



**humanistiske
data 2-82**

NAVF

**ARTIKLER
RAPPORTER
MELDINGER
SUMMARY**

**NAVF's EDB-senter
for humanistisk
forskning**

**The Norwegian
Computing Centre for
the Humanities.**

SENTERETS RAPPORTSERIE

RAPPORT nr. 1. *EDB i gjenstandsfagene*. Rapport fra en konferanse i Bergen, 18. og 19. april 1978. September 1978. 2. opptrykk november 1981. ISBN-82-7283-022-1 Pris kr. 40.

RAPPORT nr. 2. *Et norsk datamaskinelt tekstkorpus*. Rapport fra en konferanse i Bergen, 19. og 20. oktober 1978. Februar 1979. 2. opptrykk 1981. ISBN 82-7283-016-7 Pris kr. 20.

RAPPORT nr. 3. *Rapport fra den nasjonale konferanse om EDB i språk og litteraturforskning*, 4. og 5. januar 1979. Mars 1979. 2. opptrykk november 1981. ISBN 82-7283-024-8 Pris kr. 50.

RAPPORT nr. 4. *Oppbygging av EDB-katalog for folkemusea i Hordaland og kulturgeografisk registrering på Vestlandet*. April 1978. 3. opptrykk november 1981. ISBN 82-7283-000-0. Pris kr. 30.

RAPPORT nr. 5. *Rapport fra NKKM's EDB-komite*. August 1979. ISBN 82-7283-001-9 Pris kr. 15.

RAPPORT nr. 6. *Prøveprosjekt med EDB ved Norsk Folkemuseum*. Oktober 1979. ISBN 82-7283-002-7. Pris kr. 15.

RAPPORT nr. 7. *Ivar Fønnes: Norsk landbruksordbok. Prosjektrapport om databehandling og tilrettelegging for trykking*. September 1979. ISBN 82-7283-008-6. Pris kr. 25.

RAPPORT nr. 8. *SEFRAK. Rapport frå prøveprosjekt for databehandling av kulturminneregisteret*. Oktober 1979. ISBN 82-7283-003-5. Pris kr. 20.

RAPPORT nr. 9. *Jostein H. Hauge og Sigbjørn Århus: Dataregistrering i humanistiske fag med vekt på optisk lesing*. August 1978. 3. opptrykk januar 1981. ISBN 82-7283-004-3. Utsolgt.

RAPPORT nr. 10. *Roald Skarsten: Innføring i SPSS for humanister*. November 1977. 3. opptrykk november 1981. ISBN 82-7283-005-1. Pris kr. 30.

RAPPORT nr. 11. *Jostein H. Hauge og Knut Hofland: Rapport fra 4 konferanser i USA sommeren 1979*. The 17th Annual Meeting of Computational Linguistics. La Jolla Conference on Cognitive Science. The fourth International Conference on Computers in the Humanities. Data Bases in the Humanities and Social Science. November 1979. ISBN 82-7283-0078. Utsolgt.

RAPPORT nr. 12. *EDB og manuskriptregistraturer*. Oktober 1977. 2. opptrykk november 1979. ISBN 82-7283-009-4 Pris kr. 20.

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

RAPPORT nr. 14. *NOVA*STATUS HÅNDBOK*

Del 1: Søking. Brukerveiledning. 2. opptrykk januar 1982. ISBN 82-7283-011-6. Pris kr. 20.

Del 2: Fil-beskrivelser. Systemdokumentasjon. Pris kr. 20.

Del 3: Generering og oppdatering av databaser. Pris kr. 20.

humanistiske data 2-82

NAVFs EDB-senter for
humanistisk forskning

The Norwegian Computing
Centre for the Humanities

NAVF NORGES
ALMENVITENSKAPELIGE
FORSKNINGSRÅD

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

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

Senteret utgir tidsskriftet *Humanistiske Data* (4 nr. pr. år) og en rapportserie (25 er utkommet pr. 1.9.82).

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

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

Interesserte kan kostnadsfritt bestille årsmelding og *Humanistiske Data* (kr. 20,- for institusjoner).

Humanistiske Data blir utgitt av NAVFs EDB-senter for humanistisk forskning.
Redaksjonsgruppe: Jostein H. Hauge (ansv.), Eva Møller, Kristin Natvig, Svein Skotheim.
Senterets adresse: Harald Hårfagesgt. 31, Boks 53, 5014 Bergen-Universitetet.
Tlf. (05) 21 00 40, linje 2956.

Artikler, rapporter, meldinger mottas. Redaksjonen avsluttet 17. september.

Humanistiske Data is published by The Norwegian Computing Centre for the Humanities.
Editorial Group: Jostein H. Hauge, Eva Møller, Kristin Natvig, Svein Skotheim. The journal can be ordered from the address mentioned above. Contributions are welcome.

Medarbeidere i dette nummer:

Lars Sigfred Evensen, univ. stipendiat, Universitetet i Trondheim.

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

Knut Hofland, førstekonsulent, NAVFs EDB-senter for humanistisk forskning.

Eva Møller, konsulent, NAVFs EDB-senter for humanistisk forskning.

Kristin Natvig, vit. ass., NAVFs EDB-senter for humanistisk forskning.

Egil Nysæter, førstearkivar, Statsarkivet i Bergen.

Svein Skotheim, konsulent, NAVFs EDB-senter for humanistisk forskning.

Elin Solstrand, konsulent, NAVFs EDB-senter for humanistisk forskning.

Sigbjørn Århus, førstekonsulent, NAVFs EDB-senter for humanistisk forskning (permisjon fra 1.9.82).

Fotosats i kommunikasjon med Univac 1100/82

Sats: *Everts Alfabet — Bergen A/S*

Trykk: *A/S Repro-trykk*

Forsiden: Utsnitt av Riksarkivets magasin.

Innhold

Artikler:

Data om språkundervisning. <i>Lars Sigfred Evensen</i>	s. 4
EDB og publisering. <i>Svein Skotheim og Knut Hofland</i>	s. 12
Automatisk språkoversettelse i EF. <i>Eva Møller</i>	s. 18
EDB-tjeneste for språk og litteratur i Tromsø.	s. 22

Rapporter:

Senteret feirar 10 års-jubileum	s. 24
Datalogi og humanisme	s. 25
Seminar om EDB-metoder i humanistiske fag	s. 29
Hannovermessen 1982	s. 31
Datamaskinell informasjonssøking	s. 32
ALLC 1982	s. 36
COLING 82	s. 38
Datamaskinstøttet leksikografi	s. 40
Anvendelse av Artificial Intelligence i psykologi og lingvistikk	s. 41
Tre prosjektkataloger	s. 44
Tre bokpresentasjoner	s. 46
Fra bokhylla	s. 48

Meldinger	s. 50
-----------------	-------

Summary	s. 56
---------------	-------

Data om språkundervisning

Lars Sigfred Evensen

Denne artikkelen presenterer data som er samlet inn ved Institutt for anvendt språkvitenskap, Universitetet i Trondheim. Trondheimskorpuset i anvendt språkvitenskap er en manuelt lagret dokumentasjon av et stort utvalg elevers ferdigheter i lengre skriftlig framstilling i fagene engelsk og norsk.

Korpuset er samlet inn i tilknytning til den landsomfattende spørreskjemaundersøkelsen Den vet best hvor sko(l)en trykker... Her har et statistisk representativt utvalg elever og lærere angitt hva de opplever som problemer i de nevnte språkfagene.

Metoder for innsamling av språklige data har hittil vært forholdsvis lite diskutert i Norge. I artikkelen blir dette spørsmålet viet betydelig oppmerksomhet.

Anvendt språkvitenskap i Norge

Anvendt språkvitenskapelig forskning er rettet inn mot problemer og samfunnsoppgaver som har med språk å gjøre, f.eks. i forbindelse med språkundervisning (Evensen 1979). Språkvitenskapelig innsikt er en nødvendig betingelse for å løse slike problemer. Forskningsoppgavene krever imidlertid også innsikt av f.eks. pedagogisk, psykologisk og sosiologisk art. Anvendt språkvitenskap er følgelig et tverrvitenskapelig fag (Evensen 1980).

Språklig kommunikasjon er av fundamental betydning både for samfunn og individ. Ut fra dette burde forskning omkring språk, språktilegnelse og språkbruk ha høy prioritet. Slik er det imidlertid ikke i Norge. I 1979 gikk 4% av de totale forskningsbevilgningene til humanistisk forskning generelt (FOU-statistikk 1979, Oslo 1981). Bare en del av dette gikk til språkvitenskapelig forskning.

Norge er videre et land hvor store menneskelige og økonomiske ressurser settes inn i undervisning av språk. Det ville være rimelig å anta at en innsats av slikt omfang hadde et solid vitenskapelig fundament, og at virksomheten var godt analysert med henblikk på å utbedre eventuelle svakheter. Slik er det heller ikke. I det humanistiske forskningsmiljøet er skoleforskning et relativt marginalt område. Et norsk institutt for anvendt språkvitenskap ble først opprettet i desember 1977. Dette instituttet mangler fremdeles den nødvendige grunnbemanning.

Den vet best hvor sko(l)en trykker...

Undersøkelsen Den vet best hvor sko(l)en trykker... ble startet i 1979 som et ledd i arbeidet med å bygge opp anvendt språkvitenskap som universitetsfag i Norge. Undersøkelsen tok utgangspunkt i behovet for å skaffe til veie et bredt forskningsmateriale definert på fagets egne premisser. Et slikt materiale burde være problemorientert og preget av faglig bredde og statistisk representativitet.

Vi utviklet en spørreskjemaundersøkelse som kunne gi oversikt over

hva et landsomfattende representativt utvalg elever og lærere opplever som problemer i skolens språkundervisning. (1) Undersøkelsen ble avgrenset til å gjelde fagene norsk og engelsk på 8. og 9. årstrinn i grunnskolen og den videregående skolens 1. og 2. årstrinn (grunnkurset) i allmennfaglig studieretning.

En eksplorerende forundersøkelse med bruk av ustrukturerte intervjuer og essays ble brukt som empirisk grunnlag for å utvikle spørreskjema. Materialet fra forundersøkelsen viste bl.a. at det var en nær sammenheng mellom rent språklige problemer og problemer med ramme faktorer som arbeidsmåter og tilgang på ressurser. Dette påvirket utvalget av spørsmål i spørreskjemaet. På det metodiske planet viste forundersøkelsen f.eks. at intervjupersonene ofte omtalte problemene langs ordinalskalaer som meget vanskelig – meget lett og altfor mye av x – altfor lite av x. Dette kunne vi ta hensyn til ved utforminga av svaralternativene i spørreskjemaet.

Spørreskjemaet ble utviklet i samarbeid med interesserte lærere og elever som fikk tilsendt foreløpige utkast med forespørsel om kritiske kommentarer (se Evensen 1982). Undersøkelsesopplegget ble endelig utprøvd gjennom en landsomfattende pilot-studie i 1980. Erfaringene med pilot-studien var gode, og vi fikk inn svar fra alle de 16 skolene som ble spurt om å delta.

Hovedinnsamlinga til Den vet best hvor sko(l)en trykker... ble foretatt våren 1981. Da ble reviderte spørreskjema sendt til et utvalg på 128 elevklasser ved like mange skoler og de berørte skolenes lærere i norsk og engelsk. (2) Totalt fikk vi inn svar fra 109 (85.2%) av skolene. Materialet er lagret maskinelt som to SPSS-filer. (3) Elevfila har 2329 enheter og 189 variabler, lærerfila 774 enheter og 181 variabler. Filene gir informasjon om tre områder:

A: problemer som er knyttet til lærestoffet – vanskegradering av (del-)områder, eksempel på feil og vurdering av overføring mellom ulike språkssystem. Vanskegraderingen omfatter:

- ferdighetene leseforståelse, lytteforståelse, taleferdighet og skriveferdighet
- delferdighetene vokabular, idiomer, uttale, intonasjon, morfologi, syntaks, kohesjon/koherens, ortografi og tegnsetting
- situasjonstilpasset språk
- stilskriving brutt ned i vokabular, syntaks, tekstlingvistiske faktorer, ortografi og tegnsetting
- områdene kulturkunnskap, litterær og grammatisk analyse, dialekter, sidemål, språkhistorie og studieteknikk
- en totalvurdering – vanskeligste ferdighet/delferdighet/område totalt sett.

B: problemer som angår de betingelsene som læringen/undervisningen skjer under – arbeidsmåter, læremidler, motivasjon, arbeidsmiljø, målsetting, evaluering, utstyr/ressurser og lærerutdanning.

C: Visse anonyme bakgrunnsopplysninger – dialektbakgrunn, kjønn, alder, elevs karakter i faget og i læreres utdanning og undervisnings erfaring.

Materialet er lagret fullstendig anonymt og kan ikke spores til enkelt-individer, klasser eller skoler innenfor den enkelte kommune.

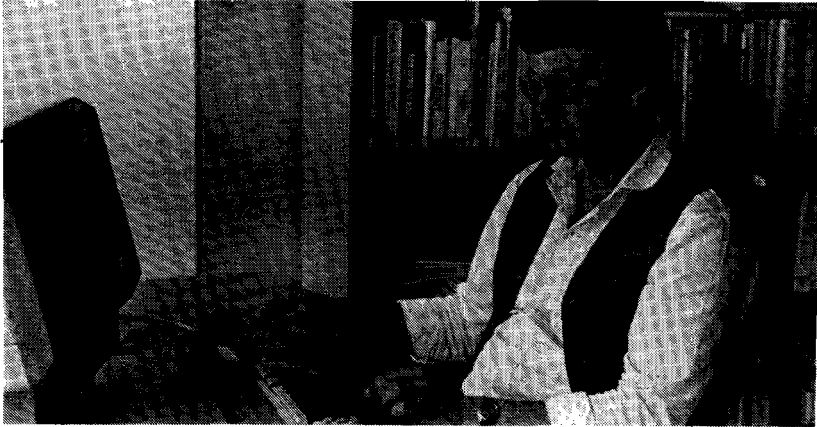
Trondheimskorpuset i anvendt språkvitenskap

Vi så fra starten av ønskeligheten av å kunne sammenholde det elever og lærere angir som problemer i et språkfag med elevenes språklige produksjon i skolen. Et språklig korpus ville kunne gi interessante analyser sett i sammenheng med spørreskjemamaterialet og tjene som kontroll av spørreskjemamaterialets validitet (jfr. Ilstad, Paasche og Hoveden 1977). Vi var imidlertid skeptiske med hensyn til mulighetene for å skaffe fram et slikt materiale med de sterkt begrensede ressursene vi hadde til rådighet. Merbelastningen på skolene ville dessuten sannsynligvis redusere svarprosenten.

Oppslutninga om pilot-undersøkelsen var imidlertid oppmuntrende. Svarprosenten var høy, særlig blant elevene (93%). Vi fikk dokumentert at spørreskjemaet virket motiverende. Mange ga skriftlig uttrykk for sin glede over å bli tatt alvorlig i forbindelse med undervisningen. Utfyllingen av skjema var gjennomgående nøyaktig og seriøs. Av 366 mottatte elevskjema måtte bare 7 forkastes på grunn av mangelfull eller ikke-seriøs utfylling. På det administrative planet var også erfaringene gode. Innsamlingsrutinene så ut til å virke effektivt. Det merarbeidet vi ville få med å inkludere et språklig materiale, virket dermed overkommelig.

Hovedundersøkelsen ble utvidet slik at vi ville få tilsendt kopier av elevenes ferskeste stiler/essays i norsk og engelsk. En bevilgning fra NAVF gjorde det mulig å refundere de merutgiftene som skolene fikk.

Vi fikk inn kopier fra 93 (72.7%) av skolene i utvalget. Totalt blir dette et materiale på 2943 skriftlige besvarelser – 1538 i norsk og 1405 i engelsk. Korpuset er foreløpig lagret manuelt. Besvarelsene er kodet etter et dobbelt system. Et skolenummer gir tilgang til den informasjon som kom fram gjennom utvalgsprosedyren – skoleslag, årstrinn, geografisk region, dialektområde, kommunens urbaniseringsgrad, skolestørrelse og skolens skriftlige målform. Vi ønsket videre å kunne sammenligne den enkelte elevs spørreskjema og skriftlige produksjon. I prosjektet Norsk i sammenholdte klasser (NISK) ble tilsvarende sammenligninger mulig ved at kontaktlærerne ble betalt for å lage klasselister med navn og tilhørende kodetall (Olav Sletta, pers. komm.). Hver elev fikk kopi av lista og førte kodetallet på alle sine besvarelser. Deretter ble listene makulert. Dette systemet er pålitelig, men relativt arbeidskrevende for den som skal være kontaktlærer. I prosjekter hvor forskeren ikke kan samle inn materialet selv, eller tilby lønn til kontaktlæreren, blir et slikt system for tungrodd til å kunne fungere godt. I norsk skole synes mange ansatte at arbeidspresset er stort, og viljen til å ta på seg ubetalt merarbeid er synkende.



Stipendiat Lars Evensen, Institutt for språk og litteratur, Universitetet i Trondheim.

Vi valgte å utvikle et enklere kodesystem, hvor den individuelle kodinga ble overlatt til elevene selv. Hver elev valgte et firesifret tall som den enkelte førte på sitt eget spørreskjema og sine egne stiler.

Innsamlingsrutiner i kritisk lys

Kodesystemet har fungert tilfredsstillende overfor elevene. Vi kunne vente et visst svinn ved at elevene ville sette ulikt tall på skjema og stiler eller ved at enkelte ville unnlate å bruke kodesystemet. Ved å trekke ut de tilfellene hvor kodinga er foretatt korrekt, har vi fått tilbake en rest for hver klasse som kunne vurderes mot de mulige svakhetene som er nevnt ovenfor. Om vi ser bort fra tilfeller hvor sammenhengen mellom stil og skjema går frem av håndskrift, eller hvor ulikt antall skjema og stiler skyldes sykefravær, får vi en rest på tolv tilfeller hvor kodetallene ikke stemmer overens. Manglende individuell koding er heller ikke noe stort problem. Her får vi en rest på atten tilfeller. Dette må betraktes som meget tilfredsstillende.

Ved flere skoler har imidlertid kontaktlæreren stilt seg skeptisk til elevenes evne til å mestre et slikt kodesystem. Ved 5 skoler er alle identitetstall påført av samme person. Ved 18 skoler har en helt unnlatt å ta systemet i bruk. Dette svekker mulighetene for representativitet ved analyser basert på individuelle sammenligninger. Vi kan imidlertid fremdeles foreta gruppevis sammenligninger ved disse skolene ettersom alle besvarelsene er påført skolenummer.

I instruksjonen til kontaktlæreren har det ikke vært klart nok presisert at en ønsket skriftlige besvarelser i begge fag fra hver enkelt elev. 20 av skolene har kun sendt besvarelser for det fag som den enkelte elev fylte ut spørreskjema for. Dette bortfallet kommer i tillegg til andre typer nevnt tidligere. Det kumulative bortfallet blir så stort at en ikke vil kun-

ne gjøre krav på statistisk representativitet ved analyser basert på individuelle sammenligninger. Vi kan foreta individuelle sammenligninger mellom stil og spørreskjema i samme fag for samme elev i 1346 tilfeller. Dette utgjør 44.5% av det totale antallet elever (n = 3025) som ble spurt om å delta i undersøkelsen og 58.8% av de elevene (n = 2329) vi mottok utfylte spørreskjema fra. Vi kan foreta individuelle sammenligninger mellom stiler i begge fag for samme elev i 838 tilfeller. Dette utgjør 35.9% av antallet elevskjema og 27.7% av det totale utvalget. Også her vil det imidlertid være mulig med gruppevis sammenligninger ut fra skolenummer.

