

Edb og talemålsforskning

Erfaringar frå prosjektet «Talemål hos ungdom i Bergen»

Helge Sandøy

1. Prosjektet TUB

Arbeidet i prosjektet «Talemål hos ungdom i Bergen» (TUB) kom i gang i 1977 og hadde som mål å undersøke utviklinga i bergensmålet og variasjonen i talemålet hos ungdommen i Bergen kommune (d.v.s. «storkommunen» etter 1972). NAVF finansierte prosjektet til ut 1983.

Etter at innsamlingsfasen var over, hadde prosjektet eit svært stort lydbandmateriale. Det var opptak av ymse typar samtalar med 104 informantar, i alt på 68 timar og 15 minutt. (Ei kort orientering om prosjektet står i *Talemål i Bergen 1/83*.) Om lag 75% av materialet blei transkribert fullt ut for å gjøre syntaktisk og tekstlig analyse lettare, for seinare å kunne gi ut dialekttekstar og for å halde materialet meir ope for nye problemstillingar.

Sjølv om ein avgrensa systematisk det materialet som skulle transkriberas, blei det svært store tekstmengder til saman. Det omfatta i alt 220.000 løpande ord. Spørsmålet blei derfor tatt opp om korleis ein kunne få ei mest mulig rasjonell behandling av materialet. Etter å ha hatt samrådsmøte med datafolk vedtok prosjektet i 1980 å dataføre transkripsjonen for å kunne utnytte betre datateknikken i analysen av materialet. Ein vurderte vinningane i analysearbeidet til å vere så store at ein kunne ta kostnadene med dataføringa. Vinningane ville ligge i raskare analyse og større nøyaktigheit i opteljinga.

2. Krav til materiale og til bruk

Transkripsjonen var fonemisk, unntatt på eit par punkt der den fonetiske variasjonen var av spesiell interesse. Ein slik transkripsjon krev ein del ekstrateikn utover skriftspråksortografien. Derfor måtte eit maskintilgjengelig teiknsett settas opp. Det blei konstruert i korrespondanse med internasjonal lydskrift (IPA) (med noen tilleggsteikn for pause, utydelig tale, svake element, og namn/ord sitert med fremmend/ utalandsk uttale).

Samtalane i lydbandopptaka var dels mellom to informantar + opptaksleiar, dels mellom fire informantar + opptaksleiar. I samtaleutskriftene var det ønskelig ikkje bare å få fram replikkvekslingane, men også å ta vare på kva som var samtidig tale, d.v.s. kva tid samtaledeltakarane snakka i munnen på kvarandre. Vi valde derfor ei oppstilling med

parallele og synkroniserte linjer, der kvar deltakar hadde si linje i linjebundelen:

2931 K1:
2932 L :
2933 M1: < de: ska vå de: ska +vå:r skol6 > / < di: så +ve: > / < n6 > /
2934 K2: vi < v6 > /
2935 M2: < jo vi fik sån fik sån "tilbud > / +ja: / < aså e:C6 +nøt > aså < fik >

2936 K1:
2937 L :
2938 M1: ja di: så +ve: /
2939 K2:
2940 M2: "tilbud6 ja / +e:g ska nå +ne: å "hø:r6

2941 K1: når^ ska "de:r6 da /
2942 L :
2943 M1:
2944 K2:
2945: M2: ka d6 +e: få nåk6 / nei kå ti +va:

Eksempel frå gruppesamtale nr. 60.

(Tala til venstre er linjenummer. L = leiar, K1 osv. er kvinnelige og mannlige informantar. < og > markerer samtidig tale. / er teikn for kort pause, for at siste segmentet er utydlig. Lydskrifta elles er det ingen grunn til å forklare nærare her.)

Alle samtaletekstane er utstyrt med samtalennummer og informantnummer. Desse numra saman med linjenummer er lagt inn som referansar i KWIC-konkordansen som er laga på grunnlag av alle tekstane. I referansedelen av konkordanslinja er det dessutan òg lagt inn informantopplysningar som vi bruker i analysen. Kvart eksempel vi finn i KWIC-konkordansen, kan derfor gi oss mange opplysningar: 1) lyden/ordet/uttrykket i språklig kontekst, 2) om det har vore samtidig tale eller avbrott i den teksten linja gjengir, 3) referanse tilbake til teksten slik at ein kan finne ein større samtalekontekst, 4) opplysningar om sosialgruppe, kjønn, alder og bydel for den informanten som har brukt ordet/uttrykket, 5) opptakssituasjon belegget er henta ifrå. I konkordansen er belegga sortert alfabetisk. Men p.g.a. lydskrifta måtte vi sette spesielle krav til alfabetiseringsprinsippa.

