer enn kvantifisering av store musikalske datamengder og beregning av korrelasjoner mellom disse. Musikalsk følsomhet, subjektiv intuisjon og egen kultur- og samfunnsbakgrunn er således alle viktige faktorer som i møtet med selve musikken er med på å gi meningsfylt kontakt og forståelse.

Men denne erkjennelsen av et primært humanistisk og hermeneutisk vitenskapsideal forhindrer samtidig ikke en viss metodisk pluralisme, der også bruk av såkalte «harde» naturvitenskapelige metoder kan inngå. Jeg ser derfor med spenning fram til hva datamaskinen, fornuftig brukt, kan tilføre vårt fag.



KONSULENTHJELP — PUNCHEASSISTANSE

EDB-spørsmål av interesse for humanistiske forskere kan tas opp med våre faste konsulenter i Bergen, Oslo og Trondheim.
Konsulentene, som har erfaring fra EDB-arbeid fra ulike anvendelsesområder innenfor humanistisk forskning, vil også være behjelpelig med å formidle kontakt med andre fagfolk der det er ønskelig.
Særlig vil det være viktig å ta seg god tid til drøftinger med en EDB-konsulent ved planlegging av nye EDB-prosjekter.
NAVF's EDB-senter tilbyr også nye brukere gratis puncheassistanse i forbindelse med prøveprosjekter innenfor de humanistiske fagområder.

Adresser:

Bergen

NAVF's EDB-senter for humanistisk forskning, Villavei 10, Boks 53
5014 Bergen-Universitetet

Oslo

EDB-konsulent Ivar Fonnes
c/o Historisk institutt,
Universitetet i Oslo
Postboks 1102 — Blindern
Oslo 3

Trondheim

EDB-konsulent Eirik Lien
Norges Lærerhøgskole
Universitetet i Trondheim
7000 Trondheim

ordning som gir terminalbrukere adgang til senterets utstyr om ettermiddagen og kvelden.

Av prosjekter med interessepartnere utenfor Bergen kan nevnes videreføring av prosjektet for Norsk Kulturråd om databehandling av opplysninger om eldre fotografisk materiale (Humanistiske Data nr. 2 1974, s. 29) og et prøveprosjekt i samarbeid med Riks-

I det forrige nummeret av Humanistiske Data, nr. 2 1974, ble det gitt en orientering om senterets arbeid i 1974. Der ble også senterets langtidsplan omtalt, og det ble uttalt at senterets viktigste arbeidsinnsats er knyttet til kurs- og informasjonsvirksomhet, prosjektassistanse og generell programutvikling. Virksomheten i 1975 har vært en videreføring av aktiviteter innenfor disse områder.

I 1975 har vi også hatt gleden av å registrere en økt EDB-aktivitet blant humanistene i Trondheim — et forhold som har direkte sammenheng med arbeidet til senterets EDB-konsulent her. Nytt i 1975 er at senteret har knyttet direkte kontakt med universitetsmiljøet i Tromsø. På litt lengre sikt må det være målet å tilby Universitetet i Tromsø konsulent- og servicetjenester på linje med de andre EDB-miljøene, men avpasset etter mottakermiljøets egne forutsetninger og behov.

Kurs- og informasjonsvirksomhet.

Ved Universitetet i Trondheim ble det i vinter holdt et kurs for humanister i NU-ALGOL med konsulent Eirik Lien som kursleder. Kurset samlet deltakere fra flere institutter. Når det gjelder slike kurs, viser det seg at det er vanskelig å tilrettelegge kurstilbudet for humanister slik at mange ser seg i stand til å følge dem, men de få som gjennomfører et slikt kurs, kommer senere ofte tilbake med egne prosjekter og blir aktive EDB-brukere.

I Oslo har konsulent Ivar Fonnes i vårsemesteret samarbeidet med vit.ass. Steinar Gil om et kurs i programpakken TEXT.

Ved senteret i Bergen har det i vår ikke vært noen regulær kursvirksomhet. Arbeidet har vært konsentrert om å gi instruksjon til tidligere kursdeltakere og nye brukere. Instruksjonen har både omfattet veiledning i programmeringsspørsmål og øving i bruk av senterets datamaskin og terminalutstyr.

I november vil det bli holdt et seminar i Bergen om litterær statistikk med Dr. D. Wickmann fra Aachen som foreleser.

I Oslo og Bergen har det vært gitt informasjon til interesserte hovedfagsstudenter ved HF-fakultetene om det program- og maskintilbud som står til disposisjon på studiestedet.

Som et supplement til bladet HUMANISTISKE DATA har senteret begynt å gi ut en bulletin kalt SEKVENS, hvor en raskt kan få ut til aktuelle miljøer slike opplysninger som det haster å kunngjøre.

For å få et dekkende inntrykk av EDB-situasjonen i Tromsø, både generelt og med tanke på humanistenes interesser, oppholdt konsulent Lien og adm.leder Hauge seg et par dager ved Universitetet i Tromsø i september. Oppholdet ble lagt til et tidspunkt da sekretariatleder Bjørn Henrichsen ved Norsk Samfunnsvitenskapelig Datatjeneste i Bergen også var der, slik at det kunne arrangeres møter hvor humanistenes og samfunnsviternes interesser kunne sees i sammenheng.

Det kom klart til uttrykk at EDB-senteret ved Universitetet i Tromsø er innstilt på et nært samarbeid med humanistiske brukere, som i dag riktignok utgjør et meget beskjedent innslag blant brukerne. Inntil Universitetet i Tromsø får sin egen datamaskin om et par år, vil Universitetet trolig fortsette sin nåværende ordning med å nytte regneanlegget ved Universitetet i Bergen via fast datalinje. Dette betyr at den programutrustning for humanistiske oppgaver som finnes i Bergen er direkte utnyttbar for brukere i Tromsø.

Det kan for øvrig nevnes at NAVF's EDB-senter i høst etter oppdrag fra EDB-senteret i Tromsø utarbeidet en redegjørelse om humanistenes brukerinteresser i forbindelse

med Universitetets planlegging av eget maskinkjøp.

Senteret har i løpet av høsten arbeidet med spørsmålet om å finne nye veier i arbeidet med å gjøre flere humanistiske forskere kjent med de muligheter som databehandling kan gi i forsknings- og utviklingsarbeidet. Det er blant annet fremmet forslag til NAVF, Fagråd A, om at senteret introduserer kortvarige EDB-stipend for humanistiske forskere som ønsker å sette seg inn i bruken av databehandling gjennom et konsentrert og tilrettelagt undervisnings- og studietilbud av ca. en måneds varighet. Ansvarlige for opplæringsprogrammet vil være senterets konsulenter i Oslo, Bergen og Trondheim.
Fagråd A har stilt seg positiv til forslaget og er innstilt på å sette ordningen igang forsøksvis under forutsetning av at det kan skaffes midler til tiltaket i 1976.

Prosjektarbeid.
Siden opprettelsen av konsulentstillingen i Trondheim i 1974 har EDB-interesserte humanister på dette stedet hatt gode muligheter til å få drøftet og planlagt EDB-tiltak. I løpet av det siste året har konsulenten deltatt i planleggingsarbeid og prosjekter ved flere institutter: Bl.a. Historisk institutt, Institutt for sosiologi og samfunnskunnskap, Engelsk institutt, Nordisk institutt. I arbeidet har det vært til stor nytte at konsulenten helt fra starten av har disponert en terminal, innkjøpt av NAVF, på sitt kontor på Norges lærerhøgskole på Lade. Situasjonen har derfor vært den noe uvanlige at det er det humanistiske miljøet som foreløpig er best teknisk utstyrt for databehandling på Lade.
Som et generelt tilgjengelig hjelpemiddel har konsulenten utarbeidet en maskinlagret katalog om prosjekter innen- og utenlands hvor datamaskinen nyttes i arbeidet med språk og litteratur. Katalogen er lagt til rette på universitetsanlegget i Trondheim og interesserte kan få opplysninger fra katalogen ved å vende seg til konsulenten.

I Oslo har vår konsulent videreført sin kontakt med EDB-tiltak på blant annet Historisk, Nordisk, Slavisk-baltisk og Britisk institutt. Karakteristisk for flere prosjekter har vært et behov for hjelp ved tilpassing/videreutvikling av programmer etterhvert som det faglige arbeid presenterer nye problemstillinger.

Den største enkeltoppgaven for konsulenten i Oslo har ellers vært knyttet til arbeidet med Norsk Landbruksordbok (se Humanistiske Data nr. 2 1974 side 29), hvor det for tiden foregår kontinuerlig punching av materiale og databehandling av det i forbindelse med et opplegg for datamaskinell fotosetting.

KONSULENTHJELP — PUNCHEASSISTANSE
EDB-spørsmål av interesse for humanistiske forskere kan tas opp med våre faste konsulenter i Bergen, Oslo og Trondheim.
Konsulentene, som har erfaring fra EDB-arbeid fra ulike anvendelsesområder innenfor humanistisk forskning, vil også være behjelpelig med å formidle kontakt med andre fagfolk der det er ønskelig.
de med Fysisk institutt, Universitetet i Oslo m.fl. om metoder for direkte innspilling fra elektronisk orgel til datamaskin av musikkmateriale, og avspilling av maskinlagret musikkmateriale for kontroll-lytting.

Prosjektsamarbeidet ved senteret i Bergen har et tosidig siktemål. En del av virksomheten er rettet mot de humanistiske brukermiljøene ved Universitetet i Bergen, men arbeidet skal også være innrettet mot utviklings- og prosjektarbeid av felles interesse for flere brukermiljøer i vårt land. I overensstemmelse med denne to-delte målsettingen har en et samarbeid igang med en rekke institutter i Bergen: Klassisk institutt, Romansk institutt, Nordisk institutt, Historisk institutt, Historisk museum. Arbeidet er knyttet til oppgaver i sammenheng med tekstbehandling og bruk av databehandling i forbindelse med arkivmateriale. For å gi flere brukere muligheter til å nytte senterets terminaler har en fra høsten etablert en

ordning som gir terminalbrukere adgang til senterets utstyr om ettermiddagen og kvelden.

Av prosjekter med interessepartnere utenfor Bergen kan nevnes videreføring av prosjektet for Norsk Kulturråd om databehandling av opplysninger om eldre fotografisk materiale (Humanistiske Data nr. 2 1974, s. 29) og et prøveprosjekt i samarbeid med Riksbibliotektjenesten om databehandling av manuskriptregistraturer. I Fotoprosjektet er det punchet detaljerte opplysninger inklusive fritekstopplysninger om ca. 2000 bilder og utviklet programmer for ulike typer sortering av data. De faglige uttestinger vil foregå i løpet av høsten og resultatene vil foreligge i en rapport ved årsskiftet.
Arbeidet med databehandling av manuskriptregistraturer er i en innledende fase. Et representativt utvalg kataloger og fortegnelser over håndskriftmateriale er skrevet av for optisk lesing og overført til magnetbånd. I løpet av høsten vil en starte arbeidet med å utvikle programmer for å omstrukturere materialet maskinelt etter nye prinsipper bl.a. med tanke på å innpasse det i større materialsamlinger.

Det overordnede mål er å finne fram til datamaskinelle metoder som kan lette forskernes adgang til kildemateriale f.eks. ved å gi adgang til å nytte flere kriterier som søkeinnganger.

Programmeringsvirksomhet.
I forbindelse med konsulentbistand til de prosjekter som benytter programpakken TEXT i Oslo har konsulent Fonnes foretatt en del justeringer av programmene. Dessuten er det skrevet et generelt program for søking i tekster.

I samarbeid med konsulent Lien i Trondheim har Fonnes siden i sommer arbeidet med å teste ut programpakken EYEBALL på CDC-anlegget ved Universitetet i Oslo, jfr. eget oppsett i dette nummeret om EYEBALL.
Konsulent Fonnes har ellers arbeidet med programmeringsspørsmål for Norsk Landbruksordbok.