Vi hadde tatt sikte på å få inn elevenes siste skriftligbesvarelser før påsken 1981. Den skoleoversikten som en del av skolene skulle trekkes ut fra, forelå imidlertid seinere enn forutsatt. Det ble ikke mulig å opprettholde tidsplanen ved alle skolene. Innsamlingsarbeidet ble spredd over et tidsrom som strakte seg fram til slutten av skoleåret. Dette innfører nok en feilkilde i materialet. Fra en av klassene i utvalget fikk vi inn stiler både fra påskententamen og årsprøven. Dermed kan vi få en indikasjon på omfanget av denne feilkilden.

Potensiale for framtidig forskning

Trondheimskorpuset i anvendt språkvitenskap gir et relativt omfattende materiale for språklige performansanalyser og feilanalyser.

Nedenfor vil jeg framstille noe av det forskningspotensialet som ligger i korpuset:

- Materialet gir et landsomfattende bilde av norske elevers skriftlige produksjon i de to språkfagene som alle norske elever får opplæring i. Materialet er hentet fra fire årstrinn i de to skoleslagene som omfatter flest elever. Utvalgsmetoden og kodinga gjør det mulig å kople inn data fra Den vet best hvor sko(l)en trykker... i språklige analyser. Et slikt bilde vil være relevant for forskning og debatt omkring sentrale områder; elevers språklige nivå og problemer på ulike trinn, spørsmål om arbeidsmåter, læremidler, motivasjon, fagplaner, evaluering, ressurser og lærerutdanning.

Skal en kunne gi et totalbilde, krever imidlertid korpusets omfang at det blir gjort tilgjengelig for maskinell behandling. Ved manuelle analyser vil materialet måtte brytes ned i små undergrupper av arbeidsmessige årsaker.

- I nyere forskning om tilegnelse/læring av fremmedspråk har mange forskere framhevet den systematiske, regelbundne karakteren av elevers språklige produksjon på alle ferdighetsnivå. Elevspråket bør etter disse forskernes mening prinsipielt betraktes som uttrykk for et eget språksystem – et mellomspråk (inter-language) (Corder 1967, Selinker 1972). Trondheimskorpuset gir materiale for tverrsnittsun dersøkelser av mellomspråk på fire etterfølgende årstrinn i to skoleslag.

I Skandinavia drives engelskundervisning på et relativt høyt nivå internasjonalt sett. I skandinavisk forskning er det følgelig et særlig behov for mellomspråkstudier ut over elementærtrinnet. Korpuset gir materiale for grunnskolens to høyeste årstrinn og to årstrinn i videregående skole.

- I et mellomspråkperspektiv blir elevers feil sett i et nytt lys. Feil blir en innfallsvinkel til systematisk studium av elevers språklige utvikling. Feilanalyser bør dermed ses i forhold til performansanalyser som omfatter alle sider ved elevers språkproduksjon. (Svartvik 1973 og Richards 1974). Trondheimskorpuset gir mulighet for å sammenligne produksjon av både morsmål og fremmedspråk. Dermed kan en få et bedre bilde av hvilke feil som er spesifikke for fremmedspråket og hvilke som er generelle for begge språk. Et bilde for to språk gir bedre bakgrunn for å utvikle hypoteser om årsaker til ulike feil/feiltyper.
- Studier av feiltyper og frekvens kan sammenholdes med hva elever og lærere angir i spørreskjema som vanlige feil. Korpuset kan dermed brukes i metodisk sammenheng som en ekstern kilde til validering av spørreskjemamaterialet. En slik validering vil være interessant sett i forhold til Trans (1975) påvisning av høyt samsvar mellom elevers selvrappotering om vanskegrad og deres språklige produksjon i en eksperimentell situasjon.
- I nyere forskning om språk har en rettet større oppmerksomhet mot tekstlingvistiske faktorer (se f.eks. Fossestøl 1980). Når en sammenligner uveide gjennomsnitt (medianer) for vanskegradering av delferdigheter i Den vet best hvor sko(l)en trykker... , ser det ut til at en meget vesentlig del av elevenes problemer i skriftlig framstilling vil bli oversett ved analyser som ikke går ut over setningsplanet.

Trondheimskorpuset gjør det mulig å analysere tekstlingvistiske faktorer i elevers lengre framstillinger (jfr. Enkvist 1974, Larson 1978, Lieber 1981). Her vil skrivesituasjonen og emne for framstilling være kjente størrelser.

- I dialektorienteerte studier kan materialet brytes ned etter geografisk region. Kommunenummer kan brukes som innfallsvinkel til finere inndelinger slik som i Kommunedatabanken (Henrichsen og Rokkan 1977), hvor dialekt er lagt inn som en variabel med 17 verdier. Inndeling er her foretatt på grunnlag av dialektkartene i Helleland og Papazian (1973).

I studier av engelsk som fremmedspråk gir slik koding mulighet for en relativt presis vurdering av elevenes språklige bakgrunn ved analyser av positiv og negativ overføring fra morsmålet. En dialektkode vil også være sentral ved studier av overføring mellom talemål og skriftspråk i norsk.

- For sosiologvistisk orienterte studier gir Trondheimskorpuset informasjon om kommunens sentralitet, næringsstruktur og befolknings-tetthet gjennom kommunenummer og en kode for urbaniserings-grad. Spørreskjemaene i Den vet best hvor sko(l)en trykker... gir an-onyme bakgrunnsopplysninger av sosiologisk art.

Korpuset har flere begrensninger. Det gir kun et skriftlig materiale innsamlet over et visst tidsrom. Vi har videre sett at vi ikke helt har greid å oppnå målet om statistisk representativitet. Det synes likevel klart at Trondheimskorpuset i anvendt språkvitenskap har egenskaper som gjør det nyttig for forskning om språkundervisning i Norge.

Det videre arbeidet med korpuset går ut over rammene for mitt enga-
jement som stipendiat og er foreløpig ikke planlagt i detalj. Materialets omfang og egenskaper gjør det rimelig å se utnyttelsen av det som et sam-
arbeid mellom forskere fra flere miljøer. Fra andre språkinstitutter ved Universitetet i Trondheim er det ytret betydelig interesse for samar-
beidsprosjekter knyttet til materialet. (4)

Noter

1. Professor Elisabeth Ingram er veileder for prosjektet. Hennes kritiske kommentarer og vennlige oppmuntring har vært av uvurderlig betydning under arbeidet. Særlig gjelder dette den første fasen da alle konturer var vage.

Andre ansatte ved instituttet har også gitt verdifull bistand under arbeidet. Aud Marit Simensen, Anders Martin Kvam og Ellen Andenæs har gått kritisk gjennom foreløpige ut-
kast til spørreskjema. Ellen Andenæs deltok også i diskusjoner som førte fram til et kode-
system som blir beskrevet i denne artikkelen.

Utenfor instituttet har omkring fire tusen personer vært koblet inn på en eller annen måte. Dette gjelder elever og lærere i utvalget, andre skolefolk og personer i akademiske institusjoner. Jeg vil rette en varm takk til alle som har bidratt til undersøkelsen.

2. Utvalget ble definert ved en prosess hvor populasjonen av elever og lærere innenfor de aktuelle skoleslag og årstrinn ble stratifisert etter geografisk region, kommunenes urba-
niseringsgrad, skoleslag og årstrinn. Innenfor hvert substratum ble en elevklasse ved hver av fire skoler trukket ut tilfeldig fra lister stilt til disposisjon av Statistisk Sentralbyrå og Rådet for videregående opplæring.

Opplegget ble klarert overfor Kirke- og Undervisningsdepartementet, Datatilsynet og berørte skolestyrer og fylkesskolestyrer. Vi søkte om godkjenning ved den enkelte skole og ba skolens ledelse finne en kontaktperson som kunne administrere gjennomføringen ved skolene. Deltakelsen var basert på full anonymitet og frivillighet for skolene, den enkelte lærer og den enkelte elev.

3. Programpakken SPSS organiserer data som en matrise hvor rekkene er enheter, ko-
lonnene variabler og skjæringspunktene mellom rekker og kolonner er verdier. Data fra Den vet best hvor sko(l)en trykker... er lagret slik at hvert utfylt spørreskjema utgjør en enhet. Hvert (under-)spørsmål utgjør en variabel som for hver enhet har en verdi. Hver verdi er en numerisk kode for det enkelte svar.

4. Edb-konsulent Eirik Lien har lest gjennom en foreløpig utgave av denne artikkelen og kommet med mange verdifulle kommentarer og forbedringsforslag. Inger Ellingsen har skrevet manuskriptet på maskin. En varm takk til begge.

Litteratur

Corder, S.B. 1967 - «The Significance of Learners' Errors». *International Review of Applied Linguistics (IRAL)* Vol. 5, No. 4, 161-170.

Enkvist, N.E. 1974 - «Några textlingvistiska grundfrågor». I Teleman og Hultman, red. *Språket i bruk*. Lund: Liber Laromedel.

Evensen, L.S. 1979 - «Applied Linguistics: Some Definitions». Stensil. Trondheim: Institutt for anvendt språkvitenskap.

Evensen, L.S. 1980 - «Applied Linguistics in Norway: Some Perspectives on Research». Stensil. Trondheim: Institutt for anvendt språkvitenskap.

Evensen, L.S. 1982 - «Forskning for morgendagens fremmedspråkundervisning». I *Lærere og skoleforskning: Rapport fra en konferanse om «Skoleforskning i humanistiske fag: behov og organisering»* arrangert av Rådet for humanistisk forskning i NAVF 4.-6. november 1981. Oslo: NAVF.

Fossestøl, B. 1980 - *Tekst og tekststruktur: Veier og mål i tekstlingvistikken*. Oslo: Universitetsforlaget.

FOU-statistikk 1979 - Oslo: Forskningsrådenes samarbeidsutvalg. 1981.

Helleland, B. og Papazian, E. 1973 - *Oss sei oss, oss*. Oslo: NRK.

Henrichsen, B. og Rokkan, S., red. 1977 - *Kommunedatabanken: En håndbok for brukere*. Bergen: Norsk Samfunnsvitenskapelig Datatjeneste.

Ilstad, S., Paasche, T. og Hovden, J. 1977 - *Survey-metoden*. Trondheim. Tapir.

Larsson, N. 1978 - *Modeller och metoder i textlingvistikken*. Lund: Studentlitteratur.

Lieber, P. 1981 - «Zero-Referent Items in ESL Student Writings». I Sigurd, B. og Svartvik, J., red. *AILA 1981: Proceedings 1, Sections and Workshops*, 147-148.

Richards, J. ed. 1974 - *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman.

Selinker, L. 1972: «Interlanguage». *IRAL*, Vol. 10, No. 3, 209-231.

EDB og publisering

Svein Skotheim og Knut Hofland

Bakgrunnen for denne artikkelen er det utviklingsarbeid som NAVFs EDB-senter for humanistisk forskning har drevet på dette feltet. Mikro-maskiner, tekstbehandlingsanlegg og stormaskin (UNIVAC 1100) har gjennom lengre tid vært utnyttet innen publisering for å vinne erfaringer. Først og fremst er publisering via edb brukt i forbindelse med produksjon av bøker, bladet *Humanistiske Data*, årsmeldinger m.v. Teknikken er spesielt interessant for humanister ettersom den kan innebære økonomiske besparelser ved publisering av store tekstmengder, noe som er særlig aktuelt nå i forbindelse med den økonomiske innstramning som skjer innen humanistisk og annen forskning.

Artikkelen er ment å være kortfattet og skissemessig og går derfor ikke særlig grundig inn på de forskjellige muligheter som foreligger. Formålet er kun å gi en oversikt over det arbeidet som har vært utført ved Senteret og å skissere en del sentrale problemstillinger med tanke på fremtiden. Det er lagt vekt på å begrense bruken av faguttrykk. Det forutsettes at leserne i grove trekk vet hva tekstbehandlingsanlegg, mikro-maskin m.v. er.

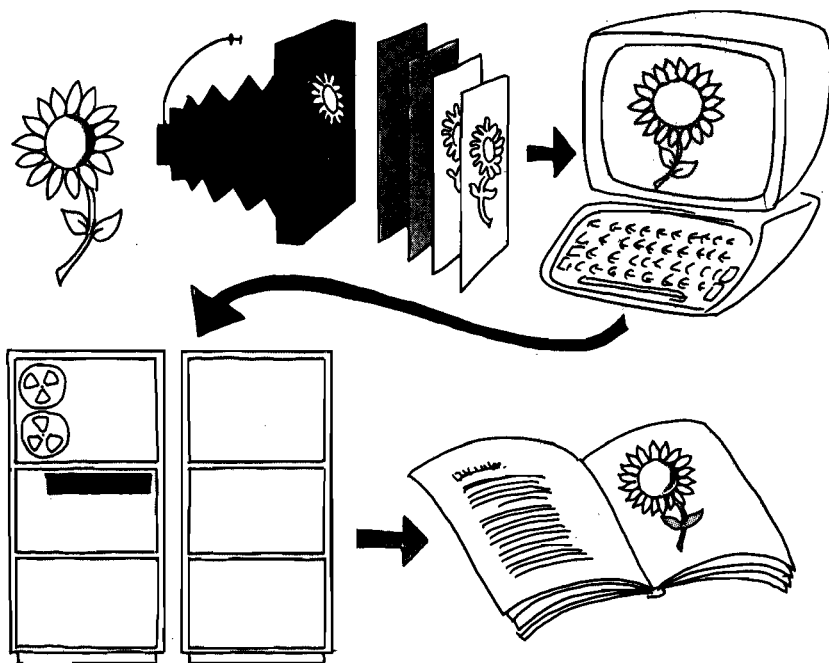
Det må først understrekes at ikke alle typer manuskripter egner seg like godt til det publiseringsopplegg som er skissert i artikkelen. På sikt vil elektronisk databehandling for eksempel få betydning i forbindelse med «on demand publishing». Det er et system der tekst og manuskripter ligger i databanker og der de forskjellige institusjoner og personer kan koble seg inn og bestille en utskrift av det manuskriptet de ønsker. Utskriften kan ha form av ferdig innbundet bok/katalog, og bestilleren må betale de produksjonskostnader som er aktuelle i forbindelse med utskriften. Systemet er i bruk i USA og flere andre land og har vist seg å være konkurransedyktig når det gjelder publikasjoner med små opplag. Universitetsforlaget planlegger et opplegg som vil være tilpasset forholdene i Norge.

Det skilles mellom satsproduksjon og trykking. Artikkelen tar først og fremst for seg produksjon av fotosats som et grunnlag for selve trykkeprosessen.

I den siste tid har laserskrivere fått en økende utbredelse. Disse skriverne fungerer etter xerografi-prinsippet som en kopimaskin, og gir utskrift på vanlig papir med en oppløselighet på opptil 300 punkt pr. tomme (fotosettere har 5 ganger bedre oppløselighet). Laserskrivere som introduseres i dag, ligger i prisklasse ca. 100.000 kr. og har en kapasitet på 10-15 sider pr. minutt. Disse er rettet mot tekstbehandlingsanlegg. Laserskrivere kan skrive forskjellige skrifttyper og størrelser og også tegne figurer og skraver flater o.l. Hva som er mulig er i stor grad avhengig av det tekstbehandlingssystemet som driver skriveren. Den neste generasjon skjermer til tekstbehandlingssystem vil ha større oppløselighet og bedre grafiske muligheter enn i dag. Brukeren vil derved få frem på skjermen teksten og figurer slik de blir skrevet ut på laserskriver eller fo-

tosats. Dette vil medføre at markedet for laserskrivere vil øke sterkt og prisene vil gå nedover. Prisen på en side fra en laserskriver vil ligge på ca 1/10 av en fotosatt side, men uten at kvaliteten er redusert tilsvarende. Laserskriveren vil særlig være aktuell ved små opplag og korrektur til materiale som skal fotosettes.

Hva er så en fotosetter og hva er fotosats? En fotosetter er i prinsippet en utskriftenhet som kan sammenlignes med vanlige skrivere eller for den saks skyld en skrivemaskin. Ved fotosats blir de enkelte tegn (bokstaver) belyst på fotografisk papir enten via en karaktermaske på en filmrull eller fra et katodestrålerør med høy oppløselighet. Fotopapiret må senere fremkalles. Dersom vi sammenligner med en vanlig skrivemaskin/skriver, gir fotosetteren en skrift med bedre kvalitet og dermed øket lesbarhet, den kan gi forskjellige tegnstørrelser, forskjellige skifter (normal, kursiv, fet m.v.) og forskjellige skrifttyper. Den kan også lett koble over til andre alfabet (gresk, kyrillisk, fonetiske tegn, matematiske tegn m.v.), og den gir mer kompakt skrift og sparer dermed plass og papirutgifter ved trykking. De første fotosettere kom til Norge i 1965. Siden har teknikken utviklet seg raskt, og fotosettere er i dag meget avanserte tekstbehandlingsanlegg og «kvalitetsprintere» i en enhet.



Fotosetteren produserer satsen ved at det legges inn koder for forskjellige funksjoner, såkalte typografikoder. Eksempler på slike typografifunksjoner er skrifttype, bokstavstørrelse, linjeavstand, innrykk første linje i avsnitt o.s.v. Kodene settes direkte på fotosetterens tastatur eller som koder i teksten som behandles av egne fotosatsprogrammer. Det skilles gjerne mellom 80 til 100 forskjellige slike typografifunksjoner. Fotosetteren kan også behandle et stort antall skrifttyper, tegne ut varemerker (logos), lage rammer, skillestreker m.v. I den senere tid er det også blitt mulig å få ut bilder og figurer direkte fra fotosetteren.

Det som kommer ut av fotosetteren, er en lang rull med tekst med sats i den spaltebredde som er kodet inn. Denne brukes så som grunnlag for videre trykking ved at den klippes opp og formes til sider (brekkes om), og fotograferes og overføres til trykkplaten.

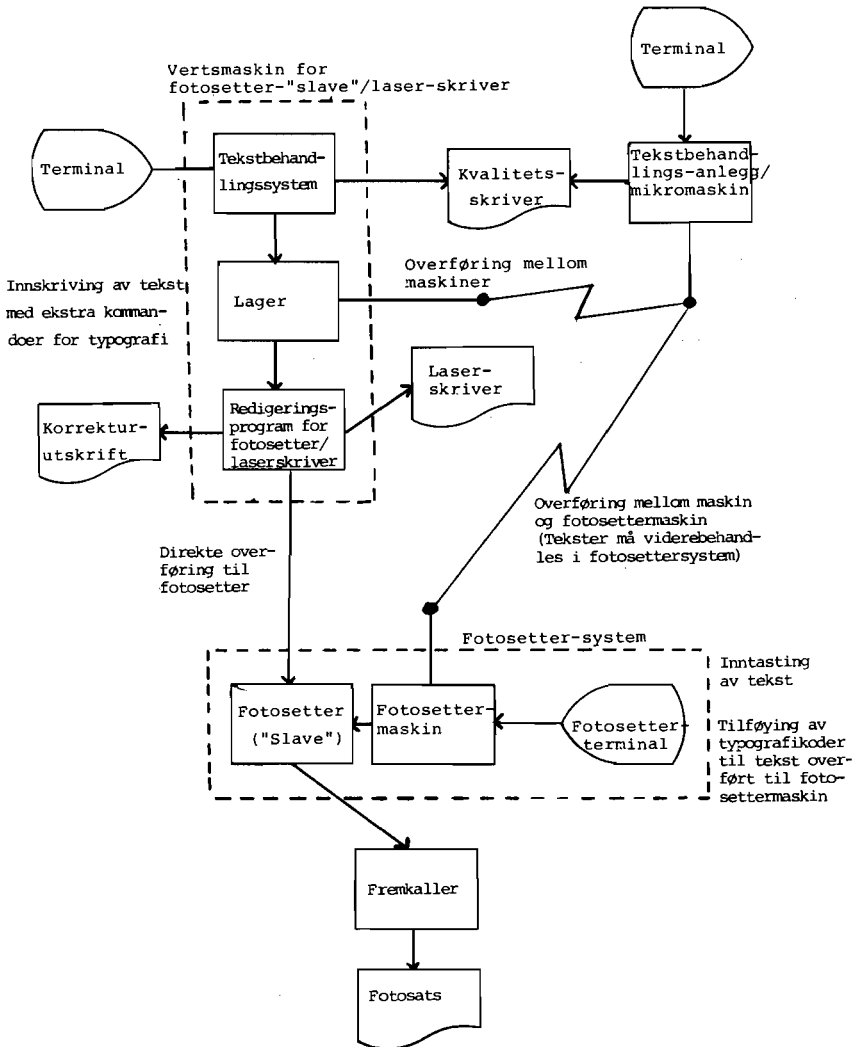
Svært mange institusjoner med egne dataanlegg har etter hvert koblet på en fotosetter. Et nærliggende eksempel på dette er Københavns Universitet som i mange år har hatt denne konfigurasjonen. Universitetet har en stormaskin, UNIVAC 1100/82, og det er koblet til en Compu-graphic 8600 til denne som utskriftsenhet. Det er også utviklet et eget program (Photodoc) for ombrekking, sidenummerering, indeksering, generering av innholdsliste m.v. Det eksisterer kommandoer for format, tittelsats, paginering, venstre/høyremarg, sentrering, splitting av tekstlinje, fast inndeling i linjer, innrykking av tekst, skifting av skriftstørrelse, linjeavstand m.v., utarbeiding av fotnoter, innholdsliste, indekser, tabelloppstillinger, overskrifter, orddeling, unntaksordbok osv. Universitetet i Bergen vurderer å kjøpe inn et tilsvarende anlegg og bruke programmet Photodoc.

