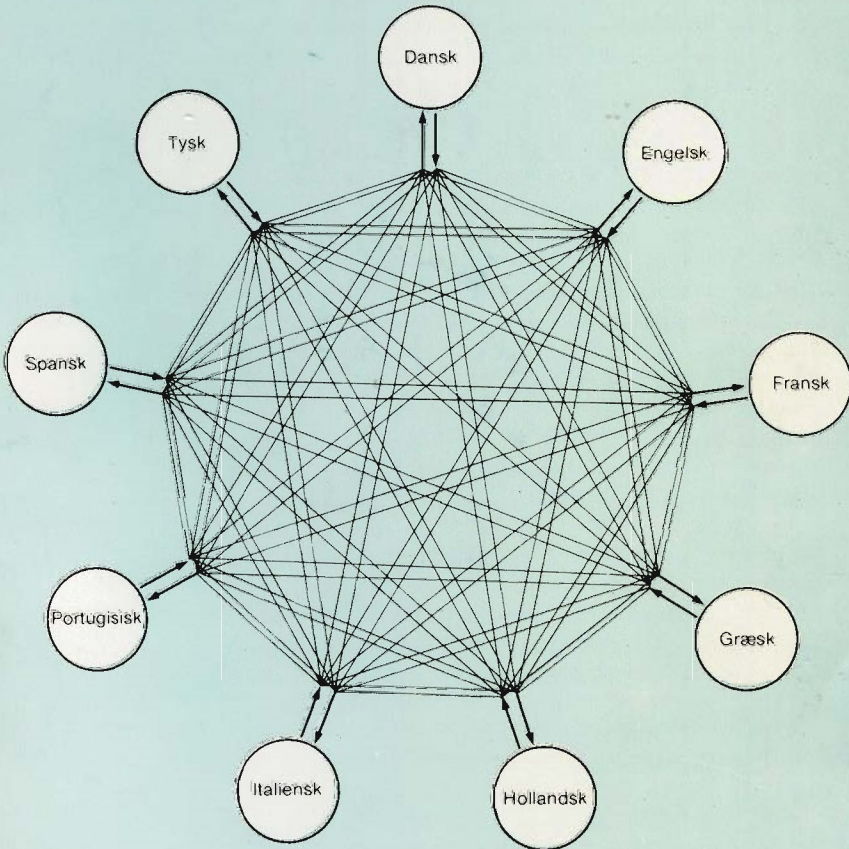


humanistiske data



SENTERETS RAPPORTSERIE

Rapporter utgitt f.o.m. 1980

RAPPORT nr. 13. *Datatjenester for og datasamarbeid mellom kunst- og kulturhistoriske museer*. Februar 1980. 2. opptrykk november 1981. ISBN 82-7283-010-8 Pris kr. 50.

RAPPORT nr. 14. *NOVA *STATUS HÅNDBOK*

Del 1: Søking. Brukerveiledning. 3. opptrykk februar 1983. ISBN 82-7283-011-6 Pris kr. 20.

Del 2: Fil-beskrivelser. Systemdokumentasjon. Utsolgt.

Del 3: Generering og oppdatering av databaser. Utsolgt.

RAPPORT nr. 16. *Årsmelding 1979*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-013-2 Gratis.

RAPPORT nr. 17. *Svein Lie: Automatisk syntaktisk analyse*. Del 1. Grammatikken. Desember 1980. ISBN 82-7283-014-0 Pris kr. 30.

RAPPORT nr. 18. *Datateknologi og humanistisk forskning*. Bidrag til en NAVF-utredning. Desember 1980. ISBN 82-7283-015-9 Pris kr. 30.

RAPPORT nr. 19. *Statistiske metoder på arkeologisk materiale*. Rapport fra et seminar på Bryggens museum, Bergen 24.-26. november 1980. Mars 1981. ISBN 82-7283-017-5 Pris kr. 35.

RAPPORT nr. 20. *EDB-prosjekter i humanistiske fag 1980*. Juni 1981. 2. opptrykk oktober 1981. ISBN 82-7283-018-3 Pris kr. 45.

RAPPORT nr. 21. *Rune Johansen: Bruk av EDB i teatervitenskapelig forskning*. Mai 1981. ISBN 82-7283-019-1 Pris kr. 35.

RAPPORT nr. 22. *Årsmelding 1980*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-020-5 Gratis.

RAPPORT nr. 23. *Stig Welinder: A program package for archaeological use*. 1981. ISBN 82-7283-021-3 Pris kr. 45.

RAPPORT nr. 24. *Rapport fra seminar om bruk av edb innen teater og teatervitenskap*. Januar 1982. ISBN 82-7283-026-4 Pris kr. 50.

RAPPORT nr. 25. *Ole Lauvskar: Diskriminantanalyse i SPSS*. Desember 1982. ISBN 82-7283-028-0 Pris kr. 55.

RAPPORT nr. 26. *Stig Welinder: Paleodemography*. Oslo 1982. ISBN 82-7283-030-2 Pris kr. 55.

RAPPORT nr. 27. *Årsmelding 1981*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-029-9 Gratis.

RAPPORT nr. 29, 30, 31, 32: *Stig Welinder et al.: STAR I-IV A program package for archaeological use*. Bergen 1983. Samlet pris kr. 180. (Rapportene kan også kjøpes enkeltvis).

nr. 29 STAR I Introduction and Star manual. ISBN 82-7283-033-7 Pris kr. 50.

nr. 30 STAR II Student textbook and STAR examples. ISBN 82-7283-034-5 Pris kr. 60.

Forts. 3. omslagsside.

humanistiske data 3-86

NAVFs EDB-senter for
humanistisk forskning

The Norwegian Computing
Centre for the Humanities

NAVF NORGES
ALLMENNVIITENSKEPELIGE
FORSKNINGSRÅD

NAVFs EDB-senter for humanistisk forskning ble opprettet av Norges allmennvitenskapelige forskningsråd i 1972. Senteret har som oppgave å arbeide på nasjonal basis for utbredelse av edb i forskningsarbeidet i de humanistiske fagene. Det er opprettet en samarbeidsavtale med Universitetet i Bergen som bl.a. gir Senteret adgang til edb-tjenester ved Universitetet.

Av sentrale oppgaver kan nevnes utvikling av programutrustning for humanistiske forskningsoppgaver, konsulenthjelp og informasjonstjenester.

Senteret utgir tidsskriftet *Humanistiske Data* (3 nr. pr. år) og en rapportserie (38 er utkommet pr. 10.12.86).

Senteret er sekretariat for International Computer Archive of Modern English (ICAME), og utgir bladet ICAME NEWS.

Senteret driver egne opplæringsprogram for vitenskapelig personale og medarbeidere i den kontor-tekniske gruppen innenfor de humanistiske fag. Det blir også holdt forskjellige kurs og seminar om edb og humanistisk forskning. Tidspunkt og emner blir kunngjort i *Humanistiske Data* og på institusjonen.

Interesserte kan kostnadsfritt bestille årsmelding og *Humanistiske Data* (kr. 70,- for institusjoner f.o.m. 1987).

Humanistiske Data blir utgitt av NAVFs EDB-senter for humanistisk forskning. Redaksjon: Einar-Arne Drivenes, Jostein H. Hauge, Kristin Natvig (red.).

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

Artikler, rapporter, meldinger mottas. Redaksjonen avsluttet 10. desember.

Humanistiske Data is published by The Norwegian Computing Centre for the Humanities. Editorial group: Einar-Arne Drivenes, Jostein H. Hauge, Kristin Natvig (ed.).

The journal can be ordered from the address mentioned above. Contributions are welcome. On request the Centre can supply the addresses of contributors to the journal.

Medarbeidere fra Senteret i dette nummer:

Einar-Arne Drivenes, Erik Fjornes, Jostein H. Hauge, Claus Huitfeldt, Kristin Natvig, Espen S. Ore, Per Vestbostad.

Fotosats i kommunikasjon med Univac 1100/82.

Sats: Universitetet i Bergen/NAVFs EDB-senter for humanistisk forskning.

Grafisk design og montasje: Kristin Natvig.

Trykk: John Grieg A/S

Forsidebilder:

Et maskinoversettelsessystem som kan oversette fra et av EF-språkene til alle de øvrige språk er målet for EUROTRA-prosjektet. Se Jostein H. Hauges rapport fra et nordisk seminar om maskinoversettelse.

Innhold

Artikler

Datakommunikasjon. <i>Espen S. Ore</i>	s. 4
Humanistiske databaser. <i>Kristin Natvig</i>	s. 13
Edb i samfunnsfag. <i>Harald Johansen</i>	s. 22
Machine-Readable Texts in English Language Research. <i>Stig Johansson</i>	s. 27
Computers and Philosophical Manuscripts. <i>Claus Huitfeldt</i>	s. 35

Rapporter

nota bene. <i>Torrey Seland</i>	s. 46
Senterets videoplateprosjekt – en statusrapport. <i>Kristin Natvig</i>	s. 53
Datasekretariatet og humanistiske fag. <i>Kristin Natvig</i>	s. 56
«Norsk stadnamnbase.» <i>Botolv Helleland</i>	s. 57
IBM-seminar om taleprosessering. <i>Kolbjørn Slethei</i>	s. 60
Kurs i design av pedagogisk programvare. <i>Kristin Natvig</i>	s. 62
Datasekretariatets seminar i Halden. <i>Espen S. Ore</i>	s. 65
Humanistiske fag og kunnskapsbaserte systemer. <i>Erik Fjornes</i>	s. 66
«Kommunikasjons- og datateknologi som infrastruktur i nord-norsk næringsliv og forvaltning.» <i>Fredrik Fagertun</i>	s. 67
Nordisk seminar om maskinoversettelse. <i>Jostein H. Hauge</i>	s. 72
Nytt fra RHF/NAVF	s. 81

Meldinger	s. 97
-----------	-------

Summary	s. 107
---------	--------

Datakommunikasjon

Espen S. Ore

1. Innledning

Denne artikkelen er ment som en orientering om en del muligheter en norsk (humanistisk) forsker har til å utnytte datakommunikasjon idag. Lesere som allerede har erfaring i bruk av oppringt samband ved hjelp av modem kan med fordel hoppe over avsnitt 2. og 3.a. Det samme gjelder avsnitt 3.b for de som kjenner Datapak. I 4. blir datakommunikasjon – først og fremst programmet Kermit – gjennomgått. I 5. blir EARN beskrevet. 6. tar opp enkelte telemøter og informasjonssentraler.

Artikkelen gir *korte* oversikter. Lesere som ønsker en grundig innføring i bruk av tjenester og maskiner, henvises til aktuelle brukerhåndbøker. I Datatid 8-11, 1985, finnes en grundig artikkelserie («Modemskolen»), som anbefales på det sterkeste. Televerket har egne veiledninger for bruk av sine dataoverføringstjenester. For Kermit finnes det grundig dokumentasjon. Denne vil i alle fall være tilgjengelig som datafiler på de stedene som distribuerer programmet (se 4.) Tjenestene som er omtalt i 5. og 6. har også egen dokumentasjon.

6. er ment som et supplement til Kristin Natvigs artikkel «Humanistiske databaser» i dette nr. av HD, som omtaler rene databaseleverandører av interesse for humanistiske forskere.

2. Nødvendig utstyr for datakommunikasjon

For å komme i kontakt med nettverk, databaser osv. må man enten disponere en terminal som er koblet til en datamaskin eller et nettverk som igjen har kontaktpunkter utad, eller man må ha en terminal som kan brukes mot et modem (se 3.a). Dette er absolutte minimumskrav. En personlig datamaskin har klare fordeler fremfor en enkel terminal, særlig fordi den vil kunne brukes til å lagre data som overføres. De som bruker hovedanleggene ved universitetene i Oslo, Tromsø og Trondheim, kan komme i kontakt med det internasjonale pakkenettet (se 3.b) via Uni-nett. Ved Universitetet i Bergen har man tilsvarende mulighet gjennom det lokale universitetsnett. Nærmere opplysninger om dette kan man få ved å henvende seg til de lokale edb-sentre.

En personlig datamaskin (les evt. terminal her og de fleste steder i artikkelen der det ikke eksplisitt er tale om datalagring eller bruk av lokale programmer) kan kommunisere med omverdenen via et modem dersom den har en serieport (også kalt RS232 og diverse andre navn), og

dersom de data som sendes ut over serieporten er lesbare for mottageren. Dette siste betyr at det må være mulig å definere kommunikasjons-hastighet og en del andre parametre. (Ta evt. kontakt med maskinleverandøren dersom dette ikke går klart frem av brukerveiledningene til utstyret.) Det trenges også en kabel mellom maskinen og modemmet. Hvilke pluggtyper som skal brukes i hver ende av kablen, og hvorledes de enkelte pinnene skal være koblet, er avhengig både av datamaskinen og modemmet.

3. Televerkets datanett (Datel og Datapak)

Televerket tilbyr tre forskjellige dataoverføringstjenester: *Datel*, *Datex* og *Datapak*. I *Datel* foregår dataoverføringen gjennom det vanlige (samtale-)telenettet. *Datex* er et eget datanett der overføringen er linjevekslet. Det vil si at når to datamaskiner skal utveksle data, etableres det først en linje mellom dem. Denne linjen eksisterer så helt til det blir gitt avbruddssignal. Det hele kan sammenlignes med en telefonsamtale. *Datex*-forbindelsene takseres også tilsvarende etter (egne) tellerskritt. *Datex* er spesielt egnet for overføring av store datamengder innen et begrenset område (Norden). Pga. denne geografiske begrensningen og relativt høye abonnementsavgifter er det tvilsomt om tjenesten er av interesse for denne artikkelens målgruppe, og den vil derfor ikke bli videre omtalt her. *Datapak* er Televerkets pakkevekslede datanett. Pakkevekslet betyr at det ikke etableres noen fast forbindelse mellom to maskiner som har kontakt. Istedet samles data i pakker som også inneholder mottager- og avsenderadresse. Pakkene blir så sendt over nettet ut fra disse adresseopplysningene. *Datapak* er den mest hensiktsmessige tjenesten for internasjonal kommunikasjon ut over et svært lavt minimumsnivå.

3.a *Datel* – Modem

Datel er strengt tatt ingen egen dataoverføringstjeneste. (Dvs. det tilbys også egne datatjenester under *Datel*, men de faller utenfor artikkelens ramme.) Det er Televerkets navn på muligheten til å bruke vanlige telefonlinjer til dataoverføring. For å få dette til, trenges det et *modem* i hver ende av kommunikasjonslinjen mellom datamaskinen (evt. terminalen) og telenettet. Modem er en forkortelse for modulator/demodulator. Litt forenklet er modemets oppgave å omforme (digitale) datasignaler til (analoge) lydsignaler som kan overføres gjennom telefonlinjene, og å overføre lyden tilbake til datasignaler på mottagersiden.

Modem kan leies eller kjøpes. Prisene varierer med hastighetene som modemene yter. Dersom modemmet skal brukes i forbindelse med *Datapak*, er det hastighetene 300 og 1200 *bps* (bit per sekund) som er aktuelle. Dersom modemmet bare skal brukes i direkte oppringt kontakt med en annen maskin, kan man bruke større hastigheter, men modem-

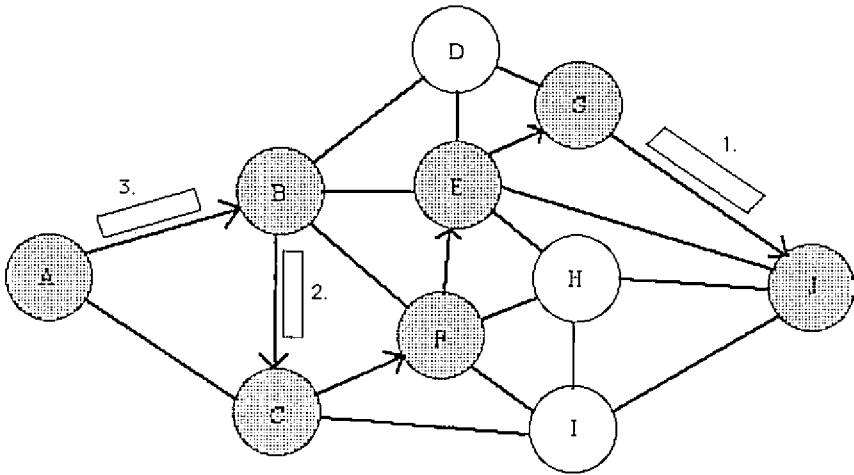
ene i hver ende av linjen må operere med samme fart. TBK tilbyr, i konkurranse med private firmaer, 300 bps modem til ca. kr. 1.000. Disse modemene virker i *full duplex*, dvs. at de kan sende og ta imot samtidig. 1200 bps modem som virker i *halv duplex* (som altså ikke kan sende og ta imot samtidig) selges også av TBK i konkurranse med private, mens TBK har monopol på 1200 bps modem som virker i full duplex og på alle typer modem med større hastighet. Bruker man et modem i halv duplex, vil man f.eks. ikke kunne bryte av en lang, uinteressant datastrøm fra dem maskinen man er i kontakt med. Dette kan både være tidkrevende og kostbart dersom man kommuniserer med en database i USA. Det anbefales å bruke full duplex.

Ved bruk av Datel foregår dataoverføringen i telefonnettet; det er altså ikke behov for noe eget abonnement. Dette kan imidlertid bli dyrt, siden tellerskritt vil løpe som ved en vanlig telefonsamtale. Det vil f.eks. ta ca. ett minutt å overføre en A-4-side ved 300 bps. Datel egner seg derfor best dersom det er relativt sjelden at man har behov for datakommunikasjon ut over det lokale telefonområde.

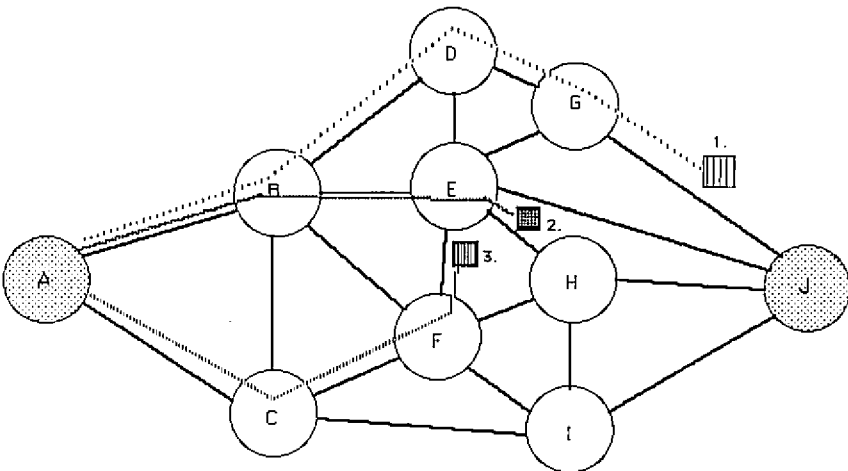
3.b Datapak

I Datapak overføres data i pakker i et eget (raskt) datanett. Pakkene er datasamlinger av en gitt størrelse (f.eks. 128 tegn). Data blir organisert i pakker på avsenderstedet og gjort om til en jevn datastrøm på mottakerstedet i en *PAD* (Packet Assembler/Disassembler). (Det finnes terminaler som er pakkeorienterte, men siden denne artikkelen er orientert mot bruk av personlige datamaskiner og vanlige terminaler, vil jeg ikke komme inn på disse.) PAD-er finnes i tilknytning til hovedanleggene ved de norske universitetene. Dersom man ikke kan bruke disse, må man kontakte Datapak via Televerkets PAD. Dette gjøres gjennom en vanlig oppkoblet modemforbindelse til telefonnr. 0165. Oppringing hit takseres som en lokalsamtale. Når kontakten er opprettet, oppgir man sin brukeridentifikasjon og et passord og deretter adressen til den maskinen man vil kontakte.

Abonnement på Datapak kan man få ved å kontakte Televerket (0140). Det koster for tiden kr 480,- pr. kvartal for første dekningsområde og kr 100,- for hvert av de øvrige. Av en eller annen grunn er landet delt inn i 7 dekningsområder med egen registrering av brukere. Dette er viktig hvis man ønsker å bruke Datapak på reiser eller midlertidige opphold andre steder i landet enn der man er registrert. Under bruk takseres Datapak etter mengden av data som overføres og etter hvor lenge man er i kontakt. Innenlands koster det kr 5 pr 1000 segmenter å 64 tegn og 22 øre for hvert påbegynt minutt. I kontakt med USA koster 1000 segmenter kr 40,- og hvert påbegynt minutt 90 øre. Både innenlands og utenlands er det en engangsavgift på 13 øre pr. anrop. Påbegynte segmenter regnes som fulle segmenter, og dette gjør at det fort kan bli



Linjevekslet nett. Pakkene (1,2 og 3) går alle samme vei fra A til J: over nodene B,C,F,E og G.



Pøkkevekslet nett. Pøkkene blir sendt fra maskin til maskin uten at veien fra A til J trenger å være fastlagt.

dyrt å bruke skjermorienterte editorer slik vi f.eks. finner på den norske EARN-noden i Trondheim.

4. Kommunikasjonsprogrammer

Utover det fysiske utstyrsbehovet, trenger man også egne programmer dersom man skal kontakte databaser/andre maskiner fra en mikromaskin. De enkleste slike programmer gjør at mikromaskinen vil bli oppfattet som en terminal av den maskinen man er i kontakt med. De fleste har også muligheter til å lagre data som er blitt vist på skjermen under en sesjon. Mer avanserte programmer har egne funksjoner for dataoverføring der det også vil bli foretatt en viss kontroll med at data blir korrekt overført.

Kermit er et eksempel på denne siste typen programmer. Det er utviklet ved Columbia University, USA. Kermit er tilgjengelig for de fleste datamaskiner det vil være aktuelt å arbeide med, både mikroer og mini/stormaskiner. Programmet kan i Norge fås bl.a. fra USE ved Universitetet i Oslo.

Når man skal overføre data mellom to maskiner, må Kermit være startet på begge. Dette gjøres ved at man setter sin lokale maskin i terminalmodus gjennom kommandoen `CONNECT` til Kermit. Når man har logget seg inn på den andre maskinen, og har startet Kermit der, er alt klart til overføringen (forhåpentligvis). Under dataoverføringen vil sender-Kermit lage pakker som i tillegg til data også inneholder kontrollkoder som brukes av mottaker-Kermit til å avgjøre om data er korrekte. Er alt i orden, sendes det en pakke tilbake, og sender-Kermit starter overføringen av neste pakke.

Før man bruker Kermit eller andre kommunikasjonsprogrammer, må man som regel «stille inn» programmet. Dvs. å oppgi endel egenskaper ved den fysiske kontakten, spesielt hastighet, og ved dataformatet som skal brukes. Vanligvis kan dette gjøres én gang for alle om man bare vil komme til å kontakte én annen maskin eller ett annet nett. Dersom det er aktuelt å kommunisere med flere forskjellige maskiner, må parametrene settes hver gang. I Kermit kan man lagre de nødvendige opplysningene slik at innstillingen kan skje med en enkel kommando.

5. Aktuelle datanett

I motsetning til kommunikasjon mellom datamaskiner slik den er beskrevet i avsnitt 4, mener jeg her med datanett maskiner som står i (tilsynelatende) fast kontakt, og der tjenester som dataoverføring og f.eks. elektroniske postkasser (se også avsnitt 6) er lett tilgjengelige uten at brukeren skal måtte bekymre seg om tekniske krav til kommunikasjonsformen.

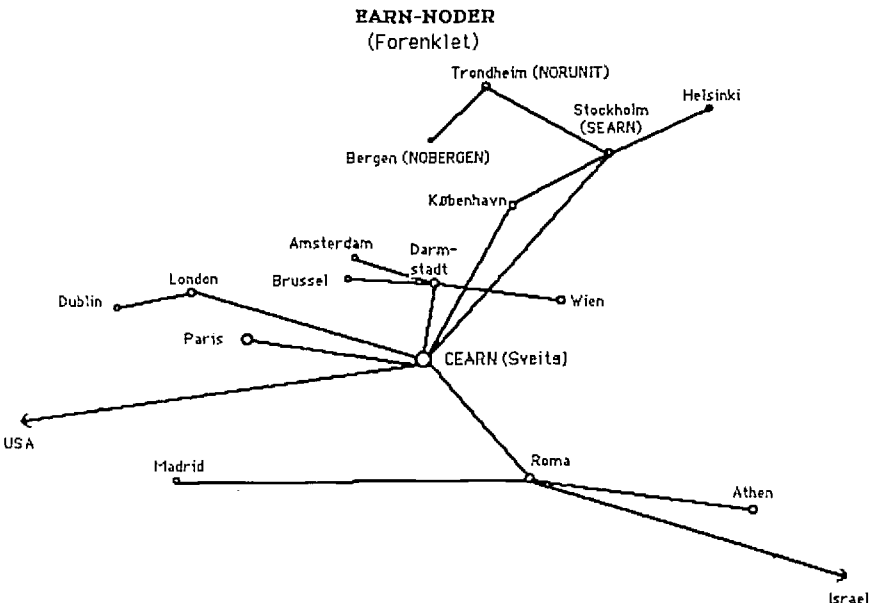
5.a EARN

European Academic and Research Network (EARN) er startet av IBM og det er først og fremst IBM-maskiner som er koblet sammen i nettet. EARN er navnet på den delen av nettet dekker Europa og Midt-Østen. De amerikanske delene av nettet er BITNET (USA) og Northnet (Canada). Man kan også kommunisere med andre nett: Arpanet og JANET i Storbritannia.

Gjennom EARN kan man sende elektronisk post og data. Det er også mulig å sende beskjeder direkte til en annen bruker på nettet dersom vedkommende er pålogget i øyeblikket.

I Norge er det (august 1986) en IBM 4361 datamaskin ved RUNIT som er nasjonal node. Maskinen er en gave fra IBM, og bruk av den i EARN-sammenheng er gratis (NB! etter statuttene er EARN kun tilgjengelig for ikke-kommersiell bruk innen forskning og undervisning). De som ønsker å bruke EARN, kan kontakte RUNIT, Ekspedisjon SB2, 7034 Trondheim - NTH. Tlf.: 07-59 30 28.

Det finnes to andre EARN-noder i Norge: En ved Universitetet i Bergen og en ved Christian Michelsens Institutt.



6. Elektronisk post og telemøter

De fleste som har arbeidet på stormaskiner, har kunnet bruke elektronisk post og andre meldingstjenester. Posten blir sendt til en adresse som

K11:17:26A

Velkommen til KOM, versjon 6(1021).

Hva heter du? (Navn og organisasjon)

= es o

espen ore (NAVF/Bergen)

Hva er passordet ditt:

= Takk!

Du har 1 ulest brev

Du har 3 uleste innlegg i EARN erfaringer

Du har 67 uleste innlegg i ARPA (Info-) IBMPC Digest (innh.fortegnelser etc.)

Du har 91 uleste innlegg i Kermit erfaringer

Du har 161 uleste innlegg

Aud H. Halvorsen (DSE/UiO) er tilstede i KOM.

Hva vil du gjøre? (Les) neste brev, (GA til) neste mote, Slutte,

(Send) brev (til), (Noe) annet.

(Send) brev (til), Personlig (svar), (Noe) annet.

= ne me

(GA til) neste mote

Møtets tittel er: EARN erfaringer

Du har 3 uleste innlegg

Tirsdag 11.18

(Tekst 44187) 86-12-01 16.25 Alf Christophersen (UiO/Ernæringslinjen)

Emne: EARN erfaringer --> VAX.

EARN erfaringer er nå opprettet på PORTACOM på INGER. Kanskje

vi skal derfor flytte aktiviteten dit fra nå av??

(Tekst 44187)-----

Du har sett alle 83 innlegg i EARN erfaringer

Hva vil du gjøre? (GA til) neste mote, Slutte, (Skrive eget)

innlegg, Kommentere (innlegget), (Send) brev (til), Personlig

(svar), (Noe) annet.

Fig. 1. Et eksempel på en sesjon med KOM på Universitetet i Oslo DEC-10. I møteinnlegget blir det gitt beskjed om at møtet nå er flyttet til det nye VAX-anlegget.

```

$ r pub:portacom
Welcome to PortaCOM (version 1.14)!
Please give your complete name.
- espen s. ore
Your name was new to PortaCOM. If you want to give another name, please
strike the RETURN key when you are asked for a password.
Please choose a personal password:
-
As a check, type your password again:
-
You have seen all news
You are in your mailbox

When replying to a menu question, type a unique abbreviation of the part of
the command outside parentheses. If you want to give the first command on
the menu, just push RETURN.

What do you want to do? List active (conferences), (Become) member (of),
Wait (for news), (Send a) letter (to), Quit, (Get) help, Other.
- li a
List active (conferences) 7
(* marks meetings you are not a member of)

```

	n	Ent.	Memb.	Name of meeting
*	2	55	45	Presentasjon (av) PortaCOM brukere
*	3	18	55	Presentasjon (av) Apne (noter)
*	4	8	39	Presentasjon (av) lukkede (noter)
*	5	3	49	Fritt forum
*	6	54	63	PortaCOM erfaringer
*	13	71	70	VMS erfaringer
*	1	88	13	USE system og driftsavdelingen
*	41	13	12	UMGCG erfaringer
*	60	23	8	USE Lukevakter
*	89	29	14	RO32 Lesesalen (USE)
*	100	4	22	EAN erfaringer
*	104	5	20	Macintosh Erfaringer
*	125	2	13	USE Operatører
*	126	1	9	Biokjemi generell (mikro celle etc)

```

What do you want to do? List active (conferences), (Become) member (of),
Wait (for news), (Send a) letter (to), Quit, (Get) help, Other.
- memb ean
(Become) member (of) EAN erfaringer
You are a member of EAN erfaringer
( 350 ) 86-10-27 15:20 Frode Flågstad (ESUto)
Receiver: Presentasjon (av) Apne (noter) /received/
Subject: EAN (erfaringer)
-----
Mote for utveksling av gløder & sorger med bruk av EAN (post/
meldingssystem). Overtar (en gang) for et tilsvarende møte
på KOM.
( 350 ) comment in ( 353 )
The title of the meeting is: EAN erfaringer
You have 4 unseen out of 4 entries
What do you want to do? (Read) next notice, (Write new) notice, (Send a)
letter (to), Quit, (Get) help, Other.
- quit
Quit
Bye!

```

Fig. 2. Inne i Portacom på VAX-maskinen. Møtet om EARN er ikke etablert ennå, men istedet er det et møte om EAN-meldingssystemet, som også virker interessant.

er et brukernavn, og når mottakeren senere logger seg inn på maskinen kommer det beskjed om at et brev venter. Slike meldingssystemer er også standard på nettverk som EARN og Arpanet. Elektronisk post har fordeler og ulemper i forhold til vanlig skriftlig kommunikasjon og telefonsamtaler. Den overføres raskere en vanlig post, og den er ikke avhengig av at mottakeren er tilstede når en beskjed skal sendes, slik som med telefonen. På den andre siden blir uttrykksformen mer muntlig enn den man vanligvis finner i et papirbrev, samtidig som den (tross alt) mer personlige kontakten man kan få gjennom en telefonsamtale, savnes.

Humanistiske databaser

Kristin Natvig

Innledning

I det følgende gis en oversikt over *on-line tilgjengelige* humanistiske databaser i Norden, samt beskrivelser av to omfattende amerikanske databaser og én fransk. Oversikten er komplett for Nordens vedkommende – opplysningene er hentet dels fra «Databaser i Norge, online og offentlig tilgjengelige august 1985», utgitt av Riksbibliotekjentesten v/ Tove Molvig, og «Nordisk databasguide 1985», NORDINFO-publikasjon 9.

De utenlandske databasene som omtales her, spenner over flere humanistiske fagområder. Samtlige inneholder i tillegg filer med data om samfunnsvitenskapelig forskning – disse er holdt utenfor i denne sammenheng, men de kan være av interesse også for humanistiske forskere. Det samme gjelder nordiske samfunnsvitenskapelige databaser (se «Nordisk databasguide»).

Generelle bibliografiske databaser – både nordiske og utenlandske – er bare omtalt hvis de inneholder spesifikt humanistiske samlinger.

I beskrivelsene gis opplysninger om databasenes innhold samt generell informasjon om oppkobling/søking. En detaljert gjennomgang av oppkoblingsprosedyrer fins i Espen S. Ores artikkel «Datakommunikasjon» i dette nummer av HD.

Pris pr. time for søking i databasene er oppgitt der dette er kjent. En må være oppmerksom på at den utslagsgivende faktor for prisen på en søkeomgang vil som regel være hvor mye informasjon som blir sendt til brukerens maskin/terminal. Ved bruk av 300 bit pr. sekund (bps) modem tar det omtrent ett minutt å overføre et skjermbilde. 1200 bps modem er fire ganger raskere.

Det fins nok flere *on-line tilgjengelige* utenlandske databaser enn de som blir omtalt her – spesielt innenfor de enkelte humanistiske fagområdene. Flere vil bli presentert i Humanistiske Data etter hvert som redaksjonen får opplysninger om slike. Redaksjonen er takknemlig for tips fra leserne om databaser som ikke er representert i denne oversikten.

