

Åtta teser om texthantering

Docent Sture Allén föreslår inrättandet av ett organ för textarkivering, förslagsvis kallat "logotek". Det skulle bli tillvarata de stora mängder maskinläsbar text i form av hålremsor som produceras vid landets sätterier men för närvarande bara kasseras efter användningen.

Docent Allén har en särskild forskartjänst i språklig databehandling vid Statens humanistiska forskningsråd och leder Forskningsgruppen för modern svenska, som med stöd av Riksbankens jubileumsfond genomför projektet "Data-maskinell undersökning av tidningsprosa".

① Den vetenskapliga och tekniska utvecklingen har lett till att den typ av tjänster som biblioteken kan erbjuda inte förslår längre. Det trycks flera miljarder ord om året i Sverige. Detta är ett av tecknen på informations-explosionen i vår tid och accentuerar kraftigt språkets roll som vårt viktigaste informationsmedel. Sedan lång tid samlas böcker, broschyrer, tidningar osv i biblioteken, som alltså utgör stora informationscentra. Dessa kommer också i fortsättningen att spela en viktig roll, men de kan naturligt nog inte motsvara alla krav som ställs i dataåldern. Det tryckta ordets användbarhet är alltför begränsad.

② Det är från flera synpunkter viktigare att ha exempelvis Gyllensten eller brottsbalken eller DN eller ett lexikon i maskinläsbar form än i klarskrift. Så fort det gäller att hantera texter på olika sätt — tex i sociologiskt, pedagogiskt, språkvetenskapligt, litteraturvetenskapligt syfte — är datamaskinen en överlägsen medhjälpare. Motvilligheten att tillägna sig denna insikt har länge varit stor. Betänkligheterna börjar emellertid rubbas i takt med att forskare och utredare i allt större utsträckning upptäcker att de kan göra uttestningen av beskrivningsmodeller, bear-

betningen av material och utvärderingen av resultat till arbetets huvudpunkter, inte själva materialsamlandet.

③ Det framställs sedan åtskilliga är mängder av viktig text i maskinläsbar form. Vid landets sätterier produceras hålremsor som styr sättmaskiner och numera också magnetband som styr apparater för fotosättning. Praktiskt taget hela detta högvärdiga material kasseras efter användningen. Om inget görs, kommer det att bli allt vanligare att en och samma text i ett visst skede kodas på ett sättteri, i ett annat skede vid en forskningsinstitution, hos en myndighet eller i ett företag.

④ De maskinläsbara texterna går förlorade därför att de faller utanför befintliga organs verksamhetsområden. För statliga myndigheter gäller arkivstadgan. De kommunala myndigheterna har sina föreskrifter i kommunallagen. Men för enskilda arkivbildare finns det ingen förordning. Det är nu inte min mening att plädara för en sådan. Det finns smidigare sätt att täcka luckan. Vad som står helt klart är att sättterierna intar en särställning bland enskilda arkivbildare genom den centrala roll de spelar i den stora informationsprocessen. Tryckeriernas

pliktexemplar av tryckta skrifter till biblioteken är en tydlig indikation på detta.

⑤ Det är nödvändigt att skapa ett organ som tillvaratar och tillhandahåller maskinläsbar text. Den frågan är inte mindre viktig än ett par andra frågor som för närvarande är aktuella. Jag syftar på dataarkiveringskommitténs arbete rörande statliga myndigheters maskinläsbara dokument och på statskontorets utredning av den framtida lagringsformen för tidningar. Det organ för textarkivering som krävs måste arbeta i nära samverkan med framför allt sättterierna. Jag kan nämna att vi genom tillmötesgående från sättterierna under de senaste åren har kunnat rädda maskinläsbara texter om tillsammans några miljoner ord. Insatsen är av mycket begränsad omfattning, men visar att arbetet är genomförbart. Och efterfrågan på det tillvaratagna materialet är stor. — Vill man ha en beteckning för det organ som skall sköta själva textarkiveringen, kan man snegla åt ordet bibliotek, som betyder "bokmagasin", och lansera ordet logotek, egentligen "ordmagasin". Det ordet ligger väl på tungan och tar sikte på en väsentlig egenskap. — Logotekets data bör vara tillgängliga i både ögonläsbar och maskinläsbar form:

det är viktigt att användaren i detalj kan kontrollera vad som faktiskt ligger på det informationsbärande mediet. Den lösning som för närvarande ser elegantast ut är att välja magnetband som maskinläsbar databärare och mikrofilm som ögonläsbar databärare. Man kan nämligen låta magnetbanden direkt generera ögonläsbar mikrofilm. När det omvända förfarandet blir tillgängligt — man kan förmoda att det inte dröjer alltför länge — kan mikrofilmen bli huvudmedium. Den uppfyller då kravet att vara både maskin- och ögonläsbar.

⑥ För att kunna fungera måste organet vara förbundet med en forskningsenhet för språklig databehandling. Inte bara den tekniska utvecklingen utan också den språkvetenskapliga och den informationsvetenskapliga sker snabbt. Ständig metodutveckling är därför av grundläggande betydelse. Textavdelningen bör rimligen rymma de arkiverade texterna plus en uppsättning standardprogram för inläsning, lagring, utskrift m.m. Forskningsavdelningens inriktning bör gälla metod- och programutveckling på språkvetenskaplig och informationsvetenskaplig grund och materialbearbetningar i samband med utvecklingsarbetet.

⑦ En institution av den skisserade typen är ekonomiskt fördelaktig. Dubbelarbete undviks som jag nämnde genom att upprepningar av den kostsamma stansningen (och den tidsödande kontrolläsningen och korrigeringen) av skrifter och dokument inte blir aktuell. Men man behöver inte ha stor fantasi för att inse att potentialen är större än så. Erfarenheterna från forskningsverksamheten kommer användarna till godo. Man kan också räkna med effektivitetsvinster. Dit hör att många undersökningar kan få en flygande start. Dit hör också att nya typer av undersökningar möjliggörs och att tvärvetenskapliga projekt underlättas.

⑧ Det är dags att slå till: förlusterna av maskinläsbar text blir bara större och större. Vi har tillräcklig erfarenhet av språklig databehandling för att kunna gå till verket. Vi har de tekniska förutsättningarna. Det borde vara självklart att vi inte skall försitta möjligheten till rationalisering och effektivisering av texthantering. Uppgiften blir inte mindre betydelsefull genom att vi internationellt sett har möjlighet att anvisa vägen.

STURE ALLÉN