For å utnytte edb-teknikken i satsproduksjon, må manuskriptet først gjøres maskinlesbart eller legges inn i et edb-anlegg. Enklest kan dette gjøres ved at det brukes et spesielt kulehode på skrivemaskinen som gjør det mulig å OCR-lese teksten. OCR betyr «Optical Character Recognition» og innebærer at en optisk leser overfører alle tegn til et tilkoblet edb-anlegg. Deretter kan den OCR-avleste teksten behandles på vanlig måte i edbanlegget. (Se artikkel i HD 1-82 om OCR.) Spesielt for mindre avanserte lesere er resultatet avhengig av en godt vedlikeholdt skrivemaskin og nøyaktighet ved skrivingen. Bruk av OCR-leser gir i seg selv ingen muligheter for redigering eller bearbeiding. Det er en mekanisk overføring av tekst til et edb-anlegg.

Dersom en har tilgang på tekstbehandlingsanlegg eller mikromaskiner/stormaskiner, kan teksten legges direkte inn på disse som et alternativ til å maskinskrive den. Dersom en kjenner typografikodene, kan disse legges inn samtidig med innskrivningen. Manus kan dermed klargjøres for trykking direkte av den enkelte forfatter. Det er også mulig å legge inn de typografiske kodene på et senere tidspunkt, f. eks. av en redaksjon eller av trykkeriet selv.

De forskjellige fotosettere har, som nevnt, forskjellige typografikoder. Det utarbeides gjerne en manual for det aktuelle edb-anlegg og den aktuelle fotosetter. En slik manual kan utbygges etter hvert som nye be-

MASKIN- OG PROGRAMVARE FOR SATSFREMSTILLING



hov dukker opp, f.eks. ønske om bruk av spesielle grafiske tegn. Det kan også bygges opp «samlekoder» for bestemte grafiske former.

Når teksten er ferdig kodet, overføres den til fotosetter. De tre vanligste overføringsmedier er magnetbånd, diskett og telelinje. Ved anvendelse av den siste overføringsmåte kan eieren av fotosetteren ringe til edb-anlegget og «hente» den aktuelle teksten ved hjelp av telefonen. Det kan også ringes fra datamaskinen til fotosetteren. Ved bruk av de to andre overføringsmetodene må teksten først legges inn på et magnetbånd eller en diskett og deretter fysisk bringes til fotosetteren og kobles til denne.

Når teksten er lagt inn i fotosetteren, produseres ferdig sats. Vi går ikke inn på en teknisk beskrivelse av produksjonen her, men bare nevner at før satsen er ferdig til videre redigering, må den fotografisk fremkalles i et eget fremkalleranlegg. Det er altså i prinsippet en vanlig fotografisk teknikk som anvendes, derav navnet fotosetter. Selve fremkallingen medfører en del driftsoppgaver og -utgifter, bl.a. til fotopapir og kjemikalier.

Økonomi

Når det gjelder priser for satsproduksjon på denne måten er forholdene i trykkeribransjen foreløpig svært ustabile og forskjelligartede. Prisen varierer fra kr. 5 til kr. 13 pr. 1000 tegn.

Hvis en sammenligner disse prisene med normale produksjonspriser for sats, d.v.s. sats produsert ved at en typograf selv taster inn tekst i fotosetter på grunnlag av manus, vil våre priser ligge på ca. 40% i gjennomsnitt.

Sammenfatningsvis kan vi si følgende om kostadene ved å fremstille et manuskript klart for kopiering/trykking:

1. Manuskriptet skrives på nytt på skrivemaskin, normal pris 7 - 10 kroner pr. tusen tegn.

2. Manuskriptet skrives på en IBM-komposer (avansert skrivemaskin). Normal pris 18 - 22 kroner pr. tusen tegn.

3. Manuskriptet skrives på nytt i trykkeriet for fotosetting. Normal pris ca. 25 til 40 kroner pr. tusen tegn.

4. Manuskriptet legges inn på tekstbehandling eller dataanlegg og kodes ferdig for fotosetting. Ferdig fotosats vil da koste ca. 5 - 13 kroner pr. tusen tegn. Dersom trykkeriet selv må kode, koster det fra 4 - 8 kroner pr. tusen tegn i tillegg.

Kopiering/trykking/innbinding

Det er mange måter å mangfoldiggjøre et trykningsklart manus på. Den vanligste i forbindelse med bøker, blader, kataloger m.v. er offsettrykk. Originalene overføres da fotografisk til en metallplate som fordeles svarte på papiret. En slik trykklar plate i A3-format koster ca. 150 kr. Denne teknikken er nødvendig dersom originalene inneholder bilder eller dersom en skal gjøre bruk av farger. Andre produksjonsalternativer som boktrykk, dyptrykk m.v. tar vi ikke opp her.

Såkalt kontoroffset kan være aktuell dersom manus ikke inneholder fotografier og dersom det ikke skal gjøres bruk av farger. Strektegninger, tabeller o.l. er imidlertid kurant.

Følgende priser gjelder ved hustrykkeriet, Universitetet i Bergen for Universitetets egne brukere: For å kunne mangfoldiggjøre en A4-side med trykk på begge sider må det lages en «master» til ca. kr. 2,50. Å lage en kopi med trykk på begge sider koster 6 øre med den mest brukte papirkvalitet, dvs. 80 gram pr. kv.m. i såkalt matt offsetpapir. I bokform i A5-format vil dermed hver trykt boksider komme på 1,5 øre pluss kr. 2,50 delt på antall eksemplarer det skal være av den aktuelle publikasjonen. Dersom vi tar utgangspunkt i en bok på 200 sider vil mangfoldiggjøring av sidene komme på ca. kr. 3. Den optimale opplagsstørrelsen for dette produksjonsopplegget vil være på rundt 200 eksemplarer. For kr. 3,63 får man altså 200 sider ferdig sortert i A5-format trykt på begge sider og med strektegninger/tabeller dersom dette er ønsket. Innbinding koster kr. 4 pr. stk.

Sammenfatning

Tekstbehandlingsmaskinene er generelt nyttige for å forenkle rettinger og få utskrift av maskinskrevne dokumenter. Den lagrede teksten kan under gitte forutsetninger leses maskinelt – enten direkte eller indirekte – av andre datasytem, f.eks. i forbindelse med automatisk fotosetting. Kostnaden og tiden for å fremstille konvensjonell fotosats fra et manus utgjøres for en stor del av kostnadene for innskriving av manus, kontroll av at innskrivingen er riktig utført, retting av feil, og gjerne rettinger mot manus etter fotosatsen er ferdig fremstilt. Kostnadene for produksjon av konvensjonell fotosats har derfor ofte blitt motvirket av fordelene med å benytte fotosats. Med maskinlesbare manuskripter fra forfatteren kan kostnadene og tiden for produksjon av fotosats i mange tilfeller radikalt endres. Det kan også gi mulighet for kvalitetshevning i den forstand at man kan koste på seg å sette tekster som ellers ville blitt publisert med annen og dårligere skriftbilde. En følge av at det er blitt teknisk lettere å få til automatisk fotosetting av tekst som ligger på tekstbehandlingsanlegg, er at grenselinjen mellom maskinskriving i tradisjonell forstand og fotosetting er blitt mer flytende. Kanskje vil den helt utviskes i publiseringssammenheng i løpet av 10 år. Man bør derfor allerede ved planleggingen av innskrivingen ta hensyn til om en del tekster skal settes og hvordan dette best skal gjøres. Den nye teknikken stiller altså krav både til forfatteren/redaksjonssiden og til den grafiske ekspertise. Til begges fordel.

Automatisk språkoversettelse i EF

Eva Møller

EF har nå 10 medlemsland med 7 forskjellige språk, og EF er forpliktet til å oversette alle viktige dokumenter til hvert av medlemslandenes språk. Dermed er behovet for oversettelser innen EF av anselige dimensjoner. Europakommisjonen besluttet derfor midt i 1970-årene å satse på utviklingen av et automatisk oversettelsessystem som skal kunne oversette mellom alle offisielle EF-språk.

Arbeidet med EUROTRA (EUROpean TRANslation system) ble påbegynt i 1978, under ledelse av *Margaret King*, Storbritannia. For å støtte forskning i ulike universitetsmiljøer, og samtidig dra nytte av tidligere erfaringer, har man fordelt utviklingsarbeidet på flere eksisterende forskningssentra. De universitetsbyer som er representert i koordineringsgruppen, er København, Leuven (Belgia), Manchester/Essex (England), Grenoble (Frankrike), Saarbrücken (Tyskland), Pisa (Italia) og Delft (Nederland).

Utviklingsarbeidet vil bl.a. bestå i forsøk på praktisk anvendelse av tidligere utført datamaskinell språkforskning. EUROTRA-prosjektet har dermed stor generell interesse for datalingvistiske miljøer siden man her kan få testet ut ulike grammatikker og lingvistiske modeller.

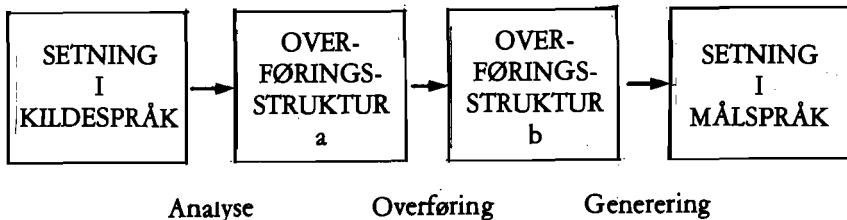
Et absolutt krav: flerspråklighet

De tidligste oversettelsessystemer var tospråklige, dvs. at setningen i kildespråket ble analysert med et bestemt målspråk for øye. Av senere utviklede systemer, som i utgangspunktet har vært tospråklige, er noen blitt utvidet til å omfatte et tredje språk – dette gjelder f.eks. SYSTRAN, som fra begynnelsen var engelsk-russisk, men som er utvidet til å oversette engelsk-fransk og engelsk-italiensk. SYSTRAN anvendes innenfor EF i dag, men gir ikke tilfredsstillende resultater.

Til forskjell fra disse tidligere systemer skal EUROTRA være FLERSPRÅKLIG, idet det skal kunne oversette fra og til alle nåværende EF-språk. Videre skal det ha innebygde utvidelsesmuligheter for evt. nye språk (f.eks. spansk og portugisisk som kan bli nye EF-språk).

Hvis EF skulle få dekket sitt oversettelsesbehov gjennom tospråklige systemer måtte en mengde systemer utvikles: n språk gir $n(n-1)$ språkpar, dvs. for 7 språk finnes 42 språkpar som det måtte utvikles systemer for å dekke. Og vokser antallet språk til 9 gir det 72 språkpar – dvs. at nye medlemsspråk ville gi uoversiktlige oversettelsesproblemer.

For å omgå dette problem er EUROTRA-prosjektet basert på en grunnleggende annerledes filosofi. Oversettelsen splittes opp i moduler, og mest mulig av arbeidet utføres med bare ett språk for øye. Oversettelsesprosessen kan illustreres ved følgende modell:



(Mægaard og Ruus, 1980)

Oversettelsesprosessen er oppsplittet i tre moduler, hhv. analyse av setning i kildepråket, overføring, og generering av setning i målpråket. Hensikten med modellen er at analyse- og genereringsdelene skal være uavhengige av hhv. målpråk og kildepråk, dvs. at den samme analyse- eller genereringsdelen vil kunne anvendes i forbindelse med alle øvrige språk. Analyse- og genereringsmodulene er altså en-språklige, mens overføringsdelen fungerer som en tospråklig bro mellom disse. Siden det for hvert språkpar må finnes en overføringsdel som er spesifikk for akkurat dette språkpar, begrenses innholdet i denne modul mest mulig.

Overføringsstrukturen

Den modulære oppbygging av systemet samt den desentraliserte utvikling ved 7 forskjellige europeiske universiteter stiller imidlertid strenge krav til en klart definert overføringsstruktur. Siden samme analysemodul skal være input til en lang rekke overføringsmoduler, og samme genereringsmodul skal likeledes motta output fra mange overføringsmoduler, må strukturen og innholdet i denne overføringsmodul nøye defineres.

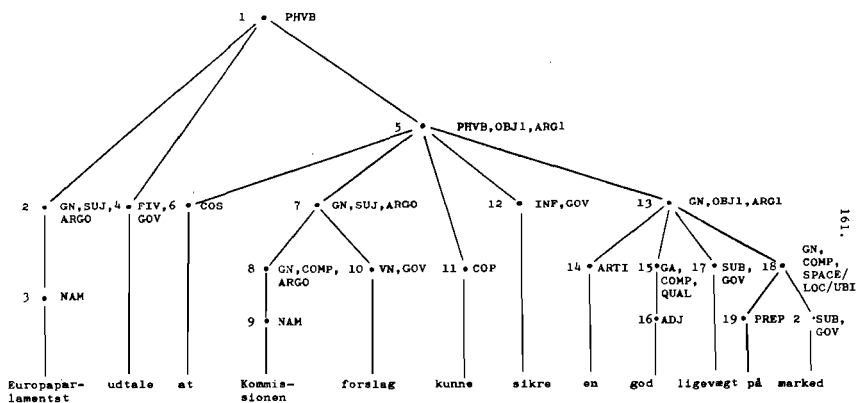
For representasjon av overføringsstrukturen er valgt dependenstrær, en annotert frasestruktur. Dvs. at analysedelen MÅ resultere i en dependens-tre-struktur, som genereringsdelen må ta utgangspunkt i. Innenfor disse rammer har utviklingsgruppene imidlertid relativt stor frihet i metodevalg, og de kan velge den type grammatikk og den strategi de måtte ønske innenfor hhv. analyse- og genereringsmodulene.

Den modulære oppbygging som er valgt stiller krav om omfattende semantisk informasjon – så store krav at prosjektets leder, M. King, anser prosjektet for svært ambisiøst når det gjelder en praktisk fungerende versjon innenfor beregnet tid (M. King og S. Perschke, 1982).

Et krav til et praktisk fungerende system må være at det ikke «gir opp» hvis det møter problemer, det må alltid produsere ett eller annet output. Systemet må derfor ha mekanismer å falle tilbake på hvis de semantiske opplysninger ikke er tilstrekkelige til å oversette en setning.

Overføringsstrukturen må derfor omfatte informasjon fra mange nivåer av lingvistisk analyse: informasjon om konstituentenes valens,

om deres overflate-syntaktiske funksjon, om morfosyntaktisk klasse, og om morfologien til de terminale elementer. Denne informasjon trekkes inn i oversettelsesprosessen etter behov for i alle tilfeller å sikre et forslag til oversettelse – et forslag som i den VERST tenkelige situasjon blir en ord-til-ord oversettelse. I tillegg kan denne omfattende informasjon ofte være nyttig når det gjelder å velge den rette form i målspråket. Nedenfor vises et eksempel på overføringsstrukturen for den danske setning «Europaparlamentet udtalte at Kommissionens forslag kunne sikre en bedre likevægt på markedet». Eksemplet er hentet fra Maegaard og Ruus (1980).



Europaparlamentet udtalte at Kommissionens forslag kunne sikre en bedre likevægt på markedet.

For en næyere gjennomgang av eksemplet henvises til artikkelen; der finnes også et eksempel på overføringsstrukturen for en fransk setning og genereringen av den tilsvarende danske. Knutenes merkelapper inneholder informasjon på 3 lingvistiske nivåer: et morfosyntaktisk, et syntaktisk og et logisk-semantisk nivå. På det morfo-syntaktiske nivå finner man opplysninger om formelle egenskaper såsom ordklasse (f.eks. NAM – egennavn) eller morfologisk klasse (FIV – finitt verb), på det syntaktiske nivå ligger opplysninger om subjekt (SUJ) og direkte eller indirekte objekt (OBJ1, OBJ2). Representasjonen på det logisk-semantiske nivå er ennå noe uavklart.

I tilknytning til dependensstrukturen ligger to andre typer av opplysninger: grammatikker og ordbøker. Disse holdes helt atskilt fra overføringsstrukturen for å oppnå størst mulig grad av fleksibilitet. I tillegg kan systemet utbygges med andre typer av opplysninger, f.eks. statistikk over ordfrekvenser e.l., i den grad dette er aktuelt.

Praktisk anvendelse

En første forsøksvis anvendelse av EUROTRA planlegges innenfor

det første året, og da begrenset til tekster som behandler landbruksspørsmål. Et bestemt emneområde og kjennskap til dette områdets terminologi vil være en viktig hjelp for å kunne generere en anvendelig oversettelse.

Imidlertid må det presiseres at man regner med at translatøren fortsatt vil være uunnværlig – den oversettelsen EUROTRA produserer vil i beste fall bare være et velegnet utkast, som translatøren må finpusse.

Litteratur

W. J. Hutchins: «Progress in documentation. Machine translation and machine-aided translation» in Journal of documentation, Vol. 34 no.2, juni 1978.

Oversiktsartikkel som presenterer en rekke systemer for automatisk oversettelse og de modeller systemene benytter. Artikkelen omhandler ikke EUROTRA-prosjektet, men er en grundig presentasjon av historien bak automatisk oversettelse, fra de tidligste forsøk i 1950-årene fram til de teorier som EUROTRA bygger på.

Alle de følgende artikler omhandler EUROTRA-prosjektet:

King, M.: «Eurotra - A European System for Machine Translation» in Lebende Sprachen, 1980.


King, M., og S. Perschke: «Eurotra and its objectives». Multilingua, april 1982.

King, M.: «A new attempt at machine translation». Upublisert.

Maegaard, B.: «The transfer of finite verb forms in a machine translation system». Abstract til Coling 82 i Prag.

Maegaard, B. og H. Ruus: «Strukturering af lingvistiske data til brug ved maskinoversættelse. EUROTRAS overførselsstruktur». Foredrag ved Nordiske datalingvistikdage, Kbh. 9-10 okt. 1979.

Maegaard, B. og H. Ruus: «Structuring linguistic information for machine translation» in Human translation - machine translation. Papers from the 10th annual Conference on Computational Linguistics in Odense, Denmark. NOK 39. Noter og kommentarer fra Romansk institut, Odense Universitet, august 1980.



*Holger Hagan,
stipendiat ved Universitetet
i Tromsø*

EDB OG HUMANIORA

EDB-tjeneste for språk og litteratur i Tromsø

I 1981 begynte *Holger Hagan* som edb-stipendiat ved Universitetet i Tromsø (UiTø) ved Institutt for språk og litteratur (ISL).

Humanistiske Data tok kontakt med ham for å få et inntrykk av hvordan situasjonen for humanistiske edb-brukere er ved UiTø.