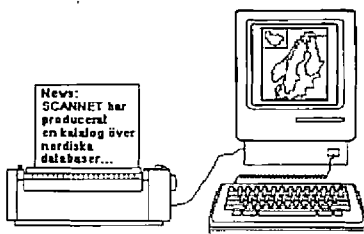
Norge

SIFs prosjektdatabaser

Innhold: NAVFs Sentral for informasjon om forskningsprosjekter (SIF)

NORDISK DATABASGUIDE 1985

Sammanställd av SCANNET
Malin Edström och Elisabet Mickos



NORDINFO-PUBLIKATION 9

driver to databaser over humanistiske og samfunnsvitenskapelige forskningsprosjekter samt publikasjoner i tilknytning til prosjektene. Foruten samfunnsvitenskapelige prosjekter omfatter materialet ca. 270 humanistiske prosjekter med støtte fra NAVF i 1984 og 1985 og 700 tilknyttede publikasjoner. I løpet av høsten '86 skal denne informasjonen oppdateres og suppleres med beskrivelser av ca. 100 nye NAVF-finansierte prosjekter og 300 publikasjoner.

NAVF-forskerne representerer ca. 120 humanistiske institutter og fag-institusjoner. Følgende fagområder er representert: arkeologi, arkitektur, filmvitenskap, filosofi, folkelivsgranskning (etnologi), folkeminnevitenskap (folkloristikk), historie, idéhistorie, kunsthistorie, litteraturvitenskap, musikkvitenskap, religionsvitenskap, språkvitenskap, teatervitenskap og teologi.

En heldekkende registrering av all humanistisk forskning i Norge vil også foreligge i databasene i løpet av høsten. Dette materialet omfatter ca. 500 prosjekter med tilknyttet litteratur som utføres ved de fire universitetene, distriktshøgskolene og ca. 120 høyskoler, museer, arkiv og andre forskningsinstitusjoner, og av forskere uten institusjonstilknytning.

Planer er lagt for å innlemme i databasen opplysningene om norsk

kvinneforskning som tidligere er publisert i katalogform av NAVFs sekretariat for kvinneforskning. Opplysningene vil være ajourførte.

Søking: Databasene ligger på hovedanlegget ved Universitetet i Bergen, SPERRY 1100/82, og kan nås via Datapak eller oppringt samband. Brukere på universitetene kan også nå basene via UNINETT. Søking foregår via fritekstsøkesystemet SIFT. Det er utarbeidet en egen veiledning om bruken av databasene.

Databasene er åpne hele døgnet. Søking er foreløpig gratis.

Adresse: Sentral for informasjon om forskningsprosjekter, Boks 53, 5014 Bergen-Universitetet.

Samisk database

Innhold: Samisk database er en bibliografisk database om samisk kultur, språk, historie og bosetting. Databasen inneholder ca. 2000 referanser, hovedsakelig til litteratur utgitt i Norge etter 1950. Se for øvrig Humanistiske Data 2-85.

Søking: Universitetsbiblioteket i Trondheim er produsent og vert for databasen, som er åpen hele døgnet. For informasjon om kostnader, kontakt UBT.

Adresse: Universitetsbiblioteket i Trondheim, Erling Skakkes gt. 47c, 7000 Trondheim.

BIBSYS

BIBSYS

Innhold: BIBSYS er en bibliografisk database for bibliotekene ved universitetene i Bergen, Trondheim, Tromsø og Norges Handelshøyskole. Følgende specialsamlinger er av spesiell interesse for humanister: UBB - kvinnelitteratur, UBT - arbeiderbevegelsens arkiv, judaica, musikkmanuskripter, norsk jernbanehistorie, norsk musikkhistorie, privatarkiv.

Søking: Databasen er åpen hele døgnet utenom mellom kl. 21-22. BIBSYS' programsystem kjøres på SPERRY 1100/62-anlegget ved Regnesenteret ved Universitetet i Trondheim (RUNIT). Kunder må være registrert som brukere ved dette anlegget. En hvilken som helst TTY-terminal med ASCII tegnkode kan brukes. Alle Televerkets sambandsmuligheter (Datel, Datex, Datapak) kan nyttes. Kunder må i tillegg betale for bruk av datamaskintid (prisen varierer med bruksmåten). Rabatter kan i visse tilfeller oppnås.

Adresse: BIBSYS, RUNIT, 7034 Trondheim-NTH.

UDOK

Innhold: UDOK er en bibliografisk database (del av referansedatabasen UBO:BOK) ved Universitetet i Oslo. UDOK inneholder prosjektinformasjon og referanser til hovedfags- og magistergradsoppgaver ved UiO innen humaniora og samfunnsfag.

Søking: UDOK nås via Datatel, Datapak og UNINETT og er åpen hele døgnet. Brukere betaler for maskintid.

Adresse: Universitetsbiblioteket i Oslo, Planavdelingen, Drammensvn. 42, 0255 Oslo 2.

Sverige og Danmark

Demografiska databasen

Demografiska databasen ble startet i 1973 og har vært en fast tjeneste siden 1978 ved Umeå universitet. Formålet med databasen er å registrere, stille sammen og administrere svenske demografiske og sosialhistoriske data for forskning, utdanning og arkiv.

Innhold: De fleste kildene som bearbeides utgjøres av kirkebokmateriale fra 1800-tallet. Materiale fra hele kommuner – som representerer ulike sosio-økonomiske miljøer – inngår i databasen. Regioner representert i basen, er Sundsvallregionen, som bl.a. belyser den tidlige industrialiseringens gjennombrudd, og Linköpingsregionen, som bl.a. speiler utviklingen i en gammel kulturbygd. Det er utgitt en brosjyre om databasen og en presentasjon av virksomheten i bokform (se «Fra bokhylla» i HD 2-86).

Søking: Brukere av databasen må kun betale for oppkoblingstid og evt. disketter/magnetbånd med statistikk eller individopplysninger. Ulike former for statistikk og sammenstillinger av individdata kan genereres ved hjelp av interaktive programmer.

For nærmere opplysninger om oppkobling og priser, kontakt Demografiska databasen ved nedenstående adresse.

Adresse: Demografiska databasen, Umeå universitet, 90187 Umeå, Sverige.

Retrieval System for Museums

Innhold: Retrieval System for Museums (REFORM) er et edb-register som inneholder 30.000 fullstendige beskrivelser av enkeltgjenstander fra museumssamlinger. Deltagende museer er for tiden Kulturarvet, Armeumuseum, bygdemuseer i Midt-Sverige og enkelte distriktsmuseer. En egen fil – BILD-REFORM – inneholder beskrivelser av ulike typer bilder i Armeumuseum og Skaraborgs länsmuseum. Se for øvrig Rune Hermanssons bidrag i dette nummeret av HD og HD 2-83.

Søking: Kontakt Rune Hermansson ved nedenstående adresse.

Adresse: Kulturarvet, Matsarvsvägen 3, S-791 74 Falun, Sverige.

Svensk konsthistorisk bibliografi

Innhold: Svensk konsthistorisk bibliografi (SKVB-BAS) er en referansedatabase for litteratur om kunst publisert i Sverige eller andre land av svenske forfattere. Databaseen dekker også litteratur om museums- og kulturpolitikk, konservering, restaurering osv. Basen blir oppdatert hver måned og inneholder for tida ca. 5000 referanser til litteratur f.o.m. 1978.

Søking: Nøkkelord gis på engelsk. For øvrige opplysninger, kontakt Charlotte Hanner eller Jan Rosvall ved nedenstående adresse.

Adresse: Svensk konsthistorisk bibliografi, Göteborgs universitet, Bebyggelsesantikvarisk avdeling, Vallgatan 22, S-41116 Göteborg, Sverige.



Det Kulturhistoriske Centralregisters Databank

Innhold: Det Kulturhistoriske Centralregister er en fulltekstdatabase over funn og anlegg av kulturhistorisk interesse fra steinalderen til nyere tid i

Danmark. I tillegg til en beskrivelse av hvert funn, gis det opplysninger om hvordan og hvor funnet er gjort, og henvisninger til i hvilke institusjoner ytterligere informasjon fins.

Søking: Informasjon ved henvendelse til Jørgen Christensen ved nedenstående adresse.

Adresse: Nationalmuseet, Frederiksholms Kanal 12, 1220 København K, Danmark.

Utlandet for øvrig

DIALOG

Innhold: Den amerikanske informasjonsgjenfinningstjenesten DIALOG driver over 220 ulike databaser – av disse er åtte humanistiske:

America: History and Life – korte sammendrag fra ulike typer informasjonskilder om amerikansk og kanadisk historie. Innholder over 200.000 poster f.o.m. 1964 og blir oppdatert tre ganger pr. år. Søkepris: \$65 pr. time.

Artbibliographies Modern – referanser til/sammendrag av litteratur om kunst og design fra bøker, avhandlinger, utstillingskataloger og 300 tidsskrifter. Inneholder 84.000 poster f.o.m. 1974 og blir oppdatert anethvert år. Søkepris: \$60 pr. time.

Historical Abstracts – sammendrag av/indekser til artikler om verdenshistorie f.o.m. 1450, utenom USA og Canada. 2000 tidsskrifter fra 90 ulike land er representert. Inneholder 221.000 poster f.o.m. 1973 og blir oppdatert hvert kvartal. Søkepris: \$65 pr. time.

Language and Language Behavior Abstracts – sammendrag av artikler i ca. 1000 tidsskrifter fra hele verden om forskning i språk og språkpsykologi. Inneholder ca. 75.000 poster f.o.m. 1973 og blir oppdatert hvert kvartal. Søkepris: \$66 pr. time.

MLA Bibliography – referanser til bøker og tidsskrifter om språk, litteratur og lingvistik, publisert årlig av the Modern Language Association. Decker britisk, amerikansk, latin, keltisk, europeisk, asiatisk, afrikansk og latinamerikansk litteratur, i tillegg til teoretiske, deskriptive, komparative og historiske studier i lingvistik. Inneholder 580.000 referanser f.o.m. 1970 og blir oppdatert årlig. Søkepris: \$55 pr. time.

Philosopher's Index – indekser til/sammendrag av bøker og over 270 tidsskrifter om filosofi og relaterte tverrfaglige fagområder. Gis også ut som trykt publikasjon. Inneholder ca. 18.000 poster f.o.m. 1940 og blir oppdatert hvert kvartal. Søkepris: \$55 pr. time.

Søking: Søking foretas via DATAPAK og videre gjennom ett av flere nettverk tilgjengelige i Storbritannia. Terminalen en søker fra må oppfylle visse spesifikasjoner. Passord er gratis og reduksjoner i søkekostnader kan oppnås på ulike betingelser.

DIALOG tilbyr en rekke søkemuligheter og tjenester i tilknytning til

søkingen. Disse er beskrevet i detalj i en fyldig informasjonspakke som fås gratis ved henvendelse. Pakken inneholder også en publikasjonskatalog.

Adresse: Dialog Information Services Europe, P.O. Box 8, Abingdon, Oxford OX13 6EG, U.K.

The Research Libraries Information Network (RLIN)

Innhold: RLIN er et amerikansk informasjonsgjenfinningssystem for bibliografisk materiale i en on-line database. Databasen inneholder mer enn 12 mill. poster fra kataloger til medlemmer av the Research Libraries Group, Inc. og andre institusjoner i USA og Canada, inkludert museer, statsarkiv, og Library of Congress. Materialet er delt inn i sju hovedfiler, basert på materialformat: bøker, filmer, kart, lydopptak, noter, serier og manuskripter/brev. Best dekning i databasen er det for materiale utgitt eller anskaffet etter 1977, men noen bibliotek har brukt systemet siden tidlig i 70-årene (Library of Congress fra 1968). Humanistiske fag som er spesielt godt representert, er kunst, arkitektur, musikk og østasiatiske emner.

RLIN omfatter også flere spesialiserte databaser. Blant disse er the On-Line Avery Index to Architectural Periodicals, en indeks over 500 arkitektur-tidsskrifter utgitt f.o.m. 1979, med vekt på design og historie. En annen er SCIPPIO the Art Sales Catalog Data Base, som omfatter amerikanske og europeiske kunstaustauksjonskataloger utgitt siden 1981.

The Eighteenth-Century Short Title Catalog (ESTC) inneholder bibliografiske opplysninger om publikasjoner trykket på engelsk i det 18. århundre. Denne databasen ble opprinnelig opprettet av the British Library, men over 330 bibliotek bidrar nå med informasjon. Ved fullførelsen er ESTC-databasen ventet å inneholde 500.000 titler av spesiell interesse for forskere i det 18. århundres engelskspråklige litteratur, kultur og historie.

Søking: Ved søking gis kommandoer i naturlig språk via 20 indekser (bl.a. forfatter, tittel og emne). Indekser kan kombineres ved å bruke boolske operatører.

Utenlandske brukere kan logge seg på RLIN via det amerikanske nettverket UNINET mellom kl. 14-06 (norsk tid). Det er planer om å gjøre alle filene (utenom «Books») og de spesialiserte databasene tilgjengelige fra kl. 10.

Ved opprettelse av konto i RLIN må et gebyr betales for å dekke de trykte søkeveiledningene. Søking i den sentrale databasen kostet t.o.m. 31. august 1986 \$63 pr. time (min. 15 min. pr. måned). For å søke i de spesialiserte databasene, må en betale \$31.50 pr. time i tillegg. (Ingen tilleggsavgift for Avery Index to Architectural Periodicals.) En omfattende informasjonsmappe fås gratis ved henvendelse.

Adresse: Library Operations Division, The Research Libraries Group, Inc., Jordan Quadrangle, Stanford, California 94305, USA.



FRANCIS

Innhold: Le Centre de documentation sciences humaines (CDSH) har drevet informasjonssystemet FRANCIS siden 1972. Blant FRANCIS' tjenester og produkter er en rekke bibliografiske databaser over humanistiske fag. Referansene gis på fransk, men mange av dokumentene er publisert på andre språk, spesielt engelsk. Informasjonen i alle databasene gis også ut i form av trykte publikasjoner. Databasene blir oppdatert hvert kvartal hvis ikke annet er oppgitt.

Amerique Latin - hovedsakelig franskspråklige publikasjoner om forskning i emner fra Latin-Amerika og De karibiske øyene. Opprettet i 1980 og inneholder ca. 2500 referanser.

Art et Archeologie - for det meste tidsskriftartikler om kunst og arkeologi (inkludert metodologi) i Midt-Østen, Asia og Nord- og Latin-Amerika. Opprettet i 1972 og inneholder ca. 26.000 referanser.

Droits Antiques - dokumenter om antikkens rettsteori- og praksis i middelhavslandene. Opprettet i 1975 og inneholder ca. 15.000 referanser. Oppdateres to ganger i året.

Ethnologie - dekker de fleste underdisiplinene. Opprettet i 1972;

f.o.m. 1976 indeks også på engelsk. Inneholder ca. 36.000 referanser.

Histoire et Sciences de la Litterature – dekker teoretiske emner og litteratur fra bl.a. Afrika. Opprettet i 1972 og inneholder ca. 78.000 referanser.

Histoire et Sciences des Religions – dekker de fleste underdisiplinene. Opprettet i 1972 og inneholder ca. 118.000 referanser.

Philosophie – dekker mange underdisipliner. Opprettet i 1972 og inneholder ca. 63.000 referanser.

Prehistoire et Protohistoire – fra menneskets opprinnelse til utviklingen av skrivekunsten. Opprettet i 1972 og inneholder ca. 39.000 referanser.

Repertoire d'Art et d'Archeologie – dekker alle kunstartene, middelalderens og moderne arkeologi, og kunstvitenskapelig teori. Opprettet i 1973 og inneholder ca. 155.000 referanser.

Science de Language – dekker lingvistikkenes hovedområder. Opprettet i 1972 og inneholder ca. 51.000 referanser.

Søking: Alle FRANCIS' humanistiske databaser er beregnet på videotekst-terminaler. ASCII-terminaler må derfor utstyres med dekodingskort. Oppkobling skjer via Datapak og videre gjennom det franske datakommunikasjonssystemet CIRCE (pris: 200F pr. time). Veiledninger til gjenfinningssystemet og til hver database er tilgjengelige.

Referanser kan skrives ut etter ulike sorteringskriterier og på flere formater, deriblant kartotekkort. En informasjonspakke om FRANCIS' tjenester og publikasjoner fås gratis ved henvendelse.

Adresse: FRANCIS, Centre de Documentation Sciences Humaines, 54, boulevard Raspail, 75260 Paris cedex 06, France.

Edb i samfunnsfag

Med konsekvenser for faget og lærerutdanninga

Harald Johansen

Nordreisa videregående skole – en kombinert videregående med seks studieretninger – er en av Data-sekretariatets forsøksskoler. Harald Johansen, lektor i geografi og samfunnskunnskap, er prosjektleder for edb-forsøkene ved sko-

len. Følgende artikkel hadde opprinnelig form av et foredrag Johansen holdt ved en konferanse om informasjonsteknologi i skole og lærerutdanning, som ble arrangert i Heimdal ved Trondheim 1.-3.12.86.

1. Erfaringer fra forsøksskolene

Arbeidet i forsøksskolene har i hovedsak bestått i å prøve ut ulike pedagogiske program overfor elevene. De mest vanlige program for samfunnsfag er laga av en database med informasjon om et spesielt emne. Deretter er det laga et eget program rundt basen slik at elevene lett kan ta fram informasjonen og behandle den på ulike måter. I tillegg er det lett å få ut et produkt på skriver.

Noen skoler har også brukt generelle verktøy, for økonomi gjelder dette f.eks. regneark eller baser. Det finnes også på markedet noen simuleringprogram for samfunnsfag. Disse simulerer ulike samfunnssituasjoner, og elevene har muligheter for å påvirke variablene og får derved en ulik utvikling.

Det finnes også utviklet spesielle verktøyprogram som letter arbeidet i samfunnsfag. Ett slikt er for spørreundersøkelser der arbeidet med å regne sammen alle verdiene er blitt svært forenkla. I tillegg kan elevene lett be maskinen rekne ut ulike mål f.eks. for samvariasjon og spredning samt å få grafiske framstillinger av resultatene. Og ikke minst kan elevene fort få et pent produkt ut på skriver.

Andre skoler har vært opptatt av å lage egne program for samfunnsfagsundervisningen, ofte rettet mot lokale forhold.

Noen få skoler har tatt i bruk Lovdata, en fulltekstdatabase, i faget Rettslære.

2. Hva vil skje framover?

a) Pedagogiske program

Skolene vil få flere pedagogiske program som gir lærerne ulike mulighe-

ter til å variere undervisninga si. Disse programmene vil dekke områder der edb har en klar fordel framfor andre hjelpemidler i faget. De enkelte samfunnsfaga vil få noen spesielle verktøy som sparer tid og/eller setter oss i stand til å gjøre noe nytt vi ellers ikke har tid til.

Både når det gjelder Datasekretariatets og forlaga sin programproduksjon er det innen samfunnsfaga det lages flest program. Dette viser at norske samfunnsfaglærere har tatt mediet i bruk og vil være med i dette arbeidet framover.

b) Telekommunikasjon

Økt bruk av telekommunikasjon vil gi skolene tilgang til fremmede numeriske og fulltekstbaser som f.eks. NTB, Statistisk Sentralbyrå, universitetsnettet m.m. I tillegg vil det tvinge seg fram muligheter til å gå inn og lese i systemene til offentlige institusjoner. Datatilgjengeligheten vil bli enorm, problemet blir å kunne velge.

Ennå er bruken av disse mediene for vanskelig. Verken leverandørene av utstyr, programskaperne eller basevertene har lagt ned nok arbeid i å lette tilgangen til alle disse kildene.

c) Masselagringsmedium

Nye former for masselagring fører til at det på skolene vil være et stort tilbud av videoplater med tekst, bilder og lyd. De data som i dag bare er tilgjengelige for noen få, vil bli tilgjengelige for mange. Læreboka vil bestå, men biblioteket vil få nye funksjoner og AV-utstyret på skolen vil endre seg.

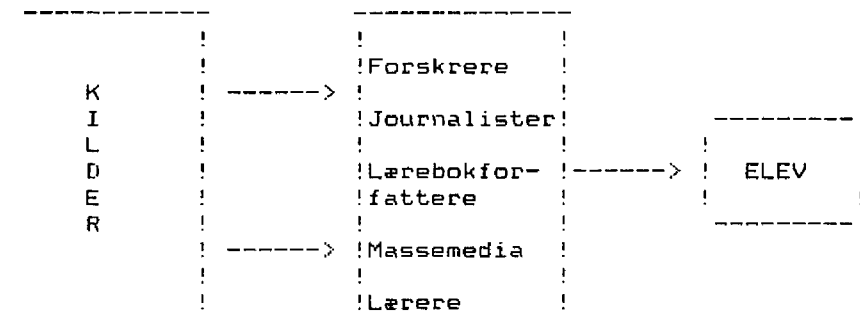
3. Hva skjer egentlig med faget?

I de siste tiåra har det vært en vedvarende debatt om popularisering i samfunnsfaga. Det har vært diskutert hvorvidt forskerne sjøl skulle popularisere eller om andre yrkesgrupper som journalister, lærere, lærebokforfattere eller andre skulle stå for arbeidet.

I dag står alle de profesjonelle mellom eleven og kildene, bare en uhyre liten del av dem kommer fram til elevene i skolens undervisning.

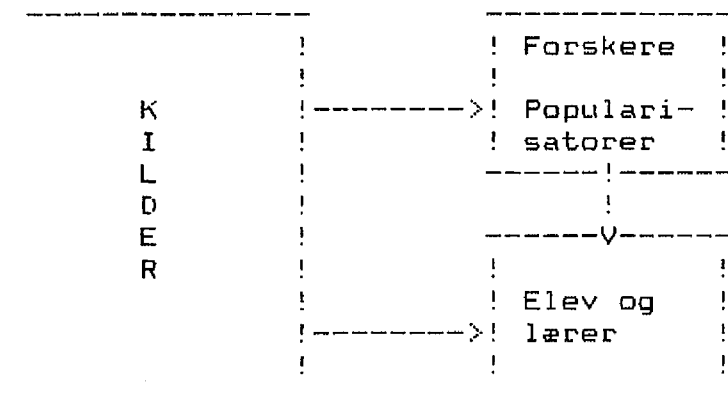
Moderne bruk av edb vil snu om denne problemstillinga. I stedet for å popularisere, vil det gjennom bruk av edb være mulig å gjøre kildene tilgjengelige for flere ved at de blir lettere å bruke, lettere å få tilgang til. Langt flere mennesker enn en utvalgt gruppe forskere ved universiteter og forskningsinstitusjoner vil kunne bruke kildene. Mulighetene for å etterprøve forskernes arbeid vil bli store, både for studenter, elever og lærere.

Mulighetene for å kontrollere popularisatorene vil bli enda større. Verken aviser, radio eller fjernsyn har plass eller tid til en slik etterprøving i kildene. Og om de gjør det, så blir ikke kildene eller bakgrunnstoffet brakt fram for publikum.



Figur 1. Situasjonen i dag.

For skolen sin del, så har vi tid til å gjøre dette. Elevene ved allmennfaglig studieretning har i dag 11 timer felles samfunnsfag over tre år. Elevene på samfunnsfaglig linje kan de to siste åra ha inntil 28 timer samfunnsfag i tillegg. Samtidig er det et krav at skolen skal trene elevenes kritiske sans overfor media. Datamaskinen vil kanskje bli det viktigste hjelpemidlet vårt i dette arbeidet, siden den gir oss økt tilgang til kildene.



Figur 2. Situasjonen i framtida.

Hovedmengden av lærestoffet vil nok fortsatt komme via forskere og lærebokforfattere. Edb vil likevel gi skolen mulighetene til direkte bruk av kildene i en helt annen grad enn nå.

Registreringsentral for historiske data har f.eks. laga et program for å lette tilgjengeligheten til folketellingsdata. Når elevene bruker dette programmet, kan vi lete oss tilbake til de enkelte familiene i bygda for året 1900. Én sak er at elevene kan finne tilbake til sin egen familie, en annen sak er at myten om storfamilien (innført av Bjørnson og Lie, holdt i live av norske samfunnsforskere) blir avlivet. De hadde ikke så mange barn, bestemor bodde ikke på gården, det var en utstrakt flytting osv.

Og når alle publikasjonene til Statistisk Sentralbyrå i løpet av et år kan bestilles til skolen på en CD-ROM, så vil mulighetene for å følge med i hva som skjer i samfunnet øke enormt.

Eller når alle diskuterer hva forsvarsministeren har stemt for på et NATO-møte og ingen i landet trykker teksten (heller pene bilder), så kan læreren snart hente ut kommunikéet fra en eller annen base her eller i utlandet. Deretter kan vi diskutere realiteter.

4. Konsekvenser for faget og lærerutdanninga

a) Edb-kunnskap

Samfunnsvitene må skaffe seg noen brukerferdigheter i edb. Det vil bli viktig å lære seg å bruke ulike portprogram for å nå fram til kildene. Slike koder, som er bærere av informasjon, har vi alltid hatt i faga. Latin, matematikk eller engelsk er tradisjonelle koder. Å bruke Teledata er bare en ny, og dette vil bli lettere med åra.

Dette skjer likevel ikke av seg sjøl. Vi må alle delta i kravene om at tilgjengeligheten til data skal bli lettere. Men for å kunne stille krav, så må vi ha noen kunnskaper og ferdigheter.

b) Ny pedagogisk problemstilling

Med en masse data og primærkilder tilgjengelig, vil ett av hovedproblema for lærerne være å kunne velge ut hva som skal brukes overfor elevene. Kunnskaper om hva som finnes og pedagogiske ferdigheter i bruk overfor elevene vil stille helt nye krav til lærerne. Lærerne i samfunnsfag vil bli mer veiledere og mindre allvitere. Det blir særdeles viktig å kunne sette opp de riktige problemstillingene, kunne velge ut det rette stoffet og så presentere dette overfor elevene.

Sannsynligvis vil det kreve mer tid til forberedelser enn i dag, og lærerne må få tid til å sette seg inn i disse nye hjelpemidlene. I tillegg vil eleven sjøl kunne bruke mediene.

c) Økt kildekunnskap

For faget sin del vil nok likevel det viktigste bli å kjenne kildene. For å nyttgjøre seg de nye mulighetene må en lærer vite oppbygning, styrke og svakheter ved f.eks. kirkebøker, folketellinger, tekstbasene til pressebyråene og tallbasene til Statistisk Sentralbyrå og Norsk samfunnsvitenskapelig datatjeneste.

Læreren må kunne bruke kildene samtidig som hun/han kan stille seg kritisk til dem. Disse faglige basiskunnskapene, ferdighetene og holdningene må lærerne tilegne seg gjennom fagutdanninga si, på grunn- og mellomfagsnivå ved universiteter og høyskoler.

Innføring av edb i skolen vil først og fremst bli en faglig utfordring for læreren. Derfor må det stilles slike nye faglige krav til utdanningsinstitusjonene.

At kildene må ut til folket, ble tidlig forstått av Martin Luther. Derfor er det i dag vår oppgave å gjennomføre en liknende reformasjon i samfunnsfaga.

Machine-Readable Texts in English Language Research

Stig Johansson

Why machine-readable texts?

Language is all around us. Speech and writing are part of everyday life. There is certainly no lack of data for those who want to study language. Why do we then need machine-readable texts? The problem is that there are innumerable aspects that we might want to study. Perhaps we are interested in the use of a word (e.g. *big* vs *great*) or a grammatical construction (e.g. the alternative negatives in *he's no teacher* and *he's not a teacher*). To study such matters systematically, it is necessary to have access to data beyond everyday experience. Traditionally, linguists collected citations from texts, often with the focus on the unusual or the peculiar. This is a laborious and time-consuming process, but it was successfully used in such monumental works as the *Oxford English Dictionary* and the classical grammars by Jespersen, Kruisinga, and Poutsma. More recently, language researchers have turned to introspection (i.e. relying upon their intuition in judging how language is used) and elicitation (i.e. testing the language use of native speakers under controlled conditions). Both sources of data are valuable, but they also have serious weaknesses. Intuition quickly becomes blurred, and it can easily be influenced by the theories of the researcher. The main problem with elicitation tests is that they are always more or less artificial, and then it is difficult to know to what extent we can generalize from the results.

The most valid and reliable source of data is language as it is actually used in speech and writing. The great advantage of machine-readable texts is that, once they have been prepared in machine-readable form and given the necessary hard- and software, they provide immediate access to data on a wide range of aspects of the language. It is no longer necessary to spend months or years collecting data. The data are there at the outset, and the researcher can proceed without further ado to the analysing stage.

Archiving of machine-readable texts

Preparing texts in machine-readable form is a laborious and expensive undertaking. For this reason, researchers have realised the importance of pooling their resources. A great number of machine-readable texts have

been deposited at the Oxford Text Archive, from where copies can be obtained at cost for use in non-profit research. During the last ten years or so Norway has played an important role in the use of machine-readable texts in English language research. The International Computer Archive of Modern English (ICAME) was set up in 1977, with the aims of:

- (1) collecting and distributing information on the use of English machine-readable texts;
- (2) compiling an archive of English machine-readable texts and coordinating the distribution of the material to academic researchers.

The Norwegian Computing Centre for the Humanities serves as a clearing-house for ICAME. It distributes texts and publishes the journal *ICAME News* (from 1987 to be called the *ICAME Journal*). The latest issue of *ICAME News*, which includes a report from the 6th ICAME Conference and a bibliography of some 300 items, gives a good picture of current work.

Three major corpora

ICAME currently distributes three major collections of texts (so-called corpora):

- the Brown Corpus (the first major collection of machine-readable texts; compiled at Brown University in the early 1960's; a million words of printed American English)
- the Lancaster-Oslo/Bergen (LOB) Corpus (the British English counterpart of the Brown Corpus; compiled in the 1970's through cooperation between the University of Lancaster, the University of Oslo, and the Norwegian Computing Centre for the Humanities)
- the London-Lund Corpus (approx half a million words of educated spoken British English; recorded and transcribed at University College London and computerised at the University of Lund)

A recent addition is the Melbourne-Surrey Corpus, a collection of Australian newspaper texts (compiled by Kurshid Ahmad and Greville Corbett, University of Surrey). The three main corpora are available both as running text and in concordance form, with a complete list of citations for each word. The material has been used in a large number of studies. Let us look at a concrete example.

English adverbs in *-ly*

In a recent article Göran Kjellmer, University of Gothenburg, asks the question 'Why GREAT:GREATLY but not BIG:*BIGLY?' (*Studia Linguistica* 38, 1984). It has been said, even by a keen observer like Jespersen, that English adverbs in *-ly* can be formed from nearly all

Brown Corpus: original version

S**U. **NMERGER PROPOSED**P *HOWEVER, THE JURY SAID IT BELIEVES **QTHE 0210E1A01
SE TWO OFFICES SHOULD BE COMBINED TO ACHIEVE GREATER EFFICIENCY AND RE 0220E1A01
DUCE THE COST OF ADMINISTRATION**U. **R**T *THE *CITY *PURCHASING *DEP 0230E1A01
ARTMENT, THE JURY SAID, **QIS LACKING IN EXPERIENCED CLERICAL PERSONNE 0240E1A01
L AS A RESULT OF CITY PERSONNEL POLICIES**U. *IT URGED THAT THE CITY * 0250E1A01
*TAKE STEPS TO REMEDY**U THIS PROBLEM. **R**T *IMPLEMENTATION OF *GEO 0260E1A01
RGIA**AS AUTOMOBILE TITLE LAW WAS ALSO RECOMMENDED BY THE OUTGOING JU 0270E1A01
RY. **R**T *IT URGED THAT THE NEXT *LEGISLATURE **QPROVIDE ENABLING FU 0280E1A01
NDS AND RE-SET THE EFFECTIVE DATE SO THAT AN ORDERLY IMPLEMENTATION OF 0290E1A01
THE LAW MAY BE EFFECTED**U. **R**T *THE GRAND JURY TOOK A SWIPE AT TH 0300E1A01

Brown Corpus: Bergen Version I.

A01 0210 #MERGER PROPOSED# However, the jury said it believes "these
A01 0220 two offices should be combined to achieve greater efficiency and reduce
A01 0230 the cost of administration". The City Purchasing Department,
A01 0240 the jury said, "is lacking in experienced clerical personnel
A01 0250 as a result of city personnel policies". It urged that the city "take
A01 0260 steps to remedy" this problem. Implementation of Georgia's
A01 0270 automobile title law was also recommended by the outgoing jury.
A01 0280 It urged that the next Legislature "provide enabling funds
A01 0290 and re-set the effective date so that an orderly implementation of
A01 0300 the law may be effected". The grand jury took a swipe at the

Brown Corpus: Bergen Version II.

A01 0210 1 #MERGER PROPOSED#
A01 0210 3 However, the jury said it believes "these two offices
A01 0220 3 should be combined to achieve greater efficiency and
A01 0220 11 reduce the cost of administration".
A01 0230 5 The City Purchasing Department, the jury said, "is
A01 0240 5 lacking in experienced clerical personnel as a result
A01 0250 4 of city personnel policies". It urged that the city
A01 0250 13 "take steps to remedy" this problem.
A01 0260 6 Implementation of Georgia's automobile title law
A01 0270 4 was also recommended by the outgoing jury.
A01 0280 1 It urged that the next Legislature "provide enabling
A01 0280 9 funds and re-set the effective date so that an orderly
A01 0290 11 implementation of the law may be effected".

adjectives. But many adjectives do not, in fact, have corresponding adverbs. What are the restrictions on adverb formation?

