

ISBN 82-7283-101-5  
ISSN 1500-499X

UTGITT AV HIT-SENTERET

REDAKSJON:  
CLAUS HUITFELDT (ansv. red.)  
KRISTINE BREVIK  
ANNE LINDEBJERG

ADRESSE:  
ALLÉGATEN 27  
N-5007 BERGEN

HIT-SENTERET                      55 58 29 54  
TELEFAKS                            55 58 94 70

E-POST:                              adm@hit.uib.no  
URL:                                    www.hit.uib.no

## INNHOLD

HIT-SENTERET	4
Programrådet	4
Personale	4
Økonomi	5
ÅRET 2000	6
FIRE UTVALGTE PROSJEKTER	6
ØVRIGE PROSJEKTER	10
Språkteknologi	10
Utgivelsesfilologi	12
Museums- og arkivprosjekter	13
Pedagogisk informasjonsteknologi	14
DELTAKELSE PÅ EKSTERNE KONFERANSER, SEMINARER O.L.	14
EGNE KURS OG SEMINARER	16
PUBLIKASJONER	17
VERV	18

## HIT-senteret

Innenfor humanistisk informasjonsteknologi har Universitetet i Bergen tradisjoner som går helt tilbake til sekstitallet. Som et av de aller første sentre i verden i sitt slag ble *NAVFs edb-senter for humanistisk forskning* etablert i Bergen i 1972. Senere ble senteret omgjort til en seksjon i forskningsstiftelsen Unifob under navnet *Humanistiske datasenter*. Også *Norsk termbank*, som ble opprettet i 1980, ble senere en seksjon i Unifob.

I 1997 opprettet Det historisk-filosofiske fakultet *Forskningsprogram for humanistisk informasjonsteknologi* (HIT-programmet). Samtidig ble Humanistisk datasenter og Norsk termbank slått sammen til én seksjon i Unifob under navnet HIT-senteret.

HIT-senteret er serviceorgan for HIT-programmet. Senteret har beholdt mange av sine nasjonale oppgaver, og mottar grunnbevilgning både fra NFR og fakultetet.

Senteret har dermed både nasjonale oppgaver og oppgaver for Det historisk-filosofiske fakultet, i tillegg til eksterntfinansierte oppdrag. Nasjonalt skal senteret bidra til å utvikle infrastruktur for bruk av IT innen humanistiske fag og fungere som kontaktskaper mellom norsk og internasjonal forskning på området. I tillegg til også å være en serviceenhet for HIT-programmet, er HIT-senterets virksomhet konsentrert om prosjekter innenfor følgende arbeidsområder:

- språkteknologi med hovedvekt på korpuslingvistikk, terminologi og leksikografi
- avansert utgivelsesfilologi
- opprettelse av omfattende digitale ressurser (avanserte web-baserte databaser)
- programvareutvikling

Dessuten har senteret de senere årene blitt mer og mer engasjert i prosjekter innenfor bruk av informasjonsteknologi i utdanning.

HIT-senterets vitenskapelig ansatte er spesialisert innenfor datalingvistikk, korpuslingvistikk, leksikografi og terminologi, utgivelsesfilologi og tekstkritikk.

Senterets fagkonsulenter har til dels meget lang erfaring fra arbeid innenfor humanistisk informasjonsteknologi. Fagkonsulentene har prosjektlederansvar for mange av de eksterntfinansierte prosjektene ved senteret.

Videre har senteret en solid administrativ stab som gjennom mange år har spesialisert seg på administrasjon av eksterntfinansierte prosjekter.

## Programrådet

Programrådet har det overordnede faglige og ressursmessige ansvar for forskningsprogrammet og for HIT-senteret. Rådet som er oppnevnt av HF-fakultetet hadde følgende medlemmer i 2000:

- Breivik, Leiv Egil, professor ved Engelsk institutt (leder)
- Forsberg, Lars, førsteamanuensis ved Arkeologisk institutt
- Jacobsen, Tove, førsteamanuensis ved Romansk institutt
- Myking, Johan, førsteamanuensis ved Nordisk institutt
- Oldervoll, Jan, førsteamanuensis ved Historisk institutt
- Persen, Marte Mjøs, studentrepresentant
- Smedt, Koenraad de, professor ved Institutt for lingvistikk og litteraturvitenskap
- Ytrehus, Line, stipendiat, Senter for Europeiske kulturstudier
- Aarseth, Espen, førsteamanuensis ved Seksjon for humanistisk informatikk
- Hofland, Knut, førstekonsulent ved HIT-senteret (ansattes representant)
- Lønning, Jan Tore, førsteamanuensis ved Institutt for lingvistiske fag, UiO (NFRs representant)
- Ore, Espen S., førstekonsulent ved HIT-senteret (Unifobs representant)

## Personale

Personalet i 2000 utgjorde til sammen ca. 19 årsverk. De vitenskapelige stillingene bestod av ca. 4,5 årsverk (inkl. faglig leder), fagkonsulentstillingene ca. 10 årsverk, de administrative stillingene ca. 3,5 årsverk. En edb-teknisk stilling utgjorde 1 årsverk.

- Thaller, Manfred (faglig leder t.o.m.28.2.)
- Huitfeldt, Claus (faglig leder f.o.m. 1.3.)
- Breivik, Kristine (kontorsjef)

- Apollon, Daniel (førstemanuensis)
- Bakke, Marianne (førstekonsulent, f.o.m. 4.11.)
- Bech, Kristin (koordinator, f.o.m. 1.8.)
- Bjørneset, Tove (førstekonsulent)
- Bruvik, Tone Merete (førstekonsulent, t.o.m. 28.2.)
- Burgess, Edle (konsulent)
- Frihagen, Anders (avd.ingeniør)
- Gunnarson, Eldbjørg (førstesekretær, 50 %)
- Hofland, Knut (førstekonsulent)
- Innselset, Kai (førstekonsulent)
- Lindebjerg, Anne (konsulent)
- Meurer, Paul (forsker)
- Myking, Johan (førstemanuensis, 25 % t.o.m. 30.6.)
- Olstad, Vemund (konsulent, 50 %, 80 % f.o.m. 1.10.)
- Ore, Espen S. (førstekonsulent)
- Pichler, Alois (forsker f.o.m. 1.11.)
- Poppe, Johan Utne (konsulent, 50 %)
- Rangnes, Odd Kjetil (førstekonsulent 1.1.–31.1.)
- Reigem, Øystein (førstekonsulent)
- Revheim, Torill (konsulent)
- Smedt, Koenraad de (forsker II, 11 %)
- Sørensen, Sindre (førstekonsulent f.o.m. 1.9.)
- Thunestvedt, Jørn (konsulent)
- Vestbøstad, Per (førstekonsulent)

## Sivilarbeidere:

- Torjussen, Lars Petter Storm t.o.m. 31.3.

## Eksternt ansatte ved Euroliterature-prosjektet:

- Bonéte, Yvonne (informasjonskonsulent og web-redaktør, deltid)
- Clinkenbeard, Andrea (korrekturleser, teknisk redaktør)

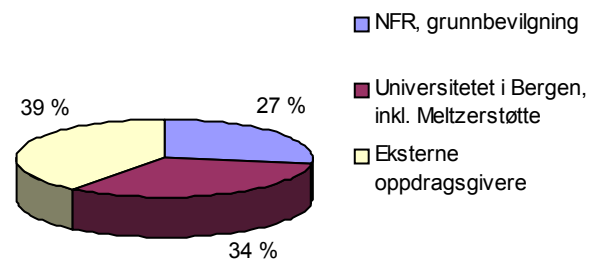
## Økonomi

HIT-senteret fikk i 2000 sine inntekter fra tre hovedkilder:

- Eksterne oppdragsgivere
- Universitetet i Bergen
- Grunnbevilgning fra Norges forskningsråd