For å møte et stigende behov for databehandling av tekster i Trondheim og Bergen har senteret arbeidet med et standard programtilbud for tekstbehandling som dels består av selvstendige programmer, men som også dels bygger på bruk av de mest anvendelige editorer på datamaskinene i Bergen og Trondheim. Fra årsskiftet vil derfor alle universitetsmiljøene ha et generelt programtilbud for tekstbehandling. Oppgavene fremover på dette feltet vil i første rekke bli knyttet til utvikling av metoder for mest mulig automatisk sammenføring av ordformer til grunnformer i de listeprodukter som blir utarbeidet til tekster (f.eks. frekvenslister).

Behovet for et tekstsøkesystem på Univac datamaskiner har ført til at senteret sammen med det samfunnsvitenskapelige datamiljø i Tromsø har undersøkt muligheten for å få et tekstsøkesystem kalt STATUS konvertert til bruk på Univac. Dette programsystemet har Statens Rasjonaliseringsdirektorat kjøpt inn fra England i første rekke til bruk i forvaltning. En venter at spørsmålet om konvertering kan avklares i løpet av høsten.

Det kan på dette punkt understrekes at senterets konsulenter generelt er opptatt av å holde seg orientert om det standard tilbud av programutrustning som foreligger på universitetenes regneanlegg, og ser det som en viktig oppgave å vurdere om programmene kan nyttes i humanistisk forskning. Som et eksempel kan nevnes at konsulentene både i Oslo, Trondheim og Bergen har satt seg inn i og utnyttet i sitt arbeid de programpakker for statistikk som i særlig grad samfunnsvitenskapen til nå har brukt. Parallelt med introduksjon av databehandling i de humanistiske fag fremstår det et behov for statistisk behandling av det aktuelle forskningsmaterialet. Både dette feltet og andre fellesområder gjør det ønskelig med et nært samarbeid mellom de samfunnsvitenskapelige og humanistiske datamiljøer.

Innføring i EDB og humaniora

En kort litteraturoversikt av Roald Skarsten

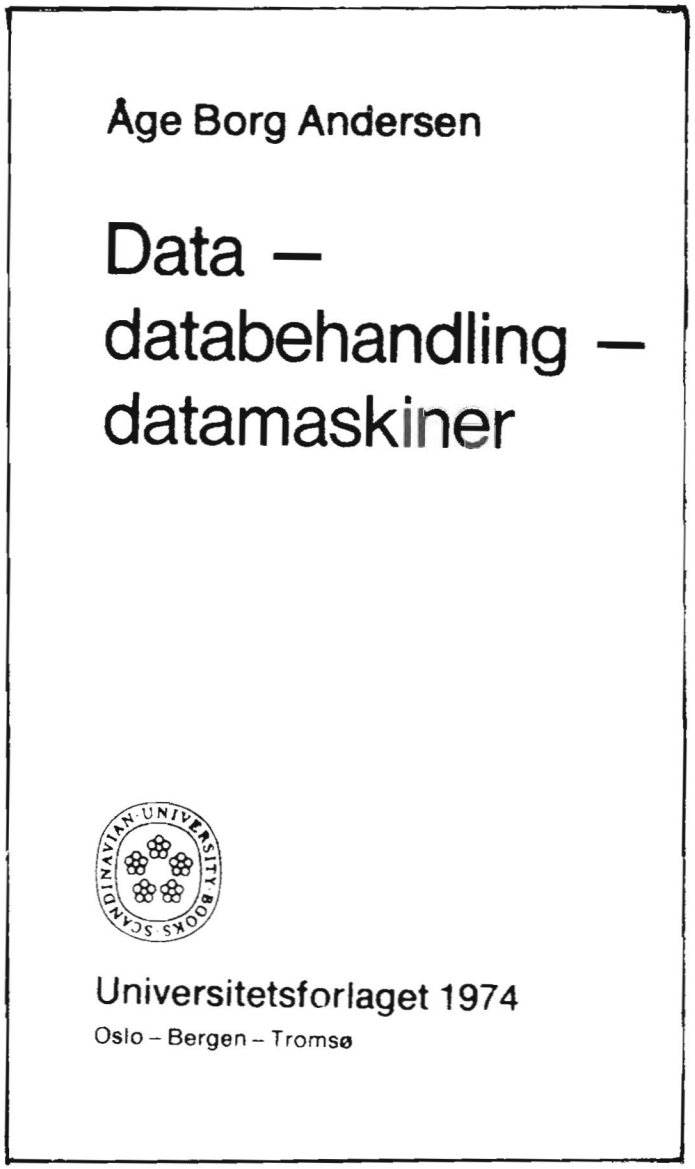
Denne lille oversikten over litteratur som kan brukes til innføring i EDB og humaniora er delt opp i tre deler. Først omtale av bøker som gir generell innføring i datamaskiner og bruken av dem, dernest bøker som gjelder applikasjon av EDB på humanistiske fagfelt og til slutt nevnes noen bøker som gir konkrete og praktiske råd og veiledning som er nyttig for de som for første gang vil ta i bruk datamaskiner innen sitt fagområde.

1. *Generell innføring i EDB.*
Bøker med det formål å gi generell innføring i EDB finnes der en del av, men ikke alle kan sies å være like velegnet. En bok som skal gi en første innføring, bør være enkelt og oversiktlig skrevet slik at man ikke legger boken fra seg før man har lest den med utbytte. Det første boklige møte med EDB kan lett gi en følelsen av å stange hodet mot veggen, og særlig hvis det ikke skjelnes mellom vesentlig og mindre vesentlig informasjon.

En nærmest klassisk introduksjonsbok på skandinavisk område er *Carl-Erik Fröberg och Bengt Sigurd, Datamaskiner. Gleerup, Lund 1967, (165 s.).* De mange opplag vitner om dens popularitet. På noen sentrale områder er det i en viss utstrekning brukt matematiske uttrykk og det kan kanskje ikke sies å være særlig velvalgt i en populær framstilling. Oversikten over de forskjellige maskinleverandørers maskintyper og den historiske oversikt med sterk vekt på svensk historikk fører til at forholdet, målt i sider, mellom vesentlig og mindre vesentlig informasjon, ikke er så god. Bokens styrke i vår sammenheng er imidlertid at den har hele 28 sider om datamaskinell språkoversettelse, og mange av de problemer som drøftes i den forbindelse står sentralt innenfor humanistisk databehandling.

Arne Sølvberg, Datamaskinen — en elementær innføring. Tapir, Trondheim 1969 (78 S.) brukes som lærebok ved NTH. I boken legges det stor vekt på grunnleggende tekniske forhold ved datamaskinens oppbygging og på programmering av datamaskinen. Dette er helt i samsvar med forordet hvor det sies at målet har vært «både å skissere datamaskinens oppbygging og å sannsynliggjøre at maskinene virkelig kan utføre det arbeidet vi setter dem til å gjøre». Boken bærer preg av at den er ment som innføring for studenter ved en teknisk høyskole, studenter som selv skal lære å programmere en datamaskin. Boken er imidlertid oversiktlig og klart skrevet slik at den med fordel kan brukes av de som bare ønsker en mer generell innføring i prinsipper og virkemåte for en datamaskin.

En bok som også oppfyller disse sistnevnte ønsker og de krav som innledningsvis ble nevnt, er *Eivind Barca, Innføring i databehandling. Tanum, Oslo 1973, 2. reviderte utgave, (111 s.).* Det er også en lærebok, for yrkesskolene, men den er ikke orientert mot opplæring i programmering, og den har en meget vellykket pedagogisk utformning, både innholdsmessig og visuelt. I vår sammenheng er det sidene 1-76 og 87-93 som er relevante. Særlig vekt bør man legge på avsnittene om sentralenheten og styringen av et EDB-system.



Av litt større bøker kan nevnes *Ole Dopping, Kort och brett om ADB. Studentlitteratur, Lund 1972 (225 s.)* og *Åge Borg Andersen, Data-databehandling-datamaskiner. Universitetsforlaget, 1974, (221 s.).* Begge disse lærebøkene er innføringsbøker. Andersen skriver i forordet at det ikke er gjort «bevisste forsøk på å popularisere fremstillingen», men at det heller ikke forutsettes noen forkunnskaper i emnet for å lese boken. Begge disse bøkene er oversiktlige og klare i sin framstilling, og egner seg godt for «videre innføring», og særlig hvis man sikter mot selvstendig programmeringsvirksomhet. En fordel med Andersens bok er at den oversetter tekniske termer fra «EDB-språket» til norsk.

Det kan i denne forbindelse pekes på behovet for ordbøker som gir korte og instruktive forklaringer på vanlige EDB-termer. Slike bøker er nyttige som oppslagsbøker for alle som får den minste befatning med EDB og «EDB-folk». Følgende bok kan anbefales: *A. Chandor,* A Dictionary of Computers, Penguin Books, 1970, (406 s.). En lignende miniordbok er utgitt på svensk av *W. N. Lansburgh, 300 Nya Termer, Almquist og Wiksell, 1972, (41 s.).*

Den største faren for en humanist som har et visst ønske om å orientere seg i EDB, er at vedkommende forsøker å begripe for mye på en gang. I slike tilfeller blir resultatet lett at motivasjonen ikke strekker til overfor mengden av stoff og den stigende følelsen av fremmedgjøring i forhold til et ukjent tema

og en vanligvis teknisk preget framstillingsform. Jeg vil derfor ikke anbefale noen å oppholde seg for lenge med EDB generelt, snarest mulig gjelder det om å komme over på et applikasjonsområde hvor motivasjonen er sterk og hvor man føler seg på trygg grunn når det gjelder selve det fagområdet som EDB anvendes på.

2. *På hvilke humanistiske forskningsfelt anvendes EDB?*

Den kanskje enkleste måten å svare på dette er ved å vise til overskriftene i tidsskriftet «Computers and the Humanities» og dets «Directory of Scholars Active» og «Annual Bibliography». Her er overskrifter som «Language and Literature», «Music», History», «Archaeology», «Visual Arts» og «General». Under den siste overskriften finner man diverse temaer som er av generell interesse i humanistisk databehandling.

Språk og litteratur.

Det utvilsomt mest omfattende felt er språk og litteratur. For dette felts vedkommende foreligger det på et skandinavisk språk en grei oversikt, nemlig *Sture Allén och Jan Thavenius (red.), Språklig databehandling. Studentlitteratur, Lund 1970, (208 s.).* Forskjellige forfattere presenterer her sine prosjekter. De spenner over så vide områder som tysk avisspråk, engelsk morfologi, svensk vokabularsystem, konkordansproduksjon og ekthetsundersøkelser. Artiklene er forsynt med til dels svært fyldige litteraturlister som kan være meget nyttige for den som vil orientere seg videre på et bestemt område.

Lignende bøker på engelsk er der mange av, noen av dem er samleverk med foredrag fra internasjonale kongresser. I samleverkene er artiklene inndelt i grupper, og for det meste samlet under overskrifter som «Lexicography», «Attribution Studies», «Stylistic Analysis», «Linguistics», «Textual Editing» og «Vocabulary Studies». Disse overskriftene viser for øvrig de fleste av de områder innen feltet språk og litteratur som det har vært arbeidet mest på med bruk av datamaskiner. Boken *The computer in literary and linguistic research, edited by R. A. Wisbey, Cambridge University Press, 1971, (309 s.)* inneholder bearbeidete foredrag fra et internasjonalt symposium i Cambridge i 1970. Etter et tilsvarende symposium i Edinburgh i 1972, utkom boken, *The Computer and Literary Studies, edited by A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith. Edinburgh University Press., 1973, (369 s.).* International Conference on Computers in the Humanities i Minneapolis i 1973 resulterte bl.a. i boken *Computers in the Humanities, edited by J. L. Mitchell. Edinburgh University Press, 1974, (318 s.).* På området stilistikk og forfatterskapsbestemmelse vil jeg for øvrig anbefale en litt eldre bok, som imidlertid fremdeles kan være svært nyttig til orientering og som er god når det gjelder metodiske spørsmål. Det er *The Computer & Literary Style,* edited by J. Leed, Kent State University Press, 1966, (179 s.).