On the basis of their frequency in the Brown Corpus, Kjellmer selected the most common adjectival suffixes: *-able*, *-al*, *-an*, etc. He turned to a reverse dictionary in order to chart the numerical relationship between adjectives ending in these suffixes and their related adverbs in *-ly*. The average adverb-forming tendency for all the types was as low as 13%. There were, however, considerable differences between the morphological types. The least adverb-forming type was *-oid* (0%), and the most

productive one was *-ful* (41%).

In explaining the difference in adverb-forming tendency Kjellmer turned to a study of the semantics of adjectives. Adjectives are characteristically stative (e.g. *black, British*), i.e. they refer to states and stable conditions, but there is a minority deviating from the normal pattern. Such 'dynamic' adjectives (e.g. *careful, abusive*) are more liable to form adverbs. The adverb-forming tendency is also related to the type of nouns which the adjectives combine with. Like adjectives, nouns are typically non-dynamic (e.g. *girl, name*), but there are also dynamic nouns (e.g. *answer, correspondent*). Kjellmer looked up a selection of stative adverb-forming (*anonymous, manual, microscopic, verbal*) and non-adverb-forming (*blonde, blue, Moslem, tall*) adjectives in the LOB Corpus. He found that the adverb-forming types cooccurred particularly with dynamic nouns.

Kjellmer can now answer the question about *great* vs *big*. While *big* is normally stative and refers to physical dimensions (as in *a big man*), *great* is either dynamic in itself (as in *a great man*) or combines with dynamic nouns (as in *a great poet*). Its normal use is outside the sphere of physical dimensions.

Analysed text

The number of texts in machine-readable form is growing rapidly. But 'raw' text is not enough. From a raw text it is only possible to retrieve examples of word forms, so-called graphic words. We cannot, for example, retrieve the occurrences of the verb *like* without at the same time getting the adjective, the noun, the preposition, and the conjunction (all mixed together). And we do not get all the other forms of *like* (*likes, liked, liking*). Nor can we search for grammatical categories such as 'adjective' or 'adverb'. For these reasons, a good deal of work has been done to 'tag' texts. There are now tagged versions of the Brown Corpus and the LOB Corpus, which make it possible to separate homographs (e.g. the different uses of *like*) and to search for grammatical patterns (e.g. sequences of nouns or adjectives).

The London-Lund Corpus is, in fact, an analysed text, with detailed prosodic marking (stress, pauses, intonation). Bengt Altenberg, University of Lund, has recently used the material in a highly illuminating study of the relationship between prosody and grammar. The study, which forms part of the project 'Text Segmentation for Speech', could have great importance for the production of natural-sounding synthetic speech.

One of the most significant advances in the course of the last few years is the development of the Linguistic Data Base at the University of Nijmegen. This is a data base system which allows sophisticated searches in analysed text.

Vertical format

```

A01 2 001 -----
A01 2 002 **   **           H
A01 2 010 VB   stop           H
A01 2 020 VBG  electing       H
A01 2 030 NN   life           H
A01 2 040 NNS  peers          H
A01 2 041 ***  ***           H
A01 2 042 .    .             H  @
A01 3 001 -----
A01 3 010 IN   by             H
A01 3 020 NP   Trevor         H
A01 3 030 NP   Williams       H
A01 3 031 .    .             H  @
A01 4 001 -----
A01 4 010 AT   a               P
A01 4 020 NN   move
A01 4 030 TO   to
A01 4 040 VB   stop
A01 4 050 NPT  \OMr           \O
A01 4 060 NP   Gaitskell
A01 4 070 IN   from
A01 4 080 VBG  nominating
A01 4 090 DTI  any
A01 4 100 AP   more
A01 4 110 NN   labour         N
A01 5 010 NN   life           N
A01 5 020 NNS  peers          N
A01 5 030 BEZ  is
A01 5 040 TO   to
A01 5 050 BE   be
A01 5 060 VBN  made
A01 5 070 IN   at
A01 5 080 AT   a
A01 5 090 NN   meeting
A01 5 100 IN   of
A01 5 110 NN   labour         N
A01 5 120 NPTS \OMPs         \O
A01 5 140 NR   tomorrow
A01 5 141 .    .

```

Horizontal format

```

A01 2  ^ '*_** stop_VB electing_VBG life_NN peers_NNS ***_*** _..
A01 3  ^ by_IN Trevor_NP Williams_NP _..
A01 4  ^ a_AT move_NN to_TO stop_VB \OMr_NPT Gaitskell_NP from_IN
A01 4  nominating_VBG any_DTI more_AP labour_NN
A01 5  life_NN peers_NNS is_BEZ to_TO be_BE made_VBN at_IN a_AT meeting_NN
A01 5  of_IN labour_NN \OMPs_NPTS tomorrow_NR _..

```

The tagged LOB Corpus is available in two formats.

a_6FW T=14/0.11/33.8, C=3/09.4, G=5/57.9, J=2/12.4,
a_AP T=438/1.94/432.4, A=26/291.7, B=30/551.0, C=17/506.6, D=8/232.6, E=44/572.1, F=43/482.6, G=59/379.8, H=14/230.4,
J=47/290.3, K=53/895.2, L=23/468.0, M=6/495.0, N=25/421.4, P=30/505.2, R=13/714.2,
a_AT T=21939/96.08/21657.4, A=7042/22908.0, B=1122/20607.2, C=912/27180.1, D=641/18640.2, E=1795/23337.1, F=2208/24982.8,
G=3420/22015.9, H=1066/17541.8, J=2997/18510.6, K=1241/20961.1, L=1184/74042.0, M=267/21229.7, N=1369/23076.2,
P=1201/20225.6, R=474/26039.7,
a_IN T=1/0.08/1.0, C=1/29.8,
a_JJ T=3/0.08/1.0, J=1/18.5,
a_JH T=159/0.74/136.0, A=6/47.3, B=6/110.2, C=6/178.8, D=4/116.3, E=14/182.0, F=1/173.5, G=24/154.5, H=4/65.8, J=14/86.5,
M=15/270.2, L=13/264.5, N=1/82.5, H=18/303.4, P=20/336.8, R=1/54.9,
a_RB T=2/0.08/2.0, D=1/6.4, K=1/16.9,
a_ZZ T=218/1.08/215.2, A=5/56.1, C=5/149.0, D=3/87.2, E=7/91.0, G=8/51.5, H=26/427.8, J=159/982.0, L=2/40.7, M=1/82.5,
R=2/105.9,
COA_6FW T=1/0.08/1.0, J=1/6.2,
COA_HH T=1/0.08/1.0, J=1/6.2,
COA_HU T=82/0.88/80.9, A=22/246.8, B=10/183.7, C=7/208.6, D=5/145.4, E=9/117.0, F=5/56.1, G=5/32.2, H=7/115.2, J=10/61.8,
K=1/16.9, H=1/16.9,
C02 140 " here in London in 1961, we are seeing " a_6FW bout_6FW de_6FW souffler*72 *** too late, of course, to level
C02 115 " the academy advertises Jean-Luc Godard's a_6FW bout_6FW de_6FW souffler*72 (translated as breathless, &
C02 155 high -- hopes it began with, " certainly " a_6FW bout_6FW de_6FW souffler*72 *** (which is almost a group
G44 111 " a substantial portion of a Haydn symphony: "non a_6FW charge*72_6FW le_6FW sentiment de la musique de la Fil*72ite
G44 113 " les paroles sont pitoyables ... l'arrangeur a_6FW coupe*72_6FW, taille*72_6FW, saire*72 les plus beaux
G41 73 human beings. " If Ulysses has little in common with a_6FW la_6FW recherche_6FW du temps perdu, still less has the
G41 174 had been as orthodox as that of, say, Galworthy, a_6FW la_6FW recherche_6FW du temps perdu would still remain a
G41 179 been even greater. " true, it is hard to imagine a_6FW la_6FW recherche_6FW without Charlus; yet it is at least
G44 211 mieux, il ne donne pas tant de fume*72er il n'y a_6FW que_6FW les_6FW auteurs qui vont de plus en plus mal ***. "
G44 112 de la musique de la Fil*72ite enchante*72e, en en a_6FW talent_1_6FW les_6FW mouvements peut approprier les airs au
G41 86 to be illuminating. " In the same year as de Coups a_6FW Satan_6FW pe_72ra published a pamphlet in English, the
G41 77 " in the ring of the poet, and such is what de Coups a_6FW Satan_6FW examiens WOC throughout the later poetry.
G44 202 not bear time, and concluded dolefully: " " il y a_6FW trop_6FW de_6FW musique sans ton Jean; c'est un
G47 53 for the irritation of the untutored viewer via a_6FW via_6FW abstract_3D painting. " he knows for what is not
G24 49 wedding feast: " " enthusiasm grew, and in a_AP few_AP more_AP points we gave a concert at which the
F32 127 outside world. " soon the head can be seen. " a AP few AP more AP pushes, and it 's half born. " and then the

KWIC concordance of the tagged LOB Corpus, with frequency information.

Great advances have been made recently in the methods of producing analysed text.

Some work relating to the LOB Corpus

My own work with machine-readable texts relates particularly to the LOB Corpus. In the last few years I have been involved in the grammatical tagging of the LOB Corpus, in cooperation with researchers at Lancaster and Bergen. Each word in the corpus has been given a tag indicating its grammatical class. A system of automatic grammatical tagging achieves approx 96% accuracy. The system, which has primarily been developed at Lancaster, is based on probabilistic principles.

In spite of the high degree of accuracy, it was necessary to spend a great deal of time in post-editing the corpus, i.e. in correcting errors made by the automatic tagging programs. A study of these errors reveals weaknesses in the programs and can be used as a basis of improvement. The post-editors had to grapple with many difficult problems of classification, e.g. the classification of *as* (conjunction vs preposition) and *so* (conjunction vs adverb), the borderline between adjective and verb with *-ed* and *-ing* forms. Classification problems are dealt with in considerable detail in the manual for the tagged corpus, and the discussion should be of interest both in future tagging projects and in English linguistics in general.

The post-edited corpus has been used to produce a homograph-separated concordance, with a complete list of citations for each word (sorted by tag) and including statistics on the distribution in different types of texts. Studies on the following levels are under way (all in collaboration with Knut Hofland, Norwegian Computing Centre for the Humanities):

- word frequencies (homograph-separated word lists)
- word combinations (a dictionary of collocations)
- tag frequencies (including the word-class distribution in different types of texts)
- tag combinations (tag-pair probabilities, word-class sequences at the beginning and the end of sentences)

The results should both be of interest in English language research in general and in work on automatic grammatical analysis and speech recognition.

Prospects

A great deal has happened since the first computer corpus was compiled some twenty years ago. We now have a wide variety of text types, both raw text and analysed text. The texts are finding an increasing number of users and uses. There is a continuing development within ICAME as well as in the computing world at large.

While the first computer corpora were produced by keyboarding, there are now other sources of machine-readable texts. In the first place, there are printing tapes produced in connection with computer typesetting; this is of course subject to approval of the publishers. Secondly, it is possible to use optical scanners. Some newspapers have introduced electronic archives, based on their printing tapes, instead of the traditional files of clippings. Though such archives are primarily meant for information retrieval, they can also be a source of observations on language use. Access to newspaper archives can be obtained on a commercial basis. An increasing number of texts are now published both in printed and electronic form (parallel publishing) or only in electronic form (electronic publishing). Exciting possibilities are offered by new storage media. Soon the linguist will be able to have access to gigantic collections of data on CD-ROM (holding as much as the equivalent of 250,000 pages on a single disc). The new disc technology also makes it possible to combine different media (e.g. sound and transcription).

Certain types of data will continue to be less easily available. The computerisation of spoken material and written unprinted texts will remain laborious, until further advances are made in automatic speech recognition and optical scanning. The preparation of tagged texts will require very much work, although advances have been made in automa-

tic tagging by computer (as used in preparing the tagged LOB Corpus). The use of less easily available sources of data can be greatly facilitated by the type of cooperation developed within ICAME.

Current work within ICAME includes the exchange of software for use with machine-readable texts, the use of networking as a means of contact, and the utilization of the new disc technology. There are plans to produce a disc with the three major corpora distributed through ICAME and some samples of spoken material in the form of both transcription and sound.

Language research in the future

Linguists in the future will no longer have to be data collectors. Nor do they need to derive their data from intuition. Machine-readable texts will be plentiful. But even very large collections of texts may not contain the examples which the researcher needs to test his/her hypotheses. There will still be a need for the selective reading and listening of traditional research and for the well-considered use of introspection and elicitation.

Last and most important, machine-readable texts provide raw material, which may mean very little without the researcher's insight and imagination. But the richer the source material, the greater are the chances of progress in the interaction between researcher and data.

Stig Johansson is Professor of English Linguistics at the University of Oslo.

Computers and Philosophical Manuscripts

The Case of the Norwegian Wittgenstein Project

Claus Huitfeldt

Introduction

When the Austrian philosopher Ludwig Wittgenstein died in 1951, he left behind a considerable amount of unpublished manuscripts.

Some parts of these manuscripts have been published posthumously, but Wittgenstein's Nachlaß as a whole is at present not easily accessible for researchers. There are roughly three main reasons for this: First, restrictions on access to the originals, second, the fact that easy reading of Wittgenstein's handwriting requires weeks and months of practice, and, third, that no complete index to the Nachlaß exists.¹

The chief objective of our project is to make Wittgenstein's Nachlaß more easily available for philosophical research. We are trying to accomplish this by transcribing the material in its entirety to machine-readable form.²

Transcription and code system

We transcribe from copies of the original by entering³ the text while at the same time inserting certain code markers. This coding is done according to a specially designed code system, which consequently is essential to all other aspects of our work.

If you start transcribing a manuscript of Wittgenstein's you soon find yourself in a position where you have to decide between several possible renderings of the same text. If you do not have any clear idea of how to handle certain peculiarities of his texts and handwriting, this decision will have to be arbitrary or based on some specific interpretation of the text, or you may try to render all the alternatives, leaving the text more or less unreadable. The code system does not eliminate all these problems, but it makes them less troublesome.

Our current code system is based on the «Indikationssystem» developed by the Tübingen Project called «Sicherung, Dokumentation und Erschließung der nachgelassenen Schriften Ludwig Wittgensteins», a project to which we owe a lot.

The three main purposes of the code system are:

- 1) to provide a system for references to the texts

- 2) to provide a means of indicating peculiarities of the original which cannot easily be rendered by the computer's ordinary character set, like different kinds of underlining, deletions, alternative expressions, etc.
- 3) to provide a means for singling out certain interesting search objects, like proper names, headings, mathematical expressions, etc.

Let us have a look at a few examples of each of the kinds listed above.

There are special codes for marking of pages, sections, and lines in the original. Each new page in the original counts as a new page – irrespective of the pagination in the original. Each sequence of one or more blank lines in the original text counts as marking the beginning of a new section. There is also a marker for the identification of end-of-line in the original. A specially designed program takes care of the numbering of pages and sections.

This gives us two ways of referring to any line in the text: Either page number and number of line on the page, or section number and number of line in the section.

Codes indicating peculiarities of the original which cannot be rendered by the computer's character set are used to mark signs, remarks and lines in the margin, indentations, section markings, deviant first-letters, handwriting in typescript, remarks not by Wittgenstein, linguistic deviations peculiar to Wittgenstein, insertions, underlining or spacing, extensions, deletions, and alternative expressions.

These are some of the «main code groups». Each of these may be divided into subgroups. The main code «underlining or spacing», e.g., requires a subcode indicating whether we are dealing with a single straight line, a double straight line, a single curved line, a double curved line, a broken (or dotted) line, or a letter-spaced word.

The codes indicating possible search objects are used to mark proper names (or derivations of proper names), headlines or subject indications, datings, «Geheimschrift», and logical or mathematical expressions. (In principle, any of the codes may of course be considered as marking a possible search object.)

All in all there are about 55 different codes. As several different peculiarities may pertain to the same parts of a text, recursive coding is both allowed for and widely used. The coding may consequently become quite complex in certain cases.

Even so, these codes do not cover all possible aspects of the texts which may be of interest to describe in a transcription. Therefore one is allowed to put comments and references to a transcription protocol directly in the text between special markers.

The transcriber produces what we call a «source transcription», containing both text and coded information. The source transcription is

reformatted and checked for coding errors by a specially designed program, which returns error messages for every code syntax error.

A text which passes the code-checking program without error messages is ready for further treatment by the rest of our software.

The texts

We transcribe from the Cornell photocopies of Wittgenstein's Nachlaß, which comprise 98 volumes of around 15,200 pages. 2650 of these pages are typescripts, the rest is handwritten material.

At the NWP, we have transcribed 2850 pages in all. In addition to these we have at our disposal a more comprehensive material which the Tübingen Group generously transferred to us.

The Tübingen transcriptions are coded according to a somewhat different code system, or, rather, code systems.⁴ Because the texts are not yet re-coded, we have not been able to find out exactly how many pages they comprise.⁵

The number of pages covered by Tübingen transcriptions about which we have fairly certain knowledge is about 4700. Assuming that all the Tübingen transcriptions of which we have no clear indications to the contrary, are complete, they comprise all in all at least 6150 pages. 1000 pages overlap with transcriptions at the NWP.

To sum up, all this means that we have at our disposal somewhere between 6550 and 8000 pages of the Nachlaß in machine-readable form.

We are working on a database which will contain exact information about both the Nachlaß itself and the transcriptions. This has not been accomplished yet, due to a lack of resources. In this work also we may benefit greatly from the work already done by the Tübingen Group.

Filtered print-outs

None of NWP's computer programs apply new or revolutionary methods of text processing or text analysis. What is special about them is that they are designed specifically to handle and make maximum use of the information of our code system, which is in turn specially developed for the transcription of Wittgenstein's manuscripts.

One of the things we can do with a text, provided it is properly coded, is to produce what we call «filtered» print-outs of each text by means of a special «filtering» program. After all, a coded text, i.e. a source transcription, is not easy to read. In fact, it is not meant to be read, at least not by the ordinary user.

The filtering program runs like this: The user is asked for the von Wright number of the text he wants filtered. He is asked whether he wants the filtered text listed on the screen, on paper, or to be stored on a disk. Finally, he is asked whether he wants a «simple» or a «detailed»

Ich kann sagen: Dieser Mann unterscheidet nicht
 Rot & Grün; ich aber sage: Wir Normalen unterscheiden Rot
 & Grün? Wir können aber sagen: "Wir sehen $\text{\$a}(\text{hier})$ 2 Farben,
 jener nur $\text{\$a}(\text{eine})$." (13)

Die Beschreibung des Phänoms
 der Farbenblindheit ist eine psy-
 chologische gehört in die Psycho-
 logie // Aber auch die des Phänoms
 des herkömmlichen Farbensehens:
 also das sieht man, das man logisch erklären kann
 Gesehen = aber das ist ja bestimmt? Sieht so
 aus wie ein Buch, aber was ist das?
 für wen ist es? Die Beschreibung
 dabei besser: welches Hilfsmittel
 bedient sie sich? Wenn ich sa-
 ge "Was setzt sie voraus?", so
 heißt das: "Wie muß ich mich auf
 sie, was reagiere ich mit
 merkmale, die ich in der Umwelt sehe, die ich mit
 Sehen. Die Beschreibung auch mit dem, was ich
 unter dem Begriff herkömmlich

Dieses Papier ist an verschie-
 denen Stellen verschieden hell;
 aber es kann ich sagen, es sei nur
 an gewissen Stellen hell, an den

*124)

*13(Ich) kann sagen: Dieser $\text{\$a}(\text{Mann})$ unterscheidet nicht
 Rot & Grün; Kann ich aber sagen: Wir Normalen unterscheiden Rot
 & Grün? Wir können aber sagen: "Wir sehen $\text{\$a}(\text{hier})$ 2 Farben,
 jener nur $\text{\$a}(\text{eine})$."

*1c(13)

#(51)

*15(Die) Beschreibung der $\text{\$a}(\text{Ph}^{\text{änomene}})$ der Farbenblindheit
 *12(ist eine psy_chologische/gehört *12(in die/zur) Psycho_logie).
 Also auch die der Ph^ano_mene des normalen
 Farbensehens? Gewi's, - *12(aber was setzt so_eine Beschreibung
 voraus, f^ur wen ist es eine Beschreibung, oder/aber was nennt man
 "Ph^anomene des norm*(alen) Farbensehens", was ihre
 Beschreibung, & f^ur wen ist sie bestimmt? Oder) besser:
 welches Hilfsmittel bedient sie sich? Wenn ich sa_ge
 "Was setzt sie voraus?", so heißt das: "Wie muß ich mich auf
 $\text{\$a}(\text{sie})$ sehen reagieren, um sie zu $\text{\$a}(\text{verstehen})$." *1a(*12(Wer
 in einem Buch die Ph^anom*(ene) der Farbenblind*(eit)
 beschreibt, beschreibt sie mit den Begriffen der_Sehenden/ Jene
 *1a(*1a(erste)) Beschreibung, an die ich dachte, war eine
 mittels_der Begriffe der Sehenden.)

#(52)

*7(Dieses) Papier ist an verschie_denen Stellen verschieden
 hell; aber kann ich sagen, es sei nur_an gewissen Stellen
 wei's, an den

1) An example of a page from the Cornell photocopy of
 von Wright no. 173.

2) Source transcription of the same page.

0 Ich kann sagen: Dieser Mann unterscheidet
1 nicht Rot & Grün: Kann ich aber sagen: Wir Normalen
2 unterscheiden Rot & Grün! Wir könnten aber sagen:
3 "Wir sehen HIER 2 Farben, jener nur EINE."

51

1 Die Beschreibung der PHÄNOMENE
2 der Farbenblindheit
3 gehört zur Psycho-
4 logie. Also auch die Phäno-
5 mene des normalen Farbensehens?
6 Gewiß, -

7

8

9 aber was nennt man "Phänomene des norma-
10 ihre Beschreibung, & für wen ist sie bestimmt?
11 Oder besser: welcher Hilfsmittel
12 bedient sie sich? Wenn ich an-
13 ge "Was setzt sie voraus? ", so
14 heißt das: "Wie muß einer auf
15 SIE schon reagieren, um sie zu
16 VERSTEHEN."

17

18 Jene erste Beschreibung, an die ich dachte, war eine mittels
19 der Begriffe der Sehenden.

52

1 Dieses Papier ist an verschie-
2 denen Stellen verschieden hell;
3 aber kann ich sagen, es sei nur
4 an gewissen Stellen weiß, an den

\ 13 \

51

\ "Die Beschreibung der *Phänomene* der Farbenblindheit 24
gehört zur Psychologie.³ Also auch die der Phänomene des
normalen Farbensehens? Gewi, - aber was nennt man
"Phänomene des normalen Farbensehens", was ihre
Beschreibung, und für wen ist sie bestimmt? Oder⁴ besser:
welcher Hilfsmittel bedient sie sich? Wenn ich sage "Was
setzt sie voraus? ", so heißt das: "Wie muß einer auf sie
schon reagieren, um sie zu *verstehen*." Jene erste
Beschreibung, an die ich dachte, war eine mittels der
Begriffe der Sehenden.⁵

52

Dieses Papier ist an verschiedenen Stellen verschieden
hell; aber kann ich sagen, es sei nur an gewissen Stellen
weiß, an den \ andern aber grau?? - Ja, wenn ich es malte,⁶ 25

1 Hinzufügung im oberen Rand, links: Ich kann sagen: Dieser Mann unterscheidet nicht Rot und Grün: Kann ich aber sagen: Wir Normalen unterscheiden Rot und Grün? Wir könnten aber sagen: "Wir sehen hier 2 Farben, jener nur eine."

2 Hinzufügung im oberen Rand, rechts: 13

3 Alternative: [ist eine psychologische / gehört [in die / zur] Psychologie]

4 Alternative: [aber was setzt so eine Beschreibung voraus, für wen ist es eine Beschreibung, oder / aber was nennt man "Phänomene des norma-
len Farbensehens", was ihre Beschreibung, und für wen ist sie bestimmt? Oder]

5 Alternative: [Wer in einem Buch die Phänomene⁶ der Farbenblindheit beschreibt, beschreibt sie mit den Begriffen der Sehenden / Jene erste Beschreibung, an die ich dachte, war eine mittels der Begriffe der Sehenden]

6 Überstrichen: so

3) Filtered print-out («simple» filtering) of the source transcription.

4) Example of what a laser print from the same source transcription might look like.

rendering of the source transcription.

In a «simple» print-out several coded peculiarities are suppressed. Deleted words are omitted, only the last of connected alternatives are rendered, etc.

The result is a «clean», easy-to-read rendering of the source transcription.

In some cases you will want to have more information about the text, for a closer study. A «detailed» print-out renders peculiarities suppressed or omitted in a «simple» print-out between special markers.

The result is a rendering of the source transcription somewhat less easy to read, but richer in information.

We hope to be able to develop this program a bit further, so that the user may choose a «specified» print-out, in which case he may specify for each coded peculiarity of the text whether he wants it suppressed, omitted, or marked.

One may perhaps feel tempted to ask: Why do we make things so complicated? There are at least two answers to this, depending on how the question is motivated.

One way of motivating the question is: We do not really need any rendering of all these peculiarities of the texts.

My answer is that sometimes you do want to know all the details of a text, but sometimes, perhaps in most cases, you do not. And then it is a great advantage to be able to strip away all these distracting details.

Besides, and more important: Several of Wittgenstein's writings comprise different «layers» of text. The most obvious example is typescripts with hand-written corrections added by Wittgenstein, perhaps years later.

By using our filtering program, you may strip away all the handwritten corrections and have a print-out of the earliest version, or strip away deleted or replaced typewritten text and have a print-out of the later version. Or you may have both texts in one.

Another way of motivating the question «why do we make things so complicated?» is the following: By means of the modern laser printers, you may render all the peculiarities of the original by typographical means, using different character sets, fonts, boldface letters, italics, enhanced writing, differently sized letters, footnotes etc. etc., and still have an easy-to-read text.

This is a possibility which we are seriously considering. We hope to write a program for automatic production of laser prints directly from source transcriptions, though at present we do not have the necessary resources.

In any case, there will still be a need for filtered print-outs like the ones produced by our filtering program. There are, in turn, two reasons for this. First: You cannot display a laser print on an ordinary screen, so you will always need the filtering program anyway. Second: At present,

very few philosophers (in Scandinavia, at least) have easy access to laser printers.

Enough for now about print-outs. What else can we do with a coded transcription? We can produce word lists and indexes.

Indexes

Our indexing program will produce the following from a properly coded transcription:

- 1) a sequential list of all headlines in the text, each with a reference to section- and section line number
- 2) a sequential list of all datings in the text, each with a reference to section- and section line number
- 3) a sequential list of references to all occurrences of mathematical and logical expressions in the text
- 4) a sequential list of references to all occurrences of «Geheimschrift» in the text
- 5) an alphabetical list of all words in the text (some common words are excluded), each with a pointer to 7
- 6) an alphabetical list of all proper names in the text, each with a pointer to 7
- 7) a sequential list of references to the text for all the occurrences of the words listed in 5) and 6), together with information about the coding of each occurrence

1)-4) are text files, so they may be listed directly on the screen or on paper

5)-7) are so-called direct access files of records. The files have to be read or printed out by means of a specially designed program. This means that the user himself may produce his own index to any text, again given the options «simple» or «detailed» listings.

A «simple» index contains references only to the words that are found in a «simple» text listing made by the filtering program. Whether you choose a «simple» or a «detailed» index, you are given one index of all the words in the text and one of proper names only.

The indexes contain all words in the text listed in alphabetical order. The sum total (irrespective of filter chosen) of occurrences of each word is indicated, and for each word all occurrences corresponding to the filtering chosen are indicated by section number and line number.

When I say «all words», one reservation has to be made: We usually exclude some «uninteresting» words: articles, pronouns, prepositions, auxiliary verbs etc. You may include them if you wish. No machine-made index will ever be perfect. Misspelt words, for example, will be indexed as words of their own. But then again, the index is a convenient means of detecting such errors.

In German, the distinction between upper- and lower case first letters is significant and therefore a problem. «Wissen» and «wissen» are two different words; the first is a noun, the latter is a verb. But what about «Wissen» as the first word in a sentence? And what about «man kann WISSEN»? We handle such complications in most cases – by a combination of suitable coding and suitable programming.

Inflection is always a problem in its own right when you have to do with machine-made indexes. «Wissen», «weiß», «weißt», «wußte», «wüßte» etc. will all be indexed as different words. Programs are available on the market, as far as I know, for handling such matters, but it is not yet clear whether they may be of any use to our project. There are also difficulties concerning text interpretation connected with this.

Searching

Now, let us have a look at our search tools. Our search program has been developed very recently. In fact, what we have now is only a prototype.

The compact lists 5)-7) mentioned earlier provide the basis for the search program, too.

The program runs like this: The user is asked for the von Wright number of the text or texts on which he wants a search to be done. (With the prototype, you can only search in one text at a time). He is then prompted for the sort of filter he wants to use in his search (like in the filtering and indexing programs).

Finally, he is prompted for the word for which he is searching. He may enter a specific word, or indicate that he is searching for all words containing a certain sequence of characters. He may e.g. search for all occurrences of the word «Wissen», he may search for all words beginning with «Wissen» or ending with «wissen», or he may search for all words containing «wissen».

The program goes through the index 5) searching for words satisfying the conditions given by the user. For each of these, the program goes through the reference list 7) to see if the occurrences of the word fall within the current filter.

Each word satisfying the conditions is listed on the screen together with its total number of occurrences and the number of occurrences within the current filter of that word. Finally, the section numbers and sum total of all sections of the text satisfying the search conditions are listed on the screen.

This number may be so high that the user decides the result is useless, closes the session, and starts searching for a new word. Or he may decide that he wants to have a look at these sections.

The program then creates a temporary file from the source transcription containing all and only those sections containing the search word, and lists them, using the specified filter, on the screen.

Having had a look at all or some of these sections on the screen the user may decide that this is still uninteresting and start searching for a new word, or he may list the result on paper and/or store it on a separate disk file.

At present, you may only search for single words (or sequences of characters) in one text at a time. You may use wild-cards and search for truncated words. We plan to extend the program for searching in any number of texts, for logical combinations of words, specifying that the words are to be found within the same sections or the same page of the text.

Later we may extend the program to offer the possibility of searching for sections resembling either a specified section or a text which you have made yourself. What the program would then do, would be to count all occurrences of indexed words in the section specified, then look for sections in the text with a similar frequency of occurrences of these words, and return e.g. a ranked list of sections having a similar frequency relationship between indexed words.

This is as close as we can realistically come to searching for topics. Pick out a section treating a certain topic, look for «similar» sections, and you will probably find something treating the same topic.

However, I suspect that on top of the ranked list of similar sections, you will, if you search in all the texts, find sections from earlier or later versions of the same text.

What we would have here is, indeed, a very powerful philosophical research tool - a research tool for the study of Wittgenstein manuscripts, at least. This tool could be used not only for searching for «topics», but also for mapping and cataloging the whole Nachlaß, enabling you to confirm or weaken your hypothesis about which texts are rearrangements or revisions of other texts, and which are not.

Future plans

Our chief objective, as previously mentioned, is to make Wittgenstein's Nachlaß more easily accessible for philosophical research. What about a text-critical «Gesamtausgabe» in regular book form? The decision is not for us to make. We have not prepared the material for editing of this kind. But as you will have noticed already, our work may facilitate the preparation of an edition in book form.

On the other hand - I, for one, am not convinced that a book edition is the most suitable way of making Wittgenstein's Nachlaß available for philosophical research in the long run. A printed edition is not flexible,

no choice of filtering is possible (except that the editor will of course «filter» the text for you in his own way, but that is something which you cannot do very much about), no searches can be done. Besides, it would be very expensive.

Having the material in the computer is a lot more convenient for research purposes, and the personal computer is on its way into the Departments of Philosophy almost everywhere. It must be admitted that there is a snag here: The material is so huge that you would have to have a hard disk of at least 40 MB to make room for all the texts and indexes. You may do work on parts of the material with a floppy-disk machine, but you would probably very soon become some sort of a disk jockey. Mainframes may be a solution for some, hardly for all.

But then new technology opens up new possibilities: The optical disks may provide a solution to the problems just mentioned.

On an optical disk information is stored as a pattern of pits in a thin metal layer – at densities far higher than today's magnetic storage media (hard disks and floppy disks). The information is written and read by a laser beam. Therefore optical disks are also called laser disks. One kind of optical disk especially suited to the distribution of large amounts of text is the Compact Disc – the new digital grammophone record introduced a couple of years ago.

A Compact Disc (or CD-ROM, which is the data version of the Compact Disc) is 12 centimetres wide and stores 550 MB of data. This means we could easily store all the texts, all the indexes, and all the software of the NWP on one single disk.

In contrast to ordinary magnetic storage media, like floppy disks and hard disks, on the CD-ROM the information cannot be altered. On the other hand, mass production of copies is cheap.