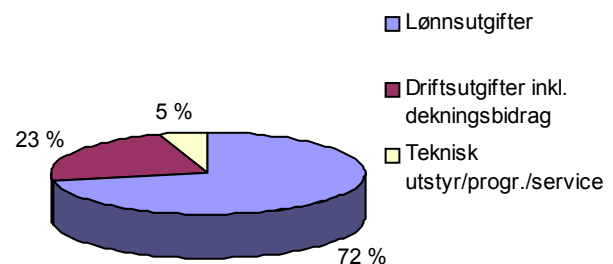
### Inntekter:

NFR, grunnbevilgning	2 380 000
Universitetet i Bergen, inkl. Meltzerstøtte	2 956 693
Eksterne oppdragsgivere	3 474 574
<b>Sum inntekter</b>	<b>8 811 267</b>
Overført fra 1999	1 189 699
<b>Disponible midler</b>	<b>10 000 966</b>



### Utgifter:

Lønnsutgifter	6 058 282
Driftsutgifter	1 928 060
Teknisk utstyr/progr./service	412 422
<b>Sum utgifter</b>	<b>8 398 764</b>
Budsjettavsetning/overført til 2000	1 602 202
<b>Disponible midler</b>	<b>10 000 966</b>



## Året 2000

Få gikk inn i år 2000 uten å føle eller registrere en viss bekymring for det såkalte år 2000-problemet. Verden gikk ikke under. Det gjorde heller ikke HIT-senteret. Både innholdet og omfanget av aktivitetene ved HIT-senteret var i 2000 omlag som året før, som vi i forrige årsmelding beskrev som en konsolidering av virksomheten etter en omfattende omorganisering i 1998.

Et mangeårig samarbeid med Engelsk institutt ved Universitetet i Bergen, COLT (Corpus of London Teenage Language), ble avsluttet med ferdigstilling av en CD-utgivelse i året som gikk.

ACO\*HUM, der HIT-senteret har hatt gleden av å være koordinator for et samarbeid mellom et hundretalls europeiske universiteter, ble avsluttet. Også Euroliterature, et annet EU-støttet prosjekt med omfattende europeisk universitetssamarbeid, ble avsluttet i 2000. Videre ble arbeidet med Meltzer-Not og pilotprosjektet Norsk Talemål avsluttet i 2000.

Innenfor arbeidsområdet språkteknologi opplevde vi en gledelig tilvekst med opprettelse av egen koordinatorstilling fra og med august 2000 og tilsagn fra NFR om støtte til et større prosjekt for Norsk språkbank i 2001–02. Det EU-støttede prosjektet MALVINE er avsluttet, men vårt engasjement videreføres i det nystartede LEAF-prosjektet. Vi er stolte over at en av våre medarbeidere ble invitert til å delta i «Editing Medieval Manuscripts», et prosjekt ved Senter for høyere studier. Helt på slutten av året lyktes vi også i samarbeid med partnere i USA og England i å etablere Text Encoding Initiative som et internasjonalt konsortium.

Nytt av året var at ansatte ved HIT-senteret fra og med 2000 gikk inn i undervisningen til fagstudiene ved Det historisk-filosofiske fakultet ved Universitetet i Bergen mer systematisk og med større tyngde. Vi føler oss sikre på at dette tiltaket er til gjensidig faglig berikelse.

Manfred Thaller, faglig leder ved senteret siden HIT-programmets begynnelse i 1997, sluttet i sin stilling 28. februar. Manfred hadde i sin relativt

korte tid ved senteret gjort en meget solid innsats, som det for undertegnede har vært en betydelig utfordring å følge opp.



Claus Huitfeldt

Faglig leder

## Fire utvalgte prosjekter

Nedenfor følger en gjennomgang av fire prosjekter utført av, eller i regi av, HIT-senteret. Det er en gammel tradisjon for HIT-senterets årsmeldinger at vi presenterer noen utvalgte prosjekter i mer detalj enn de øvrige. Tanken er at disse prosjektene samlet gir et noenlunde representativt bilde av den typen aktivitet som foregår ved senteret. De prosjektene vi har valgt ut for mer omfattende omtale fra år 2000 er Text Encoding Initiative Consortium, Halvautomatisk termekstraksjon for norsk, ACO\*HUM og Euroliterature.

### Text Encoding Initiative Consortium

Text Encoding Initiative (TEI) er en tverrfaglig internasjonal standard for representasjon av elektronisk tekst. Standarden dekker et meget stort antall teksttyper, og brukes i dag av en mengde biblioteker, arkiver, museer, utgivere og forskere verden over for lagring og utveksling av store og små tekstsamlinger. Arbeidet med TEI ble påbegynt i 1987, og gjennomført i form av et internasjonalt samarbeidsprosjekt med bred representasjon fra alle humanistiske fag. Som enhver standard må også TEI vedlikeholdes og oppdateres i takt med nye behov. For å gi dette arbeidet en solid og permanent basis er TEI Consortium nå etablert som et eget foretak. Konsortiet er dannet i samarbeid mellom Brown University, University of Virginia, Oxford University og HIT-senteret. HIT-senteret er foretakets hovedkontor.

TEIs kodesystem er basert på Standard Generalized Markup Language (SGML). Systemet er beskrevet i TEI Guidelines (P3, 1994), en 1300-siders manual som dokumenterer omlag 600 SGML-elementer som på enkelt vis kan settes sammen til brukerdefinerte dokumenttype-definisjoner (DTD). En slik DTD er kjent som TEILite, som har vist seg spesielt populær og benyttes i en rekke dokumentsamlinger tilrettelagt for forskjellige formål.

Ett trekk ved TEI som antagelig har bidratt mer enn noe annet til dets utbredelse, er dets modularitet og fleksibilitet. Systemet er bygget opp av moduler for f.eks. poesi, drama, lingvistisk koding, korpora etc. som kan kombineres på ulikt vis etter behov. De enkelte elementene i hver modul kan omdøpes, modifiseres eller utelates i samsvar med klart angitte regler, bl.a. for dokumentasjon av endringene. TEI kan dermed tilpasses de fleste behov.

TEI startet som et internasjonalt samarbeidsprosjekt i 1987. Ansvarlige for prosjektet var tre av de viktigste organisasjonene innenfor området: Association for Computers and the Humanities (ACH), Association for Literary and Linguistic Computing (ALLC) og Association for Computational Linguistics (ACL). Prosjektet ble finansiert med bidrag fra bl.a. EU (DG XIII), National Endowment for the Humanities, Andrew W. Mellon Foundation og Social Science and Humanities Research Council of Canada. Mer enn 100 forskere fra hele verden deltok i arbeidet, som ble ledet av C.M. Sperberg-McQueen (Chicago) og Lou Burnard (Oxford). Første utgave av TEI Guidelines («P1») ble publisert i 1990, nåværende utgave («P3») i 1994.

Etter at selve prosjektet var fullført hadde TEI liten finansiering, og vedlikehold og oppdatering av kodesystemet, brukerstøtte og informasjonsvirksomhet ble i mange år stort sett utført på uformell og frivillig basis. Det ble etter hvert klart at dette arbeidet burde finne sitt feste i en form for permanent organisasjon. I 1999 tok University of Virginia og HIT-senteret initiativ til å danne en slik organisasjon. Etter kort tid kom også Oxford University og Brown University med, og det ble dannet et interimsstyre bestående av representanter for disse fire universitetene samt de tre ansvarlige organisasjonene (ACH, ALLC og ACL).

Interimsstyret fullførte sitt arbeid i desember 2000 med opprettelsen av TEI Consortium som et

foretak registrert i Virginia og med hovedkontor i Bergen. Lou Burnard (Oxford) og Steve DeRose (Brown) er redaktører.

Konsortiet er organisert som et internasjonalt samarbeid mellom institusjoner og organisasjoner, med Virginia, Brown, Oxford og HIT-senteret som vertsinstitusjoner. I oppstartfasen finansierer vertsinstitusjonene arbeidet, og søker å skaffe videre finansiering og verve flere institusjoner som medlemmer av konsortiet. Konsortiets arbeid skal være drevet av medlemmene, som velger konsortiets styrende organer og råd. Enkeltpersoner vil også kunne abonnere på konsortiets tjenester og delta i dets arbeid.