kås2b306 65G015A1424 (-) +å: / (-) (-) har "dåkar "lanste:d nåk6 ste:d^a / (-) (-) (-) nei +de: ad e: nå
 kås2b306 65G015A0939 +lei av / (-) < at ic6^ d6 +e: > (-) nåk6^ sår^ særli d^ær^ ne:r6^ // (-) (-) <å>
 nås2b106 60G012C2935 (-) (-) (-) < de: ska vå de: ska +vå:r skol6 > / (-) (-) (-) < di: (-) så +ve: > /
 kås2b306 65G015A0424 kåm6r +op // (-) (-) vå +va: "dåkar skole^ op6^ i "muntli^ i fjør^ / (-) va "dåkar

Eksempel på to konkordanslinjer med nøkkelordet (key-word) *nåk6* (= 'noe') og to med *skol6* (= 'skole'). (Kolonnane i referansefeltet til venstre viser: 1: kjønn, 2-3: bydel, 4-5: aldersgruppe, 6: sosialgruppe, 7-8: språkholdning, 10-12: opptaksnummer, 13-15: informantnummer, 16: opptakssituasjon, 17-20: linjenummer i teksten som belegget er henta ifrå. I teksten står // for lengre pause, (-) tyder mellomliggende replikk frå annen samtaledeltakar.)

Bjørn Eide har gjort det nødvendige programmeringsarbeidet til samtaleoppstillingane og tilrettelegginga for konkordansutvikling.

3. Dataføringa

Dei store kostnadene med EDB-bruken låg i den omstendelige dataføringa. Det var tale om svært vanskelige tekstar p.g.a. lydskrifta, dei hadde ei innfløkt oppstilling, og analysen var avhengig av stor nøyaktighet.

Den handskrivne transkripsjonen blei maskinskriven og seinare overført til datamaskina ved optisk lesing (OCR). Denne framgangsmåten blei valt – i staden for pønsjing direkte på terminal – fordi det på denne tida var for få tilgjengelige terminalar, og dessutan for å kunne utnytte habil skrivehjelp utafør byen. På det tidspunktet var det heller ikkje aktuelt å bruke småmaskiner (tekstbehandlingsmaskiner) å skrive inn tekstmaterialet på. Kostnadene ved vår framgangsmåte var neppe større enn ved direkte pønsjing, men vi måtte plagas med store tekniske problem med den optiske lesaren, slik at det blei mye hefting i framdrifta.

Det gikk òg mye tid til korrektoren. Jamt over blei det tale om tre-fire korrekturar før alt var rett (i forhold til den handskrivne transkripsjonen).

Både transkripsjon (med kontrollar), maskinskriving, optisk lesing og korrektur i fleire omgangar foregikk samtidig (i nesten 1 1/2 år) og kravde mye organisering for at alt materialet skulle komme rett gjennom alle fasane. Til mye av arbeidet hadde vi timelønte assistentar, men prosjektmedarbeidarane måtte instruere, samordne og kontrollere arbeidet.

Sia kostnadssida er eit viktig punkt i vurderinga av EDB-bruken, kan det vere nyttig å ha ei viss konkretisering av arbeidstida som er gått med. Arbeidstimane for den timelønte hjelpa vi brukte til dataføringa, fordeler seg slik:

Maskinskriving:	440 t
Optisk lesing, programmering, korrekturinnføring	1200 t
Korrekturlesing:	700 t

I tillegg kjem den tida dei tilsette i prosjektet brukte til å organisere og kontrollere dette arbeidet. Seier vi at sjølve dataføringa tok nærare to årsverk, er det i alle fall ein viss peikepinn på omfanget.

Inn i reknestykket må ein sjølvst ta med at medarbeidarane var ukjente med EDB og terminalbruk i utgangspunktet, slik at denne tida har òg vore læretid.

4. Morfologisk merking

Ved automatisk analyse eller leiting i språktekstar volda alle homonyma problem. Ein transkripsjon i lydskrift gir i tillegg det problemet at eitt og same ord kan få svært mange former ettersom lydar lett fell bort i naturlig rask tale. Variantane eit ord kan forekomme i, er så mange at det i praksis er uråd å predikere dei. Lemmatisering (d.v.s. å gi alle orda ei «opplagsform», f.eks. i normalortografi) kan vere ein måte å løyse denne siste vansken på. Homonymiproblemet kjem ein langt på veg unna med å merke alle orda morfologisk.