Production of the so-called master disk costs about 75,000 NOK, i.e. \$10,000 today. Printing 4000 copies would mean that the costs of production would amount to about 150 NOK or \$20 for each copy.

The user would have to buy a CD-ROM player in order to be able to use the disk in connection with his personal computer. It would hardly cost him more than a printed edition of Wittgenstein's *Nachlaß*. (The price of such a player today is approximately 20,000 NOK or \$2,700, but is expected to drop to half of that very soon.) This player will enable his PC to read the information stored on the CD-ROM.⁷

CD-ROM production being so cheap, we could arrange for some sort of subscription, so that each subscriber would receive disks with improved versions of the *Nachlaß* and software as typing errors and other possible errors are detected and corrected.

What's more, there will be so much storage space left on the disk that you may even store pictures of the original on the same disk, rendering the possibility of always consulting the original if in doubt about the

transcription. Programs could provide for easy switching from a picture of the original page to a filtered transcript on the screen, you may have it both ways on a split screen etc., and you may have paper print-outs of all of this. It must be added that this last extension would require somewhat more expensive equipment on the part of the user.

Taking this last step will of course require a lot of work, and I have no idea of the total costs. Anyhow, issuing the material on a CD-ROM is possible, and it would be rather cheap. The Norwegian Computing Centre for the Humanities, at which the NWP is situated, and with which we have an agreement of cooperation, is in the forefront of optical storage technology applications in the humanities in Norway.

All this being said about future possibilities, CD-ROM and all that, I must rush to add that these are all rather airy plans. All of this does of course lie far beyond our present resources.

Today we do, in fact, have enough trouble keeping our heads above water financially, and the future of the project is obviously not as bright as this.

Before we can even think of possibilities like these, we must, of course, transcribe the rest of the Nachlaß and proof-read the 95% of it which has not yet been proof-read. That is in itself a huge task, which will require at least five more man-years of transcription and several man-years of proof-reading.

Notes

- 1 It should be mentioned, though, that professor Georg Henrik von Wright has prepared a catalogue to the material. We base much of our work on this catalogue, which seems to have become an international standard. Texts are identified and referred to by their «von Wright number».
- 2 I will not go into further detail about the history, the aims, the organization, funding etc. of the project here, as this has been treated in *Humanistiske Data* no. 1-86, pp. 64-66.
- 3 Texts have been transcribed on all sorts of computers, mainframes, PC's etc. At the moment, we use exclusively IBM-compatible PC's running under the operative system MS-DOS.
- 4 A program has been written, partly with help from the Tübingen Group, to re-code these texts. The actual work is not completed however. The rest is sheer routine, but we estimate another year for this work to be done, depending on our resources.
- 5 According to the Tübingen Group they had completed the transcription of roughly 50% of the Nachlaß. On the assumption that the Cornell photocopies are complete, this should mean that we have at our disposal transcriptions of varying qualities of about 65% of the Nachlaß, counted in pages. Unfortunately, these assumptions have recently turned out to seem less accurate than we formerly believed, so the estimate may have been somewhat misleading.
- 6 All programs referred to here are programs designed by the NWP, unless otherwise stated. All our software is currently designed for IBM-compatible PC's running under the operative system MS-DOS.
- 7 For more information about CD-ROM technology, see Øystein Reigem's article in *Humanistiske Data* no. 2-86.

RAPPORTER

nota bene

et sterkt tekstbehandlings- og «tekstbase» program

Torrey Seland

«**nota bene**» (TM) er ikke bare et sterkt tekstbehandlingsprogram, men i virkeligheten en hel integrert datapakke, spesielt konstruert for forskere innen humaniora. I denne pakken inngår et tekstbehandlingsprogram med alle de funksjoner vi er blitt vant til at et slikt program bør ha, med bl.a. funksjoner for fotnoter, indeks og bibliografi. Derneft finner vi her et tekstbase eller tekstgjenfinnings-program, som fungerer på liknende måte som en database. I pakken inngår også programfunksjoner for brevfletting (mailmerge) og for oppsetning av skjemaer (forms).

I denne anmeldelse skal vi først se litt på de forskjellige komponenter i «**nota bene**» og hvordan de fungerer, derneft kommer vi inn på en del erfaringer vi har hatt ved bruk av programmet. I en såpass kort anmeldelse som dette er ment å være, kan vi bare komme inn på noen av programmets mest sentrale og typiske funksjoner.

1) «nota bene» – et program konstruert spesielt for humanister

Programmet kommer på to disketter, men før det kan brukes, må de forskjellige komponenter installeres på i alt tre disketter. Dette er en svært nøyaktig prosess, men den er grundig beskrevet i manualen, og da denne ble fulgt nøye, oppsto der ikke noen problemer ved vår installering. Selve installasjonen går automatisk når den først er startet, en må bare følge med og skifte disketter når maskinen gir signal (bip + beskjed på skjerm) om dette. Underveis i installeringen blir man bedt om å skrive inn en del facts om ens maskinvare, slik at programmet da primært kan brukes på den maskin det er installert på. Denne prosess kan selvfølgelig gjøres om igjen dersom en forandrer f.eks. RAM kapasitet eller skrivertype.

Programmet krever to floppy-stasjoner eller en + hard-disk, og minimum 256 Kb internminne. Manualen er omfangsrik, men pedagogisk

velskreven og lettlest (bare på engelsk). Den tar også de nødvendige hensyn til både floppy-disk brukere og hard-disk brukere, der dette er nødvendig. I tillegg til manualen følger også en «Quick Reference Guide» hvor alle kommandoene er angitt, og etter ganske kort tids bruk klarer en seg med denne. Programmet har også en HELP funksjon som kan brukes underveis, men denne hører foreløpig med til de minst utbygde deler av programmet, idet den bare består av 8 skjermbilder med tekst som angir forskjellige kommandoer.

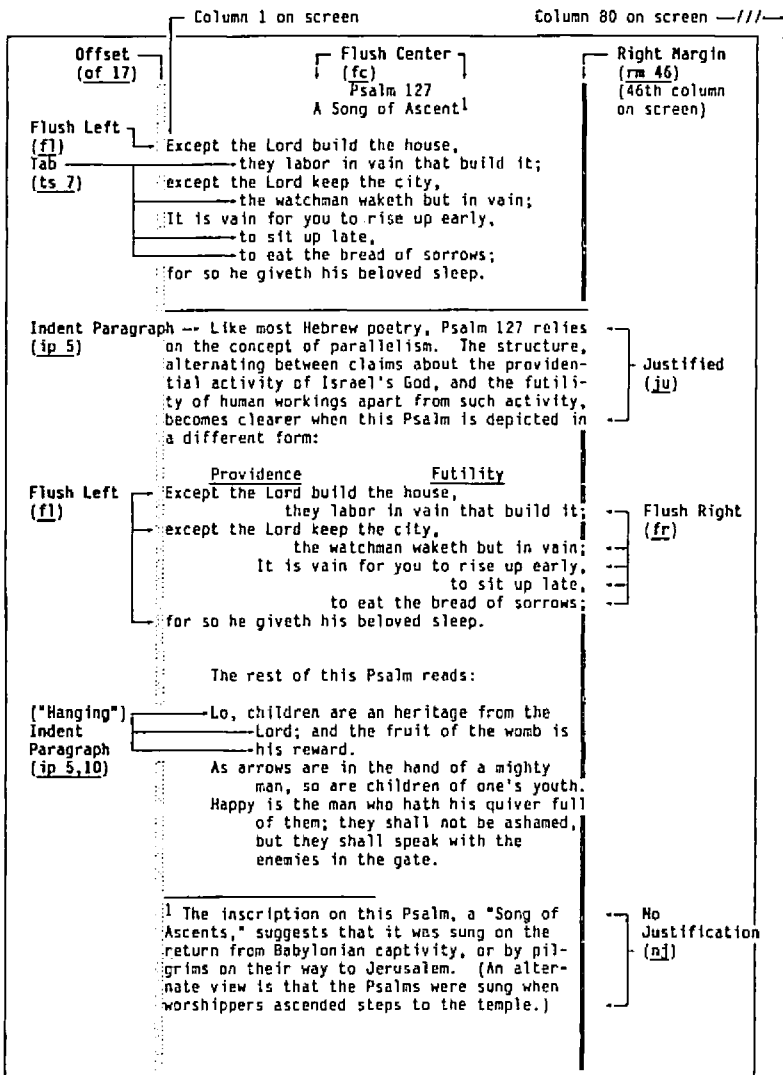
«nota bene» er såkalt «meny-styrt», dvs. at en får fram forskjellige oversikter hvor en kan velge i eller spesifisere forskjellige funksjoner f.eks. ved formattering eller utskrivning. Men i mange tilfeller kan en også gå utenom disse ved å gi direkte kommandoer. Det skal ikke så lang trening til før en er i stand til det. Det er i det hele tatt et særtrekk ved «nota bene» at så mange av dets funksjoner kan utføres både automatisk og «manuelt». F.eks. ved indeksering kan en velge om programmet skal indeksere samtlige forekomster av et ord, eller at den stopper ved hvert enkelt tilfelle og ber om bekreftelse.

I «nota bene» kommer der ingen koder i teksten på skjermen, men de enkelte koder som settes inn underveis, f.eks. ved understreking, forandring av linjeavstand osv. markeres ved en lysende trekant. Ved å plassere markøren på trekanten og «switche» over til «expanded input mode» (CTRL + F9) kan en avlese hva slags kode som er lagt inn. Det samme gjøres for å lese og redigere fotnotene etter at de er lagt inn. Første gang en fotnote skrives bruker en CTRL + F10 og åpner opp et eget fotnote-«vindu» og skriver fotnoteteksten på vanlig måte der. Etter at noten er «lukket» (ved CTRL + F10 igjen), er den markert i teksten ved et lite tall. Ved senere tillegging av nye fotnoter oppgraderes nummereringen av disse automatisk. Ved utskrift kan fotnotene enten plasseres nederst på den side de hører til, eller som slutt-noter på slutten av dokumentet.

Et annet trekk ved «nota bene» som er svært nyttig, er at en kan «feste» faste ord og uttrykk; «fraser», til i alt 35 av tastene på tastaturet. Disse frasene kan bestå av et enkelt ord eller navn som brukes mye, av en kommando eller lengre avsnitt (f.eks. kontrakt-formularer). Ifølge manualen er der ingen annen begrensning på lengden av slike fraser enn maskinens internminne, men de fleste av oss klarer seg vel med noen få ord. En slik «frase» skrives inn ved ALT + den tasten den er festet til. Enkelt og greit.

Ved bruk av «nota bene» kan en arbeide med to dokumenter samtidig ved at en splitter opp skjermen i to. Skjermen kan enten deles vertikalt eller horisontalt, eller en kan ha to hele skjermer. Dette er en svært nyttig funksjon dersom en f.eks. skal oversette et tidligere skrevet arbeide, eller en har skrevet inn notater, og så skriver disse ut til et ferdig dokument. Det er også svært enkelt å flytte tekst fra ett dokument til et

Example of Horizontal Formats (of Printed Document)



Example of Vertical Formats (of Printed Document)

	↑	The Hermeneutics of Suspicion	
Running Header	(rh)	↓	
		↓	One of the more prevalent ways of construing the relationship between the accounts provided by Marx, Freud, and Nietzsche and those offered in everyday life by mechanics, artists, executives, students, and farmers is in terms of concepts such as "depth" and "surface." To express the contrast in this way is simply to draw attention to patterns constitutive of the texts of suspicion themselves. Consider, to take one example, a famous passage from Marx:
Line Spacing	(ls 2)		
			The final pattern of economic relations as seen on the surface in their real existence and consequently in the conceptions by which the bearers and agents of these relations seek to understand them, is very different from, and indeed quite the reverse of their inner but concealed essential pattern and the conception corresponding to it (Marx, 1856, p35).
Line Spacing	(ls 1)		
			Just as pointedly, Maurice Godelier, working from within the same tradition, argues that "what is visible is a reality concealing another, deeper reality, which is hidden" (Godelier, 1972, p21). ¹
			It is claims such as these that underlie contemporary analyses of suspicion in terms of hidden, deep meanings that elucidate the "real" structure of surface accounts. This way of interpreting the relationship between conceptual frameworks has
Footnote	(fs)	↑	Separator
		↓	¹ Levi-Strauss, in <i>The Savage Mind</i> , makes much the same point when he speaks of the structural determinations underlying the colorful stories told by South American Indians.
		↓	
Running Footer	(rf)	↓	
			- 237 -
			Page Length (pl)

annet ved å bruke denne skjermfunksjonen. I versjon 2.0 av «nota bene» som kanskje allerede er kommet ut når dette leses, kan en arbeide med hele 9 skjermer samtidig. Dette gir store muligheter ved arbeide på større dokumenter.

Oppsetting av indeks-liste og bibliografi er et tidkrevende arbeid ved større manuskripter. Ved bruk av «nota bene» er denne prosess radikalt forenklet.

Ved oppsetting av indeks må alle ordene som ønskes med i indeksen markeres. Dette kan enten gjøres «manuelt» når en skriver inn teksten, eller en kan lage en liste over de ordene som en vil ha med og legge den inn på en egen tekstfil. I alt tre slike lister kan settes opp, og «nota bene» vil lage en indeks for hver. «nota bene» bruker slike lister og markerer automatisk i teksten alle de ord som listen inneholder. Deretter setter den opp en alfabetisk oversikt over indeksordene med sideangivelse. Bare én angivelse pr. side tas med i den endelige indeks. En liten svakhet ved denne funksjon er at dersom ordene forekommer i fotnoter, så anføres disse sist i indeksen i en egen liste merket med et lite tegn foran hvert ord. Men indekslisten lar seg enkelt redigere på skjermen slik at en kan inkludere fotnote-henvisningene sammen med de andre.

En liknende bibliografisk oversikt kan også settes opp med «nota bene». Men her utelates henvisninger anført i fotnoter, og for de av oss som primært anfører slike referanser i fotnoter, er dette et minus. Ulempen kan imidlertid kompenseres ved at henvisningene også markeres utenfor fotnoten, ved fotnote-angivelsen i teksten.

Det som gjorde meg personlig mest interessert i «nota bene», var at det har en slags database-funksjon, som kan brukes samtidig med tekstbehandlingsdelen. Denne delen fungerer ikke helt som en vanlig database, den kalles da også heller en «tekstbase» eller «tekstgjennfinnings-system». Kort fortalt er dens hovedfunksjoner slik: Selve basen legges inn som vanlig tekst i deler markert ved dobbelt linjeskift. Teksten kan deretter indekseres på flere måter; enten i såkalt «free form» ved at hvert enkelt ord brukes, eller at enkelte ord i avsnittene markeres. Den førstnevnte metode er fin ved f.eks. kildetekster, den andre mere aktuell ved sitater fra sekundærlitteratur. «nota bene» lager så en indeksliste i en egen fil over de ord som skal med. Selve tekstbasen kan være ganske stor og strekke seg over 4-5 disketter dersom «free form» brukes. Ved indeksering kun av enkelte termer kan flere disketter brukes. Ved bruk av harddisk er kapasiteten enda større.

Når en arbeider med et manus kan en gå direkte over i tekstbasen og søke etter materiale der. Søkingen kan skje ved angivelse av ett eller flere ord knyttet sammen med «and» eller «or». Dersom det angitte ord ikke finnes i basen, angir programmet en liste over ord som er svært like, og det kan søkes på nytt. Når ett eller flere avsnitt er funnet, kan det kalles fram på skjermen, og enten vrakes eller føres rett over i det manus en

arbeider med. D.v.s. det kan settes rett inn i teksten der markøren står, eller det kan settes inn i et eget vindu til senere bruk. Systemet krever nøyaktighet, derfor anbefales det også å ta «back-up» av de forskjellige filer, men det har fungert perfekt ved undertegnedes bruk av en mindre tekstbase basert på «free form» indeksring og søking.

Funksjoner som fletting (mailmerge) og «forms» er så velkjente at de neppe behøver noen videre kommentering her. Også disse har fungert bra.

2) Erfaringer og vurderinger av «nota bene»

Ved bruk av maskin med 256 kb internminne har en ca. 70 kb tilgjengelig til dokumenter. Dette er noe lite ved bruk av «windows»-funksjonen eller ved søking og innhenting av større mengder data fra tekstbasen. Jeg fant det derfor nødvendig å øke internminnet til 512 kb. Dersom en fremdeles skulle få «overflow», er den ene programdisketten spesielt beregnet på å ta inn slik «overflow».

«nota bene» gjør en utstrakt bruk av tastene CTRL, SHIFT og ALT i diverse kombinasjoner. Dette gir en stor mulighet for utnyttelse av forskjellige tegn og funksjoner.

Eks.: Ved sletting har en følgende alternativer: DEL sletter én bokstav, CTRL + DEL sletter et ord, CTRL + SHIFT + DEL sletter frase, SHIFT + DEL sletter setning, ALT + DEL sletter avsnitt, SHIFT + ALT + DEL sletter til linjeslutt. Liknende varierte funksjoner gis ved andre kombinasjoner. Det må innrømmes at det er ikke lett å huske alle disse kombinasjonsmulighetene, men det er heller ikke nødvendig, for der følger med et eget ark hvor de forskjellige muligheter er spesifisert.

Der er en del sider ved programmet som kunne vært bedre, og som er bedre i f.eks. Word Perfect, som det ligger nær å sammenligne med. «nota bene» har ikke automatisk side- og linjeteller. Denne kan imidlertid settes i funksjon, men da blir flere andre funksjoner, bl.a. flytting av markør og sletting, tregere. Programmet markerer dermed heller ikke sideskiftene automatisk på skjermen. «nota bene» viser også bare én linjeavstand på skjermen. På den annen side gjør dette at en får maksimal utnyttelse av skjermen til tekst. Litt verre er det at «nota bene» ikke har automatisk bindestrek-funksjon og dermed heller ikke automatisk rett høyre-marg. Rett høyre-marg kan imidlertid brukes ved utskrift (vises ikke på skjerm), men da oppførte ihvertfall min skriver (IBM 80 CPS Graphics) seg høyst unormalt, idet skriverhodet fór fram og tilbake for hvert ord det skrev.

Denne anmeldelse er skrevet på bakgrunn av erfaringer med Versjon 1. En ny forbedret versjon 2.0 skulle vært på markedet i mars 1986, men er blitt kraftig forsinket. Men når dette leses er den antagelig kommet. Ifølge den forhånds-reklame som er sendt ut angående denne versjon,

inneholder den bl.a. følgende forbedringer:

økning fra 2 «vinduer» til 9,
automatisk bindstreksfunksjon,
«avis»-kolonne system med mulighet for rett høyremarg,
forbedrede HELP funksjoner,
mulighet for 1/2 linjeavstand,
sortering og sammenligning av filer.

«nota bene» har foreløpig ikke forhandler i Norge, men firmaet har en skandinavisk forhandler i København. Prisen er ca. 6000 kr + moms. I denne ligger der visstnok også gratis oppdatering ved senere forbedrede utgaver. Alt i alt er «nota bene» et interessant program. Med den nye versjon har Word Perfect fått en sterk konkurrent, kanskje en verdig avtager til en større del av markedet enn WP kan komme til å sette pris på?

Fakta om «nota bene». Versjon 1..

Maskinvare: IBM PC, XT, AT og kompatible maskiner.

Operativsystem: PC-DOS 2.0 eller over.

Kapasitet: Minimum 256 Kb internminne, 2 DSDD diskettstasjoner eller 1 + harddisk.

Printere: «nota bene» kan brukes med over 100 forskjellige printere. Printer spesifiseres ved installering av programmet.

Kopiering: «nota bene» er kopieringsbeskyttet i den forstand at det er mulig å kopiere programmet, men ved førstegangs-installering må der skrives inn et navn som siden ikke kan slettes og som gjør at senere kopier kan oppspores.

Pris: Ca. 5800 DK + moms

Firma: Dragonfly Software
409 Fulton Street, Suite 202,
Brooklyn, N.Y. 11201
718/624-0127

Forhandler: Palle Hølge ApS Computer Services
Gl. Hovedgade 8
DK-2970 Hørsholm
Danmark

Torrey Seland er dr.art. stipendiat ved Religionsvitenskapelig institutt, Universitetet i Trondheim.

Red. ann.: NAVFs EDB-senter har avtale med forhandleren om rabattpris på «nota bene» i 1986 (Dkr 4200,-). Hvis ansatte ved universiteter og høyskoler er interessert i rabatt også i 1987, kan Senteret prøve å få i stand en ny avtale. Ta kontakt med førstekonsulent Per Vestbøstad, tlf. (05)212956.

«nota bene» er også anmeldt i *Computers and the Humanities*, Volume 20, No. 1, January-March 1986, s. 62-71.

Senterets videoplateprosjekt – en statusrapport

Kristin Natvig

Landets første videoplate med humanistiske primærkilder forelå i juni. Platen ble produsert i tilknytning til Senterets videoplateprosjekt – et prosjekt som ble startet høsten 1984, og nå nærmer seg sin avslutning.

Prosjektets hovedformål er å undersøke hvordan videoplateteknologi kan øke gjenfinningseffektivitet, bruk og formidling av store mengder humanistiske primærkilder. Videoplaten inneholder et representativt utvalg av slike kilder – fotografier, dias, film, video og manuskripter – fra 19 ulike humanistiske institusjoner, hovedsakelig museer, arkiv og universitetsinstitutter. Platen er produsert i samarbeid med disse institusjonene og Sekretariatet for fotoregistrering, Oslo. Tilknyttede bildedatabaser med tilhørende programvare er nå under utvikling.

Gjengivelseskvalitet

Materialet på platen er av høyst varierende kvalitet. Dette var ventet, ettersom mange kilder ble lagt inn på platen nettopp for å teste gjengivelseskvaliteten. Som antatt har det vist seg at manuskripter er blitt dårlig gjengitt. (Oppløsningen på video er begrenset, derfor egner *digital* lagring seg bedre for manuskripter). Det beste resultatet er oppnådd med avfotograferte malerier (tatt fra dias).

God gjengivelseskvalitet er imidlertid ikke nødvendigvis knyttet til kildetype – den er avhengig av faktorer som opptaksteknikk, original forfatning, og størrelse på det valgte utsnittet. Kvaliteten veksler også mye med den enkelte fotografers dyktighet. (Materialet ble avfotografert av ansatte ved deltakerinstitusjonene).

Datatilrettelegging

Registrering av referansedata til platens bilder – et svært tidkrevende arbeid – er nå i slutfasen. Det legges stort sett bare inn data om bilder av tilfredsstillende kvalitet. Noen av de deltakende institusjonene har foretatt dataregistreringen selv, mens Senteret har sørget for resten. Som et resultat av både dette forholdet og en viss eksperimentering med ulike dataformater, har utviklingen av databasene som er tilknyttet bilde-materialet medført et stort omformateringsarbeid.

Programvare

Senterets utstyr består av en videoplatespiller som styres fra en PC. Bilder og levende video vises på en TV som står koblet til spilleren, PC-en har sin egen skjerm. Utstyret har altså én dataskjerm og én bildeskjerm. All programvare er skrevet for utstyr med to skjermer. Det er teknisk mulig å legge tekst og grafikk fra PC over bilder fra videoplate, men Senteret har i dag ikke utstyr for en fullverdig løsning av dette.

I prosjektet har man dels tilpasset ferdige databaseprogrammer for videoplate, og dels skrevet programvare fra grunnen. Mye av programvaren er skreddersydd for de enkelte deltakende institusjoners særlige behov.

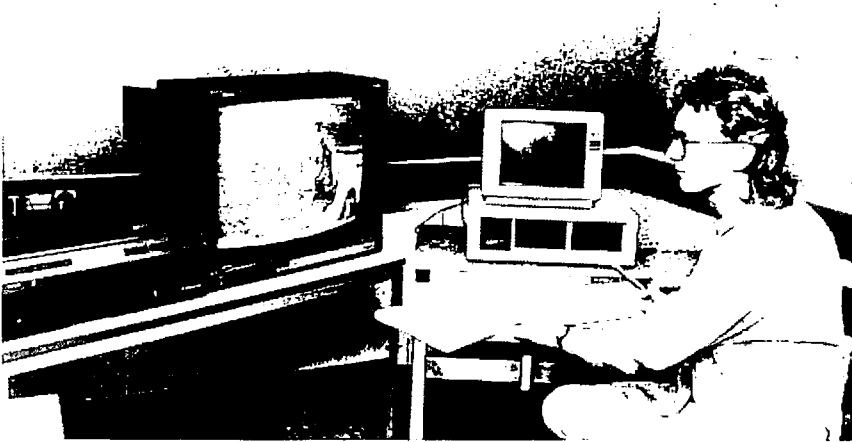
Databaseprogrammene er stort sett laget slik at brukerne skal kunne søke i referansedata på vanlig måte. Ved funn av interessante data, kan en få opp på videoplatespillerens skjerm det/de bildene som hører til det valgte datasettet. I flere av programmene har man muligheten for å «bla» i bildene direkte, uten å først måtte søke i referansedata.

Det er også utviklet programvare for å kunne ringe opp databaser på andre maskiner. PC-en blir da en terminal mot en annen maskin, samtidig som den kan styre videoplatespilleren. Med denne løsningen vil videoplatespilleren fortsatt stå lokalt, så det er ikke snakk om overføring av bilder pr. telelinje. Dette er en løsning bl.a. for de tilfeller hvor referansedataene blir så omfattende at lagringsplassen på en PC ikke blir tilstrekkelig. En annen fordel med en slik løsning er at databaseprogramvare på andre maskiner kan benyttes uten modifikasjoner.

I tillegg til databaseprogrammer er det skrevet et generelt program for «vandring» i bildemateriale, dvs. muligheten til å gå fra bilde til bilde langs forskjellige «ruter», f.eks. fra sted til sted i et landskap eller fra rom til rom i et hus. Valg av «rute» kan foretas ved hjelp av menyer.

Avslutning av prosjektet

Senterets videoplateprosjekt avsluttes til nyåret. I desember vil det bli holdt et seminar i Bergen for alle de deltakende institusjonene. På dette



seminaret vil platen med tilhørende programvare bli demonstrert, og deltakerne vil få anledning til å drøfte erfaringer med prosjektet. En rapport fra seminaret blir gitt i neste nr. av HD.

Videre arbeid med optisk lagringsteknologi

På videoplateprosjektets avsluttende seminar vil Senteret også orientere om andre optiske lagringsmedier. Demonstrasjoner vil bli arrangert i bruk av CD-ROM for lagring og søking i tekst, og innscanning av dokumentsider og bilder til WORM-plater. (Ad. CD-ROM og WORM – se HD 2-86, s. 52-53).

Senteret har gått til innkjøp av en CD-ROM-spiller og har fått låne en WORM platestasjon til utviklingsarbeid. Planer for utprøving av dette utstyret vil bli presentert i senere nr. av Humanistiske Data.

Datasekretariatet og humanistiske fag

Kristin Natvig

I oktober hadde NAVFs EDB-senter for humanistisk forskning besøk av *Jan Wibe* og *Jan Bredeveien* fra Datasekretariatet. Datasekretariatet ble opprettet i 1984 som et midlertidig organ i Kirke- og undervisningsdepartementet for å administrere forsøksvirksomhet med datateknologi i grunn- og videregående skoler. Denne forsøksvirksomheten er et resultat av stortingsmelding 39 (1983-84) om Datateknologi i skolen.

I mai i år utga Datasekretariatet en statusrapport om virksomheten. Rapporten inneholder bl.a. en oversikt over prosjektarbeid ved de 21 forsøksskolene som Datasekretariatet har valgt ut på bakgrunn av søknader. Ved de fleste av disse skolene pågår det utvikling av pedagogisk programvare i humanistiske fag som norsk, fremmedspråk, historie og forming.

I rapporten gis det også orienteringer om Datasekretariatets egne initiativ i form av programvareutvikling, kurs, veiledning m.m. Rapporten gir inntrykk av at få tiltak er satt i verk for å stimulere bruk av edb i humanistiske fag. I en samtale med Humanistiske Data fortalte Wibe og Bredeveien at grunnen til dette er at Datasekretariatets politikk utarbeides i henhold til ideer lærerne selv bidrar med. Hittil har Datasekretariatet konsentrert seg om å stimulere det som fins av initiativ fra lærernes side. Programvareutvikling f.eks. er blitt satt i gang etter forslag fra lærerne selv.

Datasekretariatet er imidlertid svært interessert i å trappe opp virksomheten rettet mot humanistiske skolefag. Det er allerede blitt arrangert kurs for lærere ved forsøksskolene hvor en har arbeidet bevisst med å trekke inn humanistisk-faglig kompetanse. Blant Datasekretariatets framtidige planer, er et kurs i utvikling av undervisningsprogrammer i tegning.

Datasekretariatet skal også prøve å få i stand en konferanse om edb-utstyr for samiske behov. Som et resultat av et samarbeidsprosjekt er det blitt utviklet programvare som kan generere de fleste samiske tegn på Tiki- og Scandis-maskiner.

Datasekretariatet har engasjert et veilederkorps på 12 personer som bl.a. assisterer lærere i design av undervisningsprogrammer. Foreløpig har bare én av disse veilederne humanistisk bakgrunn, men korpset skal utvides med flere humanister, kunne Wibe og Bredeveien fortelle.

Humanistiske Data håper i senere nummer å kunne presentere edb-prosjekter i humanistiske fag ved noen av Datasekretariatets forsøksskoler.

«Norsk stadnamnbase»

Botolv Helleland

Litt historikk

Alt i 1973 kom nordiske namnegranskarar saman til eit symposium i København for å drøfta edb-registrering av namn med tanke på å oppretta maskinleselege namnearkiv. Etter symposiet sette eit utval med representantar for dei nordiske landa seg i sving for m.a. å laga eit høveleg skjema for dataregistrering av namn. Skjemaet tok utgangspunkt i at kvart namn kunne handsamast som ein post med fleire eller færre felt som kunne fyllast ut. Dei viktigaste felta eller opplysningane som hadde plass på dette skjemaet, var oppslagsform, uttale, lokalitetstype, andre språklege og reale opplysningar, heimfesting (koordinatar), i alt 14 felt. Dette skjemaet har i prinsippet vorte lagd til grunn for seinare utgåver.

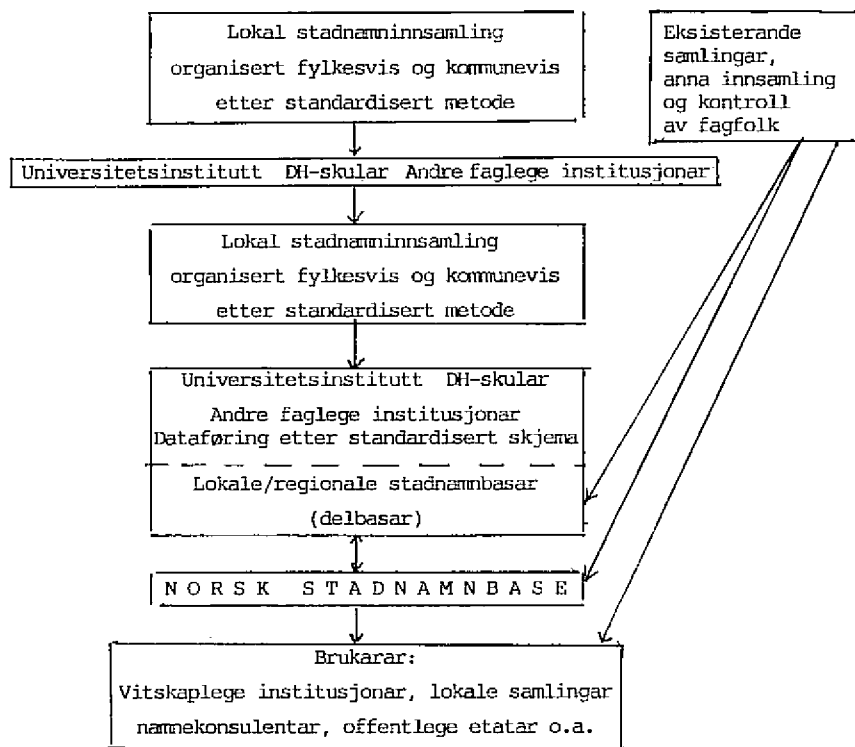
Hausten 1982 vart det halde ein nasjonal konferanse i Oslo med databehandling av namn som hovudtema. Her vart m.a. registreringsprogrammet FELTED, utarbeidd av *Håvard Hjulstad*, presentert. I kjølvatnet av denne konferansen sette Samarbeidsnemnda for namnegransking ned eit nasjonalt utval som fekk i oppdrag å greia ut om prinsipielle og praktiske sider ved dataregistrering av stadnamn. Nemnda (*Peter Hallaråker*, *Botolv Helleland*, *Oddvar Nes*, *Per-Bjørn Pedersen* – sistnemnde formann) leverte si innstilling 19.11.1983. Her har utvalet prøvd å samla dei ulike røynslene som har vore gjorde med dataregistrering av stadnamn i fagmiljøa og rår til eit registreringsskjema med til saman 17 felt samla i visse grupper, men slik at ein skil mellom ein dokumentardel som skal fyllast ut straks, og ein analysedel som kan fyllast ut seinare. Dokumentardelen skulle omfatta dei tradisjonelle opplysningane som normalt skal registrerast under innsamlingsarbeidet i marka (oppslagsform, uttale, preposisjon og eventuell kasusbruk, koordinatfesting, lokalitetstype og eventuelt andre kommentarar).