På norsk finnes det en artikkel om tekstbehandling i Norsk Teologisk Tidsskrift for 1971: Roald Skarsten, Om datamaskinell tekstbehandling, (s. 181-199).

På tysk foreligger det en innføringsbok som omhandler alle de tre deler som denne artikkelen behandler: *Winfried Lenders, Einführung in die linguistische Datenverarbeitung 1. Max Niemeyer Verlag. Tübingen 1972, (98 s.). (Nr. 8 i serien Germanistische Arbeitshefte).*

Boken gir både en orientering om de forskjellige områder innen datamaskinell tekstbehandling, med bibliografiske henvisninger, og en redegjørelse for datamaskiners oppbygning og virkemåte, foruten et interessant kapittel som med eksempler gir veiledning i å formulere programmerbare problemer.

En annen bok som skal nevnes er D. G. Hays, *Introduction to Computational Linguistics. Elsevier, 1967, (231 s.).* Det er en bok som både gir god innføring i en datamaskins virkemåte og i forskjellige former for lagring av data. Videre går den forholdsvis utførlig inn på de forskjellige teknikker som er

aktuelle i forbindelse med datamaskinell lingvistikk. Boken har egne kapitler for f.eks. datamaskinell lagring og bruk av grammatikker, parsing og automatisk syntaksanalyse, foruten automatisk språkoversettelse. Boken kan kanskje ikke anbefales til den aller første innføring fordi den bl.a. i for stor utstrekning involverer programmering (ALGOL), men med sin behandling av sentrale metodeproblemer i datamaskinell lingvistikk kan den brukes til «videre innføring».

Musikk.

På musikkens fagområde finnes det et samleverk utgitt av *H. Heckmann, Elektronische Datenverarbeitung in der Musikwissenschaft. G. Bosse Verlag Regensburg, 1967.* Boken inneholder 13 artikler om forskjellige sider av datamaskinell musikkbehandling. På engelsk foreligger samleverket *Musicology and the Computer, Musicology 1966-2000: A Practical Program, B. S. Brook, ed. The City University of New York Press, 1970, (275 s.).* Sidene 231-270 inneholder en bibliografisk oversikt. Den nyeste utvikling på området får man orientering om i artikkelen, «Use of the Computer in Music Research: A Short Report on Acomplishments, Limitations and Future Needs.» av H. B. Lincoln i september-november nummeret for 1974 av Computers and the Humanities.

Arkeologi.

Den beste introduksjon til bruk av datamaskiner i arkeologisk forskning får man ved å lese oversiktsartikler i «Computers and the Humanities». Spesielt skal det pekes på en artikkel av R. Whallon, Jr., «The Computer in Archaelogy: A Critical Survey». Artikkelen er utstyrt med svært fyldige referanser (sept. 72). Videre en bokanmeldelse i sept./nov.-nummeret for 1974 som samtidig nevner de nyeste bøker på området og aktuelle fagtidsskrift. Anmeldelsen er ved S. S. Lukesh og R. R. Holloway, og står på s. 341-343.

Historie.

Muligheter og begrensninger når det gjelder bruk av datamaskiner i historieforskningen drøftes forholdsvis utførlig av C. Tilly, «Computers in Historical Analysis», Computers and the Humanities, sept./nov. 1973, (s. 323-334). Dette er en artikkel som egner seg utmerket som innføring fordi den diskuterer prinsipielle spørsmål i forbindelse med bruk av datamaskiner i historieforskningen, samtidig som den gir konkrete eksempler, jfr. overskrifter som «Is History Computable?» og «Historical Demography as an Illustration». Nevnes bør også en oversiktsartikkel fra sept. -72 (s. 67-79) i samme tidsskrift: J. H. Silbey, «Clio and Computers: Moving into Phase 11, 1970-1972».

Innenfor de fagområder som her er nevnt spesielt, og på andre fagområder, er der en mengde artikler som kan leses med utbytte av dem som vil orientere seg, men det vil føre for langt å henvise til dem her. Vi vil derfor gi en generell henvisning til de bibliografiske oversikter i det nevnte tidsskriftet *«Computers and the Humanities»*. (Published by Pergamon Press edited at Queens College, Flushing New York). Dette er et uunnværlig tidsskrift for de som vil være orientert om humanistisk databehandling. Det kommer ut med fem nummer i året og inneholder artikler fra alle de forskjellige humanistiske arbeidsområder og ofte sammenfattende oversiktsartikler for utviklingen innen de forskjellige felter. Videre har tidsskriftet regelmessige oversikter over tilgjengelige maskinleselige tekster, tilgjengelige programmer og igangværende prosjekter.

Tidsskriftet *«Computer Studies in the Humanities and Verbal Bahavior»* utgis på forlaget Mouton & Co. i Haag, og kommer ut kvartalsvis. Den internasjonale *Association for Literary and Linguistic Computing,* som ble etablert for et par år siden, utgir en bulletin som kommer med 3 nummer i året. Foruten disse tidsskriftene, som dekker hele det humanistiske området, finnes det noen få tidsskrifter som bare dekker spesielle fagfelt, f.eks. klassiske språk, (Revue) eller middelalderstudier *(Computers and Madiaeval Data Processing).*

3. *Litteratur med praktiske råd.*

Det tredje punkt i denne oversikten, litteratur med gode praktiske råd for de som starter på bar bakke, er det vanskeligste, for her er det forholdsvis lite med velegnet litteratur. Det som først skal nevnes er fra området språk og litteratur, ettersom det i de fleste tilfeller av humanistisk databehandling vil være aktuelt med en eller annen form for tekstrepresentasjon. En god og detaljert bok er *B. Munk Olsen, Anvendelsen af elektronisk databehandling ved løsningen av filologiske oppgaver: Konkordanser, Indices Verborum. Romansk Institut, Københavns Universitet, 1968, (128 s.).* Første del av boken behandler tekniske aspekter, og her er det god grunn til å lese første kapittel. Resten av del 1 er så sterkt forbundet med et bestemt maskinmerke, som ikke er representert i universitetsmiljøene i Norge i dag, at det ikke er umiddelbart nyttig lesning. Del 2 derimot gir et godt eksempel på hvilke praktiske problemer man står overfor når man skal gjøre en tekst maskinleselig. Problemets art fremgår kanskje av følgende eksempel: Det drøftes hvordan man skal representere sitattegnene for å unngå tvetydighet m.h.t. begynnelse og slutt av sitatet. Lignende problemer i hopetall skal man være oppmerksom på i denne fasen. Det kreves en mental omstilling for å tilfredsstille en datamaskins krav på absolutt entydighet. Boken inneholder videre en god drøftelse av forskjellige løsninger for produksjon av konkordanser. Lesning av en slik bok vil neppe løse alle de problemer man står overfor, men den er med sin detaljrikdom meget nyttig for å bli klar over hvilke retningslinjer som må legges til grunn for punchearbeidet.

Tilsvarende problemer for punching av gresk tekst er behandlet i en rapport av *P. Borgen og R. Skarsten, Maskinleselig tekst til Philo av Alexandrias samlede verker. Religionsvitenskapelig institutt, UiB, 1972.*

Når det gjelder musikk kan det vises til to instituttpublikasjoner på norsk. Den ene er *Tore Simonsen, Norstil 70, et datamaskinsystem for stilanalyser av musikk. Musikkvitenskapelig institutt, Universitetet i Trondheim, 1973. (heftet 59s.).* Den andre er *Petter Henriksen og Tor Sverre Lande, Musikode, Innlesningskode til elektronisk databehandling av musikk, Institutt for musikkvitenskap, Universitetet i Oslo. 1974 (heftet 47 s.).* Forordet i sistnevnte bok angir målsettingen på følgende måte: «1) lage en kompakt datamaskinleselig kode som bevarer det semantiske innhold i standard musikknotasjon, og 2) finne en struktur som koden kan «blåses opp til» ved innlesningen i datamaskinen, og som vil egne seg for alle typer stilanalyse som musikkforskeren kan tenkes å utføre på et notebilde — harmonisk analyse, motivanalyse, melodisk analyse, formanalyse, m.m.». Begge de her nevnte bøker (hefter) vil være til stor nytte for den som trenger praktiske råd når han selv skal starte et prosjekt. Det er i det hele tatt grunn til å peke på at praktiske råd får man best hos de personer som har gjennomført prosjekter, og ofte vil verdifull erfaring være nedfelt i prosjektrapporter. Det er derfor det beste først å tale med andre innen sitt eget fagområde som har vært gjennom denne første fasen før man selv starter med punching. (Bruk «Computers in the Humanities»).

Forskjellige typer dataregistreringsutstyr er utførlig behandlet av Ben Schneider: «The Production of Machine — Readable Text: Some of the Variables.», Computers and the Humanities, september 1971, (s. 39-47).

Når det gjelder retting av punchede tekster og tilhørende problemer kan det henvises til en drøftelse av dette i forrige nummer av Humanistiske data.

Det skal for øvrig pekes på at NAVF's EDB-senter for humanistisk forskning, med konsulenter i Bergen, Oslo og Trondheim, yter forskjellige former for assistanse til humanistiske forskere som ønsker å ta i bruk EDB som et hjelpemiddel innen sitt fagområde. Interesserte er alltid velkommen til å kontakte senteret.

en programpakke for stilistisk analyse av tekster

EYEBALL

Senteret anskaffet i vår programpakken EYEBALL som er utviklet ved University of Minnesota. Den vil antakelig være av interesse også for en del filologer her i landet, selv om den er begrenset til bare å behandle engelsk tekst.

Pakken er delt i fem komponenter, beregnet vekselvis på seriell og interaktiv bruk av datamaskinen. Ved å behandle en tekst ved hjelp av de ulike typer programmer i denne pakken, vil en få stadig mer detaljerte bearbeidinger av teksten.

Fase 1 splitter opp teksten i de enkelte ord og gir referanse til dem, lager konkordans og frekvensordliste, teller opp stavinger, splitter opp sammensetninger og klargjør filer for interaktiv analyse. Programmene i denne fasen inneholder også en «funksjonsordliste» på ca. 200 ord hvor opplysning om ordklasse er lagt inn. Hvert enkelt ord i teksten blir sjekket mot denne lista, og aktuell ordklasse knyttet til ordet dersom ordet fins i lista.

Fase 2 er for interaktiv bruk og inneholder operasjoner for syntaktisk analyse av preposisjonsfraser og underordnede setninger.

Fase 3 har program som separerer de analyserte underordnede setningene fra resten av teksten.

Fase 4 er igjen beregnet for interaktiv bruk, og her blir de overordnede setningene analysert.

Fase 5 «syr sammen» resultatene i de fire første fasene, slik at hvert ord i teksten blir forsynt med følgende informasjon:

grammatisk kategori
syntaktisk funksjon
stavingslengde

Sentret har mottatt to versjoner av pakken, en for IBM og en for CDC-maskiner. Programmene er skrevet i programmeringsspråket FORTRAN. Pakken er nå under utprøving ved Universitetet i Oslo ved konsulent Ivar Fonnes (verken Bergen eller Trondheim har IBM eller CDC).