Edb er et etablert hjelpemiddel innenfor en rekke områder ved universitetet. Dette viser seg ved at datamaskiner brukes av mange personer innenfor fagområdet, og ved at man gjerne har en fast ansatt konsulent. ISL hører ennå til de fagmiljøer hvor relativt få bruker edb, og hvor man ikke har en konsulent i full stilling. Jeg ser det derfor som en av mine oppgaver å legge forholdene til rette slik at flere humanister ser de mulighetene som maskinell behandling av data gir, sier Hagan.

Hvilke muligheter er det du har i tankene?

Nærliggende eksempler er store tidsbesparelser ved forenklet behandling av vanlig tekst, og fremstilling av ordlister og konkordanser.

Du nevner at edb er sterkere etablert i andre miljøer ved UiTø. Hvilke miljøer sikter du til?

Først og fremst hele det matematisk-naturvitenskapelige området, men også innenfor samfunnsvitenskapene, museumssektoren, fiskerifag og medisin.

Du har nå vært stipendiat i over et og et halvt år. Hvilke ting er det du i særlig grad har vært opptatt av, ved siden av ditt eget forskningsprosjekt?

Man kommer ikke forbi at nye brukere må krysse en terskel når de setter i gang med å bruke datamaskiner som et hjelpemiddel. Jeg ser det som et poeng å gjøre det lettere å krysse denne terskelen. Ved ISL finnes det forskere som arbeider med tekstmateriale i flere ulike alfabeter, f.eks. kyrillisk og samisk. Det sier seg selv at datamaskinen blir mer in-

teressant for slike brukere i samme grad som den kan «snakke deres språk». Eller mer korrekt: skrive deres bokstaver. Det ideelle er at skjermen kan tilby flere tegnsett, og at skriveren kan gjøre det samme. På utskriftssiden har ISL i dag en Facit 4542 matriseskriver som gir gode muligheter for programstyrt valg fra et stort tegnrepertoar, samtidig som man selv kan få laget nye tegnsett etter egne spesifikasjoner. Matrisehøyden setter selvsagt grenser for valgfriheten, men problemet lar seg løse på programsiden. En annen begrensning ligger i at utskriftskvaliteten ikke kommer på høyde med typenhjulsskrivere. Når det gjelder mulighetene for å få skjermer med alternative tegnsett, så er jeg som filolog forbauset over at produsenter og leverandører i så liten grad er opptatt av dette området. Verden over anvendes terminaler i situasjoner hvor et fleksibelt tegnsett, gjerne med brukerdefinerte tegn, er en opplagt salgssfordel. Likevel opplever man at en større terminalprodusent forlanger to tredjedeler av terminalprisen for å forandre tegnrepertoaret, uansett hvor minimal endringen måtte være. Med en slik prispolitikk kan man neppe vente særlig interesse fra kjøpersiden.

Mitt poeng er at desto nærmere maskinens tegnsett ligger den språko-orienterte brukers behov, desto lavere blir terskelen vedkommende må krysse. Dette området har jeg viet mye tid og oppmerksomhet, og jeg kommer til å fortsette med å se meg om etter gode løsninger.

RAPPORTER

Senteret feirar 10-års jubileum

Den 3. juni arrangerte Senteret ei samkome for medlemene i Rådet for humanistisk forskning i NAVF, gjester frå Universitetet i Bergen og dei tilsette ved Senteret for å markera at det no er 10 år sidan Senteret starta si verksemd.

Det var helsingar frå ordføraren for NAVF-humanistane, dosent *Ola Kai Ledang*, og frå ass. universitetsdirektør *Bjørn Henrichsen*. Henrichsen sa seg glad for det nære samarbeidet som alltid har vore mellom Senteret og Universitetet i Bergen. Sjølv om dei materielle tilhøva ved Universitetet no er vanskelegare enn før, vonar eg at det vil la seg gjere å bygge samarbeidet vidare ut i åra som kjem, sa Henrichsen mellom anna.

- Dei humanistiske fagmiljøa ved Universitetet i Bergen var tidleg ute med å ta i bruk datamaskinelle hjelpemiddel i forskninga, sa direktør *Jostein H. Hauge*, som har leia verksemda i alle dei 10 åra. Dette gjorde det naturleg å leggje eit nasjonalt edb-senter for dei humanistiske faga til Universitetet i Bergen i 1972.



Senterets direktør, Jostein H. Hauge, ønsker velkommen. Bakerst i bildet sees fra venstre bl.a. styreformann Sofie Rogstad, tidl. styreformann Kjell Falck, ass. univ.dir. Bjørn Henrichsen og styreformann i NAVF, Inge Lønning.

Sentret vart først starta som eit 5-års program, men i 1977 vedtok NAVF å etablere det som ei permanent datateneste for dei humanistiske fagmiljøa. Gjennom åra har Sentret ytt hjelp til ei lang rekkje edb-prosjekt i dei humanistiske fagmiljøa både innanfor og utanfor universiteta. Sentret har lagt vekt på program- og metodeutvikling, og sett i verk eigne opplæringsprogram og informasjonstiltak. Hauge meinte det var viktig at Sentret heile tida arbeider vidare med å fornye sine verkemiddel med utgangspunkt i sin kunnskap om målsetting og arbeidsmåtar i dei humanistiske faga. Han kom særleg inn på korleis Sentrets arbeid vil bli endra etter kvart som edb blir eit velkjent hjelpemiddel i dei humanistiske faga. I framtida vil meir vekt bli lagt på mellom anna metodeutvikling og opplæring av forskarar.

Per-Kristian Halvorsen, som for tida er forskar-NAVF knytt til Sentret, gav på tilstellinga ei innføring i ulike arbeidsfelt innanfor datamaskinell lingvistik. Han samarbeider sjølv med Sentret om utvikling av metodar for automatisk språkanalyse i nær fagleg kontakt med kjende forskningsmiljø i USA. I framtida vil språkvitskapleg innsikt bli ei særskild viktig kunnskapskjelde når ein skal utvikle brukar-vennlege former for kommunikasjon med datamaskin, meinte Halvorsen.

Datalogi og humanisme – to uforenlige verdener?

Ny datalogisk utdanning for humanister

Datalogi og de humanistiske fag – hvilken sammenheng findes mellem disse? Er det ikke sådan at interessefeltet til teknologien og humanisten udgør to adskilte og uforenlige verdener?

Ikke nødvendigvis, mener en arbejdsgruppe som har været nedsat ved Aarhus Universitet for at udrede problemerne omkring oprettelsen af en datalogisk uddannelse for humanister. På den ene side har nemlig datalogien behov for at inkorporere en række af de fagfelter, som humanister har specialiseret sig på, f.eks. kommunikation gennem og analyse af naturlige sprog. På den anden side må humanisten inddrage viden om og forståelse af teknologien i sit arbejdsområde, både for at kunne anvende den moderne teknologi hensigtsmæssigt indenfor sit arbejde, og for at få et så vidt mulig fuldstændigt og korrekt billede af de sociale og kulturelle forhold som i sidste instans er humanistens arbejdsfelt.

Arbejdsgruppen bag udredningsarbejdet har været sammensat af repræsentanter fra både de datalogiske og de humanistiske fagområder (note 1), og der har været afholdt en seminarrække med deltagere fra begge sider. Målet med udredningsarbejdet har været at foretage en afklaring af det faglige grundlag et sådant studie kan bygge på, herunder bl.a. en begrebsafklaring. Man har forsøgt at fremdrage forskelle og ligheder mellem det datalogiske og det humanistiske fagområde og således få et fingerpeg om hvilke faglige traditioner indenfor de to fag en kombinationsuddannelse kan bygge på. Den her foreliggende præsentation

er et sammendrag af nogle af de temaer, udredningsgruppen har behandlet.

Når resultatet af udredningsarbejdet er interessant også her i landet, skyldes det at man har taget udgangspunkt i skæringspunktet mellem to traditionelt modsætningsfyldte fagområder: det datalogiske og det humanistiske. Resultatet af denne «konfrontation» har videre interesse end akkurat en studieplan isoleret betragtet, idet den indgår i den omfattende debat omkring datateknologiens konsekvenser for samfundet totalt set og for det enkelte individ.

I en konfrontation mellem mennesker og maskiner har humanisten traditionelt i kraft af sine studier bedre forudsætninger for at forstå menneskets holdninger og synspunkter, mens den traditionelle edb-teknokrats arbejde (i karikeret form) er rettet mod maskiner uden mennesker. Er det da ikke muligt, ved at sammenarbejde de to arbejdsfelter, at tilføre datalog-rollen nye kvaliteter og ny viden omkring samspillet mellem menneske og maskine?

Uddannelsens sigtemål er at uddanne kandidater som kan deltage i systemarbejde, og som er specielt kvalificerede på områder knyttet til grænsefladen mellem bruger og system. Endvidere kan kandidater naturligtvis tænkes at indgå i traditionelle humanistiske arbejdsfunktioner som anvender edb-teknologi som hjælpemiddel.

Mennesker og maskiner kræver hver sit sprog

Fagligt koncept for uddannelsen

I arbejdet med grænsefladen bruger/system skal systemarbejderen producere forskellige beskrivelser af system og arbejdsprocesser, f.eks. nogle som er beregnet på programmereren som skal instruere maskinen, andre til brugeren som skal anvende systemet i sit arbejde, og atter andre til andre edb-fagfolk som skal vedligeholde systemet. Disse forskellige modtagergrupper (og måske endnu flere) har allesammen forskellige behov for information, forskellige erfaringer og/eller interesser, og stiller forskellige krav til systembeskrivelsen.

Erkendelsen af disse forskellige krav er det som danner grundlag for det faglige koncept for uddannelsen: man kræver at datalogen skal lære at lave modtager-orienterede, parallelle men relaterede samt ligeværdige beskrivelser af arbejdsprocesser under forandring. Dette skal nærmere forklares:

a) Modtagerorienterede beskrivelser

At beskrivelsen skal være modtagerorienteret betyder at systemarbejderen må kunne analysere de forskellige og ofte konfliktfyldte modtagerbehov der findes. Først og helt grundlæggende må hun skelne mellem tekster som skal forstås af datamaskinen (dvs. programmer), og de som skal kunne forstås og bruges af mennesker i tilknytning til systemet. Denne skelnen forekommer måske indlysende, men er ikke slet så oplagt ud fra en mere traditionel datalogisk tankegang som (i karikeret form) kan udtrykkes således: «Der findes een teknisk, neutral måde at

beskrive et edb-system på. Andre beskrivelser må ses som ufuldkomne populariserede udgaver af denne »rigtige« beskrivelse.»

Vi ved at computerens software historisk set er udviklet i et naturvidenskabeligt-matematisk miljø, og det er ikke mærkeligt at dette miljø har sat sit præg på de beskrivelser, der laves af edb-systemer. Dette får især betydning i kraft af den forskellige begrebsbrug her og i andre miljøer. Mens den aristoteliske begrebstype – hvor et begreb kan karakteriseres ved et sæt af nødvendige og tilstrækkelige betingelser – spiller en væsentlig rolle i naturvidenskab og matematikk, er det vi kan kalde hverdagslivets begreber af en anden art, baseret på prototyper. Prototypiske begreber er mindre skarpt afgrænset end aristoteliske dvs. at en genstand kan falde mere eller mindre ind under begrebet afhængig af hvor meget eller lidt den ligner prototypen, mens aristoteliske begreber er skarpt og klart afgrænset.

Disse forskelle i begrebsbrug er det vigtigt at tage hensyn til når man producerer tekster i et edb-baseret system. Edb-arbejderen har traditionelt udviklet en høj grad af forståelse af den måde maskiner forstår sprog og begreber, men glemmer let at denne begrebsbrug er vidt forskellig fra den, som en anden ikke-edb-uddannet part i kommunikationen benytter. Sprogbrugsanalytikeren ved, at en teksts anvendelighed bl.a. er betinget af at den er udformet med et sprog og en referenceramme, så at modtageren kan anvende den. Det er med andre ord vigtigt at gøre sig klart hvem beskrivelsen er rettet mod – maskinen eller mennesket!

Endvidere må, som før nævnt, skelnes mellem forskellige grupper af mennesker tilknyttet systemet, hvilken baggrund de har og hvilke interesser.

b) Parallelle, relaterede og ligeværdige

I forlængelse af kravet om modtagerorientering ligger at beskrivelserne skal være parallelle, relaterede og ligeværdige: teksterne beskriver «samme sag» set fra forskellig side, og er derfor parallelle og relaterede til hinanden. Ingen beskrivelse kan imidlertid udnævnes til at være den vigtigste og den hvortil i sidste instans alle andre kan reduceres – de forskellige beskrivelser er ligeværdige. Det modsatte synspunkt er (igen i karikeret form) at «programmet er det vigtigste dokument, hvortil i sidste instans alle andre dokumenter skal relateres og ved hjælp af hvilket de defineres».

I udredningsarbejdet har man valgt at skelne mellem systemiske og ekstrasytemiske beskrivelsesformer. En systemisk beskrivelse er en beskrivelse som hviler på systembegrebet, hvor et system f.eks. kan defineres således: «Et system er en del af verden, som vi vælger at betragte som en helhed, adskilt fra resten af verden i en periode, en helhed som vi vælger skal indeholde en samling komponenter, hver karakteriseret ved en udvalgt mængde af dataenheder og mønstre, og ved handlinger, som kan indeholde den selv og andre komponenter.» (K. Nygaard). Eksempler på systemiske beskrivelser er programmer, systembeskrivelser og en

række grafiske beskrivelser.

Betegnelsen «ekstrasystemiske» beskrivelser bruges som en samlebetegnelse for beskrivelser som ikke hviler på systembegrebet. Pr. definition kan ekstrasystemiske beskrivelser ikke reduceres til systemiske, idet de indeholder information som ikke kan rummes i en systemisk beskrivelse. Vi har med andre ord to kvalitativt forskellige beskrivelser af samme system, beskrivelserne er relaterede og parallelle, samt ligeværdige – ingen af dem er «vigtigere» end den anden.

c) Arbejdsprocesser under forandring

I uddannelsens faglige koncept hedder det videre at der skal produceres beskrivelser af arbejdsprocesser under forandring. I systemudviklingsarbejdet er det ofte et problem at tilpasse eksisterende systemer til ændringer i virksomheden eller samfundet. Et af målene er derfor i videre udstrækning at indarbejde en historisk, dynamisk synsmåde i systemudviklingen.

Når man fremhæver det dynamiske, foranderlige element i systemudviklingsarbejdet møder man igen spændingen mellem den matematisk-naturvidenskabelige tænkemåde og den humanistiske. Forskellen kan belyses gennem de to yderpositioner, igen i karikeret form: Mens den naturvidenskabelige forsker vil konstruere sin model med henblik på en vis stabil brugsperiode og ser destruktion af modellen som en exceptionel hændelse, ser den humanistiske forsker i højere grad konstruktion, brug og destruktion af modellen som en evig iteration. Et mål i systemudviklingsarbejdet bliver i højere grad at bygge på den historiske synsmåde i systemudviklingen, som humanister er velkendt med fra sine fagområder.

Forskningsprofil

De forskellige videnskabsteoretiske traditioner som findes inden for naturvidenskab og humaniora betyder at forskning rettet mod undersøgelser af den nye teknologiske konsekvenser tematiseres forskelligt inden for fagområderne. Denne forskel vil man i det nye studium forsøge at udnytte på en frugtbar måde. Vi citerer fra rapporten:

«Fra datalogisk side har vægten ligget på anvendelser af teknologi og arbejdsformer ved teknologiudvikling, mens en intern kritik af de fagligt-tekniske begreber har været lavere prioriteret.

Fra humanistisk side har man væsentligst interesseret sig for de videnskabsteoretiske og teknologihistoriske sider, hvor TEKNOLOGI her både har at gøre med »selve teknikken« OG læren bag den, dvs. de naturvidenskabelige metoder og erkendelsesinteresser. Manglende teknisk indsigt har givet de humanistiske projekter DERES begrænsninger.

Virkeligheden har og vil forandre sig afgørende for befolkningen gennem »den 2. industrielle revolution«; vi vil opleve en hidtil uset sammenvævning af kultur og teknologi. En faglighed, der indeholder en videnskabsteoretisk opsplitning mellem teknologi og humaniora, vil være ude af stand til at forstå denne nye virkelighed. På den anden side er

der, som vi har set, i øjeblikket væsentlige teoretiske og metodiske forskelle. Hvad gør man? Svaret ligger i en strategi vi kan benævne »en produktiv kritisk spænding«.

Den går ud på at man vægter de to traditioner ligeværdigt i studiet, at man ikke hugger en hæl og klipper en tå for dog at få husfred, men at man i en længere periode forsøger at lade de to traditioner spille kritisk mod hinanden. Fastholdelsen af en sådan produktiv spænding er afgørende for uddannelsens karakter af eksperimenterende tværfaglighed, hvis formål ikke så meget er at forsone forskelligheder, men snarere at oparbejde en ny videnskabelighed om det nye genstandsområde, virkeligheden selv har skabt.» (Rapport nr. 3-4, s.35).

I den grad man i det videre arbejde vil være i stand til at formidle og drage nytte af spændingsfeltet mellem de to forskningstraditioner, således som det har været forsøgt i forprojektet, er der al grund til at imøde-se fremtidige resultater med stor spænding!

Note 1. Arbejdsgruppens deltagere:

Peter Bøgh Andersen, Nordisk institut.

Jens Erik Kristensen, Institut for idehistorie.

Arne Kjær, Datalogisk afdeling.

Bjarnar Sveigård, Datalogisk afdeling.

Overstående præsentation bygger på et møde med arbejdsgruppens deltagere i Århus, februar 1982, samt de udredningsrapporter som foreligger.

Eva Møller

Seminar om EDB-metoder i humanistiske fag

I tida 22.-24. mars 1982 ble det årlige «utviklingsseminar» arrangert på Finse. På disse konferansene samles Senterets konsulenter, representanter for edb-tjenestene ved de humanistiske fag ved universitetene og en del inviterte kollegaer fra universiteter og høyskoler som bruker edb i sitt fagarbeid.

Som vanlig var det utpekt et par hovedtema som foredrag og debatter konsentrerte seg om. Denne gangen ønsket man å ta for seg mikromaskiner, bruk av edb i forbindelse med publisering og analog-digital konvertering.

Under første sesjon tok *Nils Myklebust* fra Storm System A/S for seg de 16 bits mikromaskinene som nå kommer. Han nevnte en del fordeler med disse, bl.a. mulighet for flere terminaler, større programmer, raskere prosessering og at det blir enklere å flytte software fra stormmaskiner til mini- og mikromaskiner. Han orienterte ellers om hvilke programtilbud som vil være tilgjengelige for disse maskinene.

Dette siste temaet var emnet for et foredrag av *Morten Kristiansen* fra Østfold DH.

Under den andre hovedsesjonen innledet *Mauritz Fredriksen* fra firmaet *Stokes & Fredriksen A/S* og *Knut Hofland* fra Senteret om fotosats og fotosettere. Fredriksens foredrag var en detaljert gjennomgang av hele publiseringsprosessen fra manusskriving til den ferdige publikasjon. Han tok spesielt for seg fotosettere og drøftet de forskjellige tekstsystemer, maskinalternativer m.v. Knut Hofland tok for seg det utredningsarbeidet som har foregått ved Universitetet i Bergen i forbindelse med mulig kjøp av fotosetter og en eventuell tilkopling til Univac-anlegget av programsystemet Photodoc for produksjon av fotosats. Når det gjelder edb og publisering vises det for øvrig til egen artikkel i dette nummeret av *Humanistiske Data*.

Til slutt tok *Tor Sverre Lande*, Institutt for informatikk, UiO, for seg mikromaskinen PERQ som bl.a. kan tilkoples en laserprinter og lage utskrifter av meget høy kvalitet. I tillegg har denne maskinen andre fordeleler, f.eks. en avansert grafisk skjerm.

Eric Sandnes, Universitetet i Trondheim, gjennomgikk prinsippene for analog-digital konvertering og hvorledes dette var brukt i et system for digital lagring av talesekvenser og gjentatt avspilling av disse.