«Norsk stadnamnbase»

Fleire institusjonar har kome i gang med dataregistrering av stadnamn, men det har til no ikkje vore full semje om korleis eit framtidig data-basert stadnamnarkiv skulle byggjast opp eller fungera, sjølv om ein nok innan fagmiljøa lenge har innsett at ein gong i framtida ville ein slik teknologi pressa seg fram. Og målet er alle samde om – nemleg at alle stadnamn i landet skal samlast inn med dei nødvendige dokumentaropplysningane og gjerast tilgjengelege for alle aktuelle brukarar. Full

effekt vil ein fyrst kunna få av eit slikt arkiv når ein kan søkja i det maskinelt. Fram til i dag har stadnamnmaterialet vore sett opp i manuelle setelarkiv, ordna regionalt eller på landsbasis i alfabetiske eller etterleiddssorterte grupper. Ein god del av dette materialet har vorte kopiert og utveksla mellom institusjonane, men etter kvart som samlingane veks, tek det lenger og lenger tid å setja nytt tilfang inn i dei ulike registraturane.

Namnet «Norsk stadnamnbase» kom opp i samband med at innsamlingsprosjektet i Sogn og Fjordane førebudde ein konferanse om dataregistrering av stadnamn våren 1986. Konferansen vart halden på Fylkesarkivet i Kaupanger 28.-29. april, med Samarbeidsnemnda for namnegransking som formell arrangør.



Ein forenkla modell av korleis ein kan tenkja seg oppbygginga av ein «norsk stadnamnbase».

På konferansen vart ulike sider ved oppretting og drift av eit maskinelt stadnamnarkiv på nytt drøfte. Det vart særleg lagt vekt på dei faglege krava til registreringsskjemaet, og til kva einiskilddata (feltbaserte data) som bør leggjast inn i ein stadnamnbase. Med bakgrunn i røynslene frå Sogn og Fjordane-prosjektet vart det gjort framlegg om å innføra eit klassifikasjonssystem for lokalitetstype og kulturfunksjon ved hjelp av ein nærare definert nøkkel. Når det galdt lydskrifta, var det serje om å byggja på ein fonemisk lydskriftstandard som *Helge Sandøy* fyrste gongen la fram på den 4. nasjonale konferansen i namnegransking i Oslo i november 1985, og som har vorte vidare utbygd seinare.

Eit utval (*Gulbrand Alhaug*, *Peter Hallaråker* (formann), *Oddvar Nes*, *Tom Schmidt*) fekk i oppdrag å arbeida vidare med dei tilrådingane som kom fram på konferansen i Kaupanger med sikte på å leggja fram eit forslag til eit nasjonalt skjema for dataregistrering av stadnamn. Forslaget skulle leggjast fram for Samarbeidsnemnda for namnegransking hausten 1986.

Venteleg vil det nasjonale fagmiljøet kunna samla seg om eit registreringsskjema som er fleksibelt nok til at alle ynskjelege opplysningar kan få rom, samstundes som det er eit brukande grunnlag for å byggja opp ein nasjonal stadnamnbase. Dette siste spørsmålet skulle òg utvalet ta opp, men her vart det ikkje sett nokon tidsfrist. Sjølv om det sannsynlegvis ikkje er nokon datatekniske vanskar med å realisera ein slik namnebase, vil det vera ei rekkje spørsmål som må avklarast, både når det gjeld praktisk gjennomføring, økonomi, maskin- og programvare og ikkje minst lokalisering av «delbasar» og former for tilknytning. Med omsyn til maskinell koordinatfesting har ein alt hausta gode røynsler ved bruk av digitaliseringsborda til kartetatane.

Sidan ein namnebase med tida skal innehalda millionar av postar (namn) som det kan vera aktuelt å søkja gjennom for visse føremål, seier det seg sjølv at det må liggja ei nøye planlegging til grunn for oppbygginga av basen. Her vil det vera rimeleg å dra nytte av dei røynslene som er gjorde m.a. ved Prosjekt for datamaskinell språkbehandling i Bergen. I alle fall er det langt fram, og det er vel ingen i dag som kan seia når ein norsk stadnamnbase kan realiseras som eit ferdig produkt.

Sentralt offentleg stadnamnregister

Initiativet til ein norsk stadnamnbase har lege i fagmiljøa, og det er fyrst og fremst dei vitenskaplege behøva som skal dekkjast gjennom ein slik base. Men samstundes får ei slik materialsamling store praktiske verknader for namnebruken på kart og andre felt og elles som eit allmennkulturelt skattekammer.

Men også det offentlege har bruk for eit sentralt ført stadnamnregister som omfattar alle namn i offentleg bruk. Noko slikt register ligg enno

ikkje føre, bortsett frå ein 30 år gamal katalog over namn på den gamle hovudkartserien (M711) frå NGO. NGO er no i ferd med å fullføra nyutgjevinga av kartserien og har teke til å dataregistrera stadnamna på desse karta. Etter som det gjeld namneformer som er godkjende av statens namnekonsulentar, vil eit slikt register vera å sjå på som offisielt. Det treng berre endrast i den mon nye offisielle skrivemåtar vert innførte ved seinare kartutgåver.

I innstillinga frå Stadnamnutvalet om nye reglar for skrivemåten av stadnamn (NOU 1983:6) vert det gjort framlegg om eit slikt sentralt stadnamnregister der alle nye, offentleg godkjende, skrivemåtar skal registrerast. Når dette registeret vert kombinert med namneregisteret over offentlege kartnamn, vil samfunnet ha fått ein svært nyttig reiskap til å halda seg ajour med dei namna som til ei kvar tid kan reknast som offisielle.

Botolv Helleland er amanuensis ved Institutt for namnegransking, Universitetet i Oslo, og har vore statleg namnekonsulent sidan 1985.

IBM-seminar om taleprosessering

Kolbjørn Slethei

Hvert år arrangerer IBMs europeiske hovedkontor en seminarserie over en rekke sentrale emner innen elektronisk databehandling. I år var det i alt 8 slike seminarer, som alle fant sted i Oberlech i Østerrike. Blant seminaremnene var også taleprosessering, og undertegnede var den ene av to deltakere fra Norge på IBM-seminaret om taleprosessering i juli i år.

Det var få nordiske deltakere. Av i alt 68 deltakere var det to finner, en svenske, en danske foruten to nordmenn. Det var ingen foredragsholdere fra nordiske land. Det var store kontingenter fra Storbritannia, Tyskland og Frankrike, som også var de landene som hadde hovedtyngden av europeiske foredragsholdere. Dette gjenspeiler uten tvil det forhold at taleprosessering ikke er et felt hvor Norden (og spesielt ikke Norge) er stormakter. Svært mange foredragsholdere kom fra IBMs egne forskningssentra, både innen Europa og fra USA. Ellers kom foredragsholdere fra universiteter og forskningsinstitusjoner fra USA, Europa og Midt-Østen.

Det er trolig riktig å si at talere fra USA dominerte i form av faglig tyngde. Det var likevel interessant å kunne konstatere at europeisk forskning kunne hevde seg i dette selskapet.

De tre hovedfeltene innen taleprosessering, nemlig koding (inkl. data-reduksjon), gjenkjenning og syntetisering, ble behandlet fra mange forskjellige innfallsvinkler. Hovedvekten ble lagt på gjenkjenning.

Svært få av presentasjonene hadde bare teoretiske siktemål. De var imidlertid svært ulike med hensyn til hvor de befant seg på en tenkt akse mellom teoriutvikling og produktimplementering. Noen presenterte data-maskinelle modeller som simulerer menneskelig «taleprosessering», slik vi er i stand til å beskrive delkomponenter innen produksjon og gjenkjenning; andre hadde omtaler av prosesser og komponenter innen ren produktutvikling.

Jeg vil her trekke fram noen få av den store mengden av foredrag. Utvalget er nødvendigvis styrt av mine personlige interesser – hvis noen av leserne av *Humanistiske Data* skulle ønske tilleggsopplysninger eller vil vite noe om hva mer som ble behandlet, kan de få låne en del materiell som deltakerne fikk.

J.B. Allen fra AT & T Bell Laboratories presenterte en modell for såkalt perifer auditiv persepsjon. Modellen søker å framstille det som er kjent om ørets signalbehandling fram til første ganglion etter basilar-membranen. Flere av de faktorene som vi vet har betydning for forståelse av tale, som f.eks. simultan spektralmaskering og temporal maskering, var bygd inn i modellen. Modellen inkorporerte også nyere viten om den skarpe filtrering som foregår i de individuelle hårcellene i det cortiske organ. Resultatet fikk Allen fram som et «neurogram», dvs. en grafisk framstilling av neural signalstyrke i forhold til frekvens og tid. En ulempe med modellen var at den bygde på egenskaper hos hårcellene til *katter*. Det er en konsekvens av at man ikke kan utføre visse eksperimenter *in vivo* med mennesker uten å komme i konflikt med forsknings-etiske standarder.

J.S. Lim fra Massachusetts Institute of Technology hadde et foredrag om korttids Fourier-transformasjoner av tale. Teknikken gir store muligheter til manipulasjoner med et avgrenset talesignal i spektrum- og tidsplan. Dette er egenskaper som kan være svært viktige ved en eventuell videreutvikling i språkpedagogiske applikasjoner.

Det har lenge vært et ønske å redusere bitraten i et transmisjonssystem uten å svekke kvaliteten under et akseptabelt nivå. *S. Roucos* fra firmaet Bolt, Beranek and Newman presenterte hovedtrekkene i en segment-vokoder som arbeider i området omkring 300 bits/s. Hvor stor data-reduksjon dette egentlig innebærer, kommer kanskje ikke fram uten at en sammenlikner med den nå vanligste standarden for pulskodemodulasjon, som gir en transmisjon på 64 *kbits/s!* Kvaliteten var ikke særlig

god, men applikasjonsområdet var da også tjenestekommunikasjon parallelt til bredbåndsoverføring. Det er interessant å se at det i det hele er mulig å komme så langt ned i bitrate.

Hvor stor informasjonsandel i tale som er kodet i energivariasjon i *frekvens* og hvor stor del som er kodet i energivariasjon i *tidsforløp*, har vært gjenstand for mange diskusjoner opp gjennom årene. *M. Russel* fra Royal Signals and Radar Establishments i Storbritannia gjorde rede for en analyseteknikk som søkte å beregne den relative andel av spektral informasjon i forhold til temporal informasjon. Det var en ren data-maskinell, automatisert teknikk som bygde på analyse av delinformasjon, uten å gjøre bruk av psykoakustiske metoder.

Før en deltaker fra et lite språksamfunn var det av særlig interesse å se at ingen av applikasjonene var rettet mot språkundervisning. Dette gjenspeiler uten tvil egenskaper ved et internasjonalt forskningsmarked. Det er store og ressurssterke nasjoner og språkmiljøer som leverer premisser for forskningsfelt. Og når man har et verdensspråk som utgangspunkt, da er jo ikke språkpedagogiske applikasjoner noe som utgjør et synlig marked. IBM i Europa er heller ikke noe unntak fra denne regelen.

I tiden mellom seminaret og denne omtalen har imidlertid IBM åpnet sitt sjuende forskningssenter i Europa – i Bergen. Det er det første europeiske forskningssenter som IBM legger til et lite språksamfunn. Så gjenstår det å se om dataindustriens storebror vil oppdage at europeisk ikke bare er engelsk, tysk og fransk – og at Norge ikke bare er offshoret teknologi og havbruk.

Kolbjørn Slethei er amanuensis ved Institutt for fonetikk og lingvistik, Universitetet i Bergen.

Kurs i design av pedagogisk programvare

Uddevalla, 29.7.-8.8.1986

Kristin Natvig

I regi av Datasekretariatet er det i 1984 og 1985 blitt holdt kurs i programvareutvikling for lærere, de såkalte «Grimstad-kursene». Årets kurs ble arrangert i Uddevalla i samarbeid med Nordisk ministerråd

programgruppe, Dataprogramgruppen under det svenske Utbildningsdepartementet, og stiftelsen IMTEC (International Movements Towards Educational Change).

Kurset samlet 90 deltakere, de fleste lærere. Som i fjor var hovedforeleseren kanadieren *Les Green* (se Gunnar Thorvaldsens rapport fra fjorårets kurs i HD 3-85, s. 67-69). Greens hovedprinsipp for programdesign er den såkalte «torgmodellen». Kursdeltakerne arbeidet i grupper på 2-4 personer med å bruke denne modellen i utviklingen av læreprogrammer.

Humanistiske Data har spurt førstekonsulent *Per Vestbøstad* ved NAVFs EDB-senter, som deltok i kurset, om de pedagogiske prinsippene som ligger til grunn for torgmodellen.

- Hovedpoenget med torgmodellen er å skape et alternativ til de sekvensielle undervisningsprogrammene. Ved bruk av disse skal elevene tilegne seg lærestoffet trinn for trinn, men dette blir for kjedelig. Et program skrevet i henhold til torgmodellen innebærer at elevene selv kan bestemme hvilke deler av stoffet de vil øve seg i og hvor lang tid de vil bruke på øvelsene. Muligheten for lærerstyring ved bruk av programmet blir dermed begrenset – lærerens oppgave blir mer å legge til rette emneområder.

Hvordan arbeidet du med torgmodellen på kurset?

- Sammen med rektor *Ingve Markussen* ved Sørvågen ungdomsskole i Lofoten, laget jeg en modell for glosetrening i tysk. Vi kom så langt som til å utarbeide skjermbilder for et program.

Egner torgmodellen seg til design av et program av denne typen?

- Ja, det synes vi. Elever har jo ulike behov i fremmedspråkopplæringen, og torgmodellen gir mulighet for individuell undervisning. Det er en stor fordel at eleven har frihet til å velge nivå og kan repetere så ofte hun/han vil. Programmet vi arbeidet med, kan være anvendelig i alle språkfagene.

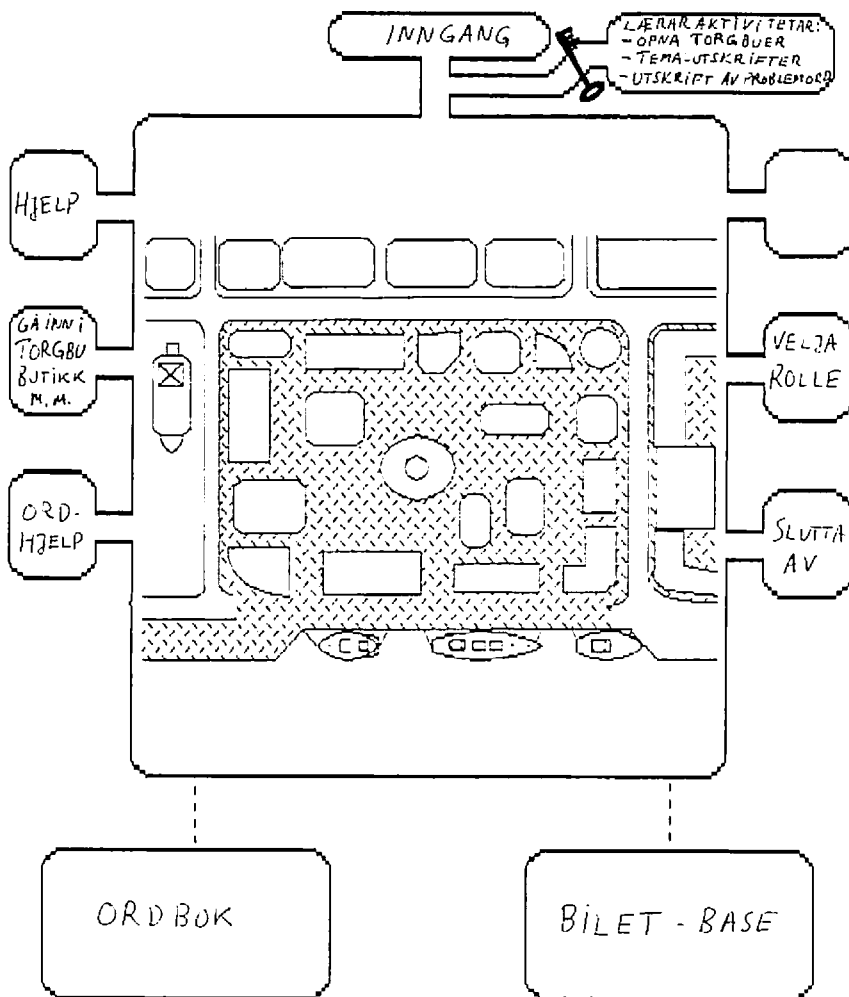
Var det noen andre av kursdeltakerne som arbeidet med programmer for humanistiske skolefag?

- Som ved de tidligere Grimstad-kursene var et fåtall av deltakerne lærere i humanistiske fag. Men det var et par grupper som arbeidet med programmer for både morsmålstrening og innøving av grammatikk i fremmedspråk.

Kan du fortelle litt om programvare-demonstrasjonene som fant sted i løpet av kurset?

- Programmer som deltakerne selv hadde laget ble demonstrert, i tillegg til hjelpeprogrammer for utvikling av undervisningsprogrammer. Blant de kommersielle programleverandørene som viste sine produkter, var Tiki Data, og Esselte og Liber fra Sverige.

Et hjelpeprogram som fikk positiv respons hos deltakerne, var «Mosaikk». Programmet tilbyr hjelp i å sette opp menyer, skjermbilder og grafikk. Det fungerer altså som en programkode-generator, og kan



Torgdiagram for glosetreneringsprogrammet *ORDTRIM*. Sideboksene viser hvilke valgmuligheter eleven har.

Eleven lærer gloser ved å vandre rundt på åpnings skjerm bildet og besøke frukthandler, avis kiosk, baker, motebutikk, skole, kirke, legekontor osv.

Interior- og situasjonsbilder fra slike besøk er tenkt hentet fra en bildebase på videoplate.

spare lærere mye tid og arbeid i utviklingen av pedagogisk programvare.

Ble det demonstrert noen programmer spesielt rettet mot humanistiske skolefag?

- «Vevplan» ble vist – programmet for design og analyse av vevede mønstre som ble presentert i HD 1-86. På meg virker «Vevplan» svært brukervennlig og effektivt, og kursdeltakere som har prøvd det selv, mener at det sparer mange timers arbeid.

Et siste spørsmål – hva er nytten ved å arrangere et kurs av denne typen på nordisk basis?

- Kurset viste at lærere og andre som arbeider med pedagogisk programvare i de nordiske land, har parallelle problemer. Derfor er det viktig å få mulighet til kontakt med hverandre for å utveksle erfaringer og diskutere løsninger på disse problemene. Erfaringen viser at det er kolossalt tidkrevende å utvikle gode læreprogrammer, så det må være fornuftig å samarbeide over landegrensene der stoff og metoder er felles.

*

Konferanserapporten kan fås ved henvendelse til: *Nordisk Ministerråd, Dataprogramgruppen, Vesterbrogade 72, DK-1620 København V, Danmark.*

Datasekretariatets seminar i Halden

Espen S. Ore

Som en oppfølging av kurset i programvaredesign i Uddevalla i sommer (se intervju med Per Vestbøstad annetsteds i dette nummer av HD), arrangerte Datasekretariatet et seminar i Halden 15.9.-18.9.86.

Under seminaret ble det fastslått hvilke målmaskiner som gjelder for prosjekter med støtte fra Datasekretariatet, og det ble også drøftet noe hvilke maskiner det var realistisk å finne på skolene. Målmaskinen ble definert som Scandis II med MS-DOS, farger og 256 Kb RAM (evt. Tiki 8-bits eller Tiki 8/16, versjon C). Man tok også opp at det kunne være naturlig å bruke større maskiner i programutviklingen selv om de endelige programmene skulle brukes på mindre maskiner. I denne forbindelse kom man også inn på problemer som kan oppstå når programmer skal flyttes fra f.eks. en Scandis til en Tiki.

Det ble også vist to prosjekter der målet var å utvikle programmeringsverktøy. Først orienterte *Lars Vavik* fra Stord Lærerhøgskole om pakken SimSim som kan brukes til å utvikle simuleringsmodeller. *Dag Svanæs* fra NTH (tidl. Datasekretariatet) introduserte pakken Mosaikk.



Med denne pakken skal det være mulig å generere f.eks. skjermbilde-sekvenser interaktivt *før* man begynner å utvikle et program for alvor.

For en som ikke har vært med på de tidligere seminarene, var det svært interessant og lærerikt å se hva som er status for Datasekretariatets arbeid. Televerket driver forsøk med et elektronisk konferansesystem (Teleboks), og deltakerne på seminaret fikk oppgitt et brukernavn og et passord slik at man senere skal ha anledning til å være i elektronisk kontakt.

Humanistiske fag og kunnskapsbaserte systemer

Erik Fjornes

22. og 23. september arrangerte NAVFs EDB-senter for humanistisk forskning et nordisk forskersymposium om kunnskapsrepresentasjon. Målet var å gi et bredt bilde av emnet for å stimulere til debatt. Symposiet ble holdt på Tyrifjorden turisthotell og samlet 28 deltakere fra Sverige, Danmark og Norge. Blant deltakerne var bl.a. fagene filosofi, logikk, språkvitenskap, sosialantropologi og informatikk representert.

Første dag var viet kunnskapsrepresentasjon generelt samt ekspertssystemkonseptet spesielt. Et hovedproblem i dagens fag ble omtalt som et konseptualiseringsproblem hvor det er behov for bedre teorier og redskaper for kunnskapsrepresentasjon. Det var enighet om at utviklingen av disse bl.a. går gjennom en konstruktiv kritikk av forutsetningene for dagens fag. Her vil utnyttelsen av kunnskapsteoretiske innsikter, formallogisk kompetanse samt kommunikasjonsteoretiske analyser av systemer i praktisk bruk, kunne føre til en avklaring av muligheter og begrensninger for representasjon av kunnskap i en datamaskin. I motsetning til en ureflektert utviklingsoptimisme vil slike arbeider kunne fungere som retningsgivere og fundament i arbeidet med å utvikle bedre teorier og redskaper.

Videre ble en implementerbar kunnskapsmodell ansett som en *fortolkning* av den aktuelle kunnskapen. Dermed åpnes mulighetene for å føre inn i faget teoretiske begreper som humanistiske forskere har utviklet for å hanskese med tradisjonelle humanistiske fortolkningsproblemer.

Andre dag var i all hovedsak viet systemer for behandling av naturlig språk. Lingvistiske og fonetiske arbeider er allerede nødvendige for utviklingen av slike systemer. Men det ble påpekt at problemer med å hanskese behandlingen av fenomener som partiell informasjon, kontekst-avhengig informasjon og ekspertinformasjon, fører til et behov for å legge større vekt på analyse og beskrivelse av kunnskap. Dette motsvarer et behov for ekspertsystemfaget til å legge større vekt på språkkommunikasjonen mellom system og bruker hvor redskapene for utvikling av naturlige språkssystemer er nødvendige.

Det var bred enighet om at mange av problemene i faget kunnskapsbaserte systemer har en tradisjonell humanistisk karakter, slik at språkforskere, filosofer, logikere og kunnskapsfortolkere har mange oppgaver å ta fatt på. Videre ble det uttrykt et generelt ønske om en oppfølging av symposiet gjennom nye symposier, formidling av emne definisjonen til internasjonale fora o.l. Som en oppsummering kan en si at emnet for symposiet ble mottatt som en ny og fruktbar innfallsvinkel i arbeidet med utviklingen av kunnskapsbaserte systemer og da som et supplement til rene informatiske og psykologiske fagmiljøer og fagmiljøer som arbeider med kunstig intelligens.

Det vil bli utgitt en konferanserapport i løpet av vinteren -87.

«Kommunikasjons- og datateknologi som infrastruktur i nord-norsk næringsliv og forvaltning»

Rapport fra en konferanse

Fredrik Fagertun

P.g.a. problemer knyttet til lockouten i vår, ble Universitetet i Tromsø tradisjonelle vårkonferanse i år til høstkonferanse. Det la ingen demper på interessen for konferansen som fant sted 16.-17. september, i år med «informasjons- og kommunikasjonsteknologi» som tema.

Vårkonferansen i Tromsø, som altså i år ble høstkonferanse, er et

årlig samarbeidsarrangement mellom Universitetet i Tromsø og FORUT (Forskningssiftelsen ved Universitetet i Tromsø). Av konferanseprogrammet kunne man lese at: «hovedmålet med konferansen er å skape et forum der forskningsmiljøene og nord-norsk næringsliv og forvaltning kan møtes og utveksle kunnskap om hverandre».

Deltakerlista, som talte bortimot 100 personer, bar da også preg av et bredt oppmøte fra nord-norske forvaltningsorganer på flere nivåer og fra særlig den del av næringslivet som allerede utnytter informasjonsteknologien eller som lever av den. Kanskje var oppmøtet fra det mer tradisjonelle nord-norske næringsliv noe skuffende.

IT-samfunnet i politisk perspektiv

Konferansens åpningsforedrag ble holdt av statssekretær *Tove Veierød* i Kultur- og vitenskapsdepartementet. Veierød konstaterte at også Nord-Norge nå står på terskelen inn i «kunnskapssamfunnet», og la vekt på at begrepet «utkant» i kunnskapsmessig sammenheng ikke måtte bli synonymt med «geografisk utkant». Hun berørte her et forhold som senere i flere sammenhenger ble tatt opp på konferansen, nemlig distriktenes behov for å bygge opp egen kompetanse innen informasjonsteknologi dersom «utkantene» skal kunne utnytte det potensiale for næringsutvikling som ligger i denne nye teknologien.

Veierød var videre opptatt av at vi ikke måtte la informasjonsteknologien skape nye klasseskiller i samfunnet. Vi må unngå klasseskiller mellom brukere og eksperter, og klasseskiller som går på kunnskap, kjønn og distrikter, mente hun.

Mens statssekretær Veierøds innlegg til dels bar i seg politikerens noe idealistiske håp og ønsker om hvordan vår IT-fremtid skal se ut, var innlegget til forsker *Kirsten Bjørnå* (FORUT) en mer konkret gjennomgang av hvilke typer planlegging og virkemidler som må til for at vi i Norge (og Nord-Norge) skal få en helhetlig datapolitikk. Bjørnå er også medlem av regjeringens datapolitiske råd.

Til tross for at informasjonsteknologien i regjeringens langtidsprogram for 1982-85 ble utpekt som et viktig satsningsområde, kunne Bjørnå slå fast at vi ennå mangler en *nasjonal datapolitikk*. Vi mangler fortsatt retningslinjer for hvordan vi skal takle den nye teknologien.

Bjørnå mente det måtte lages en datapolitisk handlingsplan for hele landet, der norsk IT-utvikling blir satt inn i et internasjonalt perspektiv. Med utgangspunkt i en slik nasjonal plan, må det lages delplaner for landsdelene. Hun var opptatt av at man i Nord-Norge nå måtte gå i startgropa, klar til å delta i dette arbeidet. Det er viktig at nord og sør ikke følger hver sin datapolitikk, men at det handles koordinert. Dette betinger imidlertid at landsdelens interesser og spesielle særtrekk og fortrinn, blir ivaretatt i den nasjonale datapolitikk og handlingsplan.

Bruk av informasjonsteknologi

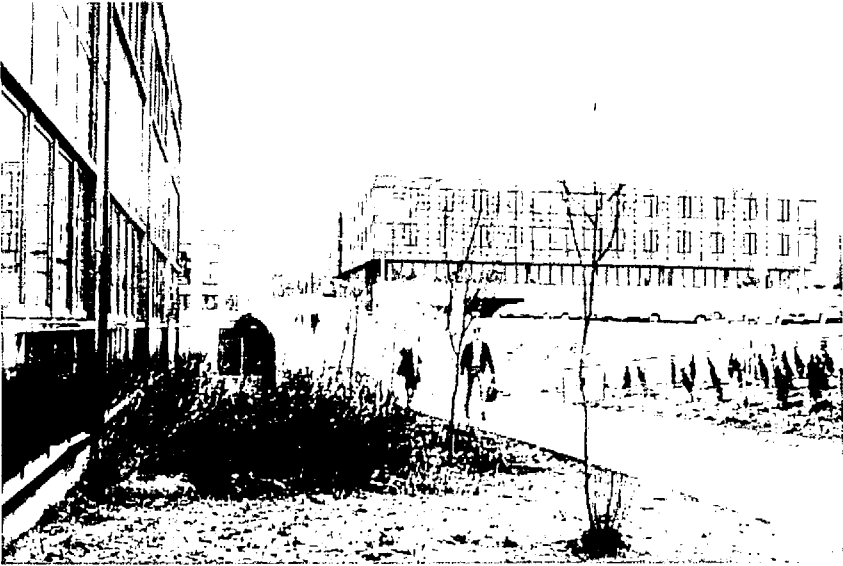
Konferansens første deltema tok for seg konkrete eksempler på anvendelse av informasjonsteknologi på forskjellige sektorer – ved universitetene, i det private næringsliv og i offentlig forvaltning. Representanter fra Oljedatabasen – FORUT, Registreringsentralen for historiske data, Fiskeriteknologisk Forskningsinstitutt, helsetaten, massemedia, bank og offentlig forvaltning, la fram sine erfaringer med bruk av teknologien. Debatten etterpå bar i noen grad preg av distriktenes håp om at den nye teknologien skal åpne for nye arbeidsplasser og en ny giv også for utkant-Norge. Samtidig var frustrasjonen stor blant de representanter for kommunal- og annen forvaltning som hadde prøvd å få til virksomheter basert på datateknologien. Bekymringen gikk først og fremst på at distriktene mangler kompetanse på feltet. I kampen om datakompetansen, er distriktene dømt til å tape, så lenge slik kompetanse er mangelvare og lønnsnivået på sektoren er kunstig høyt. Fortsetter det på denne måten, vil distrikts-Norge forbli ikke bare geografisk utkant, men også kunnskapsmessig utkant, ble det sagt med adresse til statssekretær Veierøds foredrag tidligere på dagen.

Det ble også stilt spørsmål ved bruken av offentlige ressurser på databasesektoren. Det ble beklaget at offentlige institusjoner som har lagret store mengder offentlige data, ikke har tilgjengeliggjort disse i online, søkbar form. «Telematikk til fots», ble dette kalt.

Televerket

Televerket har innledet et samarbeid med FORUT i Tromsø, og hadde samarbeidet med arrangørene om konferanseopplegget. I en bolk fikk Televerket anledning til å presentere de tjenester på telekommunikasjonssektoren som i dag finnes, og som vi kan forvente kommer fram til århundreskiftet. Jeg vil her ikke gå konkret inn på Televerkets nåværende og fremtidige tjenester – det er Televerket selv bedre i stand til gjennom sine brosjyrer.

Jeg konstaterer likevel at Televerket nå satser sterkt for å komme «inn på banen». Det satses på å gjøre næringsliv, forvaltning og undervisningssektoren kjent med de muligheter som foreligger og vil foreligge. Digitalt telenett, fiberoptikk, satelittkommunikasjon m.m., vil åpenbart gi fordeler både kapasitets- og kvalitetsmessig når stadig større informasjonsmengder skal overføres. Men det er i stor grad markedsføring med sikte på den ikke altfor nære fremtid, Televerket nå driver. En ting er at Televerket legger datakommunikasjonskanalene til rette, men det er ikke Televerkets informasjon som i hovedsak skal formidles på denne måten. Den fulle nytte av et moderne telekommunikasjonsnett får vi bare i den grad andre enn Televerket lagrer og ordner informasjon, samtidig som det finnes vilje til å gjøre den samme informasjon tilgjengelig.



Fra universitetsområdet i Breivika. (Foto: Knut Sollid).

Utdanning

Et sentralt problem som knytter seg til den videre innføring og utvikling av informasjonsteknologi på alle sektorer i samfunnet, er *utdanning*. Hvordan skaffer vi fram, helst hurtig, den nødvendige kompetanse i form av utdannet personell og hvordan bør denne utdanningen utformes? Dette spørsmålet er selvsagt viktig for utdanningsinstitusjonene selv, her konkret datafagutdanninga ved Universitetet i Tromsø, men også viktig for næringsliv og forvaltning som skal gjøre seg nytte av denne kompetansen.

Førstekonsulent ved Institutt for matematiske realfag ved UiTø, *Idun Reutz*, tok her for seg utdanningstilbudet innen informasjonsteknologi ved universitetet. På datafag vektlegges nå utdanning innen *distribuerte systemer, rastegrafikk og kunnskapsteknologi*. På faget fysikk prioriteres *signalanalyse/billedbehandling* og *elektronikk/måleteknikk*. I tillegg arbeides det med å få i gang sivilingeniørutdanning.

Ikke uventet trakk hun også fram mangelen på ressurser. Manglende utstyr og foreldt utstyr er et problem når man skal utdanne folk til arbeid i et næringsliv som har vesentlig mer avansert utstyr. På samme måte er ressursmangelen et problem for forskningen, som jo egentlig ressursmessig burde ligge i forkant av den tekniske utviklingen.