Dersom prøvingen viser seg å gi et fruktbart resultat — d.v.s. at pakken ser ut til å være et tjenlig hjelpemiddel for engelskfilologene i Norge — vil det bli laget en UNIVAC-versjon av den slik at den også kan bli implementert i Bergen og Trondheim. Dette kan i så fall antakelig skje i løpet av vinteren.

Når/hvis pakken blir driftsklar, vil vi informere om dette gjennom enten SEKVENS eller Humanistiske data.

Imens kan vi henvise interesserte til tidsskriftet «Computers and the Humanities» vol. VI pp 213 — 221 hvor det er en ganske detaljert presentasjon av pakken av de to «forfatterne» Donald Ross Jr. og Robert H. Rasche.

Eirik Lien

Steinar Gil:

Ordforråd, frekvenser og spredning

I forbindelse med en større undersøkelse av substantivene hos den russiske lyriker Anna Achmatova foretok jeg en sammenligning av hennes substantivforråd med substantivforrådet hos en del andre russiske lyrikere. Hensikten med denne sammenligningen var for det første å klarlegge hvor mange og hvilke av Achmatovas substantiver som kan sies å være typiske for den lyriske genre og hvor mange og hvilke som kan sies å være typiske for Achmatovas individualstil. For det andre ville jeg undersøke hvilken eller hvilke av de andre dikterne som hadde mest til felles med Achmatova når det gjaldt substantivforråd. Fremgangsmåten ved sammenligningen og de resultatene den ga kan muligens være av allmenn interesse.

Undersøkelsen ble utført på grunnlag av frekvensordlister og konkordanser, produsert ved hjelp av programmer i programsystemet TEXT, som er utarbeidet av NAVF's EDB-konsulent ved HF i Oslo, cand. philol. Ivar Fonnes. Det samlede tekstgrunnlag besto av i alt 220 973 løpende ord, fordelt på 8 tekstenheter av 8 forskjellige diktere. Dikterne var E. Baratynskij (1800-1844), M. Lermontov (1814-1841), F. Tjutčev (1803-1873), A. Fet (1820-1892), K. Bal'mont (1867-1942), I. Annenskij (1856-1909), A. Blok (1880-1921) og A. Achmatova (1889-1966). Tekstmaterialet omfattet samlede dikt av alle de nevnte diktere unntagen Bal'mont og Blok, som var representert med hver sin diktsyklus. Ordmassen fordelte seg på følgende måte (dikterne angitt ved initialer): E. B. 38 190, M. L. 43 315, F. T. 30 942, A F. 36 042, K. B. 8 838, I. A. 18 712, A. B. 5 404, A. A. 39 530.

Alle substantivene i A. A. ble ført opp i en tabell som i tillegg til frekvenser i A. A. også inneholdt opplysninger om forekomster og frekvenser av de samme substantivene i de andre tekstenhetene. Substantivene ble inndelt i spredningsgrupper, alle substantiver som forekom i samtlige tekstenheter, dvs. de som hadde spredning 8, for seg, så de som hadde spredning 7 osv. Tabellen var ordnet etter fallende spredning og fallende frekvens med spredningen som overordnet sorteringsprinsipp, dvs. at hver spredningsgruppe ble frekvenssortert for seg etter fallende frekvens i A.A.

I alt inneholdt substantivtabellen 2.504 forskjellige ord med en totalfrekvens på 10.025 eller 25,4% av tekstmassen. Til sammenligning kan nevnes at i E. Steinfeldts frekvensordliste for det moderne russiske skriftspråk utgjør substantivene 26,4%. Av

dette må man kunne slutte at Achmatovas substantivfrekvens ligger omtrent på gjennomsnittet for russiske tekster.

Av substantivene i A.A. var det 818 eller 32.7% som hadde spredning 1 og 1.686 eller 67.3% som også forekom i en eller flere av de andre tekstenhetene. Denne siste gruppen av substantiv fordelte seg på følgende måte med hensyn til spredning: 184 fantes i alle 8 enheter, 177 i 7, 213 i 6, 216 i 5, 238 i 4, 266 i 3 og 392 i 2. Det fremgikk tydelig av tabellen at de mest høyfrekvente substantivene i A.A. også var de som hadde størst spredning. Av de 216 substantivene i A.A. med frekvens på 10 og høyere var det hele 121 som hadde spredning 8, 41 hadde spredning 7, 25 hadde spredning 6, 21 hadde spredning 5, 5 hadde spredning 4 og de tre laveste spredningsgruppene hadde 1 substantiv hver med frekvens 10. Substantivene med spredning 1 har en totalfrekvens på 1.076. De utgjør med andre ord bare 10.9% av alle substantivforekomster i A.A. På grunnlag av disse tallene må man kunne slutte at den overveiende del av substantivene i Achmatovas lyrikk, og i særlig grad de høyfrekvente, tilhører et allment poetisk ordforråd som danner grunnstammen i russiske poesitekster uavhengig av tematikk og tidsperiode.

Analysen av spredningstallene kompliseres imidlertid av at de forskjellige tekstenhetene varierer nokså mye i størrelse (fra 43.315 løpende ord i M.L. til 5.404 løpende ord i A.B.). Det viste seg da også at de minste enhetene, K.B. og A.B. hadde adskillig færre substantivsammenfall med A.A. enn de andre enhetene. En direkte sammenligning av sammenfall i de forskjellige tekstenhetene innenfor hver spredningsgruppe vil bare være mulig, dersom enhetene er like store eller tilnærmet like store. Men ikke desto mindre vil man kunne foreta en sammenligning for å finne ut hvorvidt de observerte sammenfall av forekomster virkelig synes å være fordelt noenlunde proporsjonalt med tekstenhetenes størrelse, eller om f.eks. en liten enhet oppviser betydelig flere sammenfall enn en større enhet. I siste fall må det være mulig å tale om en markert forskjell i leksikalsk slektskap med hovedteksten (i vårt tilfelle A.A.).

For å klarlegge dette forholdet stilte jeg opp en spredningstabell som viser antall sammenfall med A.A. for hver av tekstenhetene innenfor de forskjellige spredningsgruppene. Kolonne 1 i tabellen angir spredningsgruppe, kolonne 2 det totale antall ord innenfor hver spredningsgruppe, og de 7 siste kolonner angir hvor mange av ordene innenfor hver spredningsgruppe som forekommer i de forskjellige tekstenhetene (m.a.o. hvor mange sammenfall hver enkelt enhet har med A.A. innenfor hver spredningsgruppe). Dessuten angis hvor stor prosent de observerte sammenfall utgjør av teoretisk mulige sammenfall innenfor de forskjellige spredningsgrupper.

SPREDNINGSTABELL

		M.L.		E.B.		F.T.		A.F.		K.B.		A.B.		I.A.	
S.	Tfr.	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%
8	184	184	100	184	100	184	100	184	100	184	100	184	100	184	100
7	177	174	98.3	171	96.6	172	97.2	173	97.7	118	66.7	91	51.4	163	92.1
6	213	199	93.4	190	89.2	190	89.2	195	91.5	79	37.1	65	30.5	148	69.5
5	216	172	79.6	157	72.7	161	74.5	166	76.9	58	26.9	42	19.4	108	50.0
4	238	128	53.8	136	57.1	122	51.3	152	63.9	35	14.7	43	18.1	98	41.2
3	266	90	33.8	93	35.0	79	29.7	110	41.4	19	7.1	41	15.4	100	37.6
2	392	57	14.5	61	15.6	65	16.6	65	16.6	19	4.8	25	6.4	100	25.5

Av tabellen fremgår det at sammenfallene i spredningsgruppene 7, 6 og 5 synes å være fordelt omtrent i forhold til de respektive

tekstenheters størrelse. Dog har A.F. og særlig F.T. forholdsvis flere sammenfall i gruppe 5 enn E.B. I spredningsgruppe 4 er det A.F. som utmerker seg med relativt mange sammenfall. Det samme er tilfelle for A.F. i gruppe 3, men det oppsiktsvekkende ved denne gruppen er at I.A. med bare halvparten så stor tekstmasse som A.F. oppnår nesten like mange sammenfall. Og i gruppe 2 er det I.A. som dominerer fullstendig med hele 100 sammenfall, mens F.T. og A.F. kommer nærmest med 65. M.L. som er den største tekstenheten har bare 57 sammenfall i denne gruppen. Av de større enhetene er det F.T. som har forholdsvis flest sammenfall i gruppe 2. Hva angår de minste enhetene, K.B. og A.B., så har K.B. overvekt i sammenfall i de høyere spredningsgruppene 7, 6 og 5, men i gruppene 4, 3 og 2 har A.B. klart flere enn K.B., en forskjell som er desto mer signifikant fordi A.B. er en adskillig mindre enhet enn K.B.

Hvis vi studerer tabellen vertikalt fra høy til lav spredning innenfor hver tekstenhet, ser vi at de fire største enhetene har en tydelig nedgang i antall sammenfall fra spredningsgruppe 5 til spredningsgruppe 2, dvs. at sammenfallenes antall synker med stigende totalfrekvens innenfor spredningsgruppene. For I.A. derimot ligger antall sammenfall på nokså nøyaktig samme nivå i gruppene 5, 4, 3 og 2. For A.B. kan man spore en lignende tendens, selv om det er en viss nedgang fra gruppe 3 til gruppe 2. K.B. oppviser en jevn nedgang til gruppe 4, mens gruppene 3 og 2 ligger på nøyaktig samme antall sammenfall.

Da substantiv med høy spredning ikke kan sies å være spesielt typiske for noen bestemt tekstenhet, med mindre det er tale om markerte frekvensavvik fra den ene enheten til den andre, vil de mest interessante sammenfallene være å finne i de lavere spredningsgrupper. I vårt tilfelle vil en spredning på 3 eller 2 for et substantiv bety at det må regnes som et karakteristisk element i Achmatovas og i en eller to av de andre dikternes individualstiler. Likheter og forskjeller mellom de undersøkte dikteres bruk av substantiver viser seg altså først og fremst blant substantiv med lav spredning.

Konklusjonen på denne tabellanalysen må da bli at Annenskij med hele 200 sammenfall i gruppene 3 og 2, så avgjort er den av dikterne som oppviser den største likhet med Achmatovas særpreg når det gjelder substantivforråd. Til sammenligning kan nevnes at M.L. med over dobbelt så stor tekstmasse som I.A., bare har 147 sammenfall i gruppene 3 og 2. E.B. har 154, F.T. 144 og A.F. 175. Av de store tekstenhetene ser det altså ut til å være A.F. som ligger nærmest A.A. når det gjelder karakteristiske stilelementer blant substantivene. På grunn av de små tekstmassene er det vanskelig å si noe sikkert om K.B. og A.B., men det relativt jevne nivå i sammenfall innenfor de laveste spredningsgruppene kan tyde på at det er disse to enhetene som nest etter I.A. har mest til felles med A.A. Hvis en sammenligning av Achmatova med større tekstenheter av Bal'mont og Blok bekreftet denne antagelsen, ville man med en viss rett kunne hevde at de diktere som ligger nærmest hverandre i tid, også oppviser de største likheter i ordforråd. Et moment som peker i samme retning, er at Tjutcev og Fet, hvis skapende perioder ligger nærmere Achmatovas enn Lermontovs og Baratynskijs, har forholdsvis flere sammenfall enn de sistnevnte i spredningsgruppene 3 og 2.

På det nåværende utviklingstrinn innenfor språklig databehandling innebærer en undersøkelse av den type som her er beskrevet også en god del manuelt arbeid, men uten adgang til EDB-produserte frekvensordlister og konkordanser ville den være nærmest uoverkommelig. Og etter hvert som det utvikles programmer som kan foreta automatisk analyse av de mer kompliserte strukturer i språket, er det ingen tvil om at omfattende ordforrådsundersøkelser vil kunne gjennomføres med et minimum av manuelt arbeid.

