

2. årgang, nummer 1, april 1994

[illegible]

1. Brukerkursene i januar
2. Hvilke tekster fins på Tekstlaben?
3. Programmerer
4. Kurs
5. Tekstlab-gjester i 1994
6. Kort om Tekstlaben

[illegible]

1.1 Oppsummering

PC-brukernes kurs gikk over to dager: én demonstrasjonsdag og én øvelsesdag. Mac-brukernes kurs gikk over tre: én demonstrasjonsdag, én Mac-øvelsesdag og én PC-øvelsesdag.

Etter at kurset var over, sendte vi ut evalueringsskjemaer. Her fikk vi tilbake 18, altså fra cirka halvparten av deltagerne. Det fordelte seg på 5 PC-brukere og 13 Mac-brukere. Hvordan vi skal tolke de uteblitte svarene, vet vi ikke, men her kommer resultatene fra dem som svarte.

De fleste PC-brukerne syntes kursets varighet var passe, mens ca. halvparten av Mac-brukerne syntes det var for langt. De aller fleste syntes at innholdet og kursets omfang var som ventet, og at nivået var passe høyt. De fleste PC-brukerne syntes fremdriften var passe rask, mens ca. halvparten av Mac-brukerne syntes den var for treg (noe som sannsynligvis skyldtes de tekniske problemer som oppsto). Halvparten av Mac-brukerne hadde prøvd ut kurslærdommen i de første ukene etter kurset, mens ingen av PC-brukerne hadde det. De aller fleste Mac-brukerne ville gjerne bruke lærdommen i videre forskning og undervisning, men det blant PC-brukerne var tvil blant halvparten. Nesten alle kursdeltagerne ville også gjerne følge flere kurs i fremtiden. Når det gjelder ønsker for innholdet i fremtidige kurs, var det svært varierte svar. Disse er kommentert nedenfor.

Brukerkursene var ment som et praktisk tilbud til nybegynnere som ønsket å få vite noe om, og lære å bruke, elektroniske tekster og tekstsøkeprogrammer. Det generelle inntrykket fra skjemaene, men også fra kursene selv, er at vi traff den målgruppen vi ønsket.

Noen ville ha mer kunnskap om morfologisk merking ("tagging"). Dette er noe vi helt sikkert kommer tilbake til når vi har fått de automatiske merkerne våre fullt installert. Videre har vi også tenkt å komme tilbake til Childes-korpuset. Programmer for samtaleanalyse er vi gjerne behjelpelige med, om vi får litt mer opplysninger om hva som fins. Når det gjelder tilgang til tekster, er det et problem at endel flerbrukerlisenser er svært dyre. Det kan være bedre å kjøpe inn tekstene til individuell bruk hvis det ikke er for mange interesserte. Se ellers punktene 4 og 5 for hva Tekstlaben kommer til å tilby i nærmeste fremtid.

-2-

2. Hvilke tekster finnes på Tekstlaben?

Vi får stadig flere tekster 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). La oss minne om at de fleste tekstene ikke er tilgjengelige med mindre man har skrevet under på et skjema man får hos Tekstlaben, hvor det må framgå at tekstene bare skal brukes til ikke-kommersielle forskningsformål.

2.1 Nye tekster

Barnespråk: Det store Childe-korpuset med data fra barnespråk fra mange ulike språkområder er tilgjengelig hvis man først foretar seg det som står i Les-meg-fila som ligger sammen med korpuset på Tekstlab-disken.

Finsk: Kalevala.

Norsk: I tillegg til noe tidligere materiale har vi nå også fått Aftenposten, morgen- og aftenutgave, uke 2, 1994.

Nygresk: CTI-korpus. Dette er et forholdsvis stort korpus med tekster fra avisen To Vima og noe annen sakprosa. Tekstene fremkommer på det greske alfabetet hvis man bruker Macintosh og har fonten SuperGreek (kan fås på Tekstlaben) eller bruker Unix.

Russisk: Et stort tekstkorpus fra Uppsala. Her har det vært et lite problem med fonter, som nå holder på å rettes opp.

Tyrkisk: Tyrkiske ordtak.

Walisisk: Ordbok fra Yr Odliadur, Roy Stephens, Gwasg Gomer (J. D. Lewis a'i Feibion Cyf., Llandysul, Dyfed), 1978 (ISBN 0 85088 710 0).

2.2 Gamle tekster

I forrige nyhetsbulletin fortalte vi litt om tekstene vi hadde tilgjengelige da; for engelsk, fransk, italiensk, latin, norsk, serbo-kroatisk og tysk. På noen av områdene for disse språkene finnes det en Les-meg-fil som man kan benytte for å få en oversikt over hva som finnes der. I andre tilfeller er det selvforklarende ut fra filnavnet.

2.3 Andre tekster

Tekstlaben kan være behjelpelig med å finne fram andre elektroniske tekster enn de vi selv har liggende. Det er etter hvert mange steder man kan finne tekster. Her vil vi

fremheve Oxford Text Archive, som har svært mange tekster på engelsk, og en del på andre språk, som arabisk, hebraisk og islandsk, japansk, kurdisk, portugisisk og sanskrit. En oversikt over hva de har, kan man finne på Tekstlab-disken, på filene text-project-list-1 og -2 i Info-mappen. For å få tilgang til selve tekstene krever Oxford at man fyller ut et skjema og returnerer til dem. Dette kan vi være behjelpelige med.

Andre steder hvor de har mye (eldre) engelsk litteratur, er Gutenberg-prosjektet. Også her kan vi hjelpe.