Videre innledet hovedfagsstudent *Hilde Hammes* om jenter og utdanning innen teknologiske fag. Hun stilte spørsmålet hvorfor kvinner gjerne velger bort de beste og mest utfordrende utdanningsalternativene? Selv skisserte hun tre årsaker til dette:

For det første mente Hammes at dagens familiepolitikk er den største hindringen for en bedret utvikling på dette området. For det andre pekte hun på at undervisningens form og innhold er for lite samfunnsrettet. Mens menn gladelig fortaper seg i tekniske detaljer, retter kvinner seg mer i retning anvendelse og bruk av teknologien. For det tredje mente hun det er nødvendig at kvinner skaffer seg forbilder – at de kvinner som allerede er «innenfor» formidler sine erfaringer ut i de miljøene der jenter kan rekrutteres fra.

Sekvensen ble avsluttet med en paneldebatt med deltakere fra FORUT, Statens Personaldirektorat, universitetet og næringslivet.

I denne debatten konstaterte man at det utdannes for lite folk, og at det er en geografisk skjev fordeling på de som utdannes. Det kunne også konstateres en viss uenighet mellom på den ene siden brukerne av utdannede kandidater og på den andre siden de som utdanner. Mens næringslivet ønsket seg en noe høyere grad av spesialisering inn i utdannelsen, mente representanter for lærerkrestene at datautdanningen måtte være noe generell p.g.a. bredden av behov i næringslivet. Næringslivet savnet i noen grad også evnen til analyse, prosjektledelse, gruppearbeid og kommunikasjon hos dagens kandidater. «Datafolk» evner ikke alltid like godt å se til side, ble det hevdet. Almen enighet var det imidlertid om et siste viktig punkt – med den hurtige utvikling som finner sted innen informasjonsteknologien, er etterutdanning et stort, men nesten ikke dekket, behov.

Forskning

Universitetet presenterte også et utvalg av de forskningsaktiviteter som drives. Amanuensis *Otto J. Anshus* ved datafagseksjonen, presenterte FoU-virksomheten ved seksjonen. Videre beskrev forsker *Jan P. Pedersen* arbeidet med satelittfjernmåling og billedbehandling som drives ved universitetet. Professor *Egil Arnesen* ved Institutt for samfunnsmedisin tok for seg den store hjerte/kar-undersøkelsen som har vært gjort i regi av instituttet, og fremla noen tidlige forskningsresultater herfra. Til sist tok professor *Ove Lorentz* opp nytten av datalingvistisk forskning som man håper å komme i gang med på Institutt for språk og litteratur.

Bransjeutvikling

Det siste tema på konferansen knyttet seg til bedriftsetableringer på informasjons- og kommunikasjonsektoren. Kanskje særlig interessant

for oss universitetsfolk er «programvarehuset» som er dannet av tidligere forskere ved universitetet og med utgangspunkt i forskning som har vært drevet her. IT Igloo Data A/S, med professor *Knut Skog* som leder, ble opprinnelig startet i 1984 som et aksjeselskap, der andelene delvis var bestemt ut fra den forskning de enkelte medeiere hadde bidratt med. Senere er Kloster-rederiet gått inn med kapital i foretaket.

Bedriften skal utvikle programvare for å forme dokumenter, bøker, rapporter i typografisk forstand – eller sagt på en annen måte: «legge typografens kunnskaper inn i datamaskinen». Her satses det på internasjonal markedsføring. Det norske markedet er for lite. Om denne satsingen vil lykkes, er ennå for tidlig å si. Vi ønsker professor Skog og hans kolleger lykke til, men konstaterer dessverre at jo bedre den slags satsing lykkes, jo større er faren for at universitetsmiljøene i enda større grad tappes for kompetanse.

Oppsummering

Jeg skal ikke påta meg å oppsummere nytten av denne konferansen på vegne av de andre deltakerne. Likevel tror jeg arrangørene skal være fornøyd med konferansens forløp når det gjelder informasjon omkring universitetets virksomhet på sektoren informasjonsteknologi. Det var en bred informasjon som også jeg, med mange års knytning til universitetet, hadde nytte av.

Tar vi for oss konferansens tema og spør: Hva viste konferansen oss m.h.t. informasjons- og kommunikasjonsteknologiens status og fremtid i Nord-Norge? Da blir svaret unektelig at denne teknologien i dag er sentraenes teknologi, og konferansen bar ikke preg av spesiell optimisme i retning av at dette ville endre seg særlig i fremtiden.

Fredrik Fagertun har fungert som daglig leder i Registreringssentralen for historiske data siden 1.1.85 og har vært fast ansatt i stillingen siden mai 1986.

Nordisk seminar om maskinoversettelse

København, 9.-11.10. 1986

Jostein H. Hauge

Innledning

I mer enn 50 år har spørsmålet om man kan få maskiner til å oversette fra ett språk til et annet opptatt forskere og interesserte lekfolk. Spørsmålet har hos mange satt fantasien i sving på samme måte som da man tidligere stilte spørsmålet om maskiner kunne fly. Og som i flyveforsøkene har mange prøvd seg helt fra russeren Trojanski i 1933 laget en mekanisk maskin for å slå opp i en ordbok. Særlig fart skjøt arbeidet med maskinell oversettelse etter 1946 da de første forsøkene med data-maskinell oversettelse, dvs. ordbokoppslag, kom i stand i USA. Aktivitetene øket særlig i USA (men også i Europa og Sovjet) inntil 1964 da en amerikansk forskningskomite (ALPAC-komiteen) konstaterte at «we do not have useful machine translation. Further there is no immediate or predictable prospect of useful machine translation».

Virkningene av rapporten var øyeblikkelige og katastrofale for de forskningsmiljøene som arbeidet med maskinoversettelse. Pengestrømmen til forskningsarbeid tørket inn, og prosjektmidler ble kanalisert inn i mer teoretisk orientert lingvistisk arbeid. I det tiår som fulgte, ble det imidlertid på mange felt utarbeidet et solid kunnskapsfelt innenfor data-maskinell analyse av naturlig språk som senere arbeid med maskinell oversettelse har nytt godt av.

Fra 1975 av har interessen for maskinoversettelse tatt seg opp på ny, særlig i USA, men det må også sies at maskinoversettelsearbeidet i flere europeiske land har hatt en mindre dramatisk og mer kontinuerlig utvikling, som f.eks. i Tyskland og Frankrike.

I dag er det skapt helt nye forutsetninger for arbeidet med datamaskinell oversettelse. Datamaskinene har fått enormt stor lagringskapasitet, og programvaren er blitt langt mer videreutviklet. Det lingvistiske grunnlaget er dessuten mer solid enn tidligere samtidig som teoriutviklingen omkring oversettelse som forskningsdisiplin er kommet lengre. Til dette kommer også at det i dag igjen i mange land er vilje til både i statlig og privat virksomhet å satse betydelige ressurser på utviklingsarbeid på feltet. Volumet av den informasjonsmengde som trenger en flerspråklig presentasjonform er vokst dramatisk de siste tiårene samtidig som verden

gjennom bruk av moderne høyteknologibasert industri og telekommunikasjon er blitt langt mindre og mer avhengig av språklig interaksjon og flerspråklig innsikt.

I et land som Japan, som har basert sin økonomiske politikk på utvikling av høyteknologi, har dette bl.a. ført til at språkoversettelse er et sentralt element i det nye konseptet for femtegenerasjons datamaskiner. I dag har Japan en rekke fremtredende forskningsmiljøer i datalingvistik og flere avanserte oversettelsesprosjekter. Mange mener at landet i fremtiden vil få en enda mer fremskutt posisjon på feltet.

Som leser av Humanistiske Data vil være kjent med (jf. nr. 3-84) er det i EFs regi startet et ambisiøst oversettelsesprosjekt med det mål å skape «a machine translation system of advanced design capable of dealing with all official languages of the Community». Dette prosjektet, som i dag beskjeftiger ca. 100 personer i 10 land (jf. senere i rapporten) dannet bakgrunnen for det nordiske seminaret i København.

Nordisk satsing på informasjonsteknologi

Også i Norden har det fra offentlig hold den siste tiden vært satset på sterkere profilering og utbygging av informasjonsteknologisk FoU-arbeid. I tillegg til de nasjonale forskningsfinansierende organer er også Nordisk forskningspolitisk råd gitt i oppgave å utarbeide programmer for satsing på feltet innen de nordiske land. Som et av sine forskningsprogram har Rådet satset på informasjonsteknologi og språk, foreløpig konkretisert i innsatsområdene informasjonssøking og tekstforståelse (jf. referat fra seminar i Stockholm i det kommende HD 1-87) og på maskinoversettelse.

Bakgrunnen for seminaret i København var todelt. For det første ønsket man å få presentert oversettelsesrelaterte utviklingsprosjekter som i dag pågår i de nordiske land og dernest å få drøfte hvordan man bl.a. med utgangspunkt i dette arbeidet kunne tenke seg samarbeidsprosjekter i Norden innenfor feltet maskinoversettelse. Dertil ønsket man å gi en status for det danske Eurotra-prosjektet.

Seminaret, som ble holdt på Københavns universitet, hadde lederen av det danske Eurotra-prosjektet, *Hanne Ruus*, som vert. I alt deltok ca. 40 representanter for Danmark, Sverige, Norge og Finland, hvorav ca. halvparten fra vertskapslandet. Norge var representert med 5 deltakere fra Bergen og Tromsø.

Om foredragene

I sitt innledningsforedrag stilte *Sven-Olaf Paulsen*, Århus det provoserende spørsmålet om vi stiller for strenge krav til maskinoversettelse sammenlignet med dem vi stiller til de mer konvensjonelle oversettelsesformer. Ofte vil maskinens ufullkomne oversettelser være forståelige og

brukbare for sine formål. Paulsen forsvarte verdien av slike «good-enough» oversettelser og mente at vi i vurderingen av dem ikke tok nok hensyn til at vi som språkbrukere er i stand til å foreta meningsfulle slutninger og tolkninger på grunnlag av ufullstendig og tvetydig informasjon, særlig hvis vi kjenner det saksområdet oversettelsen gjelder. I stedet for å tale om hva maskinene *ikke* kan bør vi konsentrere oss om å se hva maskinene faktisk kan produsere av forståelige oversettelser.

Det er for eksempel et faktum, hevdet foredragsholderen, at flere store foretak, f.eks. Ruhrgas, utfra en cost-benefit analyse anvender «rå» maskinoversettelser i utstrakt grad i sitt daglige arbeid. Det vanlige – og meget ærgjerrige krav – om at en oversatt tekst funksjonelt skal ekvivälere grunnteksten, går for langt. De typer tekster som det oftest er tale om å oversette, har til oppgave å informere, ikke å påvirke, og da blir oversettelseskravene naturlig nok redusert. Paulsen mente vi trenger en ny teksttypologi for å kunne bringe drøftingene om maskinoversettelse opp på et tilfredsstillende faglig nivå.

I flere foredrag ble det gitt eksempler på pågående arbeid som foregår innenfor eller som er relatert til maskinell oversettelse. *Ivar Utne* fra Universitetet i Bergen redegjorde for det arbeidet som drives ved Norsk termbank. Her er det utarbeidet databaseverktøy for mikromaskiner for raske oppslag i store fagterminologiske samlinger. På dette grunnlaget er det utviklet opplegg for maskinstøttet oversettelse fra engelsk til norsk. Det gjelder her norske korttekster som skal brukes på skilt i kontrollrom m.v. på oljeinstallasjoner hvor utgangspunktet tradisjonelt har vært engelske skilt. Oversettelsen skjer med grunnlag i en lagerkatalog som består av 68.000 enlinjers tekster. I sitt arbeid med oversettelse til norsk fra engelsk kan oversetteren slå opp i tidligere oversettelser av samme ord eller tekstdel, utføre en serie oversettelser der samme ord eller tekstdel skal ha identisk oversettelse og gjennomføre raske konsistenskontroller. Et system for halvautomatisk oversettelse gjør det mulig å identifisere tidligere utførte oversettelser og foreta interaktiv oppretting, p.t. for i alt ca. 80.000 ulike skilt.

Ved Universitetet i Bergen er det også startet et oversettelsesprosjekt med basis i HF-fakultetets edb-seksjon og Engelsk institutt. *Magnar Brekke* orienterte om prosjektet og sa bl.a. at målet er å lage en automatisert engelsk til norsk oversettelse av tekniske fagtekster.

I prosjektet brukes det programsystemet for oversettelse mellom engelsk og tysk som er utarbeidet av W. Weidner Communication Corporation i Chicago. I dette systemet skal det legges inn norsk vokabular, fagspråk og terminologi i en systemordbok. Norske bøyningsmønster, ideomatiske uttrykk, sammensatte verbaluttrykk og regler for ordstilling skal systematiseres og innarbeides. Utviklingsarbeidet, som vil foregå på en Micro VAX II, forventes å resultere i en første systemprototype i september 87.

I en serie foredrag ble det redegjort for datalingvistisk arbeid som også kan ses i sammenheng med datamaskinell oversettelse.

Fra Universitetet i Stockholm redegjorde *Gunnel Källgren* for det arbeidet som utføres med oppbygging av såkalte dynamiske leksikon. Ved bruk av leksikon som både gir generell og prototypisk informasjon og som kan ta vare på referentens diskurshistorie i teksten som analyseres, kan vi klare bl.a. å oppløse tvetydige ord og nå et mer konsistent stilnivå i oversettelser. At menneskelig forståelse av tekster på en avgjørende måte er forbundet med bruk av kontekstinformasjon er velkjent, mente foredragsholderen. I arbeid med maskinell oversettelse av tekster vil systemenes muligheter til å analysere og tolke kontekster være av fundamental karakter. Det svenske ordet «varv» kan bl.a. bety verft eller sirkelbevegelse – på engelsk wharf, lap, revolution eller rotation. Hvilket ord som skal brukes i en oversettelse vil direkte være betinget av den kontekst som ordene inngår i. Følgende eksempelsetninger som foredragsholderen brukte, vil også gi en pekepinn om de behov for kontekstkontroll og forståelse som finnes ved automatisk oversettelse: Pikene spiste biffer. De krydret dem med chilisaUCE. Mine brødre ba sine fruer på koteletter. De hadde stekt dem til dem!

Olli Blåberg fra Helsingfors universitet holdt et foredrag om den såkalte tonivåmodellen for datamaskinell morfologisk analyse. Denne morfologiske analysemodellen har med stort hell vært brukt på en rekke typologisk svært forskjellige språk. Grunnen er at man ved hjelp av modellen klarer å behandle en lang rekke ulike bøyingsparadigmer i ett og samme hovedparadigme. Det kommer bl.a. av at modellen opererer med et regelnivå og et realisasjonsnivå. Dermed kan f.eks. besværlige stammeinterne alternasjoner og allomorfer tas hånd om på en økonomisk og enhetlig måte. Det gjelder f.eks. typen enkel/enkle, rød/røtt (svensk), and/ender etc.

Helge Dyvik fra Universitetet i Bergen refererte fra sitt arbeid med syntaktisk og semantisk analyse (parsing) av norske setninger. Arbeidet bygger på systemet D-PATR som er et system for grammatikkutvikling laget av Lauri Karttunen (unifikasjonsgrammatikk). Ved å modifisere dette systemet er det bl.a. blitt mer effektivt å teste de syntaktiske regler som utvikles for bl.a. å analysere idiommer og spesielle fraser hvor den «normale» form for språkbeskrivelse kommer til kort.

Fra Universitetet i København redegjorde *Gregers Koch* for sitt arbeid med Prologbaserte logiske grammatikker ved oversettelse av enkle setninger til semantiske representasjoner.

Arendse Berenth fra samme universitet tok for seg bruk av situasjonssemantikk (implementert i Prolog) ved semantisk beskrivelse av naturlig språk. Med bakgrunn i virksomhet ved forskningsstiftelsen SITRA, Helsingfors orienterte *Kari Valkonen* og *Aarno Lethola* om automatisk morfologisk, syntaktisk, semantisk og logisk analyse av finske setninger.

Det såkalte Kielikone-prosjektet har som mål å lage en parser for finsk språk som kan være et praktisk hjelpemiddel i produksjonsomgivelser. Målet er bl.a. å kunne bruke finske setninger ved spørsmålsstilling til informasjonssystemer. I prosjektet arbeides det også med AI-baserte teknikker for kunnskapsakkvisisjon. En prototype av deres parser er utviklet for søking av informasjon fra en geologisk databank.

Den danske delen av Eurotra-prosjektet

Naturlig nok kom det danske Eurotra-prosjektet til å stå sentralt under seminaret i København. Ved prosjektet er det i 1986 ansatt 13 medarbeidere (tilsvarer 9 årsverk), fortrinnsvis språkvitere med kunnskap om datalingvistik. Denne staben vil senere bli økt til 15 heltidsansatte. Prosjektet finansieres med 80% fra EF og 20% av den danske stat, i alt ca. 25 mill. kroner for hele perioden 82/89.

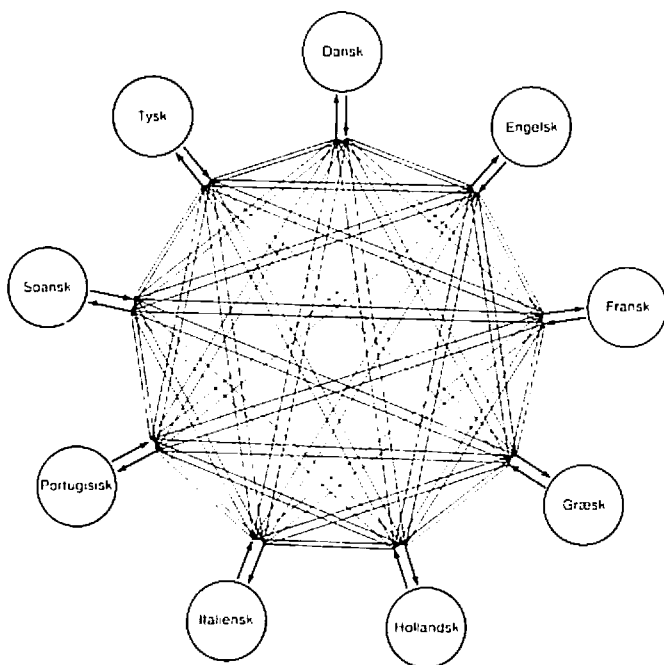
Eurotra er bygget opp med EF-kommisjonen som øverste myndighet. Leder er Sergei Perschke, Luxemburg. Den største delen av utviklingsarbeidet foregår desentralt, dvs. i de enkelte medlemsland. Derfor er det nedsatt forskjellige, tverrnasjonale utvalg for å koordinere og planlegge virksomheten.

Eurotra er et resultat av EF-kommisjonens ønske om å sette EF i stand til å møte det økende behovet for oversettelse i tiden fremover, dvs. mellom 72 forskjellige språkpar (jf. figur). Målet med Eurotra er å skape et system som kan oversette tekster innenfor *informasjonsteknologi* og som senere kan utvides til å oversette tekster innenfor andre områder. I slutfasen i 1989 skal systemet kunne behandle et ordtilfang på 20.000 ord.

På seminaret i København ble prøveteksten for testoversettelsene vist. Som prøvetekst er valgt avtaleteksten om Esprit-programmet (Proposal for a Council decision adopting the first European Strategic Programme for Research and Development in Information Technologies (ESPRIT)). Ved studium av denne teksten kan interesserte danne seg et konkret inntrykk av de ytelseskrav som stilles til det systemet som skal utvikles.

Oversettelsessystemet vil imidlertid ikke overflødiggjøre oversettere som nå arbeider i EF. De vil i fremtiden også bl.a. være nødvendige for å revidere tekster produsert av det maskinelle oversettelsessystem, men vil forhåpentligvis i økende utstrekning bli spart for rutinearbeid.

For å kunne oversette tekster ved hjelp av datamaskin, må tekstene analyseres morfologisk, syntaktisk og semantisk. Denne analysen danner utgangspunkt for oversettelsen til et annet språk. Resultatene av denne overføringen (transfer) skal så gjennom en tilsvarende prosedyre generere en korrekt setning i målspråket. Ved oversettelse av en tekst fra dansk til engelsk anvendes en dansk analysemodul. Den omfatter en dansk ordbok og grammatikk (jf. figur), en dansk-engelsk oversettelsesmodul (som



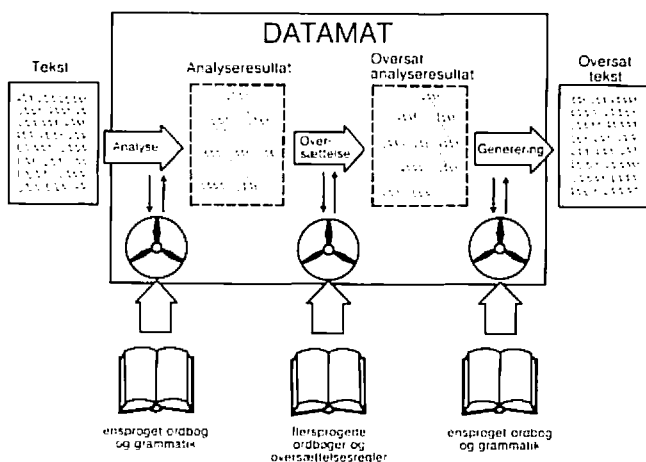
Hvert av EF-språkene skal oversettes til alle de øvrige språkene.

bruker en dansk-engelsk ordbok og oversettelsesregler) og en genereringsmodul (engelsk ordbok og grammatikk) som så genererer den engelske tekst. Dvs. den oversetter ordene og foretar en grammatisk beskrivelse som resulterer i en korrekt engelsk tekst.

Ved å bygge opp systemet modulært oppnås at de særspårlige analysedeler kan anvendes ved oversettelse til *alle* språk. Det samme gjelder tilsvarende for genereringsmodulen.

Slik sett er systemet tenkt som et helt generelt oversettelsessystem, noe som tillegges stor betydning med hensyn til de senere muligheter til å videreutvikle systemet til et kommersielt system for oversettelse.

Under seminaret i København presenterte en rekke av medarbeiderne i Eurotra ulike sider ved oversettingsprosjektet. Hanne Ruus tok for seg den modell for maskinoversettelse som Eurotra bygger på. Eurotra sikter mot en fullautomatisert datamaskinell oversettelse uten særlig tilrettelegging av de tekster som skal oversettes. Systemet skiller seg derfor fra slike hvor der er en direkte interaksjon mellom oversettere og datasystemene i løpet av oversettelsesprosessen (såkalt «Machine Aided Human Translation» eller «Human Aided Machine Translation»).



Den danske analysemodulen for maskinoversettelse i Eurotra-prosjektet.

I Eurotra baseres edb-systemet på en omfattende formell modell samtidig som systemet er strengt modulært oppbygd. For lingvistene er det utviklet formelle beskrivelsesspråk, analyseverktøy og ordbøker. De morfologiske analyseverktøy var tema for foredraget til *Anders Nygaard*. For å være i stand til å samarbeide effektivt med andre arbeidsgrupper i ti land, må alle språkgruppene bruke det samme programverktøyet for morfologisk og syntaktisk analyse, for transfer og generering. For morfologisk analyse finnes det i dag tre ulike typer, hvorav to bygger på PROLOG.

Brukerspråket i Eurotra var temaet for *Anne Lise Bechs* presentasjon. Her ble bl.a. vist hvordan oversettelse av kildetekst til måltetekst foregår trinnvis som en oversettelse mellom ulike representasjonsnivåer. Disse består igjen av ulike mengder grammatikkregler skrevet i brukerspråket. Bech tok for seg formål og virkemåte til de grammatikk- og oversettelsesregler som i dag er utviklet.

I en serie presentasjoner (*Frede Boje, Ole Togeby, Charlotte Tubro*) ble overføringsnivået i oversettelsen mellom to språk (transfer) beskrevet. Bl.a. ble det gitt en interessant presentasjon av de spesielle problemer Eurotra står overfor i og med at alle de 10 (senere 12) språk skal beskrives på en slik måte at samme type overføringsstruktur (interface structure level) kan nyttes. For Eurotra betyr det at overføringsnivået må gjøres «dypere» enn hva man trenger for oversettelse mellom to språk.

I en avsluttende presentasjon ga *Bente Maegaard* en oversikt over stadiene i Eurotra-prosjektet. Hun delte den inn i

- 1) forberedelsesstadiet
- 2) utarbeiding av et lite system pr. desember 87 (skal omfatte 2500 ord)
- 3) slutføring av et prototypisk system pr. 1989, som kan håndtere et tekstmateriale på 20.000 ord

I en drøfting etter foredragene av mulighetene for et nordisk samarbeid om maskinoversettelse opplyste *Peter Lau*, EF-kommisjonen bl.a. at Eurotra-prosjektet er åpent for en bred kontakt med interesserte parter utenfor EF i form av diskusjoner og erfaringsutveksling.

Fra starten av har det vært ansett som et viktig mål gjennom Eurotra å kunne stimulere til en forsterket innsats innen datalingvistikk og språklig basert informasjonsteknologi i Europa. Lau mente at de øvrige nordiske land har et godt utgangspunkt for et faglig samarbeid med Eurotra med grunnlag i de avtaler som foreligger mellom de ulike nordiske land og EF om kulturelt og vitenskapelig samarbeid. Det ble også nevnt at Sveits er i ferd med å opprette et kontraktfestet samarbeid om deltakelse i Eurotra-prosjektet.

I den påfølgende drøfting ble ulike nordiske oversettelsesprosjekter og



Bente Maegaard (t.v.), formann i liason-gruppen i EUROTRA, og Hanne Ruus (t.h.), leder av EUROTRA-DK.

samarbeidsmuligheter løselig drøftet. Flere så det som ønskelig at en kunne få i gang et oversettelsesprosjekt mellom et skandinavisk språk og finsk. Samtidig vurderte man hvordan man kunne knytte særlig svensk og norsk utviklingsarbeid til de aktiviteter som foregår på dansk side innenfor Eurotra.

Et visst konkret inntrykk av forskjeller og likheter mellom de nordiske språk fikk deltakerne også gjennom et gruppearbeid (ikke maskinelt, men intellektuelt!) der samme lovtekst på alle nordiske språk ble analysert, særlig med tanke på oversettelsesproblemer.

Det er grunn til å takke den danske Eurotragruppen for et godt gjennomført arrangement. Så gjenstår det å se om det man fikk høre vil gi konkrete utslag i oppfølgende prosjektplanlegging i de øvrige nordiske land.

Foredragene som ble holdt på seminaret, skal utgis i en publikasjon som kan bestilles fra EUROTRA-DK.

Nytt fra RHF/NAVF

Avlevering, oppbevaring og gjenbruk av data fra NAVF-prosjekter

I 1985 nedsatte NAVFs styre et utvalg for å utrede spørsmålene omkring avlevering, lagring og gjenbruk av data fra NAVF-prosjekter. Utvalget fikk følgende mandat:

- å utrede spørsmålet om eiendomsretten til data fra NAVF-prosjekter
- å komme med forslag til regler og rutiner for avlevering og oppbevaring av slike data
- å komme med forslag til regler og rutiner som sikrer fornuftig gjenbruk av slike data
- å avklare forholdet til det offentlige arkivverket
- å komme med forslag til avtaler med aktuelle arkivinstitusjoner
- å vurdere behovet for opprustning av aktuelle institusjoner og i den sammenheng også vurdere de økonomiske konsekvensene av en slik opprustning
- å komme med forslag til revisjon av NAVFs bevilgningsvilkår

Utvalgets innstilling forelå i juni i år. I det følgende gis utdrag fra innstillingens kapittel 2: Sammendrag og konklusjoner.

«Utvalget tar sitt utgangspunkt i en vurdering av «eiendomsretten til data fra NAVF-prosjekter». Spørsmålet har en side dels til eiendomsretten etter de regler som gjelder i dag, dels til de regler som bør gjelde etter innføringen av et generelt system for avlevering og oppbevaring av slike data. Utvalget finner ikke grunn til å gå inn på den førstnevnte

Avlevering, oppbevaring og gjenbruk av data
fra NAVF-prosjekter

Innstilling fra et utvalg nedsatt av NAVFs styre

relasjon. I den sistnevnte relasjon legger utvalget vekt på at plikten til avlevering kommer klart nok fram overfor den som søker om og eventuelt oppnår støtte til et forskningsprosjekt. Dette er nødvendig for å unngå praktiske problemer når ordningen skal settes ut i livet overfor vedkommende forsker eller forskningsinstitusjon. Det er også nødvendig ut fra forskernes grunnleggende «eiendomsrett» til materialet. Hvis de har mottatt støtte og samlet inn en større eller mindre del av de planlagte data, vil senere «vilkår» med tilbakevirkende kraft kunne støte på problemer av rettslig og praktisk art, som klart utformede vilkår fra starten av ikke vil kunne bli møtt med.

Kravet om klar presentasjon kan tenkes oppfylt både gjennom utsagn i de enkelte bevilgningsbrev og ved at brevet på vanlig måte viser til et sett

standardvilkår som ligger ved. I begge tilfelle er det viktig at bestemmelsene om plikt til avlevering i seg selv er utformet på en tilstrekkelig klar og hensiktsmessig måte.

Utvalget går inn for et opplegg der alle prosjekter som har finansiering fra NAVF skal kunne pålegges å avgi sine data til en arkivinstusjon, men gjør unntak for prosjekter innen naturvitenskap. Her foreslås det isteden at det gis et tilbud om lagring. Med finansiering menes bevilgninger til lønn eller drift. Støtte til reiser, forskningsopphold i utlandet, vitenskapelig utstyr og rene forprosjekter er holdt utenfor.

Utvalget har begrenset seg til å diskutere et system for data som er edb-basert og lagret på edb-media. Etter en vurdering av bl.a. behovet for forskningsarkiv, av økonomiske forhold, av hensynet til dem som avgir informasjon til forskere og av behovet for etterprøving av forskningsresultater, foreslår utvalget at NAVF stiller midler til disposisjon for å sikre arkivering av data.

Utvalget foreslår et system for avlevering, lagring og gjenbruk av slike data. Som hovedregel bør avlevering skje innen tre år etter den perioden NAVF har gitt bevilgning for. Utvalget presiserer at ingen vil miste rettigheter til bruk av data som de selv har samlet inn ved etablering av en slik ordning.

Utvalget foreslår at NAVF legger opp til et system for avlevering, lagring og gjenbruk av data, som i hovedsak bygger på de to NAVF-institusjonene Norsk samfunnsvitenskapelig datatjeneste (NSD) og NAVFs EDB-senter for humanistisk forskning (EDB-H). Det legges opp til følgende arbeidsdeling:

EDB-H: Data fra prosjekter innen litteratur, språkvitenskap og kultur-fag.

Personidentifiserbare historiske data fra tida før 1900. Gjenbruk av den siste type data må skje i samarbeid med Registreringssentralen for historiske data (RHD).

NSD: Samfunnsvitenskapelige og samtidshistoriske data.

Sikkerhetslagring av data fra naturvitenskapelige prosjekter.

Data fra medisinske prosjekter dersom en ikke satser på en NAVF-ekstern institusjon til dette.

(...)

For historiske individdata foreslår utvalget at disse sikkerhetslagres ved EDB-H. Utvalget forutsetter imidlertid at Registreringssentralen for historiske data (RHD) skal ha tilgang til slike data og spille en sentral rolle i organiseringen av gjenbruken av dem.

Dersom ordningen etableres fra og med 1987 og blir basert på arkivering ved NSD og EDB-H, vil det være behov for en bevilgning på ca. kr. 560.000 inklusive kr. 100.000 til utstyr. (...)

Utvalget har vurdert et opplegg for NAVF alene. Det blir imidlertid sett på som ønskelig at ordningen legges så åpent opp at også andre forskningsfinansierende organer kan benytte seg av den.»

Anbefalingen fra utvalget vil i løpet av høsten bli drøftet av de berørte fagråd.

Kristin Natvig

Behov for vitenskapelig utstyr i de humanistiske fag

I desember 1985 oppnevnte Rådet for humanistisk forskning en utredningsgruppe som først og fremst skulle lage et notat om behovet for vitenskapelig utstyr i humanistiske fag. Utredningsgruppen leverte sin innstilling i august. Gruppen bestod av følgende medlemmer:

- Direktør Jostein H. Hauge, NAVFs EDB-senter for humanistisk forskning
- Førstekonservator Kristen Michelsen, Universitetet i Bergen
- Amanuensis Kolbjørn Slethei, Universitetet i Bergen

Konsulent Claus Huitfeldt, NAVFs EDB-senter, har fungert som sekretær for utvalget.

I 1985 publiserte NAVFs utredningsinstitutt utredningen «Utstyr for forskning. Situasjonen innen humaniora og samfunnsvitenskap.» Med sin innstilling ønsker arbeidsgruppen å nyansere og utdype den situasjonsbeskrivelse som denne utredningen gir.

Gruppen har bl.a. basert arbeidet sitt på innhentede uttalelser fra 10 forskere innen særlig utstyrskrevede humanistiske fag. Disse og andre forskere har i de senere årene uttrykt bekymring over at det stilles for små midler til disposisjon for innkjøp og drift av vitenskapelig utstyr. «Mange humanister er kommet til den konklusjon at bruken av deres egen faglige arbeidskraft i dag er urasjonell og økonomisk forkastelig fordi de ikke får tilgang til tidsmessige og fagrelevante hjelpemidler», heter det i innstillingen (s. 4). Humanistiske forskere har mindre tilgang til utstyr enn både forskere på andre fagområder og deres kolleger i utlandet. Disse og flere forhold utgjør bakgrunnen for utvalgets arbeid.