Disse årlige sammenkomstene har vist seg å være meget nyttige. I tillegg til å få utvidet sine kunnskaper får deltakerne god mulighet til å sette seg inn i den virksomhet som foregår ved de respektive læresteder. Og utviklingen vil fortsette – neste seminar blir holdt vinteren 83.



Direktør Mauritz Fredriksen holdt foredrag om fotosats og fotosettere.

Reiserapport fra Hannovermessen april 82

Den årlige Hannovermessen for teknisk utstyr ble i år arrangert i perioden 21.-22. april. Årets arrangement omfattet i virkeligheten 10 messer, med ca. 5000 utstillere fra 40 nasjoner.

På 10 år har antall utstillere steget med 40%. Mens det i 1972 var 3 utstillere fra USA, deltok det i år over 100. Selv om utstillingsarealet økes hvert år, var det over 100 firmaer som ikke fikk plass på årets messe. Dette skulle gi en viss pekepinn på den betydningen bedriftene legger i å delta på Hannovermessen.



Fra CeBIT-messen, Hannover 1982

CeBIT-messen for kontor- og datateknikk la beslag på 4 haller, på til sammen 166.000 kv.m. Man fikk på CeBIT et enestående innblikk i hvor langt utviklingen er kommet på områder som software, kontor-kommunikasjon, datateknikk, tekstbehandling, grafisk databehandling, mikrofilmteknikk og mikrodatamaskiner.

Den enorme satsingen av leverandørene på mikrodatamarkedet var påtakelig. Så å si samtlige leverandører demonstrerte sine egne mikrodatamaskiner.

Mikrodataanlegg basert på en 8 bits prosessor var fremdeles dominerende, men 16 bits maskinene kommer med stormskritt. Texas Instruments viste til og med en 16 bits personal computer. CP/M og MP/M baserte operativsystemer var i klart flertall på 8 bits maskinene. MP/M sto også sentralt hos flere leverandører.

Prisutviklingen på denne type datautstyr er sett fra brukerens side svært positiv. Introduksjon av de nye 16 bits mikrodatamaskinene har ført til at flere leverandører har senket prisene kraftig. Dette gjelder særlig i den nedre delen av markedet.

Mens prisene på maskinvaren synker, fikk vi igjen bekreftet at lagringskapasiteten på platelagrene bare øker og øker. Winchester-platelagrene var dominerende, og de dekker nå spekteret fra små 5 Mb enheter opp til utstyr med en kapasitet på over 100 Mb.

Messen viste tydelig at skrivere basert på laserteknologi var kommet for å bli. De japanske leverandørene Ricoh og Hitachi demonstrerte laserskrivere på størrelse med små kopieringsenheter. Hastigheten her lå på ca. 12 A4-sider i minuttet.

En fascinerende konfigurasjon var Toshiba's arkivstasjon, bestående av laserskriver og optisk disk. Disken hadde en lagringskapasitet på 10.000 A4-sider, men kunne foruten tekst lagre bilder og grafikk. Det vil dessverre ta en tid før vi får se denne i Norge.

Temaet for CeBITs spesialutstilling 82 var «Praksis og perspektiver i kontorkommunikasjon.» Her viste man trends – spesielt nytt utstyr innen talesyntese, tekst- og bildebehandling og datakommunikasjon.

Mitt inntrykk er at man rent teknisk er kommet langt på disse feltene i dag. Men det gjenstår uten tvil mange praktiske og organisasjonsmessige problemer før hele spekteret av slike hjelpemiddel kan inngå som standard utstyr i administrativt arbeid.

Sigbjørn Århus

Datamaskinell informasjonssøking

Rapport fra Nord IoD 5 og Nordisk online brukerdag

Den 14.-16. juni ble den 5. nordiske konferanse innen informasjons- og dokumentasjonsområdet avholdt i Trondheim, og 17. juni Nordisk online brukerdag, likeledes i Trondheim.

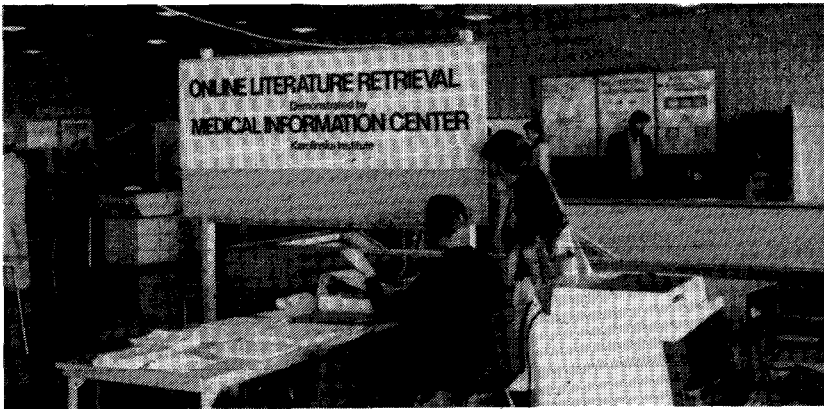
«Rett informasjon i rett tid til rett person» – dette er en hyppig brukt målsetting for et informasjonssystem.

F. Backlund fra Delegationen för vetenskaplig och teknisk informationsförsörjning, Sverige, valgte i sitt foredrag å knytte denne målsetting sammen med en vanlig modell for et informasjonssystem:

kunde – referansesystem – lokaliseringssystem – bestand

I denne kjede av systemdeler gjelder det i prinsippet at retningen fra kunde til bestand dreier seg om INFORMASJONSSØKING, mens retningen fra bestand til kunde gjelder DOKUMENTLEVERING. På bakgrunn av denne modell vurderte han anvendelsen av ulike teknologier som står til IoD-området disposisjon.

En lang rekke av de hjelpemidler IoD-området har tatt i bruk gjennom de senere år, er knyttet til første moment i kjeden, INFORMASJONSSØKING. Dette gjelder først og fremst de store bibliografiske databaser som informasjonstjenesten kan abonnere på tilgang til å søke i direkte. Slike databaser er etablert såvel i Norden som ute i Europa, USA og andre fjernere plasser. De bibliografiske databaser er for det meste knyttet til et mer eller mindre snevert definert fagområde, og gir henvisninger til referanser for litteratur innen dette område. Blant de bibliografiske databaser som ble spesielt omtalt på konferansen var f.eks.



Omkring 13 utstillere deltok ved IoD-konferansen i Trondheim. Her demonstreres mulighetene for online litteratursøking innen det medisinske området.

databaser som dekker medisinsk forskning, oljeindustri, statistisk materiale og energiområdet, samt et felles nordisk prosjekt, NOSP (Nordisk samkatalog over periodika), som er en felleskatalog over utenlandske periodika i nordiske forskningsbiblioteker.

I tillegg til de bibliografiske databaser finnes en rekke «faktadatabaser» av forskjellig art, f.eks. databaser som gir opplysninger om patenter, nye produkter osv., især til bruk for handel og industri.

Etter hvert som tilgangen til litteraturreferanser blir raskere, lettere og mer omfattende vokser problemet med den andre siden av informasjonsmodellen, DOKUMENTLEVERINGEN. Lite nytter det forskeren å få henvisning til aktuelle artikler i løpet av timer, hvis framskaffingen av artikkelen tar måneder.

Også dette problem søkes løst ved utvidet bruk av ny teknologi – men slike systemer er hittil mindre utbygget enn referansesøkingen. Et løsningsforslag representeres av Telefaxtjenesten fra Televerket, muligheten for fjernkopiering over telefonnettet. Overføringen skjer gjennom det offentlige telefonnett, og krever en spesialmaskin av omfang som en liten kopimaskin hos såvel sender som mottaker. Overføringen av en A4-side tar fra 2 til 6 minutter, og resultatet er en fullt lesbar papir-kopi.

En annen løsning som det nå eksperimenteres med, er dokumentoverføring via satellitt, som Svein Østevik fra Norsk senter for informatikk redegjorde for. I prøveprosjektet Apollo (Article procurement with online local ordering) forestiller man seg at det opprettes et lager av kilde-dokumenter som kobles til databasesystemet for on-line søking. Brukeren søker via datanettet online i databasen, og når han finner et relevant

dokument bestiller han dette (eller enkelte sider av det) via sin terminal. Dokumentet blir da skrevet ut direkte på brukerens facsimileapparat.

Endelig benyttes nyere teknologi til å framskaffe dokumenter i form av microfiche i stedet for papirkopi – en microfiche er billig og enkel å framstille hvis dokumentet allerede finnes i maskinlesbar form, og den er enkel å sende med posten.

Ovenstående bilde av bibliotekaren som via sin terminal og de interasjonale nettverk har rask og direkte tilgang til informasjon i store databaser, virker kanskje forlokkende – men er til dels en idyllisering av de faktiske forhold. Diskusjonen blant de nordiske online-brukere tyder på at det praktiske arbeid er preget av problemer, hvorav de viktigste er knyttet til nettverkets (og i enkelte tilfeller databasens) manglende stabilitet. Hvis det er umulig å koble seg inn på nettverket nytter det lite at den enkelte database er omfattende og kompleks i sine referanser. Videre oppleves det som et stort problem for brukere som har behov for å søke i flere databaser, at de ofte for hver database må lære seg et nytt søkespråk.

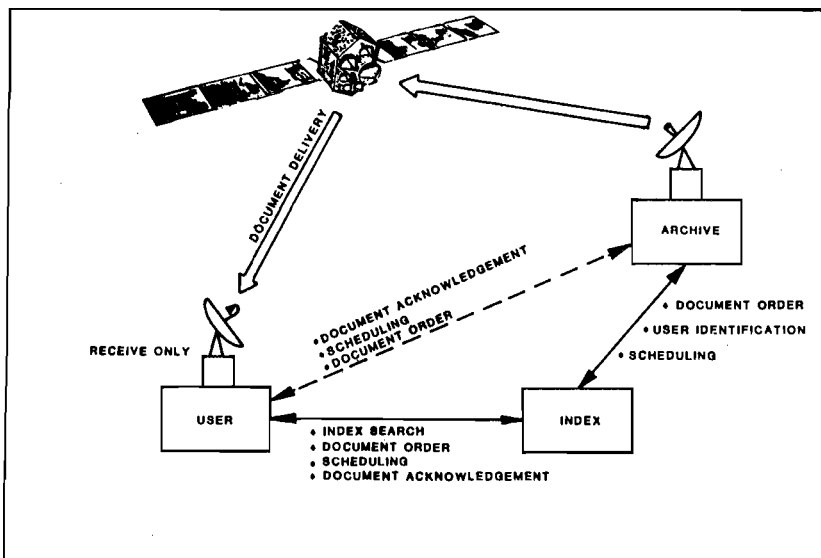
En mulighet for å avhjelpe sistnevnte problem er å definere et felles søkespråk, som kan brukes på alle databaser tilknyttet et bestemt nettverk – men dette krever et omfattende og ressurskrevende arbeid med utarbeidelse av felles standarder, ble det framhevet av representanten fra det nordiske SCANNET.

En annen mulighet for å lette dette problem redegjorde *Liv Holm*, Norsk dokumentdata, for i sitt foredrag om program-til-program kommunikasjon for distribuerte databaser. I prosjektet BIBNETT har man utviklet et system for kommunikasjon mellom databaser på ulike stor-maskiner, som betyr at brukeren bare trenger å søke i en (kjent) database. Søkningen blir da, uten at brukeren trenger å bry seg med det, automatisk utstrakt til de øvrige databaser som er tilknyttet systemet. I motsatt fall måtte hver enkelt database kontaktes gjennom en separat søking og med sitt eget søkespråk.

Enda en hjulpekilde, Userkit, ble presentert av *P. W. Williams*, University of Manchester, England. Userkit er en mikromaskin, som installeres som grensesnitt mellom brukeren og on-line-søkningen. Userkit kan f.eks. benyttes til hjelp ved oppkoplingen mot nettverket, eller den kan brukes til å senke kostnadene forbundet med søking, idet brukeren først legger sin søking inn på mikromaskinen, dernest kobler opp mot den database han ønsker å søke i, hvoretter Userkit overfører søkeordene raskere enn brukeren er i stand til manuelt.

Økonomiske aspekter har i liten grad vært berørt i ovenstående – og med god grunn. De ulike tjenester og muligheter er langt fra gratis å benytte seg av, men prisene varierer sterkt. I siste instans blir to spørsmål stående:

- Hvor mye er samfunnet eller den enkelte institusjon/organisasjon villig til å betale for informasjon – dvs. hvor høyt prioriteres informasjon?



Forsøksprosjektet APOLLO. Dokumenter overføres fra et elektronisk arkiv til brukerens terminal. (NORDINFO-NYTT 1 - vol. 5-82).

- Hvilke grupper og konstellasjoner har tilstrekkelige midler til å få tilgang til informasjon?

Ett eksempel kan belyse problemstillingen: De internasjonale litterærtjenester vedrørende petroleum er så kostbare (iflg. *Grete Schancke*, Oljedirektoratet, er prisen ca. 240.000 kr. pr. år) at norske offentlige institusjoner ikke har tilgang til dem. De multinasjonale oljeselskaper, og dermed også de norske avdelinger av disse, har derimot abonnement på disse tjenester – og dermed tilgang til informasjon som de norske styrende organer må unnvære.

Konferansen viste klart at informasjon i dagens samfunn er en omfattende vare, og en viktig ressurs som må forvaltes med omtanke. Samtidig med at den nye teknologi åpner for nye muligheter, pålegger den også IoD-området og samfunnet nye oppgaver og et større ansvar for å forvalte den viktige ressurs som informasjon er.

Eva Møller

ALLC PISA 7.-11. juni 82

Det syvende internasjonale symposium om datamaskinen i litterær og lingvistisk forskning ble holdt i Italia i år. Symposiet samlet omlag 120 deltakere fra USA og Europa.

Det foreligger ikke noe skriftlig materiale fra konferansen ennå, derfor blir det følgende ingen uttømmende rapport. Det ble holdt omlag 50 presentasjoner i 2 parallelle sesjoner fordelt over 3 dager. I tillegg var det organisert arbeidsgrupper. Foredragene omhandlet bl.a. følgende områder: forfatterstudier, konkordanser, databaser, utdanning, leksikografi, litterær statistikk, kvantitativ lingvistikk, stilistiske analyser og programvare. Jeg vil nedenfor referere fra noen av foredragene:

Fra norsk hold ble det presentert foredrag av Eirik Lien (EDB-tjenesten i Trondheim) og *Per Vestbøstad* (Norsk Tekstarkiv). Liens foredrag omtalte det arbeidet som han og *Anne Karin Ro* hadde utført i forbindelse med en datamaskinregistrering av ortografiske variantforekomster i 12 håndskrifter som inneholder *The General Prologue to the Wycliffite Bible*, med henblikk på datering og lokalisering av håndskriftene. Vestbøstad omtalte arbeidsoppgavene i Norsk Tekstarkiv.

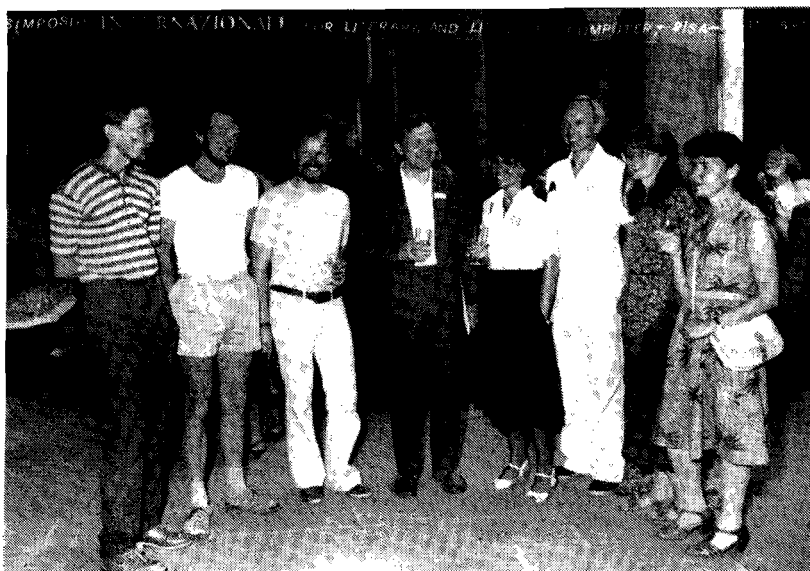
Susan Hockey (Oxford University Computing Service) og *M. D. Harris Fosberg* (Loyola University, New Orleans) hadde foredrag som omhandlet EDB-utdanning for humanister. Hockey fortalte om utdanningsopplegget ved Oxford, et opplegg som er myntet på å utdanne ferdig utdannede humanister som ønsker å bruke EDB i sitt arbeid. *M. D. Harris Fosberg* underviste imidlertid studenter under hovedfagsnivå. Hun beskrev undervisningsopplegget, som hun hadde prøvd å legge opp slik at studentene ville få tilstrekkelig innsikt i metoder og problemer knyttet til EDB-behandling i humaniora. På denne måten mente hun det var en reell mulighet for dem å kunne velge et humanistisk fag, og også bli interessert i å anvende EDB. Det var stor enighet i salen når det gjaldt dette punktet.

I. Crompton-Robert og *M. Warren* (England) hadde brukt OCP (Oxford Concordance Package, omtalt i HD nr. 1-82) for å analysere data om bilingvale barns språkbruk. De var svært fornøyd med OCP som ble omtalt som en profesjonell pakke som var lett å bruke. De savnet imidlertid en innføringsmanual.

Analysen hadde bl.a. gitt følgende resultater:

Et flerspråklig barn stiller oftere spørsmål, yngre flerspråklige barn trenger flere tenkepauser, flerspråklige barn må være mer oppmerksomme og bruker oftere ja/nei svar enn ettspråklige barn.

I prosjektet «A lexicon of Greek personal names», presentert av *T. G. Ashplant* (England), ønsket en å registrere alle greske personnavn fra de tidligste historiske tider fram til den islamske erobring i midten av det syvende århundre. Navnene er tatt fra gresk og latinsk litteratur, papyrus, mynter, gjenstander og fremfor alt inskripsjoner.



Noen av de nordiske deltakerne på ALLC-symposiet i Pisa.

Informasjonen, som vil bli publisert, inneholder følgende elementer: navn, sted, dato og referanse til hvor og hvordan navnet ble funnet og innlemmet i basen. Data er maskinregistrert og ble viderebehandlet med programpakken FAMULUS og reformatert med et SPITBOL-program. Arbeidet er gjort i nært samarbeid med Oxford University Computing Service, og data vil bli publisert v.h.a. ovennevnte senters nyinnkjøpte fotosetter (Lasercomp).

P. Bratley og *D. St-Michel* (Canada) fortalte i sitt foredrag «Scenes, Maps and Trellises» om multidimensjonal skalering, en matematisk metode for beskrivelse av handlingen i et skuespill eller en opera. Metoden er implementert i et program.

Det settes opp en matrise med en rekke for hver person og en søyle for hver scene. I matrisen markeres det med 1 når en person er til stede i en scene og 0 ellers. Deretter beregnes «avstanden» mellom hver av scenene. Avstanden mellom to scener er her definert som antall plasser der scenenes søylevektorer er forskjellige, altså simpelthen antall personer som kun er med i en av de to scenene. Deretter konstrueres en «figur» av punkter, ett punkt for hver scene, der de «sceniske» avstandene i så stor grad som mulig blir bevart. Figuren er helst i 2 dimensjoner så den kan avbildes på et papir. (Velger man 1 dimensjon i stedet, kan en også trekke inn tidsforløpet i handlingen som den andre dimensjon x).

På det ferdig utskrevne «kartet» kan man se hvordan scener fordeler

og grupperer seg, og man kan trekke linjer fra scene til scene for å se handlingsforløpet.

Foredragsholderen illustrerte forskjellene mellom Shakespeares og Verdis «Othello» ved hjelp av denne metoden.