For å kunne prøve ut og forbetre EDB-opplegget lét vi den første tjuandedelen av det dataførte materialet utgjøre eit prøveprosjekt. Der prøvde vi òg ut eit nøyaktig merkesystem med både lemma og opplysning om ordklasse og bøyingsform. Etter erfaringane frå dette prosjektet kom vi til at tidsforbruket til merkinga ikkje stod i forhold til nytta vi ville ha i analysen av dei språkvariablane vi hadde planlagt å arbeide med.

I den endelige merkinga sløyfa vi lemmatiseringa og gikk over til bare å merke for ordklasse ved verb, substantiv (her også merke for felleskjønn eller nøytrum), adjektiv og pronomen. Dessutan fikk ein del ord vi allereie hadde definert som interessante i prosjektet, ein talkode som viste kva variabel og variant dei var belegg på. Dei fleste homonymiproblema blei løyst med denne merkinga. Og til dei konkrete analysane vi hadde planlagt, viste dette seg å vere nok. Om ein ønskjer det, kan nøgnare merking føyas til seinare. (Det er alt gjort for verb og substantiv frå tre bydelar.)

```
kås2b306 65G015A1424 (-) +d: / (-) (-) har "dåkar "lanster:d nåk6 ster:d^a / (-) (-) (-) nei +de: ad e: nå p h1
kås2b306 65G015A0939 +lei av / (-) < at i06^ d6 +e: > (-) nåk6^ sår^ særli d^er^ ne:r6^ // (-) (-) <d> p h1

mås2b106 60G012C2935 (-) (-) (-) < de: ska vð de: ska +vð:r skol6 > / (-) (-) (-) < di: (-) så +ve: > / f
kås2b306 65G015A0424 kámér +op // (-) (-) vð +va: "dåkar skole" op6^ i "muntli" i fjør^ / (-) va "dåkar f
```

Eksempel på merkte konkordanslinjer
(Til høgre for teksten står kode for ordklasse (ev. genus) og ev. variabel- og variantkode).

Den forenkla merkinga blei gjort i KWIC-konkordansen og ført rett inn på skjermterminal. Vi brukte der eit program Harald Solevåg alt hadde laga for merking av språklig materiale (Solevåg 1977). I ein alfabetisert konkordans vil forekomstar av same ord (og form) stå samla etter kvarandre, og merkeprogrammet utnyttar dette til raskare innsetting av kodane. Bjørn Eide skreiv programma som la materialfilane til rette for merkeprogrammet.

Vi reknar med at merkearbeidet tok om lag 550 timar.

5. Analysearbeidet

Den morfologisk merkte KWIC-konkordansen har vore utgangspunktet for det meste av analysearbeidet. Ymse seleksjonsprogram har plukka ut materiale frå konkordansen på grunnlag av den transkriberte forma av ordet (nøkkelordet på linja) og den morfologiske koden yst til høgre på linja. Ved seleksjonen er heile konkordanslinjene som inneheld potensielle belegg på ein språklig variabel, ført over på eigen fil. I mange tilfelle har ein måtta bearbeide denne filen vidare, fordi ein har vilja legge inn tilleggsopplysningar, d.v.s. merke meir nøyaktig. Ved f.eks. variablane *då* og *no* viste det seg nødvendig å ta med opplysningar om syntaktisk funksjon (Myking 1983). Somtid har ein òg komme til at ein del av materialet i filen er uaktuelt å bruke som belegg på den meir nøyaktig definerte variabelen. Da har ein stroke dei linjene. Med meir nøyaktig merking i utgangspunktet, kunne ein sjølvsagt ha sloppe unna ein del av denne bearbeidinga etterpå. Men merkingsfasen hadde dermed tatt mye lengre tid, og trulig meir enn det ein ville vinne inn att seinare. Dessutan – og det er det viktigaste – er det svært vanlig at ein under analysearbeidet oppdagar at ein treng nye eller finare kategoriinndelingar av materialet. Da blir det aktuelt å finsikte materialet att. Den forenkla merkinga vår gav oss altså i mange tilfelle denne arbeidsmåten: først grovseleksjon or KWIC-konkordansen, deretter fininndeling ved bearbeiding i egne filar.