Utvalget definerer begrepet «vitenskapelig utstyr» som følgende: «Med vitenskapelig utstyr menes utstyr (fysisk og symbolsk) som forskere bruker til å innsamle, registrere og organisere forskningsmateriale, til å analysere og tolke det, og til å beskrive resultatet av analysen/tolkningen/ hypoteseprøvingen.» (s. 6) Faktorer som virker inn på hvordan utstyr faktisk brukes og hvordan det kan/bør brukes, er driftsmidler, personell, utdanningsmuligheter og administrative rutiner.

Et av kapitlene i innstillingen er viet til en generell analyse av utstyr for humanistiske fag. «Den faglige spredningen gjør det vanskelig å peke på noen samlende teoretisk eller metodologisk faktor som skulle tilsi et fellesskap i behov for vitenskapelig utstyr. Dette gjelder både *omfanget* av behovet, ... og det gjelder *typen* av utstyr», heter det i innstillingen (s. 8). Utvalget påpeker at humaniora «... ofte blir oppfattet som fag som pr. definisjon så å si bare trenger penn og papir.» (s. 8) Men humanistene trenger *bedre* penn og papir – «I vår tid betyr dette først og fremst datamaskinell behandlingsmulighet av de symboler – språklige, materielle, tonale o.a. – som er forskningsobjekt for humanister i svært mange tilfelle.» (s. 8-9)

Gapet mellom behov for utstyr og behovsdekning i form av bevilgninger over statsbudsjettet har økt de siste årene. Derfor har universitetene skaffet rabatter på personlige datamaskiner til forskere. Dette har bidratt til å løse den enkeltes behov for datakraft, men har den utilsiktede bieffekten at forskerne arbeider mest effektivt *hjemme*. De bevilgende myndigheter må sørge for at humanistiske forskere skaffes verktøy *på arbeidsplassen*, hevder utvalget.

Utvalget argumenterer for at det er rimelig å betrakte tekstbehandlingsutstyr som *vitenskapelig* utstyr for humanister. Det påpekes at «I mange fag foregår sentrale deler av forskningsprosessen i løpet av det omfattende skriftlige fremstillingsarbeidet. Det kan derfor hevdes at utstyr som forenkler skriveprosessen, gjør det lettere å organisere – og reorganisere – stoffmengden og beholde en intellektuell oversikt over et komplisert stofftilfang, må sees på som vitenskapelig utstyr – om enn i utvidet forstand.» (s. 10) Eksempler blir gitt på tekstbehandlingsprogrammets anvendelighet for oppgaver i forbindelse med behandling av tekster og språklig materiale.

Også undervisningsutstyr kan ses på som forskningsutstyr. I enkelte fag, f.eks. lingvistikk, og film- og dramavitenskap, er det selve *formidlingsprosessen* som er forskningsobjektet. Og i humanistiske fag er formidling av innsikt en del av selve forskningsprosessen. Utstyr for behandling av en formidlingsprosess i forskningøyemed og utstyr for formidling av forskningsresultater kan ha et nært fellesskap. Ettersom tekstlig formidling vil bli den dominerende form i de kommende 3-5 år, bør tekstbehandlingsutstyr utgjøre en del av humanistenes vitenskapelige utstyr.

Etter utvalgets oppfatning bør alle forskere ha som basisutstyr en PC med tekstbehandlingsprogram som kjerneutrustning og en lavkvalitets-skriver. Laserskrivere er fortsatt så dyre at de må sees på som en ressurs som må deles. Annet utstyr som regnes for å være ønskelig for alle humanister, er bl.a. databaseprogram, regneark eller statistikkprogram.

Ifølge utvalget tilsier prishensyn at en del typer utstyr må deles. Ettersom anskaffelser må planlegges godt på forhånd, bør fagmiljøene

selv avgjøre innkjøp. I denne forbindelse foreslår utvalget at NAVFs EDB-senter rustes videre opp til å bli en kompetansesentral, og at edb-seksjonene ved HF-fakultetene bygges ut.

Utstyr som bør bli en felles nasjonal ressurs, er en optisk karakterleser av høy kvalitet. Utvalget påpeker at en optisk leser vil spare humanister både tid og penger ved omforming av tekster til maskinleselig form. Utviklingen av utstyr som kombinerer både optisk lesing og grafisk scanning bør følges nøye.

I utredningen gis det en detaljert oversikt over status og behov for vitenskapelig utstyr innen ulike fagfelt. Når det gjelder **språkfagene**, dreier hovedtyngden av utstyrssøknader til RHF i 1985-86 seg om innkjøp av PC'er. En PC til ca. 40.000 kr. vil tilfredsstille de fleste språkforskernes behov for edb-utstyr. Miljøer for datalingvistikkk representerer fag som trenger kraftigere utstyr enn PC'er. Særlig aktuelle her er kraftige arbeidsstasjoner for bruk av LISP og PROLOG. Slikt utstyr er ikke bare nødvendig for å utføre sentrale forskningsoppgaver – det er en klar forutsetning for å tiltrekke høyt kvalifisert personell til disse miljøene. Også flerspråklige fag som religionshistorie, teologi og klassisk filologi har behov for spesialutstyr.

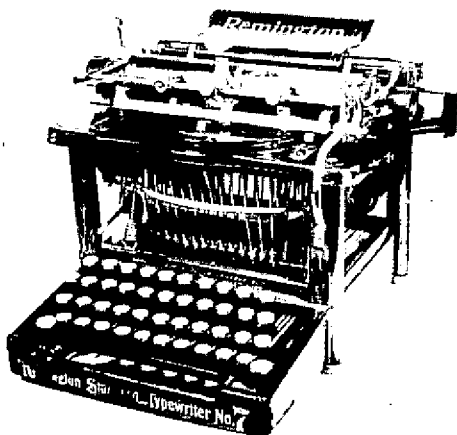
Når det gjelder **arkeologi**, hevder utvalget at «Når nye utstyrstyper tas i bruk innen arkeologien kan konvensjonelle operasjoner gjøres langt mer eksakt og rasjonelt og nytt utstyr gjør det mulig å ta opp nye problemstillinger.» (s. 22) Utstyrbehovet faller særlig innenfor tre områder: 1. Utstyr til arbeid i felten. 2. Utstyr til registrering og behandling av feltdata, og til uttegning av kart og gravningsplaner. Overføring av arkivdata til edb, edb-utstyr til statistikk, grafikk m.v. 3. Utstyr til konservering og til arkeometri.

Edb-bruken må trappes opp ved de **kunst- og kulturhistoriske museene**, slik at katalogisering kan gjennomføres av de ca. 4 millioner uregistrerte gjenstandene ved museene. Det må også satses på bruk av edb ved lagring og bearbeiding av registreringsdata. Museene vil også ha stor nytte av billedlagring på videoplate. Videoutstyr kan forenkle arbeidet med å registrere og dokumentere håndverk og folkeliv. Det er også behov for vitenskapelig utstyr til forskning og undervisning i konservering.

Det største utstyrbehovet ligger på edb-siden også for instituttene for **etnologi og folkloristikk**. I tillegg er lydutstyr og utstyr for bildeddokumentasjon viktig for forskning og undervisning i de etnofolkloristiske miljøene.

De **kunsthistoriske** fagmiljøene har særlig behov for edb-utstyr til tekstbehandling og edb-baserte kataloger og registre. Utvalget påpeker dessuten at edb-metoder for datalagring av bildeinformasjon gir helt nye perspektiver for edb-bruken innenfor de kunsthistoriske fagene.

Om utstyrbehovene til **historie- og arkivfag** hevder utvalget: «Frem-



Humanistene sliter med foreldet vitenskapelig utstyr.

over vil elektronisk utstyr i stigende grad kunne utnyttes i arbeidet med historisk datafangst og ved behandling av innsamlet kildemateriale. Som i arkeologien vil ikke minst mulighetene til å utnytte moderne videoteknikk og bærbart datautstyr gi et helt nytt grunnlag for en mer effektiv og fullstendig dokumentasjon.» (s. 32) Bruk av videobånd og digitale laserplater vil gi muligheter for en total visuell dokumentasjon og meget kompakt datalagring. Ettersom stadig flere historiske kilder alt fra starten vil foreligge på edb-medium, må alle historiske forskere ha adgang til edb-utstyr for å kunne nytte disse kildene. Avanserte leserapparater for mikrokort og mikrofilm, koblet sammen med mikromaskiner for å lette framfinningsprosessen, vil forbedre historikernes arbeidsmuligheter.

Ved Riksarkivet er det behov for å utvide grunnutstyret, og for utstyr knyttet til optisk platteteknologi og til formidlingsformål. Også statsarkivene bør få utvidet lagrings- og behandlingsskapasiteten på edb-siden.

Satsingen på utvikling av **humanistiske mediefag** medfører behov for en større grunninvestering i utstyr knyttet til de elektroniske bilde- og lydmediene. På forsknings- og undervisningsnivå er det først og fremst behov for utstyr som gir adgang til studier av medieprodukter, mens på formidlingsiden er produksjonsutstyr mest interessant.

En satsing på humanistisk forskningsformidling via film og video vil innebære store utstyrsinvesteringer, enten produksjonen foregår lokalt i instituttene eller i regi av et nasjonalt senter. I tillegg bør det opprettes en sentral institusjon for distribusjon av undervisnings- og forskningsmateriale på film og video, i tillegg til lokale videosamlinger.

Utvalget hevder at «Interaktiv video kan vise seg å bli et viktig redskap innen mediepedagogikken samt innen formidling av forsknings-

resultater i form av interaktive videoprogrammer.» (s. 38) Utvalget foreslår derfor opprettelsen av et nasjonalt laboratorium for arbeid med interaktiv video, med videoplaten som det sentrale medium.

Innenfor **musikkvitenskapen** må de nye elektroniske hjelpemidlene både nyttes som forskningsredskaper og gjøres til objekter for musikkvitenskapelig analyse. I tillegg må det tradisjonelle utstyrsnivået holdes vedlike.

Et minimum av elektronisk utstyr vil bl.a. si PC'er i nettverk, noteskriverutstyr og grafiske terminaler. Også utstyr for digitalisering og syntetisering av musikk er ønskelig.

Ettersom behovet for tekniske investeringer i de musikkvitenskapelige miljøene er svært stort, blir det foreslått at det etableres et nasjonalt senter for både forskning og for skapende og utøvende virksomhet. Et slikt forskningsstudio bør ha måleinstrumenter og utstyr for observasjon, komposisjon og fremføring.

I et eget kapittel i utredningen anslår utvalget totalbehovet innenfor en 3-års ramme for vitenskapelig utstyr i de humanistiske fagene. Behovet for edb-utstyr er størst, men også andre typer utstyr anses for å være nødvendig, spesielt innenfor de arkeologiske fagområdene og museumsvirksomhet.

Når det gjelder investering i personlig datautstyr, regner utvalget med at en «minimumspakke» til hver enkelt forsker vil koste ca. 30.000 kr. Den enkeltes andel av felles, delte tjenester anslås til å utgjøre ca. 7000 kr. Ifølge utvalget kan den nåværende behovsdekning settes til ca. 5%. Ettersom utvalget regner med et behov for personlig datautstyr hos ca. 80% av alle forskere, vil det totale investeringsbehovet ligge på 30 mill. kr. Behovet for tilsvarende utstyr for forskningsassistenter utgjør 5,5 mill. kr.

Utredningen inneholder også en diskusjon om hvilke retningslinjer som bør ligge til grunn for RHF's bevilgningspraksis, og ulike kriterier for vurdering av utstyrssøknader. Utvalget går inn for at RHF bør legge til grunn en vid utstyrsdefinisjon. Det er også viktig at de tradisjonelle utstyrsbehovene ikke blir skadelidende.

Videre blir det drøftet om RHF bør se det som sin primære oppgave å supplere universitetenes basisutstyrspark, eller gi prioritet til f.eks. spesialutstyr og utprøving av ny teknologi. Utvalget reiser dessuten spørsmål om det – særlig ved store utstyrsstildelinger – er aktuelt å be institusjonene om å prioritere sine søknader. I denne sammenheng drøftes det også om det kan være en riktigere politikk å gi store utstyrsbevilgninger til utvalgte miljøer heller enn å spre ressursene over et stort antall mindre søknader. Etter utvalgets mening bør RHF opprettholde – og om mulig styrke – kontakten med institusjonene slik at rådet til enhver tid har et dekkende bilde av utstyrssituasjonen og de planer for investeringer som finnes, særlig ved de store forskningsinstitusjonene.

Som vedlegg til utredningen gis de 10 tidligere nevnte uttalelser, i tillegg til en enkel kvantitativ analyse av søknader om og bevilgninger til vitenskapelig utstyr fra RHF 1985/86.

RHF inviterte de humanistiske forskningsinstitusjonene til et kontakt- og drøftingsmøte den 1.12. i Oslo. På møtet deltok bl.a. representanter fra universitetene, DH-skolene, Arkeologisk museum i Stavanger, Musikkhøgskolen og NAVFs instrumenttjeneste.

Etter en innledning fra RHF om bakgrunnen for de siste års utredningsarbeid, ga institusjonene en vurdering av den beskrivelse som er gitt i utstyrsinnstillingen. Det var enighet om at utstyrsbehovene er store både på universitetene, ved DH-skolene og ved de øvrige institusjoner som var representert. Samtidig er det visse lyspunkter ettersom flere av de større institusjonene har lagt opp til relativt ambisiøse planer for investering i datautstyr, særlig personlige datamaskiner.

Fra Nasjonalt råd for distriktshøgskolene ble det fremholdt at utstyrsituasjonen for de humanistiske fag ved distriktshøgskolene syntes også å være langt dårligere enn i tilsvarende fag ved universitetene.

Ved flere av institusjonene er det for tiden oppbygging av nye fag hvor den faglige virksomheten forutsetter store utstyrsinvesteringer, eksempelvis mediefag og datalingvistikk.

På møtet konstaterte man at RHF har behov for nær og kontinuerlig kontakt med de vitenskapelige miljøene for å kunne føre en fornuftig og rettferdig tildelingspolitikk når det gjelder vitenskapelig utstyr. Samtidig har de store institusjonene behov for kontinuerlig oversikt over det utstyr som bevilges fra forskningsrådene. Det ble fra RHF's side opplyst at alle utstyrstildelinger fra 1986 av blir registrert i en database og presentert i forskningskataloger. RHF understreket at det utstyr som NAVF bevilger til forskning er institusjonenes og ikke den enkelte forskers eiendom.

Selv om edb-utstyr for tiden dominerer søknadsmassen til RHF er det viktig – mente møtet – også å tilgodese forskningsmiljøenes behov for mer konvensjonelt vitenskapelig utstyr.

RHF-representantene lovet å ta med seg de synspunkter som var fremkommet på møtet når Rådet om kort tid skal utforme retningslinjer for sin tildeling av vitenskapelig utstyr til forskningsmiljøene.

Kristin Natvig

Jostein H. Hauge

Informasjonsteknologi: utfordringer for humanistiske, medisinske og samfunnsvitenskapelige fag

NAVFs styre ser det som hensiktsmessig å dele informasjonsteknologi i to programmer: et naturvitenskapelig og teknisk/industrielt område, og et program for samfunnsmessige konsekvenser og anvendelser av informasjonsteknologi innenfor humanistiske, medisinske og samfunnsvitenskapelige fag. NAVFs styre har oppnevnt et koordineringsutvalg for det andre delprogrammet som består av lederne for teknologiutvalgene under RHF, RMF og RSF: *Knut H. Sørensen*, RSF (leder), *Helge Dyvik*, RHF, *Lars Walløe*, RMF og *Knut Bjørseth*, RSF (sekretær).

I august leverte koordineringsutvalget innstillingen «Informasjonsteknologi: utfordringer for humanistiske, medisinske og samfunnsvitenskapelige fag.» I følgende sammendrag av innstillingen vil det bli lagt særlig vekt på de humanistiske fagenes forhold til informasjonsteknologi.

Utvalget påpeker at informasjonsteknologiens framvekst har gitt grunnlag for nye relasjoner mellom teknologiske fag på den ene siden og humanistiske, medisinske og samfunnsvitenskapelige fag på den andre. I prinsippet er et komplekst samspill mulig mellom disse fagene – et samspill som det er på tide å erkjenne og ta på alvor.

Etter utvalgets oppfatning virker det forskningspolitisk sett rimelig å se satsingen på informasjonsteknologi i lys av de tre øvrige hovedinnsatsområdene for norsk forskning, deriblant tradisjons- og kulturformidling. Disse hovedinnsatsområdene representerer til en viss grad de sentrale problemstillinger ved humanistisk, medisinsk og samfunnsvitenskapelig forskning på informasjonsteknologi. I tillegg til dette krysningsfeltet er det fire hovedtyper koplinger mellom informasjonsteknologi og de tre fagområdene. Den første av disse er humanistiske, medisinske og samfunnsvitenskapelige fag som *leverandører av innsikt* ved utvikling av informasjonsteknologi.

Mange problemer innenfor informasjonsteknologien angår menneskelig interaksjon og kunnskap. «Når informasjonsteknologien tas i bruk i kommunikasjon mellom mennesker, er det viktig at den tilpasses en menneskelig kommunikasjonsform. Dette forutsetter studier av menneskelig symbolbehandling, menneskelig resonnement og andre menneskelige ferdigheter, som et grunnlag for simulering.» (s. 10-11). I denne sammenheng står systemer som kan prosessere naturlig språk sentralt. Slike systemer må bygge på formelle lingvistiske studier, mens taleprosessering må bygge på fonetiske og fonologiske analyser i tillegg til naturvitenskapelige teknikker. Videre er formell logikk og logisk semantikk «grunnlag for mye av betydningskomponentene i naturligspråklige systemer, og for resonnerende komponenter i kunnskapsbaserte systemer overhodet.» (s.

Informasjonsteknologi:

Utfordringer for humanistiske, medisinske
og samfunnsvitenskapelige fag

Innstilling fra NAVFs koordineringsutvalg

11) Utviklingen av kunnskapsbaserte systemer innebærer problemområder som er felles for matematiske, filosofiske, lingvistiske og psykologiske forskningsmiljøer. Kommunikasjon i bildeform krever forskning innenfor humaniora (bildepråk og bildekommunikasjon), medisin og psykologi (visuell persepsjon) og informasjonsteknologi (redskaper for visuell kommunikasjon). Robotisering omfatter alle disse forskningsfeltene.

På disse områdene skjer utviklingen meget raskt internasjonalt. Etter utvalgets oppfatning må norske miljøer settes i stand til å holde tritt med denne utviklingen, og flerfagligheten må ivaretas.

«Forskning på anvendelser av informasjonsteknologi» er tittelen på et annet kapittel i utredningen. Utvalget hevder at studier av konsekvensene av den tiltakende anvendelsen av informasjonsteknologi må «bygge på en avklaring av en del grunnleggende problemer knyttet til *verdisyn*, *kunnskapsstruktur* og *formidlingsprosesser* i den informasjonsteknologiske kulturen.» (s. 15) Videre heter det at «Informasjonsteknologi kan formidle både *handlingskrav* og *bevissthetmessig påvirkning*. En avklaring av slike spørsmål har betydning for undersøkelser av hva innføring av informasjonsteknologi betyr for menneskers *handlingsmuligheter*, *selvoppfatning*, *moral* og *samfunnssyn*.» (s. 15)

Den økende anvendelsen av informasjonsteknologi reiser mange spørsmål som humanister kan bidra til å finne svar på. En avklaring av f.eks. hvorvidt IT innebærer mer instrumentell tenkning og større kulturell homogenisering, krever undersøkelser av informasjonsteknologiens sosiale og kulturelle egenskaper.

Informasjonsteknologien må også forstås som kulturelt fenomen. Hvorfor den gir opphav til språklige og kulturelle endringer, og hvilken betydning disse har for tradisjons- og kulturformidling, er foreløpig ubesvarte spørsmål.

Informasjonsteknologi medfører også endringer i *kulturproduksjon* og *-formidling*. «Nye lyd- og bildeformer fører til nye distribusjons- og bruksmønstre. Hva er nytt i disse mønstrene? Påvirker nye medier og nye produksjons- og distribusjonsformer innholdet i musikk, bilder og litteratur?» spør utvalget (s. 20).

Et eget kapittel i innstillingen er viet til bruk av informasjonsteknologi i de humanistiske, medisinske og samfunnsvitenskapelige fag. Etter utvalgets oppfatning vil de forskningsmessige utfordringene i denne sammenheng være avhengig av hva slags utstyr som blir tilgjengelig.

Utvalget antyder åtte hovedområder for anvendelser av informasjonsteknologi. Blant disse er simulering/modellutvikling (pågår allerede innenfor lingvistik), metoder og systemer for søking i tekst og analyse av tekstelementer (fins i mange humanistiske disipliner), bildebehandling/bildeanalyse (systemer/metoder for klassifisering, lagring og sortering av billedata), databaser (vil bli stadig viktigere i f.eks. arkiv- og museumsfaglig arbeid), framstillingsteknologi, og ekspertsystemer (til støtte ved forskningsarbeid).

Utvalget hevder også at «Bruk av informasjonsteknologi i de humanistiske, medisinske og samfunnsvitenskapelige fag bør også invitere til *kritisk selvrefleksjon*.» (s. 25). En av grunnene til dette er at IT kan innebære en kraftig omdanning av fagenes struktur eller innhold.

En fjerde kobling mellom informasjonsteknologi på den ene siden og humaniora og samfunnsfag på den andre, gjelder forskning på organiseringen av informasjonsteknologisk forskning og spredningen av denne teknologien. «Innholdet i informasjonsteknologien er basert på bestemte kunnskaps- og erkjennelsesstrukturer som har betydning for hva slags kunnskap som frambringes, og hvilke typer av løsninger som utvikles», heter det i innstillingen (s. 27). Aktuelle forskningsoppgaver omfatter bl.a. klarlegging av informasjonsteknologiens epistemologiske egenart, og drøfting av økonomiske, politiske og kulturelle forutsetninger for overføring av informasjonsteknologi.

Til sist i innstillingen foreslår koordineringsutvalget at deres arbeid blir videreført i form av et informasjonsteknologisk forskningsprogram, foreløpig for perioden 1987-1991. «Hovedsiktemålet med forskningsprogrammet skal være å styrke både forskningen, kompetanseutviklingen og

rekrutteringen på området.» (s. 32). Et slikt program må ha et bredt faglig perspektiv, og styres fortsatt av et faglig koordineringsutvalg. Utvalget gir i tillegg sin tilslutning til tanken om en felles handlingsplan for informasjonsteknologisk forskning som også omfatter RNF, og ber NAVFs styre overveie eventuelle konsekvenser for koordineringsutvalgets sammensetning og mandat.

Utvalget gir også en kortfattet redegjørelse for RHFes, RMFes og RSFes programmer for informasjonsteknologi. På bakgrunn av rådernes budsjettforslag, fremmer utvalget et forslag til totalbudsjett for perioden 1987-1991 på mellom 10.8 mill. og 14.5 mill. kr pr. år. Realiseringen av programmet etter oppbyggingsfasen vil være avhengig av at NAVF kan få ekstra midler til disposisjon som ledd i en nasjonal handlingsplan for informasjonsteknologi, eller via andre eksterne kilder.

Kristin Natvig

Humanistisk informasjonsteknologi under RHF

I forrige nummer av HD ble det gitt en omtale av RHFes innstilling om humanistisk informasjonsteknologisk forskning. Som meldt der, ønsket RHF å drøfte innstillingen med bl.a. representanter for de fire universiteter. Et møte om saken ble holdt i Oslo den 27.8.1986.

På møtet ble det gitt en orientering om RHFes initiativ i forbindelse med humanistisk informasjonsteknologisk forskning og om innstillingen. Det ble sagt at RHF ønsket å høre institusjonenes syn på innstillingen og at de forslag som ble fremmet ville bli brukt i det fortsatte plan- og budsjettarbeidet.

Institusjonene sa seg generelt godt fornøyd med den situasjonsbeskrivelsen som er gitt av relevante aktiviteter innen forskningsfeltet, men fant også grunn til å nyansere og utdype den omtalen som er gitt i innstillingen. Når det gjelder de konkrete oppfølgingstiltakene som ble foreslått i innstillingen, sa institusjonene seg generelt interessert i forfinansiering av amanuensisstillinger knyttet til informasjonsteknologiske emner. På den annen side så de samtidig formelle og økonomiske vansker med å kunne videreføre stillingene etter en programperiode betalt av RHF.

Flere av institusjonene understreket behovet for en sterk teknisk utstyrsopprustning i en første fase av en programinnsats. Behovet for studentstipend og koordinerte studietilbud ble også understreket. Det ble dessuten uttrykt velvillighet overfor forslag om en overordnet planlegging og koordinering av et program for humanistisk informasjonsteknologisk forskning.

På sitt møte 24.10.1986 har RHF vedtatt å opprette et eget program for humanistisk informasjonsteknologiforskning fra og med 1987. Det er oppnevnt en styringsgruppe for programmet bestående av:

Professor Helge Dyvik, Univ. i Bergen, formann

Professor Cathrine Fabricius Hansen, Univ. i Oslo

Professor Magne Dybvig, Univ. i Trondheim

Professor Knut Taraldsen, Univ. i Tromsø

Professor Jens E. Fenstad, Univ. i Oslo

Direktør Jostein H. Hauge, NAVFs EDB-senter for humanistisk forskning

Sekretær: førstekonsulent Tore Ellingsen, RHF

I første omgang skal gruppen lage et samlet forslag til Rådet om hvordan budsjettmidlene til tiltaket for 1987 (ca. 1.7 mill. kr) bør fordeles og disponeres med utgangspunkt i den fremlagte innstilling. Gruppen skal dessuten vurdere enkeltsøknader under programmet og gi en samlet innstilling til RHF.

Det vil i desember 86 bli foretatt utlysning av midler for humanistisk informasjonsteknologisk forskning for 1987.

Jostein H. Hauge

Konferanse om humanistisk forskning i dagens samfunn

Rådet for humanistisk forskning har siden 1983 arbeidd med et større utgreiingsarbeid for å få fram et vel dokumentert grunnlag for en debatt om forskningspolitiske og utdanningspolitiske spørsmål innenfor humanistisk forskning i Norge. Dette har resultert i tre delutredninger: (1) Ressurser, aktivitet, vilkår. (2) Eigenart, målsetjing og utviklingstendenser. (3) Humaniora i samfunnet. Den siste delutredninga kom høsten 1985. I forlengelsen av dette arbeidet har RHF laga ei utredning om «Humanistisk utdanning i Norge. Egenart. Søkning og yrkesmuligheter.» Konferansen i Oslo 3.-4. november 1986 var en oppfølging av dette omfattende utredningsarbeidet med hovedvekt på humaniorafagenes samfunnsmessige relevans.

Konferansen var delt i tre hovedseksjoner: (1) Hvor står den humanistiske forskningen i dagens forskningspolitiske bilde? (2) Humanistenes rolle innenfor kulturlivet – hva er den og hva bør den være? (3) Behovet for humanistisk kompetanse i nasjonalt og internasjonalt næringsliv.

Professor og ordfører i RHF *Atle Kittang* holdt innledningsforedrag

hvor han tok utgangspunkt i noen av dagens negative trekk for humanistisk forskning. Dette viser seg bl.a. ved at humanioras andel av de totale forskningsmidler går ned. Kittang la vekt på at humaniora ikke *bare* skal ha en tradisjonsformidlende og tilbakeskuende rolle. Humaniora forvalter også kunnskap av vital betydning for den løpende kunnskapsproduksjonen som er nødvendig for framtidens samfunn. Informasjonsteknologien er et eksempel på et sterkt voksende felt hvor sentrale humanistiske fag som filosofi og lingvistikk kan bidra med grunnleggende kunnskaper.

Kittang viste til at to høydepunkter i humaniorafagernes historie, renessansen og 1800-tallet, nettopp var karakterisert av at humaniorafagene var i fronten av samfunnsutviklinga. I renessansen skapte de et profant vitenskapsparadigme og på 1800-tallet var språk-, litteratur- og historievitenskapene en viktig del av den politiske moderniserings- og demokratiseringsprosessen.

Kittang pekte videre på den fornying som har skjedd innenfor humanistisk grunnforskning ved at forskerne har satt søkelyset på nye områder. Eksempler på dette er kvinneforskning, arbeidslivsforskning, medieforskning, samisk og kvensk språk, historie og kultur. Kittang mente at en nå stod ovenfor et problem når det gjaldt rekruttering både til forskerstillinger og til hovedfagsstudiet.

Professor *Vigdis Ystad*, som har vært leder for den siste humaniorautredninga, snakket om humanistenes rolle innenfor kulturlivet. I foredraget berørte Ystad utviklinga innen dagens massemedia hvor det stadig blir vanskeligere å formidle kulturstoff. Dagens tabloidaviser med dens kjappfotjournalistikk er heller ikke egne fora for kulturdebatten. Ystad var også bekymra for den voksende kløft mellom elitekultur og massekultur hvor vi mer og mer synes å miste et kulturelt fellesskap eller en felleskultur. I debatten som fulgte var en enige om at dagens humanister stort sett var fraværende i avisenes debatt- og kronikkspalter, men det var stor uenighet om hva grunnen til dette var. *Hans Fredrik Dahl* mente at ansvaret for dette måtte ligge hos humanistene selv. Han understreket at forsvaret for humaniora ikke måtte bygge på «kulturbekymring», som han mente låg bak Ystads resonnementer. Dahl var heller ikke enig i at vi nå har fått en kløft mellom elitekultur og massekultur.

Adm. dir. *Thorbjørn Ustaheim* A/S Ikovation holdt innledningsforedraget om behovet for humanistisk kompetanse i nasjonalt og internasjonalt næringsliv. Hans hovedkonklusjon var at næringslivet har behov for den spesialkompetanse humanistene sitter inne med. Men dette behovet er ikke overveldende stort. Ustaheim la mer vekt på at næringslivet trenger humanister fordi de representerer en annen erfarings- og utdanningsbakgrunn enn de som tradisjonelt har gått inn i næringslivet. Dagens næringsliv trenger m.a.o. humanister mer enn humanistisk kunnskap ifølge Ustaheim. Kravet om omstilling, fornyelse og innovasjon er

påtrengende i dagens næringsliv. Fornyelse kommer ofte utenfra, hvorfor ikke fra humanistene?

Kjøp humaniora!

I en kommentar til Ustaheims foredrag la utredningsleder *Otto Hauglin* fra NAVFs utredningsinstitutt fram et interessant tallmateriale når det gjaldt behovet for humanistisk kompetanse i 1990-åra. Det viser seg at den tilgangssituasjonen vi i dag har på humanister ikke kan fylle det erstatningsbehovet vi vil få i jobber som allerede i dag finnes for humanister, fram mot år 2000. Disse jobbene befinner seg først og fremst i undervisningssektoren. Hans entydige råd til dagens 15-åringer var derfor: Kjøp humaniora!

Einar-Arne Drivenes

MELDINGER

Senterrapport nr. 38: Edb-prosjektet for kunst- og kulturhistoriske museer

Edb-prosjektet for kunst- og kulturhistoriske museer ble utført innenfor rammene av et to-årig forskerstipend i regi av RHF. *Jon Birger Østby* har vært engasjert i stillingen, som var administrativt knyttet til Senteret. Østbys sluttrapport fra dette prosjektet er nylig utgitt i Senterets rapportserie.

I rapporten gjør Østby rede for prosjektets bakgrunn og målsetning, og for status for museenes edb-arbeid ved prosjektstart. Resultatene av arbeidsoppgavene knyttet til prosjektet – vurdering av egnet utstyr og programvare for databehandling ved museene, og opplegg for registrering av gjenstander, fotos og billedkunst – blir gjennomgått. Blant øvrige emner som tas opp i rapporten, er erfaringer med en prøvedatabase for gjenstander fra flere museer, og organiseringen av framtidige datatjenester for museene.

Rapporten inneholder også diverse vedlegg, inkludert veiledninger for registrering av gjenstander og fotos på edb, en søknad til Kultur- og vitenskapsdepartementet om tilskudd til etablering og drift av en «Museenes datatjeneste», og en oversikt over edb-utstyr ved kunst- og kulturhistoriske museer.

Rapporten, som er på totalt 80 sider, koster kr 60 (inkl. porto) og kan bestilles fra Senteret.

En kortfattet rapport fra prosjektet ble gitt i *Humanistiske Data* 2-86.

Programvare-omtale

Tidsskriftet *Computers and the Humanities* har ofra stor spalteplass på å omtale aktuell programvare.