Det faglige utbyttet av konferansen ble dessverre noe redusert p.g.a. dårlige lydforhold i foredragssalene og sjenerende støy fra trafikkerte gater rundt lokalene.

Det var lagt opp et godt sosialt program både for medfølgende familiemedlemmer og deltakere. Programmet inkluderte bl.a. en ekskursjon til den kommuniststyrte landsbyen Volterra, der vi ble mottatt i den ærverdige bystyresalen av en representant for byens folkevalgte. Vi fikk ved gjentatte anledninger nyte godt av italienernes store vennlighet og gjestfrihet, noe som samtlige deltakere satte stor pris på.

Elin Solstrand

COLING 82

**Ninth International Conference on Computational Linguistics
Praha 5.-10. juli**

Vel møtt i Palo Alto, California, om to år, sa den neste verten, Dr. *Martin Kay*, ved avslutningen av kongressdagene i Praha der COLING 82 ble holdt i dagene 5.-10. juli.

Praha har ærerike tradisjoner i åndsliv og vitenskap, også innenfor lingvistikk, der Prag-skolen har hatt stor innflytelse.

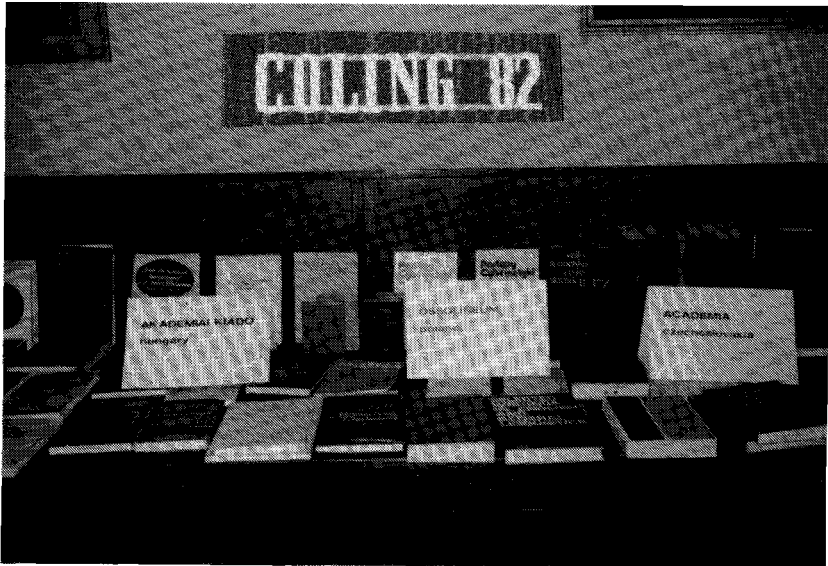
Alt i 1964 ble det holdt en større internasjonal konferanse om formell lingvistikk i Praha, og senere har interessen for datamaskinell språkbehandling vært i jevn utvikling.

Coling 82 betød en viktig stimulans for dem som arbeider med data-lingvistikk i Tsjekkoslovakia. Bl.a. må det bety meget at forskere der kunne få møte personlig flere hundre aktive forskere fra i alt 25 ulike land, noe som tsjekkiske forskere dessverre sjelden får mulighet til.

De lokale arbeidsbetingelser for disse forskere er heller ikke de beste. Som formannen i organisasjonskomiteen, professor dr. *Ján Horeký*, sa i innbydelsen: «Computer facilities may be more easily accessible in many other places, but the spirit of scientific discussions in Prague, we hope, will outweigh this as well as other possible difficulties».

Den lokale arrangementskomitee og sekretariatslederen, dr. *Eva Hajičová* fortjener stor ros for den måten konferansen ble arrangert på – ikke minst tatt i betraktning det omfattende byråkratiske apparat som arrangementet måtte samordnes med.

I løpet av de 5 dagene ble det holdt ca. 150 foredrag og kortere presentasjoner («short communications») som forelå i bokform ved åpning



COLING-82 ga de øst-europeiske land god anledning til å presentere sin forskning innenfor kvantitativ og datamaskinell lingvistikk.

gen av kongressen. Dertil kommer to workshops og flere paneldiskusjoner.

Av praktiske grunner ble det arrangert flere synkroniserte sesjoner, slik at de ca. 350 deltakere (8 fra Norge) kunne pendle mellom ulike temaer.

Konferansen ble naturlig nok en god mulighet for de østeuropeiske land til å presentere sitt arbeid – i alt ble ca. 1/3 av alle foredrag og kortere presentasjoner holdt av østeuropeere.

Det er karakteristisk for situasjonen i dag at temaene innenfor datamaskinell lingvistikk stadig differensieres. I år hadde den faglige komite forsøkt å gruppere foredragene i følgende hovedemner, her gjengitt etter det offisielle program:

- Linguistics in its relation to computational linguistics
- Machine translation
- Question-answering and artificial intelligence
- Grammatico-semantic analysis
- Mathematics in its relation to computational linguistics
- Information retrieval and linguistic data bases
- Knowledge representation

Det kan nevnes at det bl.a. ble holdt følgende tre nordiske foredrag i disse grupper: *Bente Maegaard*, Københavns Universitet: «The Transfer

of Finite Verb Forms in a Machine Translation System», *Anna Sågvald Hein*, Uppsala Universitet: «An Experimental Parser for Swedish» og *Per-Kristian Halvorsen*, NAVFs EDB-senter for humanistisk forskning, Bergen: «Order-Free Semantic Composition».

Konferansen ga inntrykk av at det er en stor aktivitet i det forskningsarbeidet som sikter mot å anvende datamaskinen som språkbruker. Særlig synes i dag arbeidet med å utvikle hensiktsmessige datastrukturer for å representere og analysere syntaktisk og semantisk informasjon å være fremtredende. Metoder fra kunstig intelligens trekkes inn for å kunne lagre «kunnskap om verden». Dette er en forutsetning for automatisk analyse av tekstinhold og for å kunne bruke naturlig språk i spørresystemer (question-answering).

Maskinoversettelse er et annet emne som får stigende oppmerksomhet, og hvor løsningsstrategiene er blitt flere. En foredragsholder skilte mellom forskningstiltak med rene forskningsformål for øye (f.eks. formalisering av grammatikker), oversettelsessystemer som bare sikter mot å gi en forståelig oversatt tekst (indicative translation) og slike som skal brukes i praktisk oversettelsesarbeid.

Fra amerikansk hold ble det rapportert om interessante prosjekter der en hadde satt oversetteren i fokus og der systemet for maskinoversettelse var å se på som ett tilbud blant flere som oversetteren kunne gjøre seg nytte av i sitt arbeid. Maskinoversettelsen ble her styrt av oversetteren, som også hadde adgang til avansert tekstbehandling, datamaskinelle ordbøker og mer omfattende termbanker. For å få automatiserte rutiner i praktisk og effektiv bruk må det menneskelige element, oversetteren, fremdeles stå i fokus, ble det hevdet.

Ikke minst i Japan er det i dag i gang et omfattende arbeid med å skape automatiserte oversettelsessystemer, og det ble på konferansen vist til meget lovende resultater. Flere kongressdeltakere sa at de med spenning møtte frem til disse konferansene for å se hva japanerne hadde greidd i mellomtiden – og mange forutså et «japansk under» også på dette felt.

Jostein H. Hauge

Datamaskinstøttet leksikografi

Rapport fra et nordisk forskersymposium

Et nordisk forskersymposium om bruk av edb i leksikografi ble holdt i tiden 12.-15. august på Sandbjerg, Sønder-Jylland, Danmark.

På symposiet deltok 22 leksikografer og edb-medarbeidere fra de nordiske land unntatt Island, herav 5 fra Norge.

Symposiet var basert på aktiv deltakelse i form av korte foredrag om pågående arbeid. I tillegg var det invitert en utenlandsk foredragsholder, *Harry Somers*, University of Manchester.

Foredragene gjaldt bruk av edb i terminologiarbeid, presentasjon av

pågående ordboksarbeid hvor edb inngår, og drøfting av metoder innenfor språklig databehandling av betydning for ordboksarbeid.

Symposiedeltakerne tok også opp hvordan de nye databehandlingsmetodene vil påvirke det leksikografiske arbeidsfeltet. Det ble bl.a. reist spørsmål om en i fremtiden mer og mer vil ha leksikografisk materiale tilgjengelig i datamaskinell form heller enn i bokform. Det ble også uttrykt ønske om et nærmere nordisk samarbeid på dette feltet.

Fra norsk side orienterte *Ivar Fønnes* om «Problems and methods in making the computer handle large character sets», *Lars Otto Grundt* om «The Use of Computers in the Preparation of a Large Norwegian-French Dictionary» og *Håvard Hjulstad* om «Terminology Projects at the Norwegian Term Bank».

Vi forstår det slik at alle innleggene (ca. 15) vil foreligge senere i en rapport sammen med en oppsummering. De som er interessert i rapporten kan henvende seg til *Henrik Holmboe*, *Institut for Lingvistik, Aarhus Universitet, Otto Rudsgade 6769, DK-8200 Århus N, Danmark*.

La oss til slutt komme med en liten parentetisk bemerkning: For utenforstående virker det underlig at man på en nordisk konferanse med kun nordiske foredragsholdere unntatt en, valgte engelsk som konferansespråk. Hvem andre enn vi i Norden kunne finne på noe slikt?

«Anvendelse av Artificial Intelligence i psykologi og lingvistik», Sverige, august 1982

Den 2.-15. august ble det avholdt et nordisk forskerkurs i Mullsjö, Sverige, med tittelen «Anvendelse av Artificial Intelligence i psykologi og lingvistik».

De nordiske forskerkurs har som mål å fremme forskerutdannelsen i Norden ved å gi mulighet for fordypelse i et nytt eller spesielt emne, og gjerne emner med tverrvitenskapelig karakter. Dette siste gjelder i høy grad Artificial Intelligence (AI), som må ses i nær sammenheng med en rekke tilgrensende vitenskaper såsom psykologi, lingvistik, filosofi, informatikk og matematikk.

Kurset hadde ca. 25 deltakere med en tilnærmet lik fordeling mellom de nordiske land (men bare 2 deltakere fra Norge) og med bakgrunn fra psykologi, lingvistik eller informatikk.

Følgende forelesere deltok på kurset: fra England kom *Margaret Boden* og *Yorick Wilks*, hhv. psykolog/filosof og lingvist/filosof, fra Canada *Raymond Perrault*, ansatt ved Department of Computer Science ved University of Toronto, og fra Holland *Harry Bunt*, som arbeider med utviklingen av dialogsystemer ved Philips. Norden var representert ved *Benny Brodda*, Stockholms Universitet, og *Sten-Åke Tärnlund*, Uppsala Universitet, hhv. lingvist og informatiker.

Kurset fokuserte primært på de aspekter av AI som har spesiell relevans for psykologi og lingvistikk. Av emner som ble behandlet i forelesningsform kan nevnes: AI i forhold til teoretisk psykologi, AI og epistemologi, forholdet mellom syntaks, semantisk og pragmatisk analyse, planleggingsstrategier, samt AI anvendt i forbindelse med menneske-maskin-kommunikasjon.

AI og lingvistikk

Det synes som om lingvistikk er en særdeles sentral del av AI – men iflg. Wilks er dette ikke alltid tilfelle. Når det virker slik, skyldes det at en lang rekke programmer arbeider med naturlige språk og språkforståelse som uttrykk for programmets «intelligens». Dette gjelder f.eks. for Winograds «klosse-verden» hvor en robot flytter rundt på klosser med ulike farger og fasonger etter anvisninger på engelsk. Representasjon av viten i «frames» utviklet av Minsky 1974-75, og andre liknende former (f.eks. «scripts» utviklet av Schank, 1975) har likeledes ytringer i naturlig språk som utgangspunkt. Utviklingen av denne type konstruksjoner bygger på en oppfattelse av at forståelsen av naturlige språk forutsetter viten om verden, eller i det minste en begrenset del av verden – dvs. at det er nødvendig å organisere viten om verden for å skape forståelse av språket.

Imidlertid stiller Wilks spørsmålet hvorvidt – eller på hvilken måte «frame-structures» kan anvendes ved analyse av naturlig språk: Det virker som om «frames» på en og samme tid skal være representasjon av ren viten (– og da må likheten med engelsk være tilfeldig) og samtidig på en eller annen måte skal kunne matches direkte mot en tekst, uten behov for kompleks grammatisk analyse. Denne siste mulighet vurderer Wilks som lite lovende, men samtidig har den vært medvirkende til å fjerne mye forskning innen AI fra annet arbeid med analyse av naturlige språk.

Imidlertid peker ideen om representasjon av viten slik den er utviklet i forbindelse med bruk av «frames» framover mot andre anvendelsesområder, f.eks. innenfor dialog-programmer og database-implementering.

AI-programmeringsverktøy

Programmeringsspråket LISP (forkortelse for LISp Processing) står sentralt i AI-arbeidet, både som frittstående språk og som basis for nye språk som PLANNER, STRIPS og LOGO. Videre ble det presentert et nytt programmeringsspråk som har fått en viss utbredelse i Sverige, PROLOG (PROgramming in LOGic). PROLOG er basert på et subsett av 1. ordens predikatlogikk; en setter opp teoremer som blir utført som et program. PROLOG virker som et kraftig verktøy i situasjoner hvor det er behov for å utføre kompliserte logiske operasjoner.

Hvilken vei går AI?

De siste årtiers forskning innen AI kan sies å angripe problemene fra to ulike synsvinkler: På den ene side skjer en del forskning som har som



*Deltakere og lærere på Nordisk forskerkurs i kunstig intelligens,
Mullsjö, august 1982.*

mål å programmere datamaskiner til å utføre oppgaver som ville betraktes som intelligente hvis et menneske hadde utført dem. Til denne kategori hører f.eks. sjakkprogrammer: disse benytter seg maksimalt av maskinens overlegenhet med hensyn til lagerkapasitet og hurtighet, men forsøker ikke å etterligne den menneskelige tenkeevne – programmer av denne art kan sies å være imitasjoner av en funksjon, men ikke en prosess. 1)

På den annen side utføres forskning innen AI som ikke alene skal ettergjøre resultatene av menneskelig tenkning, men som også ønsker å simulere de prosesser mennesker anvender når de tenker. Forskning av denne art søker å avdekke hva intelligens er og hvordan den menneskelige hjerne fungerer for med dette utgangspunkt å lage programmer som simulerer den menneskelige hjernes prosesser.

Hva som må forutsettes for å skape en «intelligent» datamaskin, og om det i det hele tatt er mulig, strides partene om – og besvarelsen av spørsmålet avhenger i høy grad av definisjonen på «intelligens». Ubestridelig er det imidlertid at det foregår vesentlig forskning innen AI som vil kunne få betydning for den framtidige anvendelse av datamaskiner.

1) En diskuterende introduksjon til emnet Artificial Intelligence finnes i heftet Källa 15: Bli'r datorn mänsklig? Utgitt av Forskningsrådsnämnden och Styrelsen för teknisk utveckling, Stockholm, 1982.

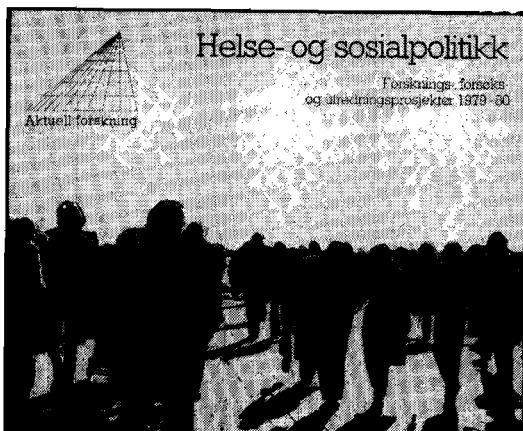
Her finnes innlegg fra Herbert Simon, Lars Kristiansson og Margaret Boden, samt litteraturhenvisninger.

*Eva Møller
Knut Hofland*

Forskningsinformasjon

Tre prosjektkataloger

NAVFs utredningsinstitutt har i samarbeid med Senteret laget kataloger over forskningsprosjekter innenfor tre ulike felt.



Helse- og sosialpolitikk

Helse- og sosialpolitikk er et forskningsfelt hvor virksomheten er stor og uoversiktlig. Mye av forskningen bygger på innsikter og metoder hentet fra flere samfunnsvitenskapelige fag, andre prosjekter ligger i skjæringspunktet mellom samfunnsfag og medisin.

Denne katalogen inneholder opplysninger om 349 prosjekter – forskning, forsøk og utredninger – som var i gang ved årsskiftet 1979/80, og en større bibliografi over litteratur som bygger på prosjektene.



Barneforskning

Denne oversikten over nyere forskning om barns livsforhold og utvikling ble finansiert av NAVF og Forbruker- og administrasjonsdepartementet. Oppdraget er et ledd i arbeidet med etableringen av NAVFs senter for barneforskning i Trondheim fra høsten -82.

De 239 prosjektene er klassifisert etter det emne og problemområde som behandles. Gjennom prosjektklassifikasjonen har man søkt å få fram barneforskningens tverrfaglige karakter og samtidig vise den tematiske spredning som eksisterer innenfor barneforskning.

Til hvert prosjekt fins bl.a. en prosjektbeskrivelse, navn, adresse og telefonnummer til en kontaktperson samt en bibliografi over publiserte arbeider utført i tilknytning til prosjektet.



Humanistisk forskning

Denne katalogen over 941 forskningsprosjekter innen de humanistiske fag er laget på grunnlag av en rundspørring til forskerne ved universitetene og de øvrige forskningsinstitusjoner. I tillegg til opplysninger om det enkelte forskningsprosjekt inneholder den en større bibliografi over litteratur med tilknytning til prosjektene. Følgende fagområder er med i oversikten: arkeologi, arkitektur, filosofi, folkelivsgranskning, folkeminnevitenskap, historie, idehistorie, kunsthistorie, litteraturvitenskap, musikkvitenskap, religionsvitenskap, språkvitenskap, teater-film-drama og teologi.

Katalogene fås gratis ved henvendelse til: *NAVFs utredningsinstitutt, Wergelandsveien 15, Oslo 1.*

Word frequencies in British and American English

Denne boka er resultat av et samarbeid mellom førstekonsulent *Knut Hofland* ved Senteret og professor *Stig Johansson*, Universitetet i Oslo. Førstnevnte har hatt ansvaret for det edb-faglige og for produksjon av boka, mens Johansson har stått for den lingvistiske siden og den generelle planlegging. Boka er basert på Brown Corpus og LancasterOslo/Bergen (LOB) Corpus som hver inneholder over en million ord fra 1000 forskjellige trykte engelske og amerikanske språkkilder.

Boka er den første større publikasjon i sitt slag med ordfrekvenser basert på moderne britisk-engelske tekster.

Den består av fem ulike ordlister og en innledning og diskusjon av noen av resultatene fra ordlistene. Hoveddelen av ordlistene utgjøres av en komplett ordliste til LOB Corpus, ca. 50000 ord. Videre er det en frekvensliste med de 7500 hyppigste ordene i korpuset. For de 100 hyppigste ordene er det også angitt rang og absolutt og relativ frekvens innen de 15 kategorier tekst som materialet består av. I den fjerde listen er de 15 kategoriene slått sammen til 4 hovedkategorier, og det er angitt absolutt og relativ frekvens innen disse kategoriene for de ca. 5000 hyppigste ordene.

Den siste listen gir en sammenligning for ca. 10000 ord mellom ordfrekvenser i hhv. LOB og Brown Corpus.

Boka har også et vedlegg på 3 mikrokort (ca. 600 sider). Her er det ranglister for hver av de 15 tekstkategoriene i LOB Corpus og en liste med ord som er over-/underrepresentert i de 4 hovedkategoriene i forhold til gjennomsnittet.