De viktigste oppgavene for TEI-C i 2001 vil være verving av medlemsinstitusjoner, avholding av valg på styre og råd, samt opprettelse av arbeidsgrupper for videre arbeid med utvikling og vedlikehold av TEI Guidelines. Konsortiet vil også organisere seminar og kurs, drive informasjonsvirksomhet og yte konsulenttjenester.

*Web-referanse:*

[www.hit.uib.no/hit/tei-pro.htm](http://www.hit.uib.no/hit/tei-pro.htm)

## Halvautomatisk termekstraksjon for norsk

Dette prosjektet har som siktemål å utvikle metoder og algoritmer for halvautomatisk termekstraksjon fra norske fagtekster, og implementere disse i programvare som kan benyttes i praktisk terminologiarbeid. Prosjektet er finansiert av Meltzerfondet.

Termekstraksjon krever tilgang til avanserte datalingvistiske ressurser (bl.a. et heldekkende leksikon, en morfologisk parser og en navnegjenkjenner). Slike ressurser har i de siste årene blitt utviklet i ulike norske universitetsmiljøer, bl.a. Scarrie-prosjektet (ved HIT-senteret) og Tagger-prosjektet (ved Tekstlaboratoriet og Dokumentasjonsprosjektet, UiO). Prosjektet tar sikte på å benytte disse ressursene og å utvikle nye programvaremoduler som lett kan gjenbrukes i andre prosjekter.

I det første prosjektåret ble det laget en fullstendig reimplementering av den Constraint Grammar-baserte parseren/taggeren som Tekstlaboratoriet og Dokumentasjonsprosjektet ved UiO har utviklet for bokmål og nynorsk (Oslo-taggeren).

Reimplementeringen kan anses som en fullverdig erstatning av Oslo-taggerens programvare som

også er betydelig raskere enn den opprinnelige programvaren med henblikk på preprosessering, sammensetningsanalyse og multitagging. Den kan kjøres på praktisk talt alle plattformer og er tilgjengelig både som CommonLisp-modul, via en nettverksokkel og via et vev-grensesnitt.

Tekstlaboratoriet har gått over til å bruke nyimplementeringen istedenfor den opprinnelige programvaren for Oslo-taggeren.

[Programvaren ble skrevet i CommonLisp og testet i Macintosh Common Lisp (Macintosh), Allegro Common Lisp (Linux og Windows NT/2000) og CMUCL (Linux). Siden programvaren er skrevet i henhold til ANSI CommonLisp-standarden, vil en port til andre CommonLisp-implementeringer være lett å gjennomføre.]

Den leksikalske databasen som benyttes av parseren (for både bokmål og nynorsk) ble komprimert ved hjelp av en egenutviklet algoritme som komprimerer leksikonet til en indeksert fil på bare 5 % av den opprinnelige størrelsen. Dette gjør det mulig å holde hele leksikonet i RAM.

For å forenkle videreutviklingen av taggeren ble det laget et vev-grensesnitt for testing og redigering av regelsettet og administrering av hele taggesystemet. Bl.a. er det mulig å «klone» taggersystemet og tilpasse kopien til spesielle behov, f.eks. i prosjektet «Navnegjenkjenner».

En av samarbeidspartene i prosjektet er Utenriksdepartementets oversettelsestjeneste, som har gitt tillatelse til bruk av både UD-basen (UDs terminologiske database) og EF-direktoratets dokumentsamling i prosjektet. Et utvalgt antall dokumenter vil danne grunnlaget for testingen av ekstraksjonsmodulen.

For å få tilgang til korpusfunksjonalitet er det satt i gang utvikling av et grensesnitt til CorpusWorkBench-programvaren (Institut für maschinelle Sprachverarbeitung, Stuttgart).

Reimplementeringen av Oslo-taggeren danner grunnlaget for programvareutviklingen i den norske delen av prosjektet «Navnegjenkjenner», et samarbeidsprosjekt mellom miljøer i Danmark, Sverige og Norge (Tekstlaboratoriet, HIT-senteret) om utvikling av automatiske navngjenkjenner for norsk, svensk og dansk. Programvaren er tenkt utvidet med grammatikkregler og filtre for gjenkjenning av bl.a. egennavn, stedsnavn og navn på organisasjoner og bedrifter.

Den nyutviklede tagger-infrastrukturen kommer også til å spille en sentral rolle i utviklingen av en «Norsk språkbank».

*Web-referanse:*

[www.hit.uib.no/hit/termekstraksjon-pro.htm](http://www.hit.uib.no/hit/termekstraksjon-pro.htm)

## **ACO\*HUM «Advanced Computing in the Humanities»**

### **SOCRATES/ERASMUS – et tematisk nettverksprosjekt**

Prof. Koenraad de Smedt har vært koordinator for dette prosjektet som startet i 1996 med 104 partnerinstitusjoner i 17 europeiske land.

Prosjektet ble etter søknad tildelt et 4. og siste prosjektår fra 1. november 1999 til 31. oktober 2000, med et budsjett på 43800 Euro. Det spesielle målet med dette siste året har vært informasjons-spredning og videre utnyttelse av resultater som ble oppnådd i de tidligere prosjektårene.

De spesifikke målene for denne fasen var følgende:

- Lage en kunnskapsbase om utdanning og ressurser i humanistisk informasjonsteknologi
- Flette sammen resultater fra ulike grupper og nettverk
- Lage retningslinjer for best praksis
- Planlegge etter- og videreutdanning
- Utarbeide forslag til studentrettede aktiviteter
- Sikre kontinuering av utnyttelse av prosjektets resultater

Aktivitetene har vært rettet mot to store områder innenfor humanistiske fag:

1. Datalingvistik, språkteknologi og talekommunikasjon (inkl. ikke-indoeuropeiske språk)
2. Tekstteknologi og humanistisk informasjonsteknologi

Prosjektets aktiviteter bestod av møter, workshops, deltagelse i konferanser og seminarer og tilrettelegging av informasjon på Internett.

De viktigste resultatene som ble oppnådd dette året består av:

- En utvidbar kunnskapsbase på Internett som inneholder:
  - (a) kurs i datalingvistik, språkteknologi og talekommunikasjon
  - (b) kurs i tekstteknologi og humanistisk informasjonsteknologi
  - (c) utdanningsressurser for datalingvistik, språkteknologi og talekommunikasjon

- En plan for etter- og videreutdanning
- Et utarbeidet forslag til studentrettede aktiviteter
- Retningslinjer for best praksis
- Avtaler med ELSNET og ALLC om kontinuering av prosjektets resultater

En fullstendig sluttrapport om prosjektet, med alle resultatene, er tilgjengelig fra nettsiden.

*Web-referanse:*

[www.uib.no/acohum](http://www.uib.no/acohum)

## **Euroliterature**

Euroliterature er et treårig EU-finansiert samarbeidsprosjekt som startet i september 1997, som videreføring av tidligere pionerprosjekter finansiert av SOCRATES (HUMANITIES I-IV, Transcult, Pegasus bl. a.) og i nært samarbeid med andre EU prosjekter (for eks. CRE-ODL koordinert av Den europeiske rektorkonferansen, Genève). Med bakgrunn i tidligere erfaringer og feltstudier innen europeisk virtuell læring (Open Distance Learning) har Euroliterature satt søkelyset på integrasjon av ODL i tradisjonelle universitetsdisipliner. Eksperimentell undervisning, evalueringer og akademiske produkter har fokusert på fagfeltet «allmenlitteratur og tekstvitenskap i møte med nye medier». Prosjektets praktiske målsetting har vært å integrere virtuell læring og multimediarkeitektur i studiesituasjon og etablere samarbeid om studieplan mellom en rekke europeiske faginstututter. Parallellt har prosjektfellesskapet gjennomført akademiske studier, evalueringer og organisert workshop og konferanser. Universitet i Bergen, HIT-senteret har koordinert prosjektet (koordinator Daniel Apollon).