Når ein enten skal merke filen nøgnare, eller ein skal reinske ut materiale som egentlig ikkje er belegg på det ein vil undersøke, kan det vere nyttig å ha sortert materialet etter visse prinsipp: alfabetisk, baklengs alfabetisk (nyttig ved fonologiske eller bøyingsmorfologiske studiar), etter bydel, kjønn o.s.v. Visse delar eller problemtypar i materialet kan da samlas for seg, materialet blir meir oversiktlig og lettare å gi seg i kast med. Slike sorteringar eller redigeringar av filane kan maskina gjøre raskt.

I dei ferdige filane kan ein telje opp belegg for kvar kategori og for alle krysskombinasjonar av kriterium. I staden for å gjøre dette for hand fikk vi i somme tilfelle laga program som både sorterte og talde. Men i mange tilfelle fann vi det nyttigast og raskast å kjøre statistisk analyse direkte på materialfilane med konkordanslinjer. Kvar linje i filen inneheld alle dei nødvendige opplysningane om språklige og

utomspråklige variablar, slik at maskina kan rekne direkte frå grunnlagsmaterialet.

For noen få språklige variablar er dette ei rask og god løysing. Men skal ein samanlikne variasjonen for fleire variablar (studere samvariasjonen), kan slike filar bli både store og uoversiktlige, ettersom linjene også inneheld så mye unødvendig materiale, og derfor er uhandterlige. For slike tilfelle har vi plotta inn tala på belegg på dei ymse variantane i ein eigen fil som er organisert som eit meir vanlig skjema.

Seleksjons- og sorteringsprogramma er skrivne av Eva Møller, Bjørn Eide og Helge Sandøy. Medarbeidarane i prosjektet har sjølve skrive og kjørt statistikkprogramma ved hjelp av SPSS-pakka på Sperry 1100. Dei har måtta rekne seg ein del tid til å sette seg inn i både dei teoretiske og praktiske sidene ved dette arbeidet. Ein del prøving og feiling har det sjølv sagt blitt ettersom språkvitarar tradisjonelt får lite utdanning på dette feltet.

6. Andre bruksmåtar

Slik materialet nå ligg på filar – både oppstilt i tekstar og sortert i KWIC-konkordans – kan det brukas til fleire formål enn det som var hovudmålet med prosjektet TUB. Vi planlegg ei utgåve av tekstutdrag frå samtalan, først og fremst til undervisningsformål. Med tilgang til fotosettar som er utstyrt med lydskrift, er det enkelt å få feilfri sats, utan ny korrektur på desse innfløkte tekstane.

Vi har òg levert tekstar til ei hovudoppgåve om replikk- og samtaleanalyse. Dessutan har vi levert materiale til ymse grammatiske analysar. Belegg på grammatiske fenomen hentar ein lettast frå KWIC-konkordansen. Det er òg planlagt ein spesiell fonotaktisk undersøkelse av TUB-materialet som skal gi data til eit prosjekt om talesyntese.

7. Vurdering av EDB-bruken

7.1 Manuell mot elektronisk databehandling

Ei vurdering av EDB-bruken må gjerast i forhold til konkrete alternative arbeidsmåtar. Ettersom utgangspunktet her var at tekstane skulle transkriberast, var alternativet å bruke dei handskrivne transkripsjonsarka når ein skulle rekne belegg til analysane. Avgrensar ein seg til dei statistiske analysane i hovudprosjektet, ville kanskje slik manuell teljing koste mindre enn heile prosessen med å dataføre materialet. Men ein kan tenke seg visse ulemper med denne arbeidsmåten: For det første det banale problemet med å halde orden på tusenvis med lause ark som ein stadig måtte bla fram og tilbake i. For det andre er ikkje eit menneske like nøyaktig som maskina til å leite. Og for det tredje – og det er det viktigaste –: I det ideelle analysearbeidet har ein klare hypotesar i utgangspunktet slik at kategoriane er veldefinerte, og teljinga kan gå beint fram med å skrive f.eks. strekar i eit skjema. I praktisk analysear-

beid er det ofte slik at ein etter ei stund kjem i tvil om kategoriane er rette. Det kan komme på tale å omdefinere både variablane og variantkategoriane. Denne revurderinga av hypotesar og variablar skjer når ein sit med materialet framfor seg. I slike tilfelle er det ein umistelig fordel å ha belegga samla i velredigerte filar (eller datautskrifter) – i staden for å ha dei spreidd på dei fleire tusen transkripsjonsark. (I verste fall har ein heller ikkje notert referansar gode nok til å finne tilbake til belegget og konteksten.)