Er man interessert i svensk, fins det flere måter man kan koble seg til Språkdata i Gøteborg på. Da kommer man inn i deres selvforklarende bruker-program, hvor man kan få søke på ord-konkordanser i forskjellige svenske kilder, bl.a. aviser.

3 Programmer

3.1 Nye programmer

• Finsk tagger for engelsk

Vi har under installering en morfologisk merker (fra Universitetet i Helsinki) for engelsk. Den skal kunne brukes til å merke engelske tekster automatisk, slik at man siden kan søke etter merker snarere enn ord og deler av ord. Vi har også en merker (fra Xerox) som skal kunne merke tekster automatisk bare v.h.a. en ordliste. Denne håper vi å kunne bruke for norsk etterhvert.

• Et Hypertekst-program for å lære om lingvistikk (på engelsk) og Karl den store (på tysk)

Tekstlaboratoriet har kjøpt inn en CD fra Zürich som inneholder to ting: et program som heter *Hyperlinguistics: An educational hypertext for English linguistics*, hvor man kan lære om lingvistikk, og et hvor man kan lære om Karl den store: *Karl der Grosse: Hypertext im Unterricht zur Geschichte des Mittelalters*. Begge programmene er svært brukervennlige, og er basert på at man ser flere vinduer samtidig, som man klikker i for å komme fram og tilbake og rundt omkring. Programmene kan benyttes på Tekstlaben, hvor det står en Macintosh med innebygd CD-spiller.

Lingvistikk-programmet gir muligheten til å lese om forskjellige emner som man velger fra en emneliste, og også til å løse oppgaver innenfor de forskjellige emnene.

Karl den store-programmet gir muligheten til å lære mye om både Karl den store selv og om samfunnet på den tiden, dvs. middelalderen generelt. Også dette programmet inneholder forskjellige emner som man kan klikke seg gjennom, og oppgaver man kan løse. Her får man reaksjon med en gang hvis man har svart galt. Siden programmet er på tysk, passer det godt for laveregradsstudenter i tysk, og for disses lærere.

Dette er et tekstsøkeprogram med mange muligheter. Vi har ikke rukket å ta det i bruk ennå, men håper å kunne gi kurs i det etter hvert, hvis det viser seg å svare til forventningene.

Tekstlaboratoriet har en rekke programmer som kan hjelpe en i tekstsøking og -behandling. For PC har vi Tact, Word Cruncher og Lexa, og foreløpig kompetanse i de to førstnevnte. For Mac har vi Conc, Free Text Browser og MacConcordance, men vi anbefaler bare førstnevnte. Tekstlaben har både kortere bruksanvisninger til utdeling og lengre manualer man kan få lese i, for disse programmene.



Den store interessen for kursene i januar gjorde at vi stort sett lot dem være forbeholdt ansatte. Imidlertid vet vi at en god del hovedfagsstudenter vil ha nytte av dem, og at en del ansatte var forhindret fra å delta i januar. Vi vil derfor gjerne opplyse om at det vil bli holdt nye kurs i juni, med samme innhold som i januar, hvor både ansatte og studenter kan delta. Vi ber ansatte som vet om hovedfagsstudenter som kunne ha

nytte av tilbudet til å gjøre det kjent overfor dem. Nærmere opplysninger om påmelding, tid og sted kommer senere.

+++++

5. Tekstlab-gjester i 1994

I tillegg til vanlige tekstlabaktiviteter håper vi å arrangere flere gjesteforelesninger i Tekstlabens regi. Vi har invitert Willard McCarthy fra University of Toronto til å komme i november. Han blir antagelig en uke, og skal snakke om bruk av maskinleselige tekster i humanistisk forskning. Han har vært spesielt opptatt av å studere litterære tekster v.h.a. TACT, og har bl.a. skrevet artikkelen "Finding implicit patterns in Ovid's 'Metamorphosis' ". TACT-programmet er jo utviklet i Toronto og McCarthy er en ekspert man kan benytte seg av.

Videre kommer Douglas Byber (University of Northern Arizona) til Oslo til høsten. Vi håper han kan gi noen forelesninger om statistikk i tekster. Vi antar at ikke så rent få vil synes språkstatistikk er viktig å få lært mer om.

Vi vurderer også andre muligheter, f.eks. Lou Burnard ved Oxford Text Archive og Michael Sperber McQueen fra Chicago. De vil kunne snakke om TEI (Text Encoding Initiative), som er et merkesystem for tekster, slik at en rekke egenskaper ved tekstene raskt kan identifiseres. Dette besøket er nok mest for spesielt interesserte, særlig for dem som selv bygger opp korpora.

Til slutt vil vi oppfordre fakultetets ansatte til selv å komme med ønsker. Siden Tekstlaben ennå er i startfasen, er vi interessert i å høre mer om hva de ulike forskerne gjerne vil høre eller lære mer om.

6. Kort om Tekstlaben

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). Vi er glade over å kunne meddele at fakultetet har sett det som viktig at ingeniørstillingen beholdes. Den er derfor utlyst som fast stilling. Forskerstillingen består til et stykke ut i 1996, med unntak av perioden april-desember, da forskeren har barselspermisjon.

Tekstlaboratoriets tilbud er ment å være tilgjengelig for brukerne, de vitenskapelig 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.

Tekstlaboratoriet, HF

P.b. 1102 Blindern

0317 Oslo

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

Fax: 22 85 69 19

E-post: jannebj@hedda.uio.no

helgeha@hedda.uio.no