Det tredje og siste prosjektåret (1.9.1999 – 31.12.2000 inkl. en tre måneders forlengelse) har gjennomført en omfattende formidlingsaktivitet med følgende høydepunkter:

- Organisering av «European Euroliterature III Integration Workshop» ved Carlos III universitet, Madrid, oktober 1999.
- Organisering og gjennomføring av Seksjon V, «The Future of European Education», ISSEI 2000 konferansen, organisert av UiB, 14.–18. august 2000 med 734 deltagere.
- Gjennomføring av et eksperimentelt europeisk kurs «The Future of Literature and Translation in the Digital Age», med deltagelse bl. a. av universitetene i Leuven, Carlos III (Madrid), Granada, Roehampton Institute (London), Mainz-Germersheim, Aix-en-Provence, Santiago de Compostela, Almeria, Tarragona, Leiden, Turku, Åbo Akademi.
- Publikasjon, produksjon og distribusjon av CDROM, Apollon, Daniel & Fure, Odd-Bjørn & Svåsand, Lars (Eds.), «Proceedings of the 7th Conference of the International Society for the Study of European Ideas (ISSEI), 14.–18. august 2000 – University of Bergen – Norway», ISBN 82-91320-20-9 © 2000. Publikasjonen inneholder mer enn 300 bidrag. Teknisk redaktør: Anne Lindebjerg og Jørn Thunestvedt, HIT-senteret, UiB.
- Publikasjon av felles europeisk bok: Austermühl, Frank & Apollon, Daniel, «Humanities Education and the Challenge of the Information Society», Bergen-Germersheim desember 2000 (under trykking) med en rekke bidrag fra europeiske forskere. Teknisk redaktør: Andrea Clinkenbeard, UiB.
- Tollef Ladehaug Johansen gjennomførte hovedfagsavhandling «Institutional conditions for learners attitudes towards Open and Distance Learning (ODL) in humanistic departments», juni 2000, Institutt for administrasjon og organisasjonsvitenskap, UiB.
- Gjesteforsker Beril Akinci, finansiert av NFR i samarbeid med Euroliterature, forsvarte doktorgraden på forskningsmateriale innsamlet og bearbeidet i Euroliterature-sammenheng. Arbeidet tar for seg holdninger til IKT blant litteraturorienterte forskere ved UiB, Granada og Carlos III. Feltarbeidsresultatene blir analysert i lys av moderne HRM (Human Resource Management) teori.

*Web-referanse:*

[www.euroliterature.uib.no](http://www.euroliterature.uib.no)

## Øvrige prosjekter

### Språkteknologi

#### Prosjektutvikling innen språkteknologi

Fra 1. august ble HIT-senteret arbeidsplass for HF-fakultetets koordinator for språkteknologi, Kristin Bech, hvis oppgaver først og fremst består i å samle og spre informasjon, fungere som kontaktledd til andre institusjoner, hjelpe faglig ansvarlige med å initiere og utvikle prosjekter, samt hjelpe til med utarbeidelse av søknader. I samarbeid med Koenraad de Smedt, Claus Huitfeldt og Helge Dyvik er noen søknader blitt utarbeidet, og andre påbegynt, for forskjellige prosjekter med språkteknologisk relevans, bl.a. til Meltzerfondet (WordNet), Bergen Universitetsfond (ICAME), Norges Forskningsråd (Norsk språkbank), Nordisk Ministerråd (dokumentasjonssenter), og EU (Marie Curie Training Site). I 2000 ble det dessuten skrevet en intensjonsavtale for kontakt og samarbeid med språkteknologimiljøet på Voss, det ble etablert kontakter med NFR og Norsk Språkråd, og arbeidet med en søknad for et Senter for Fremragende Forskning innen datalingvistik og språkteknologi ble påbegynt.

*Web-referanse:*

[www.hit.uib.no/westnode/index\\_norsk.htm](http://www.hit.uib.no/westnode/index_norsk.htm)

### Korpusprosjekter

#### Aviskorpus

Arbeidet med automatisk å samle inn tekstmateriale fra web-utgavene av utvalgte riks- og regionavisser fortsatte i 2000. Materialet passerte 100 millioner ord i oktober etter ca. 2 års drift. Etter en kort indekseringsperiode er hele materialet søkbart via et web-grensesnitt. Materialet deles nå automatisk inn i en nynorsk- og en bokmålsdel. Dette gjøres på grunnlag av forekomster av en del hyppige ord som er spesielle for hhv. nynorsk og bokmål. Hver dag blir det laget en liste med ordformer som ikke har forekommet tidligere i materialet. Systemet ble presentert på konferansen LRC2000 i Athen.

*Web-referanse:*

[www.hit.uib.no/hit/avis-pro.htm](http://www.hit.uib.no/hit/avis-pro.htm)

#### Norsk talemålskorpus

Dette er et pilotprosjekt som har utviklet og testet metoder som gjør det mulig å bygge opp et digitalisert, søkbart talemålskorpus. I prosjektet er det digitalisert og transkribert ca. 18 timer tale med informanter fra Bergen, Voss og Tromøya. Materialet er søkbart via en Web-leser og etter et søk i materialet kan en spille av den aktuelle lyden til de enkelte konkordanslinjene.

I 2000 har materiale fra pilotprosjektet blitt tagget i Oslo ved hjelp av Oslo-taggeren. I forlengelse av pilotprosjektet har det blitt lagt inn materiale for hovedfagstudenter og doktorstudenter. Også noe av dette materialet har blitt tagget i Oslo. Det er laget et forbedret søkeskjema der en kan oppgi informasjon om informanter og der en også kan søke etter tagger (ordklasser) i materialet.

*Web-referanse:*

[www.hit.uib.no/hit/tale-pro.htm](http://www.hit.uib.no/hit/tale-pro.htm)

#### UNO

##### (Ungdomsspråk i Norden)

UNO er et forskningsprosjekt og et nettverk for nordiske forskere på ungdomsspråk.

Prosjektet undersøker og sammenlikner nordiske ungdommers talespråk og uformelle skriftspråk.

Prosjektet har brukt programmet SyncWriter (Macintosh) til transkripsjon. Dette programmet hadde ingen eksportmulighet, men ved å ta vare på utskriftsfil (Postscript) og dekode denne har en kunnet få ut nok informasjon til å bruke Senterets system for søk i talekorpus.

*Web-referanse:*

[www.uib.no/uno/](http://www.uib.no/uno/)

#### ICAME

##### (International Computer Archive of Modern and Medieval English)

ICAME er en uformell organisasjon av forskere som arbeider med samlinger av engelske maskinlesbare tekster (korpus). Målet er å bygge opp et arkiv over engelske korpus og å gjøre disse tilgjengelig for andre forskere.

HIT-senteret er vert for arkivet som pr. i dag inneholder 20 korpus med tilsammen 17 millioner ord. Arkivet selges på en egen CDROM, ICAME Collection of English Language Corpora.



År 2000 var et vanlig driftsår med ordremottak, brukerveiledning og svar på innledende forespørsler angående materiale. Vi distribuerte 139 kopier av ICAME CDROM. Senteret driver også distribusjonslisten Corpora med mer enn 1500 medlemmer.

*Web-referanse:*

[www.hit.uib.no/hit/icame-pro.htm](http://www.hit.uib.no/hit/icame-pro.htm)

## COLT

### (Corpus of London Teenage Language)

The Bergen Corpus of London Teenage Language (COLT) er et talespråkkorpus basert på spontane samtaler mellom 13 – 17-åringer i London. Korpuset består av omtrent 472.000 ord transkribert tekst.

Materialet utgjør transkribert tekst av ca. 55 timer lydopptak med tilhørende digitaliserte lydfiler. Tekstene finnes også i et tagget format der hvert ord har fått en ordklassekode.

Det er laget et web-sted hvor brukere kan søke i materialet og spille av lyd. Det er også mulig å bla kronologisk i filene og spille av tilhørende lydsegment.

I 2000 ble hele materialet publisert på CDROM. Systemet ble presentert på ICAME-konferansen i Sydney og ALLC/ACH-konferansen i Glasgow. Under åpningsforedraget i Glasgow ble systemet nevnt som et eksempel på en interessant utvikling innen korpuslingvistikk.