Steinar Gil:

OM TO KONKORDANSER

Produksjon av konkordanser er etter hvert blitt en dagligdags affære overalt hvor det drives språklig databehandling, og mange språk- og litteraturforskere har benyttet seg av muligheten til å få konkordanser i form av utskrift fra datamaskinens printer. Disse konkordansene er beregnet på intern bruk, noe som fører til at hver konkordans brukes av et fåtall forskere og at forskere som arbeider med de samme tekster rundt om i verden, sitter med hver sine «hjemmelagede» konkordanser. Denne situasjonen er imidlertid i ferd med å forandre seg. I løpet av de siste årene har det blitt utgitt et stort antall konkordanser i bokform.

Som eksempler på hvordan EDB-produserte konkordanser kan se ut og hvilke opplysninger de kan gi forskeren, skal jeg kort omtale to nylig utkomne konkordanser: *La Defence et Illustration de la Langue Francoyse,* Concordance etablie par Suzanne Hanon, Odense University Press, Odense 1974, og *A Concordance to the Poems of Osip Mandelstam,* ed. Demetrius J. Koubourlis, Cornell University Press, Ithaca and London 1974. Den førstnevnte konkordansen er basert på prinsippet KWIC (Keyword-in-context), som i korthet går ut på at oppslagsordene står i alfabetisk orden midt på siden med like lange kontekster til høyre og venstre. Innenfor hvert oppslagsord er kontekstene ordnet alfabetisk etter høyrekonteksten og deretter etter kronologisk forekomst i teksten. En slik oppsetning har den store fordel at den skiller ut mange stående uttrykk i språket og ordkombinasjoner som er typiske for forfatterens stil. Hver kontekstlinje er forsynt med henvisning til side- og linjenr. i den tekstutgave som ligger til grunn. På minussiden kan noteres at konkordansen hverken er lemmatisert eller homografseparert og at den heller ikke inneholder ordfrekvenser. Hvorvidt teksten i seg selv er så interessant og sentral at den rettferdiggjør utgivelse i bokform, får det bli opp til franskfilologene å avgjøre.

Hva Mandel'stam-konkordansen angår, så skulle den være sikret en ganske stor leserkrets. Mandel'stam er en sentral skikkelse i det 20. århundres russiske lyrikk, og interessen for hans diktning er større i dag enn noensinne. Alle Mandel'stamforskere vil utvilsomt hilse konkordansen velkommen, så meget mer som den bærer preg av

pålitelighet og godt redigeringsarbeid. Konkordansen har et instruktivt forord, hvor det bl.a. gjøres rede for hvordan problemer som lemmatisering og homografseparering er løst. Tekstgrunnlaget for konkordansen er bind I av den reviderte utgave av *Collected Works in Three Volumes,* edited by G. P. Struve and B. A. Filipoff, som er den beste og fyldigste utgave til dags dato. Konkordansen er en såkalt «cluster concordance», dvs. at alle ordformer som hører til samme lemma er oppført etter hverandre, noe som selvsagt letter ordsøkingen i vesentlig grad. Kontekstenheten er én verselinje med en begrensning på 59 karakterer (inklusive mellomrom). Linjer på mer en 59 karakterer er altså avkappet. Hver side består av to spalter. I venstre spalte står først oppslagsordene (dvs. ordformene slik de forekommer i teksten) for seg med frekvensangivelse, deretter følger kontekstlinjene med angivelse av sidenr. i tekstutgaven. I høyre spalte står første linje av diktene hvor oppslagsordene forekommer, med angivelse av linjenr. for forekomsten. Opplysningene om første linje i diktene er av stor verdi for forskere som ikke er i besittelse av den Mandel'stamutgave som er lagt til grunn for konkordansen.

Redaktøren har valgt å ta med alle diktvarianter og linjevarianter i tekstgrunnlaget. Ord som forekommer i variantene er merket (v). Det har selvfølgelig sin interesse hvilke ord som er brukt i variantene og på hvilken måte de er brukt, men samtidig gjør innlemmelsen av variantene det vanskelig å foreta frekvenssammenligninger med andre konkordanser, hvor variantene ikke er tatt med. Man må i så fall trekke fra alle kontekstlinjer merket (v). En bedre løsning ville ha vært å samle variantene i en egen «minikonkordans». Til slutt i konkordansen finnes en frekvensordliste over alle ordformer hos Mandel'stam, sortert etter fallende frekvens. Denne listen egner seg igjen dårlig som grunnlag for frekvenssammenligninger, fordi variantfrekvensene er med også her. Et annet minus ved konkordansen er at man bare får en omtrentlig angivelse av det totale antall verselinjer, ord og ordforekomster. I en fotnote står det at teksten i alt besto av litt mindre enn 9000 linjer, variantene medregnet, ca. 16000 ord og ca. 41000 ordforekomster.

En konkordans vil alltid kunne gjøres bedre. Man kunne f.eks. tenke seg konkordanser som inneholdt opplysninger om ordklassefrekvenser, ordenes spredning i teksten osv. Men også konkordanser av den type jeg har omtalt her, kan være et uvurderlig hjelpemiddel for forskeren, eller som Koubourlis sier det i forordet til Mandel'stam-konkordansen: «the concordance user need walk no more: he is able to fly».

KURS OG SEMINAR

BERGEN
NAVF's EDB-senter gir følgende undervisning i vårsemesteret 1976: Et innføringskurs i databehandling for humanistiske forskere og studenter (februar). Et kurs i bruken av de tekstbehandlingsprogrammer som er utviklet ved senteret (mars).

TRONDHEIM
NAVF's EDB-konsulent i Trondheim, Eirik Lien, vil i løpet av vårsemesteret 1976 gi et kurs i NU-ALGOL for humanister med tilhørende øvingsopplegg.

OSLO
NAVF's EDB-konsulent i Oslo, Ivar Fonnes, vil i første halvår av 1976 gi et elementært innføringskurs i programmering og samarbeide med en hjelpelærer om et kurs i programpakken TEXT.

.....

Nærmere opplysninger om kursene blir gitt ved oppslag på instituttene.

I Bergen, Trondheim og Tromsø vil det dessuten bli gitt brukerveiledning i tekstsøkesystemet STATUS.

LOGOTEKET

En text-och ordbank i språkbehandlingens tjänst

Logoteket.

Logoteket (egentligen «ordmagasin») har till uppgift att samla in, lagra och tillhandahålla maskinläsbara texter och textbearbetningar och att bygga upp en svensk ordbank. Denna verksamhet etablerades vid halvårsskiftet 1975 som ett nationellt serviceorgan med placering vid Göteborgs universitet. Serviceverksamhet efter dessa linjer hade då pågått i ett tiotal år inom den forskningsgrupp som kom att bilda stommen till Logotekets moderorgan, avdelningen för språklig databehandling (Språkdata).

Vad kan Logoteket erbjuda?

Logoteket disponerar över dels texter omfattande bland annat skönlitteratur, facklitteratur och tidningsartiklar, dels ordboksmaterial. Dessutom finns dels program för bearbetning av dessa texter och ordsamlingar, dels maskinkapacitet vid avdelningens datoranläggning, som är specialutrustad för avancerad textbehandling. Logoteket kommer att förfoga över ett växande och mer varierat lager av maskinläsbara texter (innsamling av sättremsor och magnetband från landets tryckerier pågår kontinuerligt). Förteckningar över tillgängliga texter publiceras med jämna mellanrum. Texterna kommer i sin tur att bilda underlag för en ordbank som registrerar den moderna svenskans skiftande ordförråd. Grunden till detta arbete har redan lagts vid avdelningen genom en större undersøkning av svenskt tidningsspråk, publicerad i Nu-svensk frekvensordbok (hittills tre volymer). Flera andra lexikaliska material har senare tillkommit. Uppsättningen av standardprogram för textbehandling - ordindex, konkordanser, fraseologiska listningar, frekvenslistor m.m. i olika sorteringsordning - vidgas efterhand. I Logotekets service ingår slutligen också råd och anvisningar i samband med uppläggning av textundersøkningar.

Vem kan utnyttja Logotekets service?

Logotekets material och tjänster står till förfogande för interesserade användare inom forskning och utbildning, förvaltning och näringsliv. Texter och bearbetningar tillhandahålls kostnadsfritt för icke-kommersiella ändamål. För maskintid, material och eventuell specialprogrammering måste dock Logotekets omkostnader täckas (vilket inom det vetenskapliga fältet normalt kan ske med hjälp av anslag). Råd och anvisningar ingår inom rimliga gränser i Logotekets allmänna service. Utnyttjandet av Logotekets material regleras i ett standardavtal som bland annat tar hänsyn till reglerna för upphovsrätt.

Vem vänder man sig till?

Föreståndare för Logoteket är Sture Allén, professor i språklig databehandling. Den fortlöpande serviceverksamheten sköts av docent Martin Gellerstam. Ansvarlig för databehandlingen är driftledare Rolf Gavare.

GRANADA

Et verktøy for kontroll og oversetting av tekst

Kristen Rekdal, Thorild Wessel

1. Muligheter for enklere programmering

Mange som ønsker å bruke datamaskinen som et verktøy i sitt arbeid, opplever at løsning av tilsynelatende enkle problemer krever uforholdsmessig mye programmering. Programmeringsinnsatsen begrenser dermed hvilke oppgaver maskinen kan brukes til.

I forhold til programmering i maskinkode, har eksisterende høynivå programmeringsspråk, som ALGOL og FORTRAN, alt gitt en vesentlig forenkling. Ved Regnesentret ved Universitetet i Trondheim (RUNIT) har det lenge vært arbeidd for å forenkle programmeringen enda mere. GRANADA-systemet er et resultat av dette arbeidet.

2. GRANADA — et erklærende språk

GRANADA står for GRAmatikalsk Non-Algoritmisk DAtabeskrivelse. Språket er rettet mot det å beskrive struktur i tekster og oversetting mellom tekster. Ved oversetting f.eks., beskrives bare hvordan inn-data og ut-data ser ut, ikke hvordan algoritmen for oversetting skal være. Beskrivelsen er ikke-algoritmisk eller statisk.

GRANADA har ikke som ambisjon å kunne beskrive naturlige språk fullstendig. Det er derfor ikke mulig å lage en fullstendig oversetter fra f.eks. russisk til norsk med dette systemet. Men det er likevel nok av enklere tekststrukturer som er interessante å beskrive.

En tekstbeskrivelse består av et sett definisjoner eller erklæringer. Disse kan også kalles en grammatikk. Hver definisjon består av ei høyreside som definerer en tekst eller mengde av tekster og ei venstreside som navngir den definerte teksten.

Eksempel:

```
hovedstad='Oslo';
```

Denne definisjonen sier at begrepet eller navnet *hovedstad* er definert ved den teksten som består av bokstavene 'O' 'S' 'L' 'O' etter hverandre.

Dersom flere tekster faller inn under dette begrepet, kan vi skrive:

```
hovedstad='OSLO'/'STOCKHOLM'/
'HELSINKI';
```

skråstreken leser 'eller'. Definisjonen sier at en *hovedstad* er enten OSLO eller STOCKHOLM eller HELSINKI.

3. Tekstdefinisjoner brukt til datakontroll

En anvendelse for slike tekstdefinisjoner er å lage program for kontroll av data. For å fortsette eksemplet fra pkt. 2 anta at vi har ei lang liste over navn på hovedsteder. Denne blir stanset på hullkort eller hullband og vi ønsker å kontrollese dataene etterpå.

En grammatikk som definerer et slikt datasett er:

```
Hovedsteder=xhovedstad;
hovedstad='OSLO'/'STOCKHOLM'/
'HELSINKI';
```

'x' betyr 'en sekvens av'. GRANADA-systemet kan ta en slik grammatikk og direkte generere et kontrollprogram. Programmet vil lese dataene og kontrollere at bare ordene 'OSLO', 'STOCKHOLM' eller 'HELSINKI' forekommer. Dersom det ved en feil er blitt stanset f.eks. 'OLSO', vil det bli gitt feilmelding.