4) LOB Corpus: Rank List

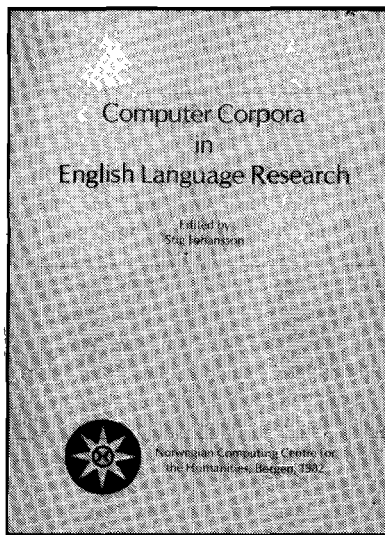
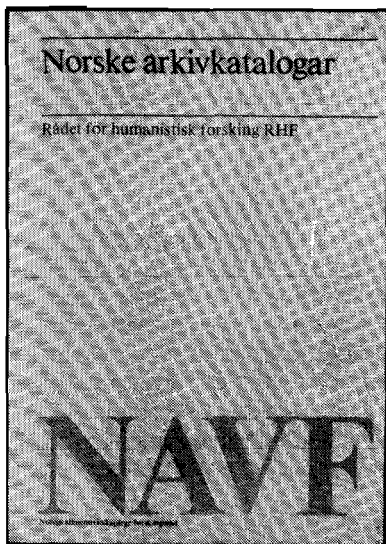
Word	absolute frequency	number of categories	number of samples
the	68315	15	500
of	35716	15	500
and	27858	15	500
to	26760	15	500
a	22744	15	500
in	21108	15	500
that	11188	15	500
is	10678	15	493
was	10499	15	479
it	10019	15	488
for	9299	15	500
he	9776	15	441
as	7337	15	500
with	7197	15	500
be	7188	15	497
on	7027	15	500
i	6896	15	382
his	6286	15	441
at	6043	15	498
by	5798	15	499
had	5391	15	443
this	5287	15	498
not	5142	15	498
but	4956	15	499
from	4698	15	500
are	4592	15	498
are	4544	15	456
which	4467	15	484
her	4030	15	287
she	3812	15	255
or	3781	15	484
you	3580	15	287

WORD FREQUENCIES IN BRITISH AND AMERICAN ENGLISH

Knut Hofland Stig Johansson

NAVF The Norwegian Computing Centre for the Humanities
NAVF's EDB-senter for humanistisk forskning

For det samlede LOB og Brown Corpus er det en komplett baklengsordliste (ca. 75 000 ord) og en liste ordnet etter en indeks som gir ulikhet/likhet i frekvens i LOB og Brown korpora (ca. 10000 ord).



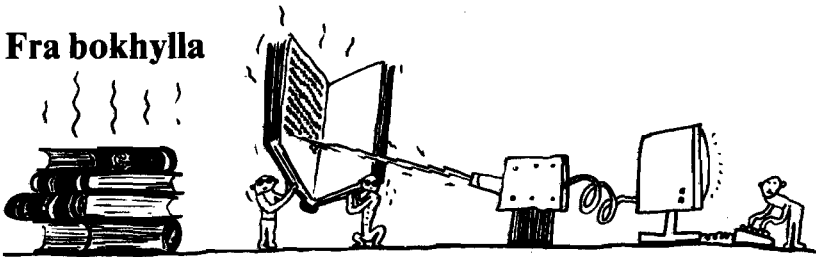
Norske arkivkataloger

Boka gir en oversikt over kataloger, registre, mikrofilmer, kildeutgaver m.m. i offentlige norske arkivinstitusjoner. Den er utarbeidet av førstearkivar *Egil Nysæter* ved Statsarkivet i Bergen i samarbeid med en styringsgruppe. Kontorfullmektig *Anne Hauge* og konsulent *Øystein Reigem* ved Senteret har stått for den datamaskinelle tilrettelegging og klargjøring for trykking. NAVF har bl.a. dekket lønnsutgiftene i prosjektet og trykking av boka. Interesserte kan henvende seg til Riksarkivet.

Computer Corpora in English Language Research

Boka er redigert av professor *Stig Johansson* og bygger på konferansen «Computer Corpora in Research and Teaching» som Senteret og Johansson sto som arrangør for 1.-3. juni 1981. I tillegg til 7 foredrag på konferansen, delvis omarbeidet med sikte på publisering, er det to artikler skrevet spesielt for boka. Samlet gir boka et godt bilde av det arbeid som drives med utgangspunkt i engelskspråklige corpora. Boka koster kr. 55 og kan bestilles hos *Studia Universitetsbokhandel, Parkveien 1, 5014 Bergen-Universitetet*.

Fra bokhylla



Cook, Michael

Archives and the Computer, Butterworths 1980, 152 s.

Denne handboka tar for seg bruk av edb i arkivinstusjonar. Ho er meint å vera til hjelp for arkivfolk som ikkje har noko erfaring med datamaskinar frå før. Forfattaren, som har erfaring frå arkivarbeid og undervisning i arkivfag ved University of Liverpool, opnar derfor boka med ei allmenn innleiing om databehandling, for deretter å presentera forholdsvis detaljert ulike system som har vore eller er i bruk i britiske og amerikanske arkiv og arkivskapande institusjonar. Eit mindre kapittel tar for seg problema med å ta vare på arkivmateriale i maskinleseleg form. Ellers finst det lister over tekniske uttrykk og arkivsystem saman med ein 9 siders bibliografi.

Forfattaren har medvite utelate omtale av edb-system som gjeld overføring av kjeldemateriale (slik som folketeljingar og kyrkjebøker) til maskinleseleg form.

Boka er ei lettlesen og pedagogisk god innføring i emnet edb og arkiv. Frå eit arkivsynspunkt er det spesielt verdifullt at framstillinga heile tida er nær knytt til dei vanlege arbeidsoperasjonane ved eit arkiv, med hovudvekt på produksjon av katalogar. Dei system som vert drøfta er tilpassa større datamaskinar, medan bruk av mikromaskinar og tekstbehandlingsutstyr berre så vidt er nemnt. Når det gjeld bruk av mikromaskinar kan interesserte visast til ein artikkel om eit amerikansk prosjekt som kom i gang i 1980: *R. M. Kesner and D. Hurst: Microcomputer Applications in Archives: A Study in Progress* (Archivaria 12 (1981), s. 3-19).

Egil Nysæter

Herman Ruge Jervell og Kai Olsen:

Hva datamaskiner ikke kan. Universitetsforlaget, Tromsø, 1982. 181 sider, kr. 105.

Oppgaven de to forfattere har satt seg er, som tittelen antyder, å trekke grensene klart opp for hva en datamaskin IKKE kan! En høyst prisverdig oppgave, ikke minst når den som her løses av to personer som øyensynlig har god innsikt også i de operasjoner som datamaskinen KAN utføre.

Når dertil kommer at boken er skrevet i et lettlest språk, og med mange eksempler til å belyse de teorier boken bygger på, er det all grunn til å anbefale den som en innføringsbok.

Bokens første del, kapitel 1 til 3 (ialt 34 sider) gir litt oversikt over datamaskinens virkemåte og programmeringsarbeid, og er tenkt å gi bakgrunn for å forstå argumentasjonen i bokens andre del.

Analysen i bokens andre del bygger på et skille mellom ÅPNE og LUKKETE sammenhenger, hvor det er forfatterens poeng at datamaskiner bare kan brukes i lukkede sammenhenger.

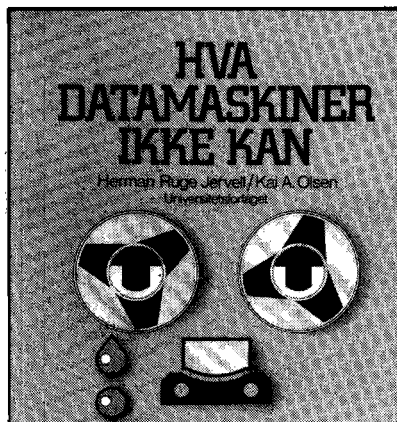
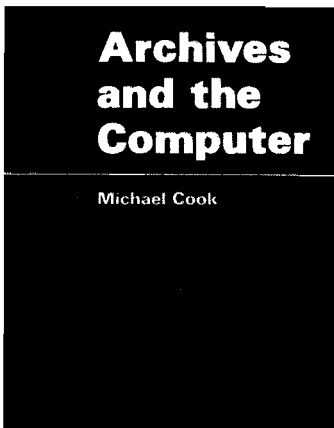
Analysen tar utgangspunkt i språk (naturlige språk), og fortsetter via eksempler på maskiner som «forstår» språk til «tenkende» maskiner, f.eks. sjakk-programmer og programmer for medisinsk diagnostisering.

Problemet åpne og lukkede sammenhenger knyttes sammen med en diskusjon av FORMALISERING, idet formalisering forklares som det å «fryse fast og lukke deler av våre omgivelser» (s.102). Formalisering ses først og fremst som et allment trekk i vårt samfunn, og dernest settes det i sammenheng med datamaskinens funksjonsmåte.

Bokens tre siste kapitler tar opp diskusjonen om konsekvenser av bruken av datamaskiner. Kapittel 9 ser litt på hva som skjer på organisasjonsnivå ved utviklingen av store systemer, kapittel 10 ser kritisk på de framtidsvyer som kommer til uttrykk i betegnelser som «det papirløse samfunn», «det pengeløse samfunn» osv., mens kapittel 11 går inn på individet i forhold til datamaskinen.

Boken når vidt omkring innen emnet, og bruker beskrivelser av mange typer av systemer for å belyse sine synspunkter. Det gjør boken til spennende lesning, men er samtidig dens svakhet. Det sier seg selv at det ikke er mulig å gå i dybden innenfor et så omfattende emne i løpet av 180 sider. Imidlertid finnes en omfattende referanseliste, ordnet etter kapitellinnhold, for den som vil fortsette lesningen.

Eva Møller



MELDINGER

Forskerstipend i automatisert arkivinformasjon

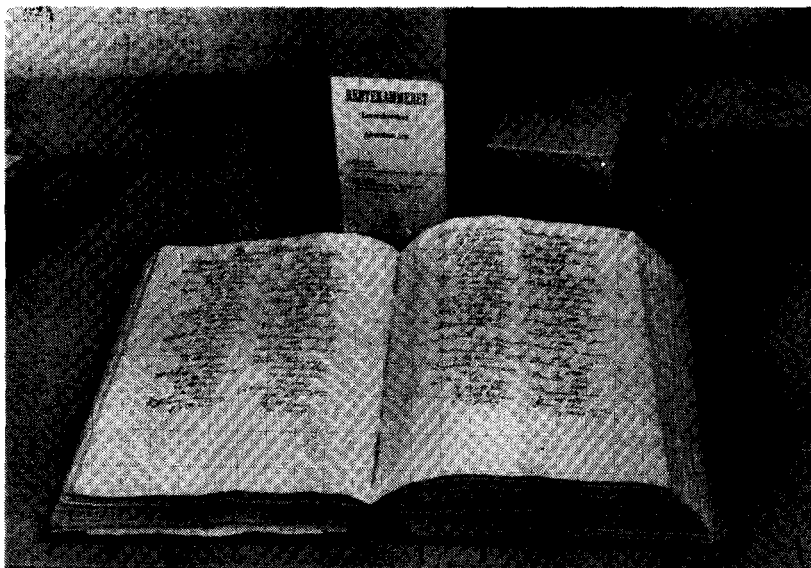
Rådet for humanistisk forskning, Norges almenvitenskapelige forskningsråd, har opprettet et forskerstipend fra 1.1.83-30.6.84 knyttet til edb-metoder for lagring, fremsøking og presentasjon av arkivmateriale. Særlig vekt ønskes lagt på de behov som forskere har for bedre adgang til kildemateriale.

Stipendordningen er knyttet til NAVFs EDB-senter for humanistisk forskning, Universitetet i Bergen. Senteret vil inngå et samarbeid med forskeren om den edb-faglige del av prosjektet.

Det forutsettes at forskeren har nær kontakt med Senteret, enten ved arbeidsplass der eller faste opphold etter nærmere avtale.

Det vil bli utarbeidet en skriftlig avtale som basis for samarbeidet.

Stipendet er i første rekke rettet mot fagmedarbeidere med forskerkompetanse i en faglig relevant disiplin. Det forutsettes at forskeren har grundige kunnskaper om edb.



Jordebok for Akershus len, 1598-1599 (foto: Riksarkivet).

Det foreligger en mer utførlig orientering om stipendordningen hvor det også blir gitt en omtale av de utviklingsområder Rådet ønsker å prioritere.

Stipendet vil bli utlyst på ordinær måte gjennom Norsk lysingsblad, dagspressen og oppslag ved aktuelle institusjoner.

Hospitantopphold ved Senteret

NAVFs EDB-senter for humanistisk forskning skal etter sin målsetting arbeide for en fornuftig bruk av edb i humanistisk fag- og forskningsarbeid. Som et ledd i dette arbeidet tilbyr Senteret de humanistiske fagmiljøer bistand i form av hospitantopphold i Senteret både for teknisk/administrativt og vitenskapelig personale.

Det har til nå vist seg at driftsmedarbeidere kan ha stort utbytte av å oppholde seg i et humanistisk edb-miljø i forbindelse med bruk av nytt edb-utstyr på arbeidsplassen, starten av edb-prosjekter o.l.

For vitenskapelig personale i de humanistiske fag gis det også tilbud om kurs for ulike fagområder.

Under oppholdet i Senteret vil hospitantene få anledning til å sette seg inn i arbeidsoppgaver og edb-emner som er av særlig interesse. Instruksjon og løpende veiledning inngår i opplegget.

Reise- og oppholdskostnadene må dekkes av de deltakende institusjoner.

Nærmere opplysninger om hospitant- og kursordningen gis ved henvendelse til Senteret.

Senterets oppdragsordning

NAVFs EDB-senter for humanistisk forskning har de siste årene utført edb-oppgaver for humanistiske fagmiljøer som betalte oppdrag.

Frem til i dag har vi utført en rekke oppgaver, f.eks. utredningsarbeid, utvikling av programmer og databehandling av forskningsmateriale for prosjekter og institusjoner. I det siste har den økende bruk av mikrodatamaskiner gitt dette arbeidet en ny dimensjon.

Vi er nå interessert i å klarlegge behovet fremover for denne typen tjenester i de humanistiske miljøer, og vil gjerne ha kontakt med institusjoner som har planer om å benytte vår oppdragsordning.

Våre edb-tjenester tilbys innen på forhånd fastsatte tidsfrister. Dette gjør det imidlertid ønskelig å få en oversikt over de aktuelle oppgaver på et så tidlig tidspunkt som mulig.

Ved henvendelse til Senteret vil det bli gitt nærmere opplysninger om samarbeidsformer og betalingsvilkår.

OUCS står til tjeneste for universitetene med programvare, maskiner og faglig veiledning ved edb-anvendelse innen de humanistiske fagområder.

OUCS's tilbud omfatter bl.a. optisk innlesing av data, anvendelse av egnede programpakker (f.eks. Oxford Concordance Package, presentert i Humanistiske Data 1-82) og administrasjon av maskinlesbare tekster (f.eks. de tekster som finnes i Oxford Archive, Brown Corpus og Lancaster-Oslo/Bergen Corpus). Videre tilbyr OUCS fotosetting av brukernes tekster bare disse finnes i maskinlesbar form og oversendes OUCS på bånd.

Nærmere opplysninger om OUCS's virksomhet finnes i en folder som kan fås ved henvendelse til Senteret.

Conference on Applied Natural Language Processing

Santa Monica, California, 1.-3. februar 1983

Sponsorer for konferansen: Association for Computational Linguistics, Naval Research Laboratory.

Konferansen vil spesielt presentere datalingvistiske teknikker for PRAKTISK problemløsning innen følgende områder:

- data, informasjon og vitenbearbeiding
- undervisning
- språkovsettelse
- talegjenkjenning og -syntese.

Ytterligere opplysninger: *Beatrice Oshika MD 72-27 System Development Corporation, 2500 Colorado Ave., Santa Monica, CA 90406*

International Conference on the Application of Mini- and Microcomputers in Information, Documentation and Libraries

Tel Aviv, Israel, 13.-18. mars 1983

Konferansen vil bl.a. behandle eksisterende systemer og software, nettverk, systemdesign og -vedlikehold samt applikasjoner innen ulike spesialområder.

Konferansen har følgende sponsorer:

International Federation for Documentation (FID)

Gesellschaft für Information und Dokumentation mbH (GID), Hamburg

Israel Society of Special Libraries and Information Centres (ISLIC)

Association for Computers and the Humanities (ACH)

ACH er en internasjonal organisasjon med følgende målsetning: The purpose of the Association shall be to encourage by suitable means the appropriate uses of computers and related technologies in the study of humanistic subjects. The Association shall further this goal by sponsoring regular International Conferences on Computers and the Humanities, and by otherwise facilitating contacts and discussion among scholars interested in the goal of the Association.

ACH utgir ACH Newsletter som utkommer med 4 nummer i året. Her orienteres det om konferanser, litteratur og programvare som kan være spesielt relevant for humanister som anvender edb i sitt arbeid.

ACH står også for utgivelsen av tidsskriftet Computers and the Humanities, som likeledes utkommer med 4 nummer pr. år.

Medlemskap og abonnement på ACH Newsletter koster \$ 15 i året.

ACH's adresse:

Association for Computers and the Humanities, 216 Lind Hall, University of Minnesota, Minneapolis MN 55455, USA.

Europeisk avdeling av ACL

Association for Computational Linguistics vedtok i juni å opprette en europeisk avdeling - «European Chapter of the Association for Computational Linguistics», eller ACL-Europe.

For europeiske medlemmer av ACL har det vært et problem at ACL var spesielt knyttet til amerikanske forskningsmiljøer, mens man har savnet et tilsvarende forum for europeisk datalingvistisk forskning. Dette savn skulle den nyopprettede europeiske avdeling av ACL råde bot på.

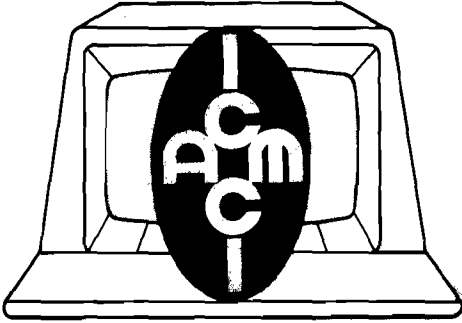
ACL-Europe har samme målsetning som hovedorganisasjonen, men skal spesielt vareta de europeiske medlemmers interesser. Videre vil avdelingen lette de praktiske problemer ved økonomiske transaksjoner og distribusjon av publikasjoner.

Henvendelser kan rettes til: *Harry Somers, Centre for Computational Linguistics, U.M.I.S.T., P.O. Box 88, Manchester M60 1QD, England.*

16.-18. juni 1982 ble ACL's 20. årlige møte avholdt i Toronto, Canada. På programmet var bl.a. en sesjon med presentasjon av modeller for semantisk representasjon, samt bruk av naturlige språk i kommunikasjon med databaser.

Nasjonal edb-tjeneste for humanister i England

Oxford University Computing Service (OUCS) er blitt opprettet i England som en nasjonal edb-tjeneste for humanister.



National Center of Scientific and Technological Information (COSTI),
Israel

Ytterligere opplysninger ved henvendelse til Senteret eller direkte til:
*The Secretariat, International Conference on the Application of Mini-
and Micro-computers in Information, Documentation and Libraries.*
P.O. Box 3054, 61030 Tel Aviv, Israel.



Sixth International Conference on Computers and the Humanities.
North Carolina State University, Raleigh, North Carolina, USA, June
6-8, 1983. Sponsored by the Association of Computers and the Human-
ities.

Purpose

The Sixth International Conference on Computers and the Human-
ities is aimed at providing a broad and rich scholarly exchange of ideas
on the uses of computing in all areas of the humanities. The formal ses-
sions and related activities will benefit both experienced scholars as well
as those new to the field.