*Web-referanse:*

[www.hit.uib.no/hit/colt-pro.htm](http://www.hit.uib.no/hit/colt-pro.htm)

## Santa Barbara Corpus

Dette er et amerikansk talespråkkorpus og den første delen er tilgjengelig som tre CD-plater med lyd- og tekstfiler. Hver tekstlinje har markering for start- og sluttid i den tilhørende lydfil. Materialet er konvertert og gjort søkbart med systemet som er utviklet ved HIT-senteret for søk i talekorpus.

## Fransk-norsk parallellkorpus

Et materiale bestående av nyhetsartikler på norsk og fransk (Norinform) har blitt parallellstilt og indeksert for søking med WebTCE. Det er laget en spesialdatabase over noen tekster som finnes på fransk, norsk og engelsk.

## Terminologiprosjekter

### NOT-basen

Denne termdatabasen er HIT-senterets viktigste terminologiske produkt. Den er resultatet av mange års terminologiarbeid i forbindelse med ulike prosjekter. Grunnlaget ble lagt tidlig på 80-tallet da det skulle utvikles norsk fagterminologi for oljeindustrien. Databasen er nå blitt til en mer allsidig fagspråksdatabase med omlag 90 000 termer fordelt på nærmere 30 000 begreper innenfor 38 fagområder (foruten oljeindustri bl.a. jus, administrasjon, optikk, miljø, biologi, kjemi og kvalitetssikring). I 2000 ble basen gjennomgått med tanke på generell kvalitetssikring og utvelging av oljerelatert terminologi for levering til Nordterm-banken. Det har vært naturlig å se disse oppgavene i sammenheng. Det meste av den oljerelaterte terminologien (dvs. selve grunnstammen i NOT-basen) er gjennomgått og godkjent. Rundt 3000 begrepsposter av det godkjente materialet er merket ut med henblikk på en mulig framtidig leveranse.

*Web-referanse:*

[www.hit.uib.no/hit/not-pro.htm](http://www.hit.uib.no/hit/not-pro.htm)

## Norsk terminologisk database.

### Struktur og applikasjoner

#### («Meltzer-NOT»)

Sentralt i dette prosjektet, som ble avsluttet i 2000, stod utviklingen av en fleksibel databasearkitektur og et sett av verktøy for terminologiske databaser, spesielt HIT-senterets terminologibase NOT.

Et program for terminologisk redigering er utviklet og ferdigstilt. Programmet har vært i bruk både innenfor NOT-prosjektet og i NORD-LEXIN-prosjektet. I tillegg er programmet tatt i bruk i Utenriksdepartementets oversettertjeneste. Databasearkitekturen og redigeringsprogrammet er altså fleksible nok til å hanskes både med terminologiske og leksikografiske databaser. Det er også utviklet en Internett-tjener for terminologiske databaser. Tjeneren bruker den samme databasearkitekturen som redigeringsverktøyet.

Resultatene av prosjektet bærer i seg viktige arbeidsoppgaver som kan føres videre, og som det er stort behov for. En av de viktigste oppgavene vil være å utvikle maskinelle metoder for semiautomatisk ekstrahering av termer fra tekst. Da må det

finnes et utbygd elektronisk fagtekstkorpus, som er påbegynt, og en datalingvistisk ordbok, som det arbeides med i andre fagmiljø i Norge.

*Web-referanse:*

[www.hit.uib.no/hit/meltzer-not.htm](http://www.hit.uib.no/hit/meltzer-not.htm)

## Halvautomatisk termekstrasjon for norsk

Se side 7.

*Web-referanse:*

[www.hit.uib.no/hit/termekstraksjon-pro.htm](http://www.hit.uib.no/hit/termekstraksjon-pro.htm)

## NORDLEXIN-N

Ordboksprosjektet *NORDLEXIN-N* tek utgangspunkt i dei svenske LEXIN-ordbøkene. Dette er ordbøker som er utvikla spesielt for minoritetspråklege innvandrarakar med lita erfaring i bruk av ordbøker. Mellom anna er rundt 1600 av orda illustrerte. I Sverige er desse ordbøkene no omsette til rundt 20 språk og utgjevne i både trykt og elektronisk form. Sidan 1996 har HIT-senteret på oppdrag frå Kyrkje-, undervisnings- og forskingsdepartementet og Læringssenteret (sidan 1999) arbeidd med å utvikle tilsvarende ordbøker for norske forhold.

I 1996–1999 arbeidde prosjektgruppa med å overføre det svenske ordboksunderlaget til norsk. Ekvivalensen var større enn venta, og rundt 80 prosent av det svenske ordboksmaterialet kunne overførast.

I år 2000 har pilotprosjektet norsk–engelsk ordbok vore hovudaktiviteten, og dette arbeidet vil halde fram også i 2001. For mange minoritetspråklege fungerer verdsspråket engelsk som eit andre morsmål, og det finst eit mangfald av ordbøker mellom verdsspråka og mindre språk som kan vere til god hjelp for personar som skal lære seg norsk. Den norsk–engelske *NORDLEXIN-N*-ordboka skal etter planen leggjast ut på Læringssenteret sine nettsider for minoritetspråklege i 2001.

Det er laga nye fargeillustrasjonar til ordbøkene. Arbeidsomfanget vart estimert til eitt årsverk, og sju illustratørar tok del i arbeidet. HIT-senteret administrerte kontraktsinngåing og utbetalingar, og har fungert som kontaktledd mellom partane.

Gjennom året har det vorte etablert mange interessante og verdfulle kontaktar med mange per-

sonar og fagmiljø. Dette gjeld ikkje minst morsmålslærarakar og andre ressurspersonar innan ulike minoritetspråk, mellom anna tamilsk, som etter planen skal vere det neste målspråket i prosjektet.

Tre leksikografar, ein programmerar og sju illustratørar har vore engasjert i prosjektet dette året.

*Web-referanse:*

[www.hit.uib.no/hit/nordlexin-pro.htm](http://www.hit.uib.no/hit/nordlexin-pro.htm)

## Utgivelsesfilologi

### Text Encoding Initiative Consortium

Se side 6.

*Web-referanse:*

[www.hit.uib.no/hit/tei-pro.htm](http://www.hit.uib.no/hit/tei-pro.htm)

### Wittgensteinarkivet

I løpet av 1999 fullførte Wittgensteinarkivet (WAB) produksjonen av den maskinleselige versjonen av Ludwig Wittgensteins Nachlass. I samarbeid med Oxford University Press har WAB i november 2000 publisert «Wittgenstein's Nachlass: The Bergen Electronic Edition», som er en CDROM-utgave og baserer seg på Wabs maskinleselige versjon. WAB befinner seg nå i vedlikeholdsfasen.

I september 2000 ble WAB omorganisert til et samarbeidsprosjekt mellom Filosofisk institutt og HIT-senteret ved Universitetet i Bergen. Prosjektet er lokalisert til HIT-senteret. Prosjektets formål er vedlikehold av den maskinleselige versjonen og «Bergen Electronic Edition», veiledning av gjesteforskerne som kommer til WAB og oppfølging og utvidelse av internasjonalt samarbeid. F.o.m. 1.11. disponerte WAB en fast halv stilling, som var finansiert av HF-fakultetet og HIT-senteret.

*Web-referanse:*

[www.hit.uib.no/hit/wab-pro.htm](http://www.hit.uib.no/hit/wab-pro.htm)

## Henrik Ibsens skrifter

Prosjektets mål er å tilrettelegge en ny vitenskapelig og tekstkritisk utgave av Henrik Ibsens samlede arbeider. Utgaven vil bli publisert både i bokform og i elektronisk utgave. Den elektroniske utgaven vil bygge på et elektronisk arkiv som man regner med vil bli vedlikeholdt også etter dette prosjektets levetid. Prosjektet startet våren 1997 og er planlagt ferdigstilt i 2008. I 2000 ble den første Ibsen-teksten publisert som en pilotutgave – et manuskript til Ibsens *Catilina*.

Prosjektet finansieres av NFR og er et samarbeid mellom universitetene i Oslo, Bergen og Trondheim, Nasjonalbiblioteket og HIT-senteret.

