

Per Vestbøstad  
NORSK TEKSTARKIV  
UNIVERSITETET I BERGEN

## NORSK TEKSTARKIV ETTER HALVTANNA ARS DRIFT

### 1. Litt historikk:

Tanken om eit arkiv for maskinlagra norsk tekst vann tilslutning på konferansen om "eit norsk datamaskinelt tekstkorpus" i Bergen, oktober 1978.<sup>1)</sup> Etter konferansen vart det skipa ei plangruppe med representantar frå fagmiljøa i Bergen, Oslo og Trondheim og frå Norsk Språkråd. Arbeidet i denne gruppa førte fram til at Rådet for humanistisk forskning (i NAVF) gav løyving til ein prosjektmedarbeidar og driftsmidlar for 1980 med lovnad om løyvingar for åra 1981-1984. Prosjektet var då formalisert som eit samarbeid mellom PDS ved Nordisk inst. i Bergen og NAVFs edb-senter same stad. Planleggingsgruppa vart omgjort til fagleg referansegruppe, under namnet Fagleg råd. I planleggingsfasen arbeidde også ei gruppe med å avklara eit standardformat for tekstlagring.

### 2. Oppstartinga:

1. mars 1980 tok underskrivne til som prosjektarbeidar ved Norsk tekstarkiv, som fekk kontorplass ved NAVFs EDB-senter for humanistisk forskning ved Universitetet i Bergen. Når det gjeld datamaskinkraft, har Tekstarkivet fungert som eit prosjekt under PDS (Nordisk inst., UiB).

Den mest nærliggjande arbeidsoppgåva var å skaffa seg oversyn over kva for maskinlagre tekstar, som alt fanst i forskningsmiljøa landet over. Etter ei grundig rundspørjing (med god hjelp frå EDB-tenestene ved HF-fakulteta i Oslo og Trondheim) kunne oversynet spreidast i septembar 1980. Oversynet er sjølv sagt halde ajour og skal sendast ut på nytt denne hausten.

### 3. Inntak av avistekst

Som ei vidareføring av eit initiativ teke av Norsk leksikografisk inst. (Oslo) og PDS for å samla inn nynorsk avistekst, kontakta eg sommaren 1980 alle nynorskavisene i landet og laga eit oversyn over den sats tekniske utrustninga deira. Ein del tekst vart samla inn, men avvikande diskettstandard i den grafiske industrien har til no hindra tekstinntak frå dei fleste mindre nynorskavisene, og frå mindre trykkeri elles. Ingen redaktørar hadde noko imot at Tekstarkivet fekk kopiera maskinlagra avistekst.

### 4. Inntak av bok-tekstar:

Den Norske Forleggerforening har rådd medlemmene sine til å gje Tekstarkivet høve til å kopiera maskinlagra tekstar frå bøker som dei har i produksjon, vederlagsfritt så lenge det gjeld ikkje-kommersiell bruk. Arbeidet med å få oversyn over kva bøker som finst maskinlagra, og kvar dei finst, har likevel vore komplisert.

Førebels har det vore relativt få bøker som har lagt godt til rette for å overføra til Univac-anlegget i Bergen. Vi har berre kunna teke inn frå større trykkeri, som kan levera magnetbandkopiar, men vi arbeider med å finne overføringsmåtar for dei vanlegaste fotosetjarane, som berre har diskettlager.

Velviljen hos trykkeria har vore stor. Når leveringstida likevel kan verta 3 månader, heng det truleg saman med at slike oppdrag fell utanom dei normale arbeidsrutinene. Prisen på slike kopiar ser ut til å kunna variera frå kr 300,- til kr 1.000,- pr. bok!

Teksten som kjem frå sats-systemet, er full av typografiske kodar. Ein del av desse skal følgja teksten vidare som spesialtegn eller format, medan andre berre kan lesast forbi. Ved hjelp av programsystemet CROMWELL (utvikla ved PDS) styrer vi omsetjing og utval av trykkerikodar. Det er sjølv sagt tidkrevjande å spesifisera alle kodealternativ i eit sats-system, men har ein først gjort det, er det lett å ta imot fleire tekstar frå same systemet.

I samarbeid med Bjørn Eide, PDS, har eg gjort ferdig spesifikasjonane til Comtecs CRT-13 og Bergens Tidendes sats-system, og arbeidet med Nortext-systemet er godt i gang.

Sjølv om vi etter kvart vil kunne standardisera tekstar frå enno fleire sats-system, vil Tekstarkivets eine konsulent ha av-grensa kapasitet og fleire oppgåver. Vi reknar difor med å lagra "rått" mykje av det vi får tak i, og først når det vert konkret spørsmål etter ein slik tekst, vil vi "foredla" han til standardformat.

### 5. Tekstproduksjon i eigen regi

For å auka tilgangen på skjønnlitterære tekstar, har Tekstarkivet i eit samarbeid med Norsk språkråd gått i gang med å registrera 10 bokmålsromanar frå 1977, som Språkrådet skal nytta til å studera påverknaden frå offisiell norsk skriftnorm. Det er meininga å registrera like mange nynorskromanar frå same året, og dessutan ønskjer ein å undersøkje romanar utkomne i 1937 og 1917 på liknande måte.

### 6. Tekstformidling

Inntak, standardisering og produksjon av maskinlagra tekst har såleis stått sentralt i Tekstarkivets arbeid denne perioden. Etter kvart som tekstmengda veks, og Tekstarkivets tenester vert vidare kjente, reknar eg med at formidlingsarbeidet vil krevja meir tid. Til no har det særleg vore spørsmål etter det som finst av ferdige konkordansar. Dette bør få oss til å vurdere om ikkje Tekstarkivet også bør foredla nokre sentrale tekstutval til konkordans- og listeprodukt, slik ein t.d. ved PDS har gjort med bokmålsaviser frå 1968-73.

### 7. Oppsummering

Norsk tekstarkiv har ikkje til no makta å skaffa fram større mengder ny maskinlagra tekst, men ein har gjort viktig grunnlagsarbeid, som truleg har gjort det monaleg lettare å finna ut kva som finst, og å få tak i det som finst av slik tekst. Det er også oppretta kontaktar med forlag og trykkeri, som vil gje oss innsyn i den årlege produksjonen på eit tidspunkt då dei fleste bøkene enno ligg maskinlagra i trykkeriet. Ein har også gjort seg godt kjent med den sats-tekniske utrustninga hjå små og store trykkeri. Komande arbeidsår vil dette nokså sikkert gje utbytte i form av ein større auke i tilgangen på moderne norsk tekst for data-maskinell bruk i forsknings- og utviklingsarbeid.