ICCH/83 has the following objectives:

To bring together in one forum all of the major research and trends,
both national and international, on the uses of computers in the huma-
nities

To provide for a maximum of interaction among individuals at all levels in the various aspects of this research

To produce a proceedings documenting the coming of age of computers in the humanities

Further information: *Sara K. Burton, Department of English, P.O. Box 5308, North Carolina State University, Raleigh, NC 27650, USA*

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Conference on Data Bases in the Humanities and Social Sciences (CDBHSS), New Jersey, USA. 10-12 June 1983

Konferansen blir holdt på Rutgers, The State University of New Jersey, Campus at New Brunswick, New Jersey, USA, og har til formål å samle humanistiske og samfunnsvitenskapelige forskere som gjør bruk av datamaskinelt forskningsmateriale.

Nærmere opplysninger kan gis av konferansens koordinator: *Professor Robert F. Allen, Room 427, Alexander Library, Rutgers - The State University of New Jersey, New Brunswick, New Jersey 08903, USA* eller NAVFs EDB-senter for humanistisk forskning, Bergen, som vil sende interesserte konferansebrosjyren. Nærmere opplysninger om konferansen vil bli gitt i et senere nummer av *Humanistiske Data*.

The 7th World Congress of Applied Linguistics

AILA (International Association of Applied Linguistics) avholder sin 7. verdenskongress i Brussel, Belgia, 5.-10. august 1984. Kongressens hovedtema blir: «Anvendt lingvistikk bidrag til internasjonal forståelse». I tillegg vil det bli avholdt symposier og seminarer innen følgende emner:

- språkproblemer i utviklingsland
- sosiolingvistikk
- psykolingvistikk
- språkundervisning
- kommunikasjon og interaksjon
- kvantitativ lingvistikk.

Ytterligere informasjon vedr. påmelding til konferansen følger senere.

SUMMARY

The objective of Information Science is to improve the management and flow of information. However, the variety of initiatives and innovations which mark the rapidly expanding character of Information Science generates its own «information problem». We offer our English abstracts to show non-Scandinavian colleagues how the complexity of these developments is reflected in the journal of our national centre for ADP in the Humanities.

Data om språkundervisning

Data on language teaching

In this article Research Fellow Lars Evensen presents data he has collected at the Dept. of Applied Linguistics, University of Trondheim.

The Trondheim Corpus in Applied Linguistics is a manually stored documentation of pupils' proficiency in written English and Norwegian. The corpus was prepared in 1979 in connection with a nation-wide questionnaire in which a statistically representative selection of teachers and pupils reported what they considered to be problems in the mentioned languages. The questionnaire was developed with the co-operation of the participants and tested in a pilot study in 1980. The principal collection of data via the questionnaire took place in 1981. The data is machine stored and gives anonymous information on the following: a) problems connected with teaching material, b) problems related to the conditions under which the learning/instruction takes place, c) the participants' backgrounds.

In order to compare the stated problems with written production, the study was extended to include the pupils' latest compositions/essays in the two languages. Evensen gives a critical evaluation of the routines used in the preparation of the corpus. Despite some limitations the corpus consists of a relatively comprehensive material for research in linguistic performance and for error analysis.

EDB og publisering

ADP and publishing

The background for this article by Consultants Knut Hofland and Svein Skotheim is the developmental work carried out by the Centre concerning the use of ADP in the publishing of Humanistiske Data, books, annual reports, etc.

The article deals mainly with photo typesetting as the basis for printing. This method involves illumination of characters on photographic paper, which must be developed. A photo typesetter turns out better quality print than a typewriter or printer.

In order to utilise ADP technology in the production of composition the manuscript must first be made machine readable or inputted directly via a screen terminal. One way of doing this is to typewrite the text for optical reading. If text processing equipment is used typography codes for types of print, size of letters etc. can be entered at the same time for later treatment by special photo composition programs. These codes can also be edited directly on the photo typesetter's keyboard.

The three most common means of transmission to a photo typesetter are magnetic tape, disk and telephone link. When using a telephone link, the owner of the photo typesetter «fetches» the text in question via a telephone call (or vice versa). Where the other two methods are involved the registered text must be physically brought to the photo typesetter in order to be hooked up.

The end product of the photo typesetting phase is a long roll of text composed in the coded column width. This roll is cut and made up before being photographed and transferred to the printing sheet.

Automatisk språkoversettelse i EF

Automatic language translation in the EEC

In this article Consultant Eva Møller at the Centre describes the development of an automatic translation system which is to translate each official EEC language to the others.

The work on EUROTRA (EUROPEan TRANslation system) was started in 1978, under the leadership of Margaret King, UK. The developmental work has been assigned to research centres in every membership country in the EEC.

Older translation systems are bilingual, whereas EUROTRA is to be multilingual. Furthermore, it must be easily extendable to new languages.

In order to meet these requirements the translation process is executed on the basis of only one language and it is split into three modules: analysis of the sentence in the source language, transfer, and generation in the target language. The purpose of this model is to enable the analysis and generation modules to be used in all the languages. These modules are monolingual whereas the transfer module functions as a bilingual bridge.

As a special transfer module is required for each pair of languages its contents and structure must be strictly defined. Dependency trees have been chosen for the representation of the transfer structure. In other words, the analysis component must result in a dependency tree structure, which the generation component must originate within. However, the development groups are relatively free to choose methods, grammar and strategy within the analysis and generation modules.

In the dependency trees the nodes are labeled with information on three linguistic levels: morpho-syntactical, syntactical and logico-semantic.

EUROTRA has not been formally tested yet but one thing is certain: the translations the system will be able to produce can never be more than, at best, suitable drafts which translators must improve.

EDB-tjeneste for språk og litteratur i Tromsø

Linguistic and Literary computing service in Tromsø

In 1981 Holger Hagan received a fellowship in ADP at the Institute of Languages and Literature, University of Tromsø. Relatively few researchers at this institute use ADP, so in Hagan's view one of his tasks is to spread information about the advantages of automatic processing of data, e.g. concerning processing of texts and production of vocabularies and concordances.

At the Institute several researchers work with texts in different alphabets, e.g. Lappish. However, suitable screens and printers are not available for this type of work. Hagan is trying to find solutions to this problem so that it will be easier for new users to «cross the threshold» when they start using computers.

Senteret feirar 10-års jubileum

The Centre celebrates its 10th anniversary

On June 3rd the Centre arranged a social gathering for the members of the Council for Research in the Humanities, guests from the University of Bergen and the employees of the Centre for the celebration of the Centre's 10th anniversary.

Speeches were made by the Assistant Director of the University, Bjørn Henrichsen, and the spokesman for the NAVF-humanists, Reader Ola Kai Ledang. Per-Kristian Halvorsen, currently a researcher at the Centre, gave an introduction to computational linguistics.

Jostein H. Hauge, who has been the Centre's Director from its inauguration, outlined the Centre's activities during the past 10 years. The Centre was started at the University of Bergen in 1972 as a five-year program. In 1977, however, the Norwegian Research Council for Science and the Humanities, its main sponsors, decided to establish it as a permanent national ADP centre for the Humanities. Throughout the years the Centre has assisted in countless ADP projects in the Humanities, both within and outside of the universities. The Centre has emphasised the development of ADP programs and methods, and has arranged training programs and information services. In Hauge's opinion the Centre's tasks will change as ADP becomes a familiar tool in the Humanities. In the future more emphasis will be given to the development of methods and to the training of researchers.

Datalogi og humanisme – to uforenlige verdener?

Computer Science and the Humanities – two irreconcilable worlds?

Eva Møller, Consultant at the Centre, gives a summary of reports issued by a committee appointed by the University of Aarhus, Denmark. This committee has outlined the problems concerning the establishment of a training program in ADP for Humanities students. It consisted of representatives from both Computer Science and the Humanities. The goal of their work was to clarify the basis of the training program and to define its terminology. They have tried to point out differences and similarities between the two fields of study in order to select scholarly traditions and views on which to base a combined education.

The goal of the new study program is to train graduates to take part in systems development, who will be specially qualified in areas connected with users/system interface. Naturally, the graduates will also be able to work in traditional Humanities fields that use ADP as a technical aid.

The recognition of the different requirements of programmers and users is the basis of the training program's concept: the systems analyst is to make recipient-orientated, parallel but also related and equivalent descriptions of systems, incorporating a historical dynamic view of work processes.

Seminar om EDB-metoder i humanistiske fag

Seminar on ADP methods in the Humanities

At Finse in March the annual «development seminar» was arranged for the consultants at the Centre, representatives of the computer services of the Arts Faculties at the Universities, and a number of ADP-users at universities and colleges. The main themes discussed this year were microcomputers, the use of ADP in publishing, and analog-digital conversion techniques.

Reiserapport fra Hannovermessen april 82

Report from the Hanover Exhibition

Sigbjørn Århus, Senior Consultant at the Centre, reports from the Hanover Exhibition of technical equipment, held in April. This year's arrangement actually consisted of 10 exhibitions, among them the CeBIT exhibition of office and data technology.

Nearly every manufacturer demonstrated its own microcomputer. Microcomputers based on an 8 bits processor with CP/M and MP/M based operative systems still dominated the market. However, 16 bits microcomputers have now been introduced, and as a result the prices of this type of equipment have been greatly reduced. On the other hand, disk storage capacity is steadily increasing. The exhibition also proved that printers based on laser technology are here to stay.

The theme of CeBIT's special exhibition this year was «Practice and perspectives in office communication». New equipment was shown for speech synthesis, text and picture processing, and data communication. Great technical progress has been made in these fields, but undoubtedly many practical and organisational problems must be solved before the whole range of facilities can become standard provision in administrative work.

Datamaskinell informasjonssøking

Automatic data retrieval

Eva Møller, Consultant at the Centre, reports from the Fifth Nordic Conference on Information and Documentation, and the Nordic On-line User Day, both held in Trondheim in June.

An assessment was made of the use of different types of technology in the fields of information and documentation. Technology is mainly connected with information retrieval, e.g. the large bibliographical data bases to which information services can subscribe, and the so-called «fact data bases».

Access to references to literature is increasingly fast and easy but the problem of document delivery is growing. A solution to this problem is represented by the Telefax system, developed at the Norwegian Telephone Company, which involves copying via the telephone network. Another solution, currently at an experimental stage, is document transmission via satellite. New technology is also used to produce documents in the form of microfiche.

The discussion among the Nordic on-line users indicates that librarians' practical work is full of problems – the most important being the networks' lack of stability. Another problem is that of having to learn a new query language for each data base. One solution for this problem is to define a common query language which can be used on all data bases connected to a network, but this would require an extensive amount of work. A different solution is program-to-program communication for distributed data bases. Yet another aid is Userkit, a microcomputer which is installed in the user/on-line retrieval interface.

ALLC PISA 7.-11. juni 82

ALLC symposium

Elin Solstrand, Consultant at the Centre, reports from the Seventh International Symposium on the Computer in Literary and Linguistic Research, which was held in Italy in June. The symposium gathered 120 participants from USA and Europe. 50 presentations

were given in the following fields: studies of authorships, concordances, data bases, education, lexicography, literary statistics, quantitative linguistics, stylistic analysis, and software.

Norwegian contributions were made by Eirik Lien (University of Trondheim), and Per Vestbøstad (Norwegian Text Archive). Lien's speech dealt with the work he and Anne Karin Ro have done in connection with machine registration of orthographical variant occurrences in 12 handwritten manuscripts which contain The General Prologue to the Wycliffite Bible. Vestbøstad presented the work being carried out at the Norwegian Text Archive.

COLING 82

The Ninth International Conference on Computational Linguistics took place in Prague in July. In addition to two workshops and several panel discussions, 150 lectures and short communications were given before the 350 participants. One third of all of the contributions were made by Eastern Europeans.

The lectures were grouped in the following main subjects: linguistics in its relation to computational linguistics, machine translation, question-answering and artificial intelligence, grammatico-semantic analysis, mathematics in relation to computational linguistics, information retrieval and linguistic data bases, and knowledge representation.

Per-Kristian Halvorsen, Fellow at the Centre, spoke on «Order-Free Semantic Composition».

The conference gave the impression that work on developing suitable data structures for the representation and analysis of syntactic and semantic information is prominent.

Solution strategies for machine translation is another subject which is receiving more attention, especially in Japan, and with promising results. In the US machine translation is developing into one of many tools the translator may use in his/her work, along with advanced text processing, computer-based dictionaries, and data bases of terminology.

Datamaskinstøttet leksikografi

Symposium on ADP-supported lexicography

A report is given from a Nordic symposium on the use of ADP in lexicography, which was held in Denmark in August. The symposium was based on short presentations of work in progress.

The speeches dealt with the use of ADP in terminological and lexicographical work, and with methods in linguistic data processing which are of importance to lexicography. The participants also took up the matter of how the new data processing methods will influence the field of lexicography.

Norwegian contributions to the symposium were: Ivar Fonnes, University of Oslo: «Problems and methods in making the computer handle large character sets», Lars Otto Grundt, University of Bergen: «The Use of Computers in the Preparation of a Large Norwegian-French Dictionary», and Håvard Hjulstad: «Terminology Projects at the Norwegian Term Bank».

All of the contributions will be printed in a report in English along with a summary of the discussions.

«Anvendelse av Artificial Intelligence i psykologi og lingvistikk»

«The use of Artificial Intelligence in Psychology and Linguistics»

This is the title of a Nordic course for researchers held in Sweden in August, from

which Consultants Eva Møller and Knut Hofland at the Centre report. Twenty-five participants with backgrounds in Psychology, Linguistics and Information Science took part in the course.

Lecturers at the course were: psychologist/philosopher Margaret Boden and linguist/philosopher Yorick Wilks, both from England; Raymond Perrault, Department of Computer Science, University of Toronto; Harry Bunt, who works with the development of dialogue systems at Phillips, Holland; linguist Benny Brodda, University of Stockholm; and systems analyst Sten-Åke Tärnlund, University of Uppsala.

Subjects dealt with were AI in relation to theoretical psychology, AI and epistemology, the relations between syntax, semantic and pragmatic analysis, planning strategies, and AI used in connection with *man-machine communication*.

Forskningsinformasjon

Research information

The Institute for Studies in Research and Higher Education at The Norwegian Research Council for Science and the Humanities, in co-operation with the Centre, has produced three catalogues of research projects.

One of the catalogues gives detailed accounts of 239 newer research projects on the life conditions and development of children. It was made in connection with the establishment in Trondheim of the Council's Centre for Research on Children.

A catalogue concerning health and social politics contains information on 349 projects – research, experiments and committee reports – along with a large bibliography of literature connected to the projects.

The third catalogue covers 941 research projects and relevant literature on the following Humanities subjects: Archaeology, Architecture, Art History, Ethnography, Ethnology, Film Science, History, History of Ideas, Linguistics, Literature, Music, Religion, Theatre and Drama, and Theology.

Fra bokhylla

Books

Senior Archivist Egil Nysæter at the Regional State Archive in Bergen reviews Michael Cook's *Archives and the Computer*. There is also a presentation of *Norske Arkivkataloger* which outlines catalogues, registers, microfilms, source editions etc. in Norwegian institutions which hold archives. The material was collected by Nysæter, and the Centre was responsible for the production of the data and the preparations for photo typesetting.

Computer Corpora in English Language Research is edited by Professor Stig Johansson, University of Oslo. It is based on the conference «Computer Corpora in Research and Teaching» which the Centre and Johansson arranged in Bergen last year. In addition to the 7 lectures given at the conference the book includes two specially written articles. It costs 55 N.kr. and can be ordered through Studia Universitetsbokhandel, Parkveien 1, 5014 Bergen-Universitetet.

The Centre has recently published *Word Frequencies in British and American English*. Prof. Johansson was responsible for the linguistic analyses and general planning, while Senior Consultant Knut Hofland of the Centre undertook the technical ADP work which included preparation for the production of the book.

The book is based on the Brown and Lancaster-Oslo/Bergen (LOB) corpora. This is the first major publication of word frequencies based on modern British English texts. It consists of an introduction, five different word lists, and a discussion of some of the results obtained from the lists. Three microfiche (about 600 pp.) containing lists are supplied with the book. The price of the book is 300 N.kr. (including postage), and it may be ordered from the Centre.

Meldinger News

In June the Association for Computational Linguistics decided to establish a European chapter. ACL-Europe has the same objective as the main organisation, but it is to attend to the special interests of its European members. Inquiries can be made to: Harry Somers, Centre for Computational Linguistics, U.M.I.S.T., P.O. Box 88, Manchester M60 1QD, England.

Also in June, ACL held its 20th annual meeting in Toronto, Canada. The program included a presentation of models for semantic representation and a discussion of natural languages in communication with data bases.

Association for Computers and the Humanities publishes ACH Newsletter, which is issued 4 times a year. Information is given on conferences, literature and software that is specially relevant to Humanities scholars who use ADP in their work. Membership and subscription costs \$ 15 a year. ACH also publishes the quarterly journal *Computers and the Humanities*. ACH's address: Association for Computers and the Humanities, 216 Lind Hall, University of Minnesota, Minneapolis MN 55455, USA.

Oxford University Computing Service (OUCS) has been established in England. OUCS serves universities with software, computers and expertise in connection with the use of ADP in the Humanities. OUCS offers optical reading of data, suitable program packages, computer corpora, and photo composition. More information can be found in a folder obtainable from OUCS.

Forthcoming conferences

ICCH 83: *see page 54.*

Conference on Applied Natural Language Processing, Santa Monica, California, 1-3 February 1983. This conference will present techniques in computational linguistics aimed at practical problem solution in the following fields: data, information and knowledge management; training and education; language translation technology; and speech recognition and synthesis technology. Further information: Beatrice Oshika MD 72-27 System Development Corporation, 2500 Colorado Ave., Santa Monica, CA 90406, USA.

International Conference on the Application of Mini- and Microcomputers in Information, Documentation and Libraries is to be held in Tel Aviv, Israel, 13-18 March 1983. The conference will deal with existing systems and software, networks, system design and maintenance, plus application in various specialty fields. More information can be obtained from: The Secretariat, International Conference on the Application of Mini- and Micro-computers in Information, Documentation and Libraries, P.O. Box 3054, 61030 Tel Aviv, Israel.

Conference on Data Bases in the Humanities and Social Sciences, New Jersey, USA 10-12 June 1983. This conference is aimed at researchers in the Humanities and Social Sciences who use computer-based research material. More information can be obtained from the co-ordinator of the conference: Professor Robert F. Allen, Room 427,

Alexander Library, Rutgers – The State University of New Jersey, New Brunswick, New Jersey 08903, USA.

AILA (International Association of Applied Linguistics) will hold its 7th World Congress in Brussels, Belgium, 5-10 August 1984. The main theme of the congress will be: «The contribution of Applied Linguistics to international understanding.» Further information will be supplied in a later issue of *Humanistiske Data*.

Activities at the Centre

The Council for Research in the Humanities, The Norwegian Research Council for Science and the Humanities, has established a fellowship at the Centre from 1 January 1983 to 30 June 1984 in connection with ADP methods for storage, retrieval and presentation of material held in archives.

This autumn the Centre will arrange courses for both technical/administrative and scientific personell employed in research work in the Humanities.

RAPPORT nr. 15. *Ivar Fønnes: Tekstsøking på tegnnivå*. Januar 1980. ISBN 82-7283-012-4. Utsolgt.

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

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

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

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

RAPPORT nr. 20. *EDB-prosjekter i humanistiske fag 1980*. Juni 1981. 2. 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*. Juni 1982. ISBN 82-7283-028-0 Under utgivelse.

RAPPORT nr. 26. *Stig Welinder: Paleodemography*. Oslo 1982. ISBN 82-7283-030-2 Under utgivelse

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

Vi takker
NAVFs EDB senter
for oppdragene
med:

- WORD FREQUENCIES IN
BRITISH AND AMERICAN ENGLISH
- HELSE- OG SOSIALPOLITIKK
- HUMANISTISK FORSKNING

Bøkene er på tilsammen over 1200 trykte sider.
Teksten er overført fra datamaskinen til
Universitetet i Bergen, via telefon til vår fotosetter.

bergmann

BERGMANNS BOKTRYKKERI A/S
Vestre Torggate 22, 5000 Bergen
Telefon 31 65 20