

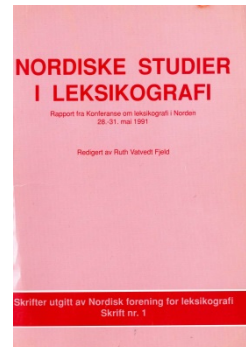
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Om SAOB:s material och en bibliografisk databas

Forfatter: Lars Svensson

Kilde: Nordiske Studier i Leksikografi 1, 1992, s. 369-379
Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Lars Svensson

Om SAOB:s material och en bibliografisk databas

SAOB:s excerptsamling omfattar ca 7,5 miljoner språkprov. Med excerperingen, som är selektiv, har man eftersträvat kronologisk spridning av materialet och samtidigt velat täcka så mycket som möjligt av svenskt språk (olika stilsfärer och olika språk i språket). Excerptsamlingen, som omfattar ca 20.000 källor, innehåller inte bara tryckt material utan också ett omfattande handskriftsmaterial. Stickkontroller i bibliografiska arbeten visar att SAOB väl täcker svenskt boktryck från 1500-talet till våra dagar. — Sedan några år tillbaka pågår vid SAOB-redaktionen ett projekt med syfte att skriva och optiskt läsa in SAOB:s källförteckningar i en bibliografisk databas med en mängd sök- och sorteringsmöjligheter. Bibliografin kommer att kunna länkas till den databas som