Her ser ein altså tydelig at om EDB-bruken er kostbar, gir han svært mange praktiske fordelar.

7.2. Transkripsjon mot ekserpering direkte frå band

Kan ein tenke seg alternative løysingar enda lenger tilbake i arbeidsprosessen i prosjektet, er ein aktuell utveg å ekserpere belegg rett frå lydbandopptaka i staden for å transkribere tekstane. I dei delane av TUB-materialet som ikkje var transkribert, ekserperte vi belegg på ein del variablar (der det transkriberte materialet gav for få belegg). Ekserperinga skjedde ved at ein lytta igjennom bandopptaka og skreiv ned alle belegg på ein variabel. Dessutan noterte ein òg ein kontekst på eitt ord både føre og etter ordet og ein referanse til teljeverket på avspelareren. Belegg og kontekst blei notert (i lydskrift) både for å lette kontrollen og for å ha eit visst grunnlagsmateriale om ein trong å revurdere variabel- og variantdefinisjonane. Desse to mulighetene ville ein miste om ein bare talde belegg – noe som sjølv sagt ville gå raskare.

Vi har prøvd å undersøke kor nøyaktig slik ekserpering er kvantitativt, og har komme til at med éi førstelytting og éi kontrollytting (av ein annan person) greier ein å få med seg mellom 95 og 100% av faktisk eksisterande belegg (Lødrup 1982). Det er i praksis fullt tilfredsstillande, så lenge ein kan rekne dei ikkje oppskrivne belegga som tilfeldige.

EDB-bruken ved denne ekserperingsmetoden gjeld bare den statistiske behandlinga. Ein må legge tala på belegg inn i ein fil, som ein bruker til statistikkprogramma. I visse tilfelle kan det sjølv sagt òg vere mest praktisk å skrive inn sjølve orda som er belegg.

Fordelane med denne arbeidsmåten kan vere fleire: Først og fremst sparer ein mye tid (d.v.s. kostnader). For det andre kan beleggmaterial-et vere kvalitativt meir nøyaktig enn eit transkripsjonsmateriale av di ekserperaren konsentrerer seg bare om det som er relevant for variabelen han arbeider med. Ein transkriptør kan ikkje konsentrere seg om alle dei mulige problemstillingane materialet skal settas inn i.

Men vi har òg registrert ulemper med denne arbeidsmåten: I det praktiske analysearbeidet blei den noterte konteksten for liten når variabelen måtte revurderas. Men dersom ein skulle skrive ned mye kontekst og ekserpere belegg til svært mange variablar, ville tida ein sparte i forhold til full transkripsjon, krype fort saman. Dessutan er det

ein viktig forskjell at det ekserperte materialet stort sett bare kan brukas til den eine variabelen det er tiltenkt, mens ein fullstendig transkripsjon kan ha eit stort bruksområde. Han kan brukas både til problemstillingar ein planlegg alt i utgangspunktet, og slike ein kjem på seinare. (Største avgrensinga gjeld fonologiske språkvariablar, som blir låst til det transkripsjonssystemet ein valde i utgangspunktet.)

7.3. Analysen må vente lenge

Det rasjonelle med den måten TUB brukte EDB på, ligg i rask og nøyaktig utnytting av materialet når det først er dataført. Den fordelen veks med størrelsen på materialet. Men skal ein utnytte denne fordelen, forutset det at arbeidsfasane er skilt: først innsamling, deretter transkripsjon og dataføring, og til slutt analysearbeid. At analysen må vente heilt til den mødesame transkripsjons- og dataføringsfasen er avslutta, er den største ulempa med opplegget TUB valde. Det førte til ein lang og slitande arbeidsbolk der hypotesane bare «låg på hylla og venta». I denne tida fikk ein ikkje oppleve noe fruktbart møte mellom hypotesar og data/resultat. Arbeidet gav ein få utfordringar ein kunne utvikle seg faglig på. Og dessutan var det lite rom for revurderingar av prosjektet ettersom så mye var låst frå starten av, p.g.a. den store materialmengda.

Om TUB-opplegget bør brukas om att i nye prosjekt, må sjølvsagt vere avhengig av formålet med prosjektet. Eg ser dei mange fordelane i det store materialet vi nå har tilgjengelig om bergensmålet. Det er alt brukt til mange analysar, og mange nye formål kjem det til å bli brukt til vidare framover òg. Legg ein stor vekt på å ha tilgjengelig ein slik «database», kan ein rettferdiggjøre å dataføre så mye materiale. Det kan ein om ein reknar kartleggingsarbeidet som viktigare enn arbeidet med teorien.