HIT-senteret deltar i prosjektredaksjonen, og har spesielt ansvar for alt overordnet IT-relatert arbeid samt tekstkoding. Senteret har også ansvar for tilpassing av programmer for redigering og formidling i tillegg til løpende IT-assistanse.

*Web-referanse:*

[www.hit.uib.no/hit/hi-pro.htm](http://www.hit.uib.no/hit/hi-pro.htm)

## Editing Medieval Manuscripts

### Prosjekt ved Senter for høyere studier

Prosjektgruppen «Redigering av middelaldermanuskripter», som ledes av Odd Einar Haugen, vil fokusere på tilrettelegging av elektroniske utgaver av gammelnorske og islandske tekster fra middelalderen. Tre hovedområder har utkrystallisert seg:

1. Tekstrepresentasjon: hvilke typer paleografisk og ortografisk informasjon i manuskriptene bør kodes?
2. Tekstrekonstruksjon: bør tekster analyseres og behandles som primærkilder, eller bør tekstvarianter fra andre kilder inkluderes i kodingen?
3. Tekstpresentasjon: hvordan kan man presentere (og distribuere) de kodede tekstene i en form som er i tråd med tradisjonelle utgaver?

HIT-senteret bidrar med tekstkoding, DTD-utvikling og generell IT-støtte.

*Web-referanse:*

[www.shs.uio.no/Groups/EdMa2000](http://www.shs.uio.no/Groups/EdMa2000)

## Museums- og arkivprosjekter

### Musved-prosjektet

I dette prosjektet vedlikeholdes og videreutvikles med databaseprogram for katalogisering av kulturhistorisk kjeldemateriale: Foto, gjenstandar, boksamlinger, kunstindustrigjenstandar m.m.

Me har kring 250 installasjonar av databasen winRegimus: flest ved kulturhistoriske museum, ein del i kommunale fotosamlingar pluss nokre i andre sektorar: konsulentar, vegkontor, sjukehus og andre verksemder som har historiske samlingar.

Det er halde sju kurs i bruken av winRegimus ulike stader i landet. Prosjektmedarbeidarar: Øystein Reigem (programmering) og Per Vestbøstad (prosjektleiar).

*Web-referanse:*

[www.hit.uib.no/hit/winregimus-pro.htm](http://www.hit.uib.no/hit/winregimus-pro.htm)

### European Cultural Heritage Technician

«European Cultural Heritage Technician» er et EU-finansiert prosjekt, der HIT-senteret samarbeider med yrkesutdanningsinstitusjonen Consorzio Provinciale per Formazione Professionale di Ravenna og Central European University, Budapest. Prosjektet tar sikte på å utarbeide og gjennomføre et undervisningsopplegg som kombinerer multimediaferdigheter med grunnleggende kulturarvskunnskap. Denne utdannelsen skal kvalifisere for en karriere innen kulturarvsarbeid, dog på et lavere nivå enn det vanlige akademiske kurs legger opp til. HIT-senteret har bidratt til dette prosjektet med å utarbeide pensumlitteratur til kursene i vår- og høst-semesteret 2000. I tillegg har HIT-senteret vært med i utviklingen av et digitalt billedarkiv til bruk i del 2 av undervisningen.

*Web-referanse:*

[gandalf.hit.uib.no/non/eht/](http://gandalf.hit.uib.no/non/eht/)

### Malvine

Manuscripts and Letters via Integrated Networks in Europe

Dette er eit EU-støtta prosjekt som samkøyrer katalogane i europeiske manuskriptsamlingar.

Bibliotek, arkiv, museum og andre dokumentasjonsentra over heile Europa har til dels store mengder upubliserte manuskript og brev frå vidgjetne og sentrale personar i europeisk kultur- og samfunnsliv. Slike dokument er vanlegvis berre registrerte (og i beste fall beskrivne) der dei er lagra, og registreringa følgjer ingen vedteken standard. Katalogane over desse samlingane er oftast berre tilgjengelege ved dei aktuelle institusjonane, sjølv om enkelte katalogar er publiserte (på trykk og/eller på Internett).

11 europeiske manuskriptsamlingar, 3 utviklarar og eit par faglege rådgjevarar frå universitets- og biblioteksmiljøa i EU, Sveits og Norge har gjennom Malvine-prosjektet vist korleis ein kan betra tilgangen til eigne manuskriptsamlingar ved å tilby felles katalogsøk frå *eitt* Internett-punkt. Der kan interesserte forskarar, studentar og andre slå opp i utvalde katalogar frå alle dei 11 samlingane ved hjelp av eitt, felles søk frå nettsida til Malvine. Prosjektet vart avslutta 23.1.2001, etter at dei lovande resultatane vart presenterte på ein Malvine-konferanse i Berlin 4.–5. desember.

Prosjektet ser ut til å munna ut i ein avtale om permanent drift og utviding av søkjekonseptet, slik at ein kan tilby ei permanent løysing, som har kapasitet til å ta hand om katalogar frå mange fleire samlingar enn dei som er med i Malvine-prosjektet. Statsbiblioteket i Berlin har teke på seg drifts- og administrasjonsansvaret for Malvine-programma vidare framover.

**LEAF-prosjektet** (Linked European Authority Files) er eit EU-støtta prosjekt med stort sett dei same partnarane som Malvine, og skal føra vidare arbeidet med å letta tilgangen til europeiske manuskript- og brevsamlingar via Internett. Hovuddinnhaldet er ei samordning av namneinformasjonen i ulike samlingar, for at slik standardisert informasjon skal kunna brukast som inngangsport ved søk etter manuskript og brev på tvers av ulike samlingskatalogar. LEAF startar opp 1.3.2001.

*Web-referanser:*

[www.hit.uib.no/hit/malvine-pro.htm](http://www.hit.uib.no/hit/malvine-pro.htm)

[www.hit.uib.no/hit/leaf-pro.htm](http://www.hit.uib.no/hit/leaf-pro.htm)

## Pedagogisk informasjonsteknologi

### ACO\*HUM

«Advanced Computing in the Humanities»

Se side 8.

*Web-referanse:*

[www.hd.uib.no/AcoHum/aco-hum.html](http://www.hd.uib.no/AcoHum/aco-hum.html)

### Euroliterature

Se side 9.

*Web-referanse:*

[www.euroliterature.uib.no/](http://www.euroliterature.uib.no/)

## Annet

### ISSEI 2000 – Proceedings på CDROM

Den 7de ISSEI konferansen «Approaching a New Millennium: Lessons from the Past – Prospects for the Future» ble arrangert av Universitetet i Bergen, 14.–18. august. HIT-senteret fikk oppdraget med å publisere en CDROM med HTML-grensesnitt som inneholder innlegg, taler og avisartikler fra konferansen.

Ansvarlig for produksjonen var Anne Lindebjerg og Jørn Thunestvedt.

*Web-referanse:*

[www.hit.uib.no/hit/issei-cdrom.htm](http://www.hit.uib.no/hit/issei-cdrom.htm)

## Deltakelse på eksterne konferanser, seminarer o.l.

Apollon, Daniel:

- En dags workshop, Rektorkollegiet i Hordaland (videregående skoler), «Utdanning for fall» – Samfunnsmessige og kulturelle utfordringer – Konsensusøvelser etter SWOT metoden (65 deltakere), Solstrand Fjord Hotell, Os, 9.–10. november.

Bjørneset, Tove:

- Fagspråkskonferansen 2000, Institutt for språk, Norges Handelshøyskole 7. april: «Hvilke utfordringer stiller språkteknologien oss overfor i det nye årtusen?»