4. Tekstdefinisjoner brukt til oversetting.
Ved tekstbehandling er vi sjelden fornøyd med bare å kontrollere at en tekst er i samsvar med en definisjon. Vi skal ikke bare lese data, men vi må gi dem ut igjen, gjerne i en annen form enn de ble innlest i. Vi trenger oversetting.

Et typisk eksempel er når tekst skal registreres på utstyr med for lite tegnsett. F.eks. tekst med store og små bokstaver, spesialtegn og diakritika skal stanses på en vanlig hullkort stans. Da må de ikke eksisterende tegn representeres ved kombinasjoner av de tilgjengelige tegn. Vi kan f.eks. angi stor bokstav ved å sette tegnet '#' foran. Utropstegn kan vi representere ved $UTROPSTEGN.

Skal denne teksten så skrives ut på en skriver med fullt tegnsett, ønsker vi å oversette for å få ei pen utskrift.

Teksten 'Han sa: "Gå!" ' må innkodes som '#HAN SA$KOLON$APOSTROF#GÅ $UTROPSTEGN$APOSTROF'.

Vi ønsker nå å lage et program som oversetter den kodete teksten tilbake til opprinnelig form. Dette gjør vi ved å lage to samhørende grammatikker, en for kodet tekst som leses inn og en for klartekst som skal skrives ut. For lettere å se samsvaret mellom grammatikkene er de satt opp ved sida av hverandre.

```
Inngrammatikk                    Utgrammatikk
tekst=                                 tekst=
x('#' storbokstav/                     x(storbokstav/
'$' spesialtegn/                         spesialtegn
litenbokstav/                          litenbokstav/
mellomrom);                            mellomrom);
storbokstav='A' — 'Å';           storbokstav='A'
                                         — 'Å';
spesialtegn=                           spesialtegn=
  'UTROPSTEGN'/                          '!'/
  'KOLON'/                               ':'/
  'APOSTROF';                            ''''
litenbokstav=                          litenbokstav=
  'A' — 'Å';                             'a' — 'å';
mellomrom=' ';                         mellomrom=' ';
```

Dette er alt programmereren må gjøre. GRANADA-systemet kan nå lese disse grammatikkene og lage oversetterprogrammet.

5. GRANADA tilgjengelig på UNIVAC 1100 datamaskiner
De enkle eksemplene foran demonstrerer grunnprinsippene i GRANADA. Det vil føre for langt å beskrive alle mulighetene her. Interesserte kan henvende seg til RUNIT. Arbeidet med systemet pågår fortsatt for å øke fleksibiliteten og uttrykksmulighetene.

GRANADA-systemet er programmert for UNIVAC 1100 datamaskiner. Foreløbig er det bare tilgjengelig på UNIT's maskin i Trondheim.

TEKSTSØKESYSTEMET STATUS

Siden meldingen om tekstsøkesystemet STATUS ble skrevet (se side 11) har arbeidet gått videre:
Jan Olav Hauge, Universitetet i Tromsø og Sigbjørn Århus, NAVF's EDB-senter har i løpet av desember foretatt en implementering av tekstsøkesystemet på UNIVAC 1110, Universitetet i Bergen. Systemet blir nå uttestet. En brukerveiledning vil foreligge i løpet av vinteren.
Nærmere presentasjon av tekstsøkesystemet blir gitt i neste nummer av Humanistiske Data.

Svein Lie

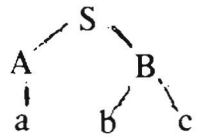
DATAMASKINELL SYNTAKTISK ANALYSE

Det har i lengre tid vært drevet forsøk med å analysere setninger automatisk ved hjelp av datamaskin. Vi skal her ta for oss et par ulike måter å angripe dette problemet på.

En analyse forutsetter en modell. Den modellen som i dag faller en lingvist først i tankene, er nok transformasjonell generativ grammatikk (TG-grammatikk). Denne grammatikken har *frasestrukturregler* (eller *omskrivningsregler)* som f.eks.:

S → A B
A → a
B → b c

S er startsymbol, og a, b og c er terminale symboler. Strukturen kan framstilles grafisk i form av et tre:



Reglene kan også være rekursive, f.eks.:

S → A B
A → a S
. . .

Transformasjonsgrammatikken har også transformasjonsregler som omgjør en syntaktisk struktur til en annen, f.eks.:

A + b + c ⇒ c + A + d + b

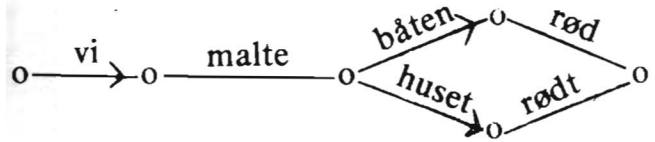
I tillegg finnes også et leksikon (ordliste), hvor vi bl.a. finner opplysninger om ordklassetilhørighet.

Vi skal se på noen problemer som melder seg når vi vil bruke denne grammatikken til analyse. Vi kan tenke oss to måter å bruke grammatikken på. For det første kan vi løpe gjennom frasestrukturreglene og transformasjonsreglene (og leksikonet) og produsere alle tenkelige setninger for deretter å se om noen av disse er identisk med den setningen vi skal analysere. Men siden antall setninger som kan produseres av en slik grammatikk, er uendelig, vil en slik strategi være ubrukbar av både teoretiske og praktiske grunner.

En annen måte å gå fram på kunne være å starte med den setningen vi skal analysere og så arbeide oss «bakover», fra de terminale symbolene gjennom transformasjonsreglene og frasestrukturreglene og se om vi endte opp i startsymbolet (i vårt eks. S). I så fall måtte vi snu reglene på hodet. Å gjøre det kan høres trivielt ut, men er ikke så enkelt som en skulle tro. Vanlige transformasjonsregler tar en struktur og omdanner til en annen struktur. Men det som vi har som utgangspunkt og skal analysere, er bare en streng med ord som er ustrukturert. Skal vi ha noe å anvende de «omvendte» transformasjonsreglene på, må vi altså vite noe om strukturen i setningen vi har foran oss. For å få vite noe om det må vi ha en slags overflategrammatikk som gir denne strukturen. Når vi har funnet den, kan vi løpe gjennom de omvendte transformasjonsreglene og frasestrukturreglene for å se om vi ender opp i et startsymbol.

Et annet problem med omvendte transformasjoner er at de som regel er frivillige. Vanlige transformasjonsregler er derimot ofte obligatoriske, slik at bare én operasjon er mulig. For om en gitt struktur er slik at den er resultat av en obligatorisk transformasjon, kan vi ikke vite om den kunne ha oppstått også på annen måte. Derfor må de omvendte transformasjonene være frivillige, noe som fører til at det blir en kraftig økning i antall veier å gå i søkningen.

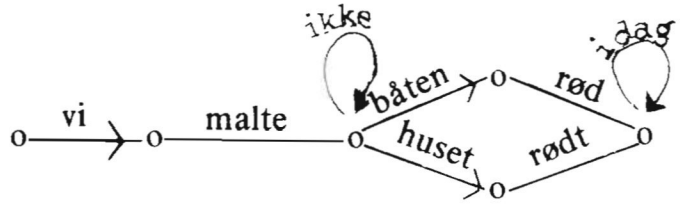
En modell som er mer innrettet på analyse, er en finite-state-grammatikk. En slik grammatikk kan representeres grafisk slik:



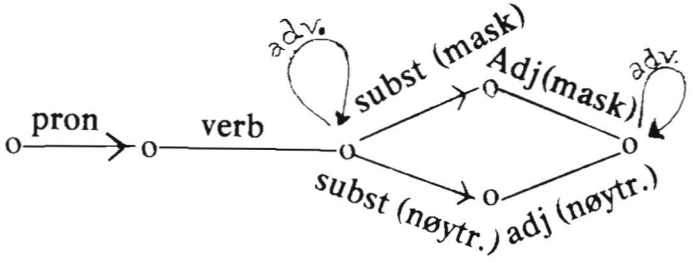
Dette er en grammatikk for et språk med to setninger:(1) Vi malte båten rød
og
(2) Vi malte huset rødt.

Vi kan tenke oss grammatikken brukt til analyse slik:
Vi tar første ord i setningen og ser om det er identisk med første symbol i grammatikken (*vi*). Hvis så er tilfelle, går vi videre til neste ord. Ved tredje ord, har vi to veier å gå, Hvis vår setning har *båten* som tredje ord, følger vi øverste vei i diagrammet. Hvis fjerde ord er *rød*, kommer vi til sluttsymbolet. Vår setning er altså en grammatisk setning som er beskrevet i grammatikken. Hvis fjerde ordet i vår setning hadde vært *rødt* derimot, vil analysen stoppe opp, for det går ingen vei fra *båten* til *rødt* i grammatikken.

Denne enkle grammatikken kan kompliseres ved at en legger inn løkker for valgfrie elementer i setningen:



I stedet for ord vil en i praksis heller bruke ordklasser eller andre symboler i grammatikken:

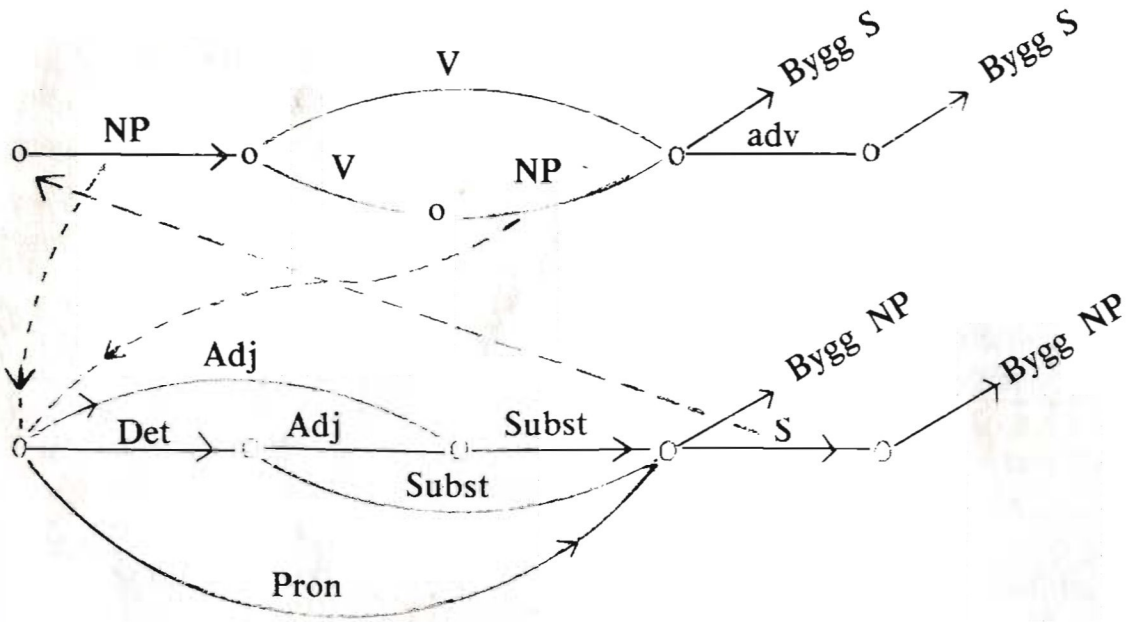


En slik finite-state-grammatikk er, som påvist bl.a. av Chomsky *(Syntactic Structures*, 1957), ikke tilstrekkelig til å beskrive naturlige språk. En vesentlig svakhet er at vi ikke kan stoppe analysen underveis og så bruke den samme grammatikken til å analysere en innføyd konstituent, f.eks. en leddsetning. Men en slik rekursiv analyse kan en få til ved å bygge ut grammatikken. En kan tillate som navn på elementer ikke bare navn på ordklasser, men også navn på komplekse ledd (som f.eks. NP = nominal) som må være til stede før analysen kan gå videre. Om vi har et slikt komplekst ledd, avgjør vi ved å «kalle opp» et annet nettverk (eller det samme). Bygger vi ut modellen på denne måten, blir den ekvivalent med en kontekstfri grammatikk (eller en «pushdown store automat«). Men den kan bygges ut til å bli like kraftig som en kontekstsensitiv grammatikk (eller en Turing-maskin) hvis en legger inn både betingelser for å gå videre med analysen og spesielle byggeprosedyrer.