Nr. 1-86 omtalar såleis grundig to tekstbehandlingsprogram som er veileigna for tyngre tekstproduksjon (avhandlingar, rapportar o.l.):

T³ («Multilingual word processing system», TCI Software Research, \$595) og

NotaBene («Academic word processing and text retrieval system», Dragonfly Software, \$495)

Nr. 2-86 omtalar eit dusin nokså ulike program:

Smooth Talker («Disk-based speech synthesizer for the Macintosh», First Byte Software, \$100)

Fancy Font («Text formatter, pronter enhancement», SoftCraft, \$180)

E-Z-DOS-IT («Concurrent operating system shell for MS/PC-DOS», Hammar Computer Systems, \$60)

Proportional Star («Daisywheel print enhancer for Wordstar», Writing Consultants, \$75)

Arts and Humanities Search («Online bibliographic data base», Institute for Scientific Information)

BRS/BRKTHRU, («Online bibliographic literature retrieval service», BRS Information Technologies, \$75 + \$17,50-\$125/time)

Word Perfect («for Apple II», SSI Software, \$180)

Expert Choice (Decision Support Software, \$498)

Oxford Concordance Program (Oxford University Computing Service, \$200)

Vocalink («Speech Recognition System», Interstate Voice Products, \$1600)

Thor («Thought organizer», Fastware Inc, \$295)

Ascii Pro (Communications program, United Software Industries, \$85)

Computers and the Humanities blir utgitt kvartalsvis og koster \$81 pr. år. Adresse: *Paradigm Press, Inc., P.O. Box 1057, Osprey, Florida 33559, USA.*

Undervisning i edb for humanister ved University of Southampton

Sebastian Rahtz, lektor i edb for humanister, har skrevet en rapport med tittelen *Humanities Computing Teaching in the Department of Electronics and Computer Science, University of Southampton*. I rapporten presenteres i detalj både det generelle edb-kurset for humanister som ble holdt i undervisningsåret 1985-86 og kursene som holdes i nåværende undervisningsår. I tillegg omtales et nytt B.A.-studium i *Modern Languages with Computing*, som omfatter undervisning i ett fremmedspråk, edb, lingvistikk og en del filosofi.

Bruken av edb i undervisningen av humanistiske fag øker stadig ved University of Southampton. Instituttene for arkeologi og historie har i flere år hatt kurs i kvantitative edb-metoder. Ved «The Language Centre» anvendes mikromaskiner i språkundervisning og planer foreligger

for å integrere edb i de tradisjonelle språklaboratoriene. Ved instituttet for musikk får studentene instruksjon i edb-assistert komposisjon. Planer er lagt for en utvidelse av edb-bruken innenfor instituttene for moderne språk, og et av de to planlagte prosjektene ved instituttet for historie går ut på å utvikle et ekspertsystem rettet mot studenters forståelse av originalkilder.

I rapporten legges det stor vekt på både de positive og negative erfaringene med fjorårets generelle edb-kurs, og studentenes respons på kurset. Årets kurstilbud er lagt opp med bakgrunn i disse erfaringene. Arkeologi-studenter får flere skreddersydde tilbud. Nye humaniora-studenter tilbys et innføringskurs i edb som dekker tekstbehandling, databaser og generelle edb-hjelpemidler. Viderekomne studenter kan ta et tre-semesterst fortsettelseskurs som tilbyr bl.a. programmering i Prolog. Viderekomne litteratur-studenter kan ta et eget kurs som heter «Literary Computing».

Rapporten inneholder også resultatene av en spørreundersøkelse foretatt blant det vitenskapelige personalet ved Faculty of Arts om deres bruk av edb.

Rapporten gir et nyansert bilde av det høye aktivitetsnivået innenfor edb-undervisning for humanister ved University of Southampton. Opplysningene og erfaringene som presenteres er ikke bare interessante i seg selv – de har også relevans for norske fagmiljøer. Rapporten kan fås ved henvendelse til: *Sebastian Rahtz, Department of Electronics and Computer Science, University of Southampton, Southampton, England.*

National Centre for Computer Assisted Language Learning (NCCALL)

NCCALL ble opprettet i april ved School of Language Studies, Ealing College of Higher Education i England. NCCALLs aktiviteter er rettet først og fremst mot undervisning på høyskolenivå i engelsk, fransk, tysk, spansk og russisk. Disse aktivitetene omfatter: innsamling og formidling av informasjon om data-assistert språkundervisning, evaluering og utvikling av programvare, opplæring av personale og samarbeid med andre institusjoner.

Hittil har NCCALL foretatt en spørreundersøkelse blant undervisningspersonale ved høyskoler i England om bruk av edb i språkopplæring. I tillegg gir senteret ut meldingsbladet CALLBOARD. Både bladet og mer informasjon om senteret kan fås ved henvendelse til NCCALL ved nedenstående adresse.

NCCALL er også i ferd med å bygge opp både et programvarebibliotek

og en samling av relevante bøker og artikler for forskningsformål. Dessuten er et skjema for programvare-evaluering utviklet, som kan fås ved henvendelse. Utfylte skjemaer vil være allment tilgjengelige.

Adresse: *Graham Davies, NCCALL, Ealing College of Higher Education, Grove House, 1 The Grove, Ealing, London W5 5DX, England.*

CMR

Center for Music Research

Center for Music Research

The Center for Music Research (CMR) ble opprettet i 1980 ved School of Music, Florida State University. Senterets formål er å bruke avansert edb-utstyr i forskning i musikk, både som kunstform og undervisningsfag.

Lærere og studenter ved senteret har adgang til et system for utvikling av musikk-programmer som består av seks kraftige arbeidsstasjoner av merket Sun koblet sammen i et nettverk, og et klasserom med 25 mikromaskiner koblet til nettverket. De fleste av mikromaskinene kan arbeide som synthesizere. Av øvrig utstyr kan nevnes et laboratorium for spesialbygging av elektroniske kretser (for sanntids databehandling), og to studioer for komposisjon av elektronisk musikk med til sammen fire synthesizere (inkludert to 16-stemmers digitale).

CMR tilbyr et ett-årig grunnprogram i edb i musikk som består av åtte kurs, ved siden av mer forskningsorienterte kurs og et eget program for personer som har doktorgrad i musikk.

Flere opplysninger kan fås fra: *Center for Music Research, Florida State University, School of Music, Tallahassee, Florida 32306-2098, USA.*

Humanities Research Center, Washington State University

Et ressurs- og referansesenter for forskning i humanistiske fag drives av English Department ved Washington State University i USA. Senteret tilbyr utstyr for tekstbehandling og mikrofilm, referansemateriale, og terminaler og skrivere for bruk av edb-baserte bibliografier og tekstredigeringsystemer som er tilgjengelige ved universitetet. Hovedfagsstuderenter har anledning til å anvende edb i litteraturstudier på følgende områder: tekstredigering, utarbeidelse av konkordanser og bibliografier, og diktanalyse.

Mer informasjon kan fås fra: *Professor John Elwood, Chairman, Department of English, Washington State University, Pullman, Washington 99164-5020, USA.*

HumaNet

HumaNet er et nyopprettet, internasjonalt on-line nettverk for humanister. Dette nettverket utgjør en del av ScholarNet, som har hovedkvarter ved North Carolina State University i USA.

Abonnenter på HumaNet tilbys bl.a. elektronisk post, on-line versjoner av meldingsblad og tidsskrifter, inngang til databaser (inkl. DIALOG), mulighet for overføring av tekster og datafiler, gratis programvare, mulighet for telekonferanser og adgang til elektroniske oppslags-tavler. HumaNet er foreløpig rettet først og fremst mot fagområdene historie, engelsk, filosofi og religion, men spekteret av både fag og tjenester er stadig under utbygging.

Medlemskap i HumaNet koster \$29.95 for enkeltpersoner, \$100 for institutter/organisasjoner. Søkeprisen er enten \$6 eller \$16 pr. time, avhengig av tidspunkt. Brukere betaler kun pr. minutt søking. Norske abonnenter kan få adgang til HumaNet via Datapak. Flere opplysninger kan fås fra: *Richard W. Slatta, ScholarNet Director, North Carolina State University, Box 8101, Raleigh, North Carolina 27695-8101, USA.*

Association for the Development of Computer-based Instructional Systems

ADCIS er en internasjonal organisasjon som har som hovedformål å fremme forskning i og bruk av edb-basert undervisning, med vekt på

ADCIS



Association for the
Development of
Computer-Based
Instructional Systems

utveksling av informasjon. Medlemsmassen består av lærere i undervisningsinstitusjoner av alle slag og på alle nivåer, i tillegg til administrativt personale, edb-eksperter, forskere og andre som arbeider med undervisningsteknologi.

Blant ADCIS' aktiviteter er en årlig konferanse (se separat omtale i dette nr. av HD), og en rekke publikasjoner. Disse omfatter bl.a. *Journal of Computer-Based Instruction*, som består av faglige artikler og bokanmeldelser, *ADCIS News*, et medlemsblad som utgis annenhver måned, og årlige «Conference Proceedings». «ADCIS-Net» er et elektronisk nettverk som brukes bl.a. til periodiske on-line konferanser.

14 ulike «Special Interest Groups» drives i regi av ADCIS, hver med egne meldingsblad. Emnene for noen av disse er edb-anvendelser i grunn- og videregående skoler, edb-anvendelser for funksjonshemmede, audiovisuelle undervisningssystemer, edb-basert musikkundervisning, bruk av forfattersystemet PLATO, og teori om edb-basert undervisning.

De ulike medlemskapskategorier i ADCIS koster mellom \$55 og \$100. En fyldig brosjyre om ADCIS fås ved henvendelse til: *ADCIS International Headquarters, 409 Miller Hall, Western Washington University, Bellingham, Washington 98225, USA.*

Nytt i biblioteket

Annual meeting of the Association for Computational Linguistics, 23.

8-12 July 1985, Chicago. Proceedings. Association for Computational Linguistics, 1985. 332 s.

Artwick, Bruce A.: Microcomputer displays, graphics and animation. Englewood Cliffs, N.J.: Prentice-Hall, 1985. 374 s.

Bjaanes, Michael: Bli kjent med Kmann. Oslo: Gyldendal, 1985. 120 s.

Brown, Douglas L.: From PASCAL to C. An introduction to the C programming language. Belmont, California: Wadsworth, 1985. 153 s.

Information Resources Management. 6. nordiska IoD-konferansen, 19.-22.8. 1985, Helsingfors. Samfundet för Informationstjänst i Finland. Helsinki: 1985. 457 s.

James, Mike: PASCAL. Programmering på mikrodatamaskiner. Rud: NKI-forlaget, 1985. 213 s.

- Light, Richard B./Roberts, D. Andrew/Stewart, Jennifer D.: Museum documentation systems: development and applications. Museum Documentation Association. London: Butterworth, 1986. 332 s.
- Opphavsrett og EDB. Delinnstilling III fra et utvalg oppnevnt ved kongelig resolusjon 21. januar 1977. Oslo: Universitetsforlaget, 1986. 37 s.
- Presno, Odd de: Amatørens håndbok om datakommunikasjon. Oslo: TANO, 1986. 152 s.
- Samordnad dokumentation av museernas föremåls och bildsamlingar. Statens kulturråd, Stockholm: 1986. 83 s.
- Vliet, J.C. van.: Text processing and document manipulation: proceedings of the international conference, University of Nottingham, 14-16 April 1986. Cambridge, Cambridge U.P.: 1986. 277 s.
- Wolverton, Van: Slik virker MS DOS. Oslo: Tano, 1986. 396 s.
- Wood, Steve: Using Turbo Pascal. Berkeley, California, 1986. 304 s.

Konferanser

CAA '87 - foredragsinnbydelse

Konferansen «Computing Applications in Archaeology» skal holdes ved University of Leicester 27.-29. mars 1987. Første dag blir viet til «tutorials» og «workshops» om edb-grafikk i arkeologi. Foredrag og demonstrasjoner blir gitt de andre to dagene.

De som ønsker å holde foredrag/demonstrasjoner på konferansen, kan sende et kort sammendrag/spesifikasjoner til: *Clive Ruggles, Computing Studies Dept., University of Leicester, University Road, Leicester LE1 7RH, England.* Påmeldingsskjema til konferansen fås ved henvendelse til samme adresse.

Third Conference of the European Chapter of the Association of Computational Linguistics

Denne konferansen - den tredje i en serie som holdes annethvert år - skal finne sted 1.-3. april 1987 i København. Programmet kommer til å dekke bl.a. følgende emner: morfologi; leksikalsk semantikk, datamaskinelle modeller for analyse og generering av språk, både muntlig og skriftlig; datamaskinell leksikografi og leksikologi; syntaks og semantikk; maskinoversettelse og oversettelseshjelpemidler; naturlig språk grensesnitt, kunnskapsrepresentasjon og ekspertsystemer. Demonstrasjoner og en utstilling er også planlagt.

Flere opplysninger fra: *Bente Maegaard, Institut for Anvendt og Matematisk Lingvistik, Københavns Universitet, Njalsgade 96, DK-2300*

København S, Danmark.

CALICO International '87

CALICO (Computer Assisted Language Learning & Instruction Consortium) skal arrangere sitt fjerde årlige symposium 6.-10. april 1987 i Monterey, California. Med dette symposiet vil CALICO skape et forum for presentasjon og diskusjon av edb-teknologi med tanke på å tilpasse teknologien til en mer effektiv undervisning i og læring og behandling av språk. Symposiet er åpent for lærere på alle nivåer, edb-folk, forskere og leverandører av maskin- og programvare.

Programmet er ikke fastlagt ennå, men vil bestå av en kombinasjon av foredrag, demonstrasjoner, paneldiskusjoner og «workshops». Flere opplysninger fra: *CALICO Symposium - 1987, 3078 JKHB, Brigham Young University, Provo, Utah 84602, USA.*

ICCH '87

Den neste International Conference on Computers in the Humanities (ICCH '87) skal holdes 10.-12. april ved University of Carolina. Flere opplysninger fra: *Bob Oakman, Department of Computer Science, University of South Carolina, Columbia, South Carolina 29208, USA.*

Use of Computers in Language Teaching

The School of Modern Languages and Culture ved Hull University arrangerer denne konferansen 22.-23. april 1987. Formålet er å gi universitetsansatte i lingvistikk og moderne språk en innføring i potensialet til data-assisterte metoder. Programmet skal omfatte egne sesjoner for både nybegynnere og mer avanserte edb-brukere, demonstrasjoner av programvare og en utstilling over læremidler.

For mer informasjon, kontakt: *Graham Chesters, Computers in Teaching Conference, Department of French, The University, Hull HU6 7RX, England.*

ICAME 8th

Den 8. internasjonale ICAME Conference on English Language Research on Computerized Corpora skal holdes i Helsinki 21.-24. mai. I tillegg til foredrag, skal programmet omfatte en fortsettelse av en «workshop» holdt ved fjorårets konferanse om «Coordination and Exchange of Programs», og en demonstrasjon av arbeid ved Research Unit for Computational Linguistics ved universitetet i Helsinki.

Det endelige programmet fås ved henvendelse til: *Merja Kytö, Department of English, University of Helsinki, Hallituskatu 11-13, 00100 Helsinki 10, Finland.*

Fourteenth ALLC Conference

Den 14. internasjonale ALLC konferansen om edb i språk- og litteraturforskning skal holdes ved Göteborgs universitet 1.-5. juni 1987. Konferansens hovedtema blir lingvistiske databaser. I tillegg til foredrag skal det holdes paneldiskusjoner, demonstrasjoner av ferdige systemer, m.m.

Påmelding innen 1. april til: *Gudrun Magnusdottir, Dept. of Computational Linguistics, University of Gothenburg, S-412 98 Gothenburg, Sweden.*

Nordisk konferanse om datastøttet undervisning i universitets- og høyskoleutdanningen

Denne konferansen arrangeres av Kungliga Tekniska Högskolan (KTH) i Stockholm 2.-5. juni 1987. Målet for konferansen er å samle ansatte ved nordiske universiteter og høyskoler hvor forsøksvirksomhet med datastøttet undervisning forekommer eller planlegges. Hensikten er å gi deltakerne et noenlunde samlet bilde av forsknings- og utviklingsarbeidet innenfor feltet og forberede muligheter for et nordisk samarbeid.

Konferansen kommer til å fokusere på pågående og planlagt forskning og utviklingsarbeid. Programmet fås fra: *Nordisk konferens om datorstødd undervisning, Göran Karlsson, Inst. f. mekanik, Kungliga Tekniska Högskolan, S-100 44 Stockholm, Sverige.*

25th Annual Meeting of the Association of Computational Linguistics

ACL 1987 skal finne sted 6.-9. juli ved Stanford University, Stanford, California. Arrangørene søker foredrag om alle aspekter ved datalingvistik, både teoretisk og anvendt. Bidrag må være arrangørene i hende innen 12. januar.

Foruten foredrag skal konferanseprogrammet bestå av såkalte «applied tutorials», utstillinger og demonstrasjoner. Adresse: *Candy Sidner, ACL-87 Program Chair, BBN Laboratories Inc., 10 Moulton Street, Cambridge, MA 02238, USA.*

Før og etter konferansen skal den 54. LSA Summer Linguistic Institute holdes ved Stanford University, hvor hovedtemaet også skal være datalingvistik. Programmet omfatter bl.a. 10 8-timers kurs for ACL-

medlemmer. Flere opplysninger fra: *Ivan Sag, Department of Linguistics, Stanford University, Stanford, California 94305, USA.*

Uka før ACL-konferansen (30.6.-3.7.) arrangeres 9 ulike intensivkurs i datalingvistikkk ved Stanford University. Pris: \$125 ved påmelding før 1.3., ellers \$175. Kursene inngår i det omfattende programmet til sommerskolen «1987 Linguistic Institute». Påmelding og flere opplysninger om sommerskolen ved henvendelse til: *1987 Linguistic Institute, Department of Linguistics, Stanford University, Stanford, CA 94305, USA.*

ICDBHSS '87

The International Conference on Data Bases in the Humanities and Social Sciences skal finne sted i Montgomery, Alabama 11.-13. juli 1987. Programmet skal bestå av ca. 100 foredrag à 10-20 min. i fem parallelle sesjoner, i tillegg til tre «keynote addresses».

Deltakeravgiften kommer på ca. \$180 (inkl. kost og losji). 30-40% rabatt på innenlandske flyruter kan oppnås. Flere opplysninger fås fra: *Dr. Lawrence J. McCrank, Dean, AUM Library and Resource Center, Auburn University at Montgomery, Montgomery, Alabama 36193-0401, USA.*

ADCIS - 29th International Conference

Call for Participation

Den 29. konferansen til Association for the Development of Computer-based Instructional Systems skal holdes i Oakland, California 9.-12. november 1987. Konferansen vil gi nybegynnere og eksperter i edb-basert undervisning anledning til å utveksle ideer, erfaringer og forskningsresultater når det gjelder pedagogisk programvare, utdanningsstrategier osv. Arrangørene søker foredrag som fokuserer på emner innenfor ADCIS' 14 «Special Interest Groups» (se egen melding om ADCIS i dette nr. av HD).

I tillegg til foredrag, vil konferansen omfatte demonstrasjoner av utstyr og programvare, paneldiskusjoner, en utstilling m.m. Bidrag søkes til alle disse aktivitetene.

Påmelding av foredrag og andre typer presentasjoner må være arrangørene i hende innen 1. mars 1987. Flere opplysninger fås fra: ADCIS Headquarters, 409 Miller Hall, Western Washington University, Bellingham, Washington 98225, USA.

SUMMARY

Datakommunikasjon

Computer communication

Senior Computing Officer Espen S. Ore of the Centre reviews the possibilities a Norwegian humanities scholar has for computer communication. Ore starts out by outlining computer equipment necessary for communication and the Norwegian telephone company's computer networks. Use of communication programs, in particular the American program Kermit, is accounted for, as well as IBM's European Academic and Research Network (EARN) and the American Defense Department's ARPANET. In conclusion, electronic mail systems and Norwegian and Swedish teleconferencing services are dealt with.

Humanistiske databaser

Humanities data bases

Information Officer Kristin Natvig of the Centre gives an overview of some on-line humanities data bases in Scandinavia, France, and USA.

In Norway there are data bases on humanistic research projects and connected publications (run by the Research Council); references to literature on Lappish culture, language, and history; references to university library holdings of women's literature, an archive of the labour movement, judaica, music manuscripts, on Norwegian railway and music history, and private archives; and references to graduate theses in the humanities at the University of Oslo.

Swedish data bases encompass data on demography and social history; a register of complete descriptions of museum artefacts; and references to literature on art published by Swedish scholars.

In Denmark a full text data base has been established of archaeological finds from the Stone Age and onwards.

Humanities data bases run by the American DIALOG and the Research Libraries Information Network (RLN) are presented, along with the French information system FRANCIS's bibliographical data bases in the humanities. FRANCIS covers Latin American topics, art, archaeology, antique law, ethnology, literature, religion, philosophy, ancient history, and linguistics.

Edb i samfunnsfag

Computing and social studies

The author of this article, Harald Johansen, teaches geography and social studies at one of the high schools selected to take part in government-funded trial projects on computing in education. Johansen outlines work being carried out at

these schools in connection with computing and social studies. He also discusses future trends in this field concerning the development of courseware, telecommunications, and mass storage media. In addition, the consequences of information technology for the discipline are analysed. One of the examples given is how direct access to historical data can contribute to dispelling myths about pre-industrial society.

Johansen points out that information technology also poses problems to teachers of social studies. These teachers must acquire computing expertise, reassess pedagogical methods, and acquaint themselves with the organisation, strengths, and weaknesses of computer-stored data such as church records and censuses.

nota bene

Research Fellow Torrey Seland, Department of Religion, University of Trondheim, reviews the program package «nota bene», which has been designed especially for humanities scholars. «nota bene» incorporates a word processing program, a text retrieval program that can be operated in a similar way as a data base, and program functions for mail merge and the design of forms.

Seland describes the various features of «nota bene» and discusses how they function in practice. «nota bene» is menu-driven and allows work to be carried out on two documents at the same time. The text retrieval program can be used simultaneously with the word processing program. When using this program package, the production of indexes and bibliographies is radically simplified.

In Seland's experience, «nota bene» is a useful program package, despite some (minor) weaknesses.

Senterets videoplateprosjekt – en statusrapport

The Centre's videodisc project – a status report

Information Officer Kristin Natvig of the Centre reports that the first Norwegian videodisc of humanistic primary sources was produced by the Centre in June. The videodisc project was started in the autumn of 1984 and is now near completion.

The main goal of the project is to investigate how videodisc technology can increase the efficiency of retrieval, use and mediation of large amounts of humanistic primary sources. The videodisc contains a representative selection of such sources – photographs, slides, film and video sequences, and manuscripts – from 19 humanistic institutions, mainly museums, archives and university departments. Connected picture data bases and software are currently being developed.

The Centre's equipment consists of a videodisc player controlled by a PC. All software has been written for two screens.

The data base programs have been designed mainly for traditional retrieval of reference data. When interesting data have been retrieved, the picture(s) that belong to the dataset can be displayed on the videodisc player's screen. Some programs allow «leafing through» the pictures without prior retrieval of reference data.

Software has also been developed for communication with data bases in other computers. This solution was chosen for cases where the amount of reference data is too large to be stored on a PC.

In addition, a general program for «roaming» in the pictorial material has been written. The user may «skip» from picture to picture along different «routes», chosen by using menus.

Datasekretariatet og humanistiske fag

The Computer Secretariat and the humanities

Information Officer Kristin Natvig of the Centre reports that the Computer Secretariat was established in 1984 as a temporary, government-funded body in order to administer trial projects on computing in schools. The Computer Secretariat supports projects at 21 schools over the whole country. At most of these schools educational software is currently being developed and tested in humanities subjects, including Norwegian, foreign languages, history, and art.

The Computer Secretariat wishes to take more initiatives to stimulate the use of computing in the teaching of humanities. Among future plans are a course on the development of educational software for teaching art, and a conference on computer equipment adapted to the Lappish alphabet.

«Norsk stadnamnbase»

Norwegian data base of place names

Lecturer Botolv Helleland, Department of Name Studies, University of Oslo, starts this report with an outline of the work carried out since 1973 in connection with data processing of names. Several institutions carry out data entry of place names, but up to now the organisation and functions of a national data base for place names has not been agreed upon.

These issues were discussed at a conference held in April, and are being analysed further by a committee appointed at the conference. The committee has also drafted a joint form for data entry of place names that can be tailored to individual needs while at the same time forming a basis for a functional data base.

IBM-seminar om taleprosessering

IBM seminar on speech processing

Lecturer Kolbjørn Slethei, Department of Phonetics and Linguistics, University of Bergen, reports on a seminar on speech processing arranged by IBM in June in Oberlech, Austria.

The three main fields within speech processing - coding, recognition, and synthesis - were dealt with from different angles. The main emphasis was on speech recognition.

Very few presentations were on a merely theoretical level. Some speakers presented computational methods that simulate «speech recognition», others reported on processes and components within product development. Slethei gives details of some of the presentations.

Kurs i design av pedagogisk programvare

Course in the design of educational software

This Nordic course was arranged in July-August in Uddevalla, Sweden, for 90 participants. The main speaker was Les Green from Canada, who has worked out the so-called «marketplace model» for the design of educational software.

Senior Computing Officer Per Vestbøstad of the Centre took part in this course. In an interview with Information Officer Kristin Natvig, Vestbøstad explains that the «marketplace model» allows pupils to choose exercises themselves and the length of time spent on each exercise.

Together with another participant of the course, Vestbøstad designed a model for vocabulary training in German. In Vestbøstad's view, the «marketplace model» is well-suited to this type of program, and can be applied to other foreign languages.

Demonstrations were arranged of both software developed by the participants and commercial authoring systems. «Vevplan», a program for the design and analysis of woven patterns, was also shown.

Datasekretariatets seminar i Halden

The Computing Secretariat's seminar in Halden

Senior Computing Officer Espen S. Ore of the Centre reports on a seminar arranged in Halden in September by the Norwegian Computing Secretariat, a government-funded support service for computing in schools.

At the seminar a discussion took place of the computers used in projects supported by the secretariat. Two projects on developing programming tools were also shown.

Humanistiske fag og kunnskapsbaserte systemer

The humanities and knowledge-based systems

In September the Norwegian Computing Centre for the Humanities arranged a Nordic symposium on knowledge representation. Erik Fjornes of the Centre reports that 28 representatives of the fields of philosophy, logic, linguistics, cultural anthropology, and computer science took part in the symposium.

The first day dealt with knowledge representation in general and the expert system concept in particular. The theme of the second day was systems for processing natural languages.

The participants agreed that many of the problems in the field of knowledge-based systems are of a traditional humanistic character.

«Kommunikasjons- og datateknologi som infrastruktur i nord-norsk næringsliv og forvaltning»

Communication and computing technology as an infrastructure in Northern Norwegian business and administration

This is the title of a conference arranged by the University of Tromsø in

September. Fredrik Fagertun, Manager of the Norwegian Historical Data Archives (RHD), provides summaries of many of the papers given at the conference. Information technology in a political and regional perspective was the theme of one of the sessions. In another session, examples were given of applications of information technology in various sectors, including RHD. The University of Tromsø presented a selection of research activities connected to information technology, including plans for research in computational linguistics.

One of the issues discussed in a session on computing education, was why girls don't aim at careers in technology. The closing session dealt with the establishment of businesses in the information and communication sector in Northern Norway.

Nordisk seminar om maskinoversettelse

Nordic seminar on machine translation

This seminar took place in Copenhagen in October. Director Jostein H. Hauge of the Centre reports that the aim of the seminar was twofold: To present Nordic translation-related projects as a point of departure for a discussion of cooperative ventures in machine translation, and to give a status report of the Danish Eurotra project.

On-going work within or related to machine translation presented at the seminar included two systems currently being developed at the University of Bergen for translation from Norwegian to English. In addition, a series of accounts were given of work within computational linguistics relevant to machine translation.

The aim of the Eurotra project is to create a system that can translate texts between all of the EEC member countries' languages. The Danish Eurotra project employs 13 people who have designed a modular translation system from Danish to English that can be applied to all languages. In the course of the seminar the Danish Eurotra staff presented the various aspects of this project.

Nytt fra RHF/NAVF

News from the Council for Research in the Humanities

In June a committee appointed by the Norwegian Research Council for Science and the Humanities presented a proposal on the deposit, storage and re-use of data from Council-funded projects. Recipients of Council funds are obliged to hand over research data to the Council. In their report the committee suggests the establishment of a computer-based system for storage and re-use of such data. This system should be located at the Norwegian Computing Centre for the Humanities and the Norwegian Social Science Data Service. The Centre should store data from projects within literature, linguistics, cultural disciplines, and pre-20th century historical data related to individuals.

In August a proposal appeared on the need for scientific equipment within the humanities. In this proposal a committee appointed by the Council for Research in the Humanities documents humanities researchers' concern about lacking funds for the purchase and management of equipment. In a general analysis of

equipment for the humanities, the committee points out that the gap between the need for equipment and the funds provided has increased in the past years.

According to the committee, basic equipment for each researcher should consist of a PC with a word processing program and a printer. Expensive equipment such as laser printers should be shared facilities. A high-quality optical character reader should be purchased and used as a joint national resource.

In the proposal a detailed overview is given of the status and need for research equipment within each discipline. The committee also assesses the total need for equipment within the next three years, and discusses the Council's policy concerning grants for equipment.

Another proposal appeared in August, on the challenges of information technology to the humanities, social sciences, and medicine. A coordinating committee with representatives of the Councils for the three disciplines is responsible for this proposal.

The committee points out that the growth of information technology has formed the basis for new relations between these three disciplines and technological disciplines. The humanities, social sciences, and medicine can give insight into the development of information technology, e.g. in solving problems related to human interaction and knowledge. In addition, studies of the applications of information technology and their consequences raise problems that the humanities can contribute to solve.

The use of information technology in the humanities, social sciences and medicine is also discussed. An outline is given of research tasks within these three disciplines in connection with studies of research in information technology and the spread of this technology.

The coordinating committee suggests that their work be continued in the form of a 5-year program for research in information technology. An account is also given of each of the three Councils' own programs for information technology.

The Council will establish a program for humanistic information technology from 1987. Researchers may apply to the Council for funding of projects and equipment within this field.

Meldinger

News

The Centre has recently published the final report on the research project «Computing methods for museums of art and cultural history.» This project was carried out within the framework of a two-year research fellowship held by Jon Birger Østby, who worked in close cooperation with the Centre.

Sebastian Rahtz has written a highly interesting report called *Humanities Computing Teaching in the Department of Electronics and Computer Science, University of Southampton*. In addition to giving an account of the use of computing in the teaching of humanities at the University of Southampton, Rahtz discusses the various types of computing courses offered to humanities students.

The National Centre for Computer Assisted Language Learning (NCCALL) was established in April at the School of Language Studies, Ealing College of

Higher Education, England. NCCALL's activities are aimed primarily at the teaching of English, French, German, Spanish, and Russian to college students. NCCALL collects and disseminates information on computer-assisted language teaching, evaluates and develops software, and trains teaching staff. The centre also publishes the newsletter CALLBOARD.

The Center for Music Research (CMR) was established in 1980 at the School of Music, Florida State University. The centre's objective is to employ advanced computing equipment in music research, both as an art form and an educational discipline. CMR offers various types of courses and programs and an impressive selection of computing facilities to researchers and students alike.

A well-equipped resource and reference centre for research in the humanities is run by the Department of English, Washington State University, USA. The centre's computing facilities are available to both researchers and graduate students.

HumaNet is a recently established, international on-line network for humanities scholars, primarily within the fields of history, English, philosophy and religion. This network forms part of ScholarNet, which is based at North Carolina State University, USA. Subscribers to HumaNet have access to a number of on-line services.

The Association for the Development of Computer-based Instructional Systems (ADCIS) is an international organisation, the main objective of which is to promote research in and use of computer-based instruction. ADCIS arranges an annual conference, issues various types of publications, and runs the on-line network «ADCIS-Net» in addition to 14 different «Special Interest Groups», each with its own newsletter.

Forts. fra 2. omslagsside.

nr. 31 STAR III Archaeology for statisticians. ISBN 82-7283-035-3
Pris kr. 60.

nr. 32 STAR IV STAR algorithms. ISBN 82-7283-036-1 Pris kr. 30.

RAPPORT nr. 33. *Årsmelding 1983*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-038-8 Gratis.

RAPPORT nr. 34. *Jostein H. Hauge: Tutorial on Machine Translation*. Rapport fra en konferanse i Lugano 2.-6. april 1984. ISBN 82-7283-039-6. Pris kr. 60.

RAPPORT nr. 35. *Ole Lauvskar: Bruk av statistiske metoder i språk- og litteraturforskninga*. Rapport frå ei spørjeundersøking. September 1984. ISBN 82-7283-041-8 Pris kr. 50.

RAPPORT nr. 36. *Årsmelding 1984*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-042-6. Gratis.

RAPPORT nr. 37. *Årsmelding 1985*. NAVFs EDB-senter for humanistisk forskning. ISBN 82-7283-043-4. Gratis.

RAPPORT nr. 38. *Jon Birger Østby: Edb-prosjektet for kunst- og kulturhistoriske museer*. Oktober 1986. ISBN 82-7283-045-0 Pris. kr. 60.

C

institutt
itetet i Bergen
gen-Universitetet

Av innholdet:

Espen S. Ore: Datakommunikasjon

Kristin Natvig: Humanistiske databaser

Harald Johansen: Edb i samfunnsfag

Stig Johansson: Machine-Readable Texts in
English Language Research

Claus Huitfeldt: Computers and Philosophical
Manuscripts

Returadresse:
NAVFs EDB-senter for humanistisk forskning
Boks 53
5014 Bergen – Universitetet