- The 10th International Symposium on Lexicography, Københavns Universitet, 4.–6. mai. Innlegg under publisering i *Lexicographica Series Maior*. Max Niemeyer Verlag Tübingen: «The Dictionary Project NORDLEXIN–N».
- ISSEI 2000 (Seventh Conference of the International Society for the Study of European Ideas), Bergen, 14.–18. august. Innlegg publisert på CDROM: «Dictionaries for Immigrants and Refugees in Norway: NORDLEXIN–N».
- SKIKT-konferanse: Kunnskapens og dialogens vilkår i informasjonssamfunnet. SKIKT = Samfunnsmessige og kulturelle forutsetninger for informasjons- og kommunikasjons-teknologi, Oslo, 24. oktober.
- ITU-konferansen 2000: «So what?» Arrangør: ITU – Forsknings- og kompetansenettverk for IT i utdanning, Oslo, 2.–3. november.
- LREC 2000 2nd International Conference on Language Resources & Evaluation, Athen, Hellas, 29. mai – 2. juni, «A Self-Expanding Corpus Based on Newspapers on the Web» (foredrag) ICAME CDROM, COLT CDROM (demo) [www.xanthi.ilsp.gr/lrec/default.htm](http://www.xanthi.ilsp.gr/lrec/default.htm)
- ALLC / ACH 2000, Glasgow, Scotland, 21.–25. juli, The Second version of the ICAME CDROM (demo), Working with alignment of text and sound in spoken corpora (poster), [www2.arts.gla.ac.uk/allcach2k/](http://www2.arts.gla.ac.uk/allcach2k/)

#### Huiffeldt, Claus

- «Texts and Text Encoding», forelesning ved IATH, UVA, Charlottesville (VA), USA, 31. mars.
- «Editorial Principles of the Wittgenstein Archives», forelesning ved Filosofisk institutt, Universitetet i Bologna, Italia 21. juni.
- «The TEI Consortium», leder for panel ved konferansen Digital Rescours for the Humanities ([www.shef.ac.uk/~drh2000/](http://www.shef.ac.uk/~drh2000/)), Sheffield, England 9.–13. september.
- «GODDAG – A Data Structure for Overlapping Hierarchies», foredrag sammen med C.M.Sperberg-McQueen ved PODDP Workshop, München, Tyskland 14. september. [www.cs.uwm.edu/~poddp00/](http://www.cs.uwm.edu/~poddp00/)
- «Editorial Principles of the Electronic Wittgenstein Edition» og «Text Encoding: MECS and its relationship to XML/SGML», to foredrag på «Informatica humanistica: filosofia e risorse digitali», Filosofisk institutt, Univ. i Bologna, Italia 22.–23. september.

#### Breivik, Kristine:

- How to Write a Competitive Proposal and Manage an EU R&D Contract. Arr. Forskningsavd. UiB. Foreleser Dr. Sean McCarthy. Bergen, 12. januar.
- ITU-konferansen 2000: «So what?» Arrangør: ITU – Forsknings- og kompetansenettverk for IT i utdanning, Oslo, 2.–3. november.
- Seminar «Service og kvalitet i administrasjonen i Unifob» Voss, 8.–9. november.

#### Bruvik, Tone Merete

- EU-prosjektet LEAF (=Linking and Exploring Authority Files): Kontraktstingingar i Luxemburg, 12. september. Prosjektet skal startast 1. mars 2001, og går over 3 år.
- EU-prosjektet MALVINE (= Manuscripts and Letters via Integrated Networks in Europe): Sluttkonferanse og siste prosjektmøte, Statsbiblioteket i Berlin 4.–6. desember. Innlegg «XML in MALVINE and LEAF».

#### Gunnarson, Eldbjørg:

- How to Write a Competitive Proposal and Manage an EU R&D Contract (del 5–8). Arr. Forskningsavd. UiB. Foreleser Dr. Sean McCarthy. Bergen, 12. januar.
- Seminar «Service og kvalitet i administrasjonen i Unifob» Voss, 8.–9. november.

#### Hofland, Knut:

- ICAME 2000, Sydney, Australia, 21.–25. april, «Working with alignment of text and sound in spoken corpora» (foredrag) ICAME CDROM (demo) [www.ling.mq.edu.au/icame2000/](http://www.ling.mq.edu.au/icame2000/)

#### Innselset, Kai:

- Fagspråkskonferansen 2000, Institutt for språk, Norges Handelshøyskole, 7. april: «Hvilke utfordringer stiller språkteknologien oss overfor i det nye årtusen?»

#### Lindebjørg, Anne:

- Grunnleggende og videregående kurs i Unix, Getronics, Oslo, 19.–23. juni.

#### Ore, Espen. S.:

- Stiftelsesgeneralforsamling i XML-Scandinavia, Oslo, 16. mai.
- Workshop i regi av ALLC/ACO\*HUM: Computerizing Texts, University of Macedonia i Thessaloniki, 25.–26. mai, Foredrag/presentasjon.

- Ibsenkonferanse, Bergen, 6.–7. juni. Presentasjon av kodearbeidet og elektroniske utgaver i forbindelse med den nye Ibsenutgaven. Ansvarlig for praktiske arrangementer rundt program og datademonstrasjoner.
- Lansering av en diplomatarisk utgave av et manuskript til Catilina ble lansert i bokform og med medfølgende CD, Oslo, 16. juni, Demonstrerte deler av CDen.
- DRH 2000, Sheffield, 9.–13. september, Møter om utviklingen av TEI Consortium. Deltager i en panelpresentasjon av TEI Consortium.
- NNE (Nordisk nettverk for edisjonsfilologi), Oslo, 20.–22.10. Ansvarlig for IT-aspekter ved konferansen.
- XML-Norge, Hafjell, 21.–22. november.

Pichler, Alois:

- «Two Readings of Wittgenstein's Philosophical Investigations». Gjeste forelesning ved Universitetet i Wrocław (Polen), Filosofisk institutt og Polsk filosofisk selskap. 12. juli.

Reigem, Øystein:

- Kurs i OpenInsight, Sprezzatura UK Ltd, London. 20.–22. september.

Smedt, Koenraad de:

- Teaching Humanities in the Information Age (Some reflections on studies in humanities computing). Foredrag: What's all the Hype in Hypertext About? A Humanities Computing Colloquium. Dublin, Irland, 10.–11. april.
- 2nda Cumbre del Milenio sobre Informática. Publisert: Human Sciences and Information Technology. I: 2nd Millennium Summit on Information Technology. Gobierno de Puerto Rico; San Juan, Puerto Rico, 9.–11. august.
- Modeling flexibility in generation architectures. Foredrag: Tagung der DFG-Schwerpunktes Sprachproduktion. Wadern, Tyskland, 4.–7. september.

Thunestvedt, Jørn:

- Kurs i CIW Foundations (Internett, html, nettverk), Getronics, Bergen, 25.–28. april.

Vestbøstad, Per:

- MCN2000: Museum Computer Network sin årskonferanse i Las Vegas, Nevada, 6.–9. september.

- EU-prosjektet MALVINE (= Manuscripts and Letters via Integrated Networks in Europe.): Sluttkonferanse og siste prosjektmøte, Statsbiblioteket i Berlin 4.–6. desember.

## Egne kurs og seminarer arrangert av eller i samarbeid med HIT-senteret

### HIT-seminarer

21. januar: *Tankens form – visualiseringar som humanistisk verktyg*, professor Jan von Bonsdorff, Institutt for kunsthistorie, Universitetet i Tromsø.
24. mars: *Major stages and aspects of corpus linguistics*, gjesteforsker Jianxin Wang, (Beijing University of Post and Telecom), HIT-senteret.
25. mai: *Current trends and reform in Chinese universities and higher education system*, gjesteforsker Jianxin Wang. Seminar i samarbeid med Senter for internasjonalt universitetssamarbeid.
15. juni: *Some new progress in computational and corpus linguistics in China*, gjesteforsker Jianxin Wang, HIT-senteret.
27. oktober: *Enspråklige ordnett fra flerspråklige data*. En metode for å avlede semantiske nettverk fra oversettelseskorpora. Professor Helge Dyvik, Inst. for lingvistikk og litteratur, Seksjon for lingvistiske fag, UiB.
15. november: *Tverrfaglig nettbasert databasesystem*. Faglig leder Christian-Emil Smith Ore, Museumsprosjektet, UiO.
24. november: *Informasjonsfiltrering for tematiske avgrensede dokumentsamlinger*. Amanuensis Brit Helle Aarskog, Institutt for informasjonsvitenskap, UiB.

