

Βυλλετιν

1. årgang, nummer 1, 28. desember 1993

[illegible]

1. Kort om Tekstlaben
2. Oppsummering fra møter med instituttene
3. Nye tekster for flere språk
4. Nye programmer
5. Kurs-annonsering
6. Adresse



Tekstlaboratoriet ved HF har nå vært i full gang siden 1. september 1993. Det er to ansatte: Janne Bondi Johannessen (lingvist og forsker) og Helge Hauglin (informatiker og Tekstlab-ingeniør). Dette håper vi vil være situasjonen også i 1994, men det kommer an på bevilgninger fra fakultetet.

Videre har Tekstlaboratoriet sett det som sin oppgave å utvide tilbudet av elektroniske tekster som kan brukes av språkforskerne ved fakultetet. Vi synes det er viktig at særlig norske tekster er godt representert, og har prøvd og prøver kontinuerlig å få tak i slike fra forskjellige kilder. Men vi ser også at det er viktig å ha et utvalg tekster på andre språk, og til ulike formål. Vi ønsker derfor å utvide mengden også på dette området, men vil her gjøre det mer etter ønske fra brukerne selv.

Andre prosjekter som vi ser som viktige på litt lengre sikt, er bl.a. å kunne ha et tilbud innenfor statistikk, både ekspertise om emnet og et programvaretilbud. Videre ser vi det som svært ønskelig at blir mulig å få til morfologisk merking av norske tekster.

Dette er svært krevende å gjøre manuelt, så Tekstlaboratoriet har flere planer om automatisk merking. Vi ønsker å få istand et prosjekt med vit.ass. hvor vi kan få laget et automatisk merke-program. Dessverre ble hele Informasjonsteknologi i humaniora-prosjektet til NRF strøket fra bevilgningssiden i år p.g.a. de ekstraordinære nedskjæringene, men vi prøver igjen neste år. Det er mulig at vi kan finne andre løsninger i mellomtiden.

Tekstlaboratoriets tilbud er ment å være tilgjengelig for brukerne, de vitenskapelige ansatte og høyeregradsstudenter, over nettet. For at man skal kunne benytte seg av det, må man ha en konto på Hedda. For at man skal kunne bruke de ulike elektroniske tekstene, må man i tillegg fylle ut et skjema hvor man lover å bruke dem til ikke-kommersielle formål. For dette må man henvende seg til Tekstlaboratoriet.

2. Oppsummering fra møter med instituttene

Vi har i høstesemesteret hatt kontaktmøter med de ansatte ved de seks språkinstituttene (GI, IBA, ILF, INL, IØO og KRI). Oppmøtet har stort sett vært over all forventning, og innspillene og ønskene fra det vitenskapelige personalet likeså. Mange har hatt konkrete ønsker om spesielle programmer og tekster, og dette har gitt oss en pekepinn om hva slags tilbud man trenger. For noens vedkommende må vi nok gå videre med individuelle samtaler for å klarlegge behovene mer i detalj. Vi regner med at disse møtene har gjort at fremtidige brukere tar kontakt når de finner det nødvendig.

3. Nye tekster for flere språk

For øyeblikket har vi disse tekstene elektronisk tilgjengelig. Man kan finne flere detaljer ved å gå inn i MacTekstlab på MacHedda (Mac-brukere) eller bruke dir-kommando på g-området (PC-brukere, når man har fått initialisert PCen sin i henhold til bruksanvisningen i "Tekstlaboratoriet ved HF, av Helge Hauglin, desember 1993).

Engelsk: Blant annet: Susanne Corpus (morfologisk og syntaktisk merkede utdrag av Brown-korpuset), Pennsylvania Treebank (gir en syntaktisk tre-analyse av diverse tekster), Brown-korpuset (amerikanske tekster i utdrag), Lancaster-Oslo-Bergen-korpuset (britisk-engelske tekster i utdrag, merket), Kolhapur-korpuset (indisk-engelske tekster i utdrag), London-Lund-korpuset (muntlig engelsk), EØS-avtalen.

Brown, LOB og Kolhapur er bygget over samme lest, med utvalg fra de samme genrene, for direkte sammenligning.

6. Adresse

Tekstlaboratoriet, HF

P.b. 1102 Blindern

0317 Oslo

Tlf.: 22 85 68 14 / 22 85 78 35

Fax: 22 85 43 48

E-post: jannebj@hedda.uio.no

helgeha@hedda.uio.no