Men reknar ein dei teoretiske spørsmåla for å vere hovudoppgåva i talemålsforskninga i dag, løyser ein ikkje dei først og fremst med eit stort materiale. I den greina som kallas sosiolingvistikk, er det viktigare f.eks. å klette ut gode eksperimentelle språkbrukssituasjonar, som kan auke innsikta i og forbetre teorien om korleis og korfor menneska utnyttar språkvariasjonen. I arbeidet både med slike og andre problemstillingar, der ein føler teorien i dag er ufullstendig eller lite tilfredsstillande, er det viktig å gjøre avstanden kortare mellom hypotesedanning og ev. avkrefting/styrking. Det vil seie at analysen må komme fortare, slik at hypotesane kan presiseras og forskaren kan modnas. For å få til det må lydbandmaterialet vere mindre, men heller meir relevant for spesielle problemstillingar, og arbeidsmåten må vere raskare. Den skisserte måten med ekserpering direkte frå lydband og med EDB-bruk bare til statistisk behandling av beleggtala er den mest rasjonelle arbeidsmåten til det formålet.

Litteraturliste

- Kort orientering om prosjektet «Talemål hos ungdom i Bergen». I: Talemål i Bergen 1/83. Bergen.
- Lødrup, Helge 1982. Om ekserpering i sociolingvistiske undersøkelser. (Prosjektnotat.)
- Myking, Johan 1983. Adverba *då* og *no* i bergensmålet. Lingvistisk og sociolingvistisk variasjon. I: Talemål i Bergen 2/83. Bergen.
- Solevåg, Harald 1977. Dokumentasjon av TAGGING-systemet. (Nordisk institutt.) Bergen.

Helge Sandøy – amanuensis i norsk talemål ved Nordisk institutt, Universitetet i Bergen – er dagleg leiar for prosjektet «Talemål hos ungdom i Bergen.»

Personregisterloven og behovet for datavern

Thore Gaard Olaussen

Lovens utforming

Lov om personregistre m.m. trådte i kraft 1. januar 1980. Bortimot et tiår før ble det første utvalget med oppgave å vurdere reguleringer på bruk av persondata, nedsatt. Utvalget med professor *Tore Sandvik* som formann la fram sin utredning i april 1974 med forslag om lovreguleringer for privat sektor innenfor områder som kredittopplysning, databehandling, (databehandlingsforetak), adresserings- og distribusjonsvirksomhet og markeds- og opinionsundersøkelse¹. Ettersom mandatet i første rekke rettet søkelyset mot bruk av persondata i kredittopplysningsvirksomhet, ble utvalget i ettertid ofte kalt «kredittopplysningsutvalget».

Det andre utvalget med *Helge Seip* som formann ble oppnevnt i 1972 med oppgave å vurdere bruk – og mulig misbruk – av persondata og personregistre i offentlig sektor. Utvalget la fram sin utredning i februar 1975 med lovforslag som bl.a. innebar generell konsesjon på personregistre i offentlig forvaltning². Imidlertid ble det fra utvalgets side foreslått at man burde sammenføre de to lovutkastene fra Sandvik-utvalget og Seip-utvalget. Vi fikk således en lov som i tråd med den Sverige innførte allerede i 1973, omfattet konsesjonsplikt på alle personregistersystem som medfører bruk av edb. I tillegg fikk vi i Norge også konsesjonsplikt på manuelle registersystem som inneholder særlig sensitiv informasjon³.

Loven som ble behandlet og vedtatt av Stortinget 1978, kan vi grovt oppsummere i følgende punkter:

1. Et datatilsyn opprettes til å behandle konsesjonssøknader ved etablering av personregister og å utøve kontroll med slike registre. Datatilsynet skal også behandle konsesjonssøknader fra kredittopplysningsinstitusjoner, databehandlingsforetak, adresserings- og distribusjonstjenestefirma og opinionsinstitut og utøve kontroll med disse.
2. Lovens personbegrep gjelder «juridiske personer» – dvs. at den foruten fysiske personer også omfatter «sammenslutninger» og «stiftelser»⁴.
3. Konsesjonsplikten i loven dekker som nevnt både registersystem som medfører bruk av edb og manuelle registre som inneholder
 - informasjon om rase, eller politisk eller religiøs oppfatning
 - informasjon om at en person har vært mistenkt, tiltalt eller dømt i straffesak