### Undervisning

- Undervist på alle nivåer, delfag, grunnfag, mellomfag (påbygning) ved Seksjon for humanistisk informatikk, HF, UiB. Foreleser: Daniel Apollon.
- Forskerkurs i dataanalyse i samarbeid med Institutt for fonetikk og lingvistikk. Datalingvistikk, Hovedfag. Foreleser: Daniel Apollon.

- Gjesteforelesning ved Institutt for kulturstudier, Seksjon for kulturarv. Foreleser: Daniel Apollon.
- Forelesninger ved Seksjon for humanistisk informatikk, korpuslingvistikk, 6 ganger. Foreleser: Knut Hofland.
- Forelesninger på grunnfag språkteknologi, Voss, korpuslingvistikk, 3 dager. Foreleser: Knut Hofland.
- Høsten 2000: Undervisning i Ling105: Særpråk georgisk. Foreleser: Paul Meurer.
- Forelesninger på Humanistisk informatikk, programmering, 11 ganger. Foreleser: Øystein Reigem.
- Forelesninger på Humanistisk informatikk, grunnfag høsten 2000, 6 ganger. Foreleser: Claus Huitfeldt.
- Forelesning på grunnfag språkteknologi, Voss, tekstkodning. Foreleser: Espen S. Ore.

## Gjæsteforskere og studenter

Februar – juli Jianxin Wang (Professor i engelsk fra Beijing University of Post and Telecom).

3.8.–31.12. Anna Bortoletto (Hovedfagstudent fra University of Padova).

15.5.–15.11. Peter Keicher (Dr.gradsstudent fra University of Karlsruhe and University of Paris VIII).

## Publikasjoner

Austermühl, Frank & Apollon, Daniel, (Eds.): «Humanities Education and the Challenge of the Information Society», Bergen-Mainz, December 2000 (in press).

Apollon, Daniel: «The Parameters of E-learning», i: Austermühl, Frank & Apollon, Daniel, (Eds.), «Humanities Education and the Challenge of the Information Society», Bergen-Mainz, December 2000 (in press).

Apollon, Daniel & Fure, Odd-Bjørn & Svåsand, Lars (Eds.): «Proceedings of the 7th Conference of the International Society for the Study of European Ideas (ISSEI), 14.–18. August 2000 – University of Bergen – Norway», ISBN 82-91320-20-9 © 2000.

Apollon, Daniel: «Kulturelle og samfunnsmessige forutsetninger for «interaktivitet»», i: Haugsbakk, Geir & Fritze, Yvonne, «Workshop: Interaktivitet, teknologi og læring», Forsknings- og kompetansenettverket for IT i utdanning, ITU, Pedagogisk Forskningsinstitutt, Universitetet i Oslo, Oslo, 2000, ISBN 82-7947-010-7.

Apollon, Daniel: «Utdanning i det tredje årtusen», Kronikk, Bergens Tidende 17. august 2000.

Bjørneset, Tove: «The Dictionary Project NORD-LEXIN-N», i: *Lexicographica Series Maior*. Max Niemeyer Verlag Tübingen (under publisering).

Bjørneset, Tove: «Dictionaries for Immigrants and Refugees in Norway: NORDLEXIN-N», i: ISSEI 2000 CDROM, HIT-senteret, Universitetet i Bergen 2000.

Hofland, Knut: 2000. «A Self-Expanding Corpus Based on Newspapers on the Web.» I: Proceedings of the Second International Language Resources and Evaluation Conference. Paris: European Language Resources Association.

Johansson, Stig and Knut Hofland: 2000. «The English-Norwegian Parallel Corpus: Current work and new directions.» I: S. P. Botley, A. McEnery, and A. Wilson (eds.), *Multilingual Corpora in teaching and research*. Amsterdam. Rodopi.

Ore, Espen S.: Henrik Ibsen: *Catilina* av Brynjolf Bjarme, utgitt etter manuskriptet NBO Ms4<sup>o</sup> 936. Redaktør: Jon Gunnar Jørgensen. Oslo: Gyldendal. Medlem i redaksjonen, ansvarlig for IT-delen av prosjektet. Laget en medfølgende CDROM med tre tekstversjoner + faksimiler og skrevet en introduksjon om IT-arbeidet.

Ore, Espen S., Fiona J. Tweedie, and Craig Dougan, «Computers, statistics, and the grouping of rune forms» i DRH 98. *Selected Papers from DRH 98*, Glasgow 1998, eds. Marilyn Deegan, Jean Anderson, and Harold Short, pp. 117–128, Office for Humanities Communication, King's College, London, 2000.

Rosén, Victoria og Smedt, Koenraad de: «\*Er korrekturlesningsevnen di god?» Resultater fra SCARRIE. Nordlyd 2000; 28:214–228. MONS 8. Tromsø, 18.–20. november 1999.

Smedt, Koenraad de og Rosén, Victoria: Automatic proofreading for Norwegian: The challenges of lexical and grammatical variation. I: Nordgård, Torbjørn, red. NODALIDA '99. Trondheim: NTNU; s. 206–215. NODALIDA '99. Trondheim, 9.–10. desember 1999.

Smedt, Koenraad de: Human Sciences and Information Technology. I: 2nd Millennium Summit on Information Technology. San Juan: Gobierno de Puerto Rico; 2nda Cumbre del Milenio sobre Informática. San Juan, Puerto Rico, 9.–11. august.

## Verv

Apollon, Daniel: Ekspert for CRE, Conférence des Recteurs Européens/ Association of European Universities, Genève: utarbeidelse av intervjustrategi og evaluering av IKT strategier ved europeiske universitetet.

Apollon, Daniel: Medlem av ACONTE, Permanent Advisory Committee on New Technologies in Higher Education, Coimbra Group of Universities, Bruxelles.

Apollon, Daniel: Medlem av European Culture Committee, Coimbra Group of Universities, Bruxelles.

Apollon, Daniel: Medlem av referansegruppen for Scenario 2010, Scenarier for IKT i høyere utdanning, Fondet for Digitale Læremidler, KUF/USIT-Oslo og Preview AS.

Apollon, Daniel: Medlem av Scientific Committee og Arrangementskomité, ISSEI 2000 konferanse, UiB, 14.–18. august 2000.

Apollon, Daniel: Invitert evalueringsekspert, Minervaprogrammet, DG EAC, EU, Bruxelles.

Apollon, Daniel: Ekspertutforming av eLearning programmet, DG EAC, EU, Bruxelles.

Apollon, Daniel: Invitert ekspert ved det europeiske symposiet arrangert av Den autonome regionen Andalusia, Universitetet i Granada og Consortio Fernando de los Rios i anledning Karl V jubileet. Tema: «Det virtuelle universitet».

Hofland, Knut: Medlem ICAME advisory board.

Huitfeldt, Claus: Medlem av gruppe som utredet fremtidig profilering av samlingene ved Nordisk institutt for Det historisk-filosofiske fakultet.

Huitfeldt, Claus: Medlem av gruppe som utredet organisering av den eksterntfinansierte forskningen ved Det historisk-filosofiske og Det samfunnsvitenskapelige fakultet på oppdrag fra begge fakulteter og Unifob.

Huitfeldt, Claus: Medlem av gruppe som utredet administrativ struktur i forbindelse med omorganiseringen av Unifob.

Huitfeldt, Claus: Styremedlem og regnskapsfører for Text Encoding Initiative Consortium.

Myking, Johan: Leiar i Nordterm arbeidsgruppe 1, Terminologisk forskning og utdanning.

Myking, Johan: Styremedlem i Rådet for teknisk terminologi.

Ore, Espen S.: Medlem i ALLC (Association for Literary and Linguistic Computing) Committee (styre).

Ore, Espen S.: Fra mars 2000 styremedlem i XML/SGML Norge.

Smedt, Koenraad de: UiBs kontaktperson hos ELSNET og medlem i programkomiteen for ELSNET sommerskole.

Smedt, Koenraad de: leder for HF-fakultetets IT-utvalg.

Smedt, Koenraad de: leder for Seksjon for lingvistiske fag.

Smedt, Koenraad de: ekspert for Den europeiske kommisjonen.

Smedt, Koenraad de: leder for Nasjonalt fagråd for lingvistiske fag.