La oss si at vi skal analysere setningen
(3) Den engelske boka du gav meg, ligger her (jfr. første eksempel på side 28).

Vi starter i det øverste diagrammet og ser om vi har en NP (nominal). For å finne ut det må vi kalle opp nettverket for NP. Vi ser der at første ord i en NP skal være et adjektiv, et bestemmerord (Det(erminer)) eller et pronomen. *Den* er et bestemmerord, og vi fortsetter. Neste ord er et adjektiv eller substantiv, *engelsk* er et adjektiv. Så skal vi ha et substantiv, noe vi har *(boka)*. Nå kan vi «bygge» en NP eller fortsette med en S (setning) før vi bygger NP. Hvis vi bygger NP nå og hopper tilbake til det øverste diagrammet, ser vi (av det øverste diagrammet) at neste ord skal være et verb. Men i vår setning er neste ord et pronomen. Analysen stopper opp, og vi må tilbake og prøve en annen vei. Vi fortsetter da i nettverket for

Et eksempel:

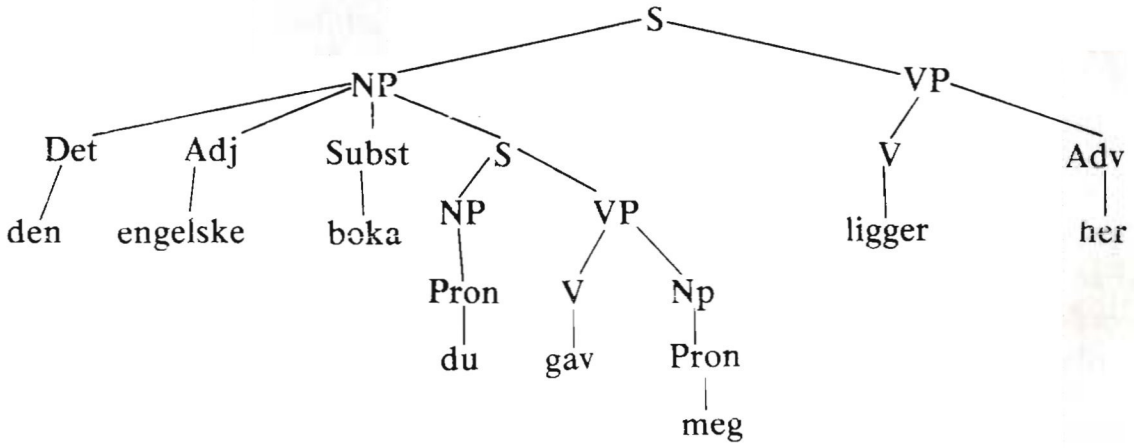


NP og ser om vi har en S (setning) etter substantivet. For å se det må vi opp igjen i det øverste nettverket. Vi ser der, uten å gå i detaljer, at *du gav meg* tilfredsstiller kravet til en setning. Vi bygger da en setning, hopper ned igjen og bygger en NP av *den engelske boka du gav meg*. Så hopper vi opp igjen. Vi sammenlikner neste ord i setningen *(ligger)* med neste ledd i skjemaet (verb). Etter verbet kan vi så ha NP eller adverb, eller setningen kan være slutt. Vi har adverb (*her*). Vi kan da gå til ordren *Bygg S*, og analysen har lykkes.

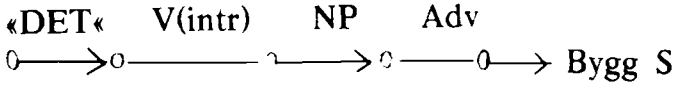
Den bygge-ordren vi har sett eksempler på, har som oppgave å angi strukturen i det som er analysert. I vårt eksempel kunne vi tenke oss den formulert slik at vi ville få bygd:

S(NP (Det den) (Adj engelske) (Subst boka) (S (NP (Pron du) (VP (V gav) (NP (Pron meg))))) (VP (V ligger) (Adv her))

eller i form av et tre:

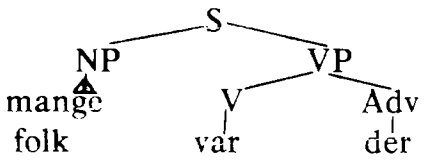


Et annet eksempel på hvordan bygge-ordren fungerer:
En setning som
(4) Det var mange folk der
vil bli analysert av flg. nettverk:



Skal vi få fram dypstrukturen her, må bygge-instruksjonen være omtrent slik: (noe forenklet):

Bygg S (NP mange folk) (VP (V var) (Adv der))
som tilsvarer flg. trestruktur:



Vi ser av disse eksemplene at vi kan få fram dypstrukturen direkte gjennom bygge-ordren uten å gå veien om en egen transformasjonskomponent med omvendte transformasjoner. I bygge-instruksjonen kan vi nemlig både flytte, stryke og legge til elementer.

Denne siste modellen er utformet av W.A. Woods (MIT, USA) og kalles gjerne Augmented Transition Network (ATN). Den er beskrevet bl.a. i W.A. Woods: *Transition Network Grammar for Natural Language Analysis* i: Communication of the ACM, vol. 13, Oct. 1970. Det er et av de svært få automatiske analysesystemer som er brukt med hell i praksis (i forbindelse med månelandingsprosjektet). Samme grunnmodell, men med visse modifikasjoner og utbygginger, har vært brukt av andre, bl.a. Martin Kay og Ron Kaplan (begge USA).

DATAMASKINEN I HISTORISK FORSKNING

Tidsskriftet *Historical Methods Newsletter* utga i juni 1974 (vol. 7 no. 3) en spesialutgave om datamaskinen i historisk forskning.

I bladet presenteres foredrag som ble holdt på konferansen «History and the Computer» i Uppsala i juni 1973, og det inneholder ifølge utgiveren «articles dealing with theory and conceptualization, essays on particular techniques, and papers featuring the results of substantive research».

Historical Methods Newsletter utgis av *University Center for International Studies and the Department of History. University of Pittsburgh.*

Det 4. internasjonale symposium om datamaskinen i språk- og litteraturforskning, Oxford, England 5. til 9. april 1976.

I tiden 5. til 9. april 1976 vil det bli holdt et 4. internasjonalt symposium i Oxford i England om datamaskinen i språk- og litteraturforskning.

Emneomfanget vil avhenge av det foredragstilbud som en får i stand.
Av relevante emner kan nevnes: klassiske studier, konkordanser, informasjonssøking, inn-og utdataproblemer, leksikografi, lingvistikk, metriske studier, moderne og orientalske språk, programutrustning, stilistisk analyse, syntaktisk analyse, tekststudier, tematisk analyse, vokabularstudier.

Nærmere opplysninger om symposiet og påmeldingsblankett for deltaking og eventuelt foredrag kan fås ved henvendelse til NAVF's EDB-senter.

KURS I KVANTITATIV HISTORIE

DANMARK 4.-16. AUGUST 1975.

Under «Nordiske forskerkurser» ble det i august i år arrangert et kurs i kvantitativ historie i Danmark — en uke på Sandbjerg i Sønderjylland og en uke ved Odense universitet. Leder for kurset var professor Hans Chr. Johansen ved Odense universitet. Kurset hadde ca. 30 deltakere fra Norge, Sverige, Danmark og Finland.

Kursopplegget fulgte et fast skjema:

- — en daglig dobbeltime i statistisk teori
- — en daglig dobbeltime i gjennomgåelse av historisk litteratur som anvender kvantitative metoder
- — en daglig dobbeltime i elementær EDB

I statistikken fikk man gjennomgått mål for gjennomsnitt og spredning, utvalgsteori og estimering, regresjonsanalyse og såvidt hypoteseprøving. Til tross for den korte tid som var til rådighet, fikk deltakerne en forholdsvis grundig innføring i de temaer som ble behandlet. Det er grunn til å anta at kurset i så måte ga et godt grunnlag for videre arbeid med kvantitative metoder. Dette skyldes ikke minst foreleserens (H.C. Johansens) pedagogiske framstillingsmåte og hans sikre grep på muligheter og begrensinger i metodenes anvendelse på historiske data.

Tekstgjennomgåelsen av historisk litteratur (ved W.P. Kennedy, England og Jørgen Elklit, Århus) ga et godt innblikk i problemer som man støter på ved bruk av kvantitative metoder på historisk materiale. I en del tilfeller var det imidlertid nokså kompliserte modeller som ble behandlet, slik at deltakerne nok av og til hadde problemer med å trenge inn i stoffet.

EDB-undervisningen bestod i et elementærkurs i FORTRAN og øvelser i bruk av programpakker, særlig da statistikkpakken SPSS. Det var et pedagogisk meget godt innføringskurs i FORTRAN (Johansen), men man kunne kanskje samordnet kurset, som ble holdt første uke, noe bedre med de praktiske øvelser som var lagt til den siste halvdel av oppholdet.

Totalt sett er det grunn til å regne med at deltakerne satt igjen med et betydelig utbytte av kurset.

Ivar Fonnes

Litterær statistikk

Rapport fra internasjonal sommerskole i litterær statistikk, Cambridge.

Den første sommerskole i litterær statistikk ble i sommer avholdt i Cambridge. Literary and Linguistic Computing Centre, University of Cambridge var arrangør, og arrangementet var støttet av den internasjonale forening for litterær og lingvistisk databehandling (ALLC) som ble startet i 1973. Litterær statistikk er en forholdsvis ny forskningsaktivitet og den voksende aktivitet henger tydelig sammen med den økende bruk av datamaskiner i språk- og litteraturforskning. Men i likhet med EDB er statistikk ikke vanlig i humanistisk fagutdannelse. På samme måte som man i lengre tid har drevet med sommerskoler i språklig databehandling, har man nå startet med sommerskole i litterær statistikk.

Behovet for et slikt kurs i litterær statistikk ble demonstrert ved at det kom deltakere både fra Canada, USA, Tyskland, Frankrike, Nederland og Skandinavia, foruten England. Tilsammen var det ca. 35 deltakere. Temaet var forberedt ved at ALLC Bulletin i 6 nummer hadde hatt en artikkelserie om Literary Statistics av Norman Thomson. Thomson var også hovedforeleser på kurset. Han hadde bl.a. laget et øvelseshefte som ble brukt i undervisningen med mange gode, faglig relevante eksempler.

Dagsprogrammet var lagt opp slik at det vekslet med forelesninger og gruppearbeid under veiledning av lærerne. Dette er en fin læremåte forsåvidt som man raskt får anledning til å anvende den teoretiske kunnskap som blir ervervet gjennom forelesningene. Programmet var tett, og man hadde også belagt en del av kveldene med forelesninger.

Foruten Thomson, som ga den grunnleggende teoretiske undervisning, var kurset lagt opp med utstrakt bruk av gjesteforelesere, minst 1 hver dag, som skulle gi mer konkret forståelse for anvendelsen av statistikk i

litterær sammenheng. Gjesteforelesningene var på en pedagogisk riktig måte samordnet med de vanlige forelesningene på den måten, at de som krever mest statistisk innsikt kom til slutt. Mange av disse gjesteforelesninger var særdeles verdifulle, spesielt vil jeg for min del fremheve forelesningene av prof. K.W. Kemp (Cardiff), «Statistics and Linguistic Analysis in Perspective: A Personal view», og Dr. D. Wickmann (Aachen), «Authorship Identification».
Til glede for interesserte som ikke kunne delta på dette kurset, vil disse gjesteforelesningene bli trykket etter tur i ALLC Bulletin. Kemp talte om metodiske spørsmål og i forbindelse med bruk av statistiske tester viste han hvordan man, i forhold til tidligere praksis, kunne få en forbedret anvendelse av disse i litterær sammenheng. Wickmann redegjorde for statistisk behandling av ordstilling som forfatteridentifiserende kriterium.

Andre gjesteforelesere var professor Y. Radday (Haifa), «The use of the «The» in Authorship Identification», prof. R. W. Bailey (Michigan), «Statistical Methods», prof. S. Michaelsen (Edinburgh), «Justification by Faith» og Mr. H. Sykes-Davies (Cambridge), «Structure of Vocabularies». Sykes-Davies som beskjedent betegnet seg selv som «an amateur word-watcher» kom bl.a. med en advarsel mot ukritisk bruk av frekvensordlister og illustrerte det med bruken av ordet «naked» hos Wordsworth. Ordet hadde en påfallende høy frekvens hos Wordsworth, og man kunne kanskje fristes til å trekke visse slutninger om arten av hans forfatterskap ut fra dette. Ved nærmere inspeksjon viser det seg imidlertid at adjektivet «naked» vanligvis ble anvendt sammen med helt andre substantiver enn det som ville være vanlig i dag, som man kan se av følgende eksempel: «he saw the woman in the naked room».

Gjesteforelesningene var med på i betydelig grad å øke utbyttet fra denne sommerskolen. Noe referat fra selve undervisningen har det liten hensikt å gi seg inn på, forsåvidt som det var regulær statistikkundervisning i tilknytning til R.M. Cormack's bok «The Statistical Argument». Oliver & Boyd, Edinburgh, 1971. Vanlige statistiske tema som populasjon og utvalg (samples), sannsynlighet, de vanligste matematiske modeller, og deres anvendelighet, og forskjellige former for hypotesetesting ble gjennomgått. Fra en pedagogisk synsvinkel kan man kanskje stille spørsmåltegn ved omfanget av stoff som ble presentert, det ble nemlig aldri tid for foreleseren til å utdype stoffet. Det kunne det ikke bli tid til når man ville gjennomgå en hel lærebok, samtidig som man bare hadde knapt en time på hvert kapittel. Man la imidlertid mer vekt på å skape forståelse for muligheten og begrensningen i statistisk argumentasjon enn på opplæring i statistisk teknikk. Deltakernes bakgrunn var også svært forskjellig, og det gjorde det heller ikke særlig enkelt for foreleseren.

I gruppearbeidet ble imidlertid deltakerne inndelt etter forhåndskunnskaper i statistikk og det var en klar fordel i forbindelse med det praktiske arbeid med statistikken, og det arbeidsheftet som Thomson hadde utarbeidet viste seg svært nyttig.

For de som var interessert i programmeringsspråket APL ble det arrangert et ettermiddagskurs i bruken av det. Thomson mente at nettopp APL var uovertruffent som programmeringsspråk i forbindelse med litterær statistikk. Forøvrig var det omvisninger på det lokale EDB-senter og vi fikk en presentasjon av det arbeidet som pågikk ved Literary and Linguistic Computer Centre i Cambridge.

Denne sommerskolen i litterær statistikk var den første i sitt slag og den bør absolutt bli en regelmessig foreteelse for den viste seg å dekke et behov blant humanistiske forskere, og inspirerte deltakerne til videre studium og bruk av statistikk i litterær sammenheng.

Roald Skarsten

Ivar Fonnes.

KVANTITATIV INNHOLDSANALYSE

Noen inntrykk fra International workshop on content analysis. Pisa, september 1974.

.....

red.anm.:

Som en vil se er dette en rapport fra en konferanse som ble holdt for over 1 år siden. Beklageligvis kom den ikke med i forrige nummer av Humanistiske Data. De opplysninger som gis om kursets emner og opplegg, har imidlertid fremdeles aktualitet. En rapport fra arrangøren er ventet i høst. Interesserte kan kontakte Fonnes.

.....

Den engelske termen «content analysis» brukes vanligvis om det vi kunne kalle *kvantitativ innholdsanalyse*. Hensikten er å gi kvantitative utsagn om innholdet i en tekstmasse. Slike oversikter gis gjerne i form av enkle tabeller, men ofte brukes det også avansert og komplisert statistikk. Et *skjematisk* eksempel kan være som følger: Tekstmaterialet deles inn i *analyseenheter*, analyseenhetene tildeles *kategorier* på grunnlag av bestemte kriterier. Man får så talt opp antall forekomster av de ulike kategorier. Slike frekvenser, stilt opp i en tabell og f.eks. fordelt på forskjellige enheter i materialet, danner grunnlaget for forskerens analyse (et eksempel på en slik undersøkelse finnes i Computers and the Humanities No. 1/74, pp.5-19).

Den internasjonale konferansen i Pisa, som ble arrangert av ISSC (International Social Science Council), hadde som formål å drøfte problemer knyttet til kvantitativ innholdsanalyse — metoder og databehandling. Konferansen hadde samlet ca. 70 deltakere fra forskjellige kanter av verden — USA, Venezuela, Australia, India, Øst- og Vest-Europa. De fleste deltakerne hadde sin faglige bakgrunn i innholdsanalyse, og mer perifere eller ingen kunnskaper i databehandling. Konferansen var da også lagt opp med henblikk på dette. Det meste av tiden var avsatt til diskusjon av metodiske problemer i innholdsanalyse. Videre ble det holdt små, elementære kurs i databehandling — ferdigprogrammer og programmeringsspråket PL/1 — med muligheter for terminalkjøring om kveldene. Endelig opererte man med en såkalt «teknisk gruppe» for de deltakerne som arbeider med utvikling av programmer/systemer for innholdsanalyse.

Vi kan altså dele inn aktivitetene i 3 hovedområder, og jeg skal gi en kort rapport fra hver av dem.

1. Drøfting av metodiske problemer
var hovedsaken på konferansen og foregikk både i plenum og i arbeidsgrupper. I arbeidsgruppene presenterte de enkelte deltakere egne undersøkelser som grunnlag for diskusjon. I plenum ble mer generelle metodiske problemer behandlet.

Spørsmålet om validitet er et av de sentrale problemer innen denne form for innholdsanalyse, og det stod også sentralt på konferansen. Det ble bl.a. redegjort for omfattende validitetstester i tilknytning til utarbeidelsen av den nyeste versjon av den såkalte «Harvard IV dictionary». Denne versjonen er utarbeidet i Australia og er en videreføring av tidligere kategoriordlister som er knyttet til den amerikanske programpakken General Inquirer (jfr. nedenfor).

Et annet sentralt spørsmål gjaldt tilnærmingsmåte/analyseretning: på den ene side analyse basert på en teori, vanligvis nedfelt i en kategoriliste (Dictionary), på den annen side en rent empirisk tilnærmingsmåte med utgangspunkt i selve tekstmaterialet uten noen apriori teori, og oftest med bruk av avansert statistikk (cluster-analyse, faktoranalyse). Begge disse retninger hadde sine talsmenn på konferansen, men hovedvekten lå nok på den første.

Det ble også pekt på at sosiolingvistisk forskning kan bidra til å løse validitetsproblemer innen innholdsanalyse. Forøvrig ble det understreket at man bør kunne nyttiggjøre seg resultater av den forskning som foregår innen syntaktisk og semantisk analyse, samt såkalt «kunstig intelligens».

En redaksjonskomité fikk som oppgave å utarbeide en sluttrapport fra konferansen med utgangspunkt i framlagte dokumenter og drøftinger i arbeidsgrupper og plenum. Foreløpig er ikke denne rapporten kommet.

2. Databehandlingskursene
— som gjaldt bruk av enkle ferdigprogrammer og programmeringsspråket PL/1 — var lagt opp med sikte på terminalbruk, og det var 6 terminaler til disposisjon for deltakerne hver kveld. Dette var naturligvis en relativt gunstig arbeidssituasjon, og deltakerne benyttet seg av tilbudet i nokså stor utstrekning. Det er derfor grunn til å regne med at de fikk et visst innblikk i en del muligheter og problemer ved datamaskinell tekstbehandling.

Det som imidlertid gjerne mangler på slike kurser — og som så vidt jeg kan bedømme det også manglet i Pisa — er et skikkelig perspektiv på de kunnskaper man tilegner seg. Man læres opp i bruk av et bestemt system, knyttet til en bestemt installasjon, men man får lite eller ingen informasjon om i hvilken grad det man lærer er avhengig av en bestemt maskin og bestemte programsystemer, og hvilke komponenter som eventuelt vil kunne brukes på andre installasjoner. Det er derfor grunn til å frykte at deltakerne kan bli både forvirret og frustrert om de forsøker å nyttiggjøre seg den tilegnede kunnskap når de kommer hjem.

3. I «teknisk gruppe»
fikk man presentert forskjellige programsystemer som er utviklet med sikte på innholdsanalyse og mer generell tekstbehandling. Det finnes en rekke slike programpakker både i USA og i Europa. Sentralt blant disse står fremdeles *General Inquirer*, som er utviklet spesielt med henblikk på innholdsanalyse ved hjelp av kategoriordlister. I tilknytning til General Inquirer er det bl.a. utviklet rutiner for homografseparering («disambiguation routines») i engelsk. Disse rutinene ser ut til å kunne behandle en meget stor del av homografene i engelsk, i størrelsesorden over 90% av løpende ord.

Det ble planlagt å lage en samlet oversikt over eksisterende programpakker med deres viktigste spesifikasjoner.

Lederen for konferansen, Phil Stone, la dessuten fram en skisse til et nytt opplegg for bruk av datamaskin i innholdsanalyse. Han hadde her nyttiggjort seg en del prinsipper fra enkelte av de eksisterende programsystemer. Skissen ble diskutert i gruppen og vil trolig bli lagt til grunn for videre utviklingsarbeid.

Hvilket utbytte har man så av en konferanse av denne art? Det synes klart at deltakere som selv arbeider med undersøkelser der man anvender kvantitativ innholdsanalyse, hadde stort utbytte av å komme sammen og drøfte metodiske problemer. Dessuten fikk de en viss innføring i hvilke muligheter datamaskinen representerer i sammenhengen.

Stort sett er det samme type problemer man arbeider med i de forskjellige miljøer, selv om man velger noe ulike løsninger. Nivået varierer heller ikke så mye, og det synes klart at det for tiden ikke finnes noe universalsystem som peker seg ut foran alle andre.

MEDARBEIDERE I DETTE NUMMERET:

JON-ROAR BJØRKVOLD, universitetslektor ved Institutt for musikkvitenskap, Universitetet i Oslo.

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

STEINAR GIL, vitenskapelig assistent ved Slavisk-baltisk institutt, Universitetet i Oslo.

SVEIN LIE, forskningsstipendiat ved Nordisk institutt, Universitetet i Oslo.

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

SVEIN NORDBOTTEN, professor i informasjonsvitenskap ved Universitetet i Bergen.

KRISTEN REKDAL, forsker ved Regnesentret, Universitetet i Trondheim.

ROALD SKARSTEN, universitetsstipendiat ved Religionsvitenskapelig institutt, Universitetet i Bergen, p.t. konsulent ved NAVF's EDB-senter.

THORILD WESSEL, forsker ved Regnesentret, Universitetet i Trondheim.

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

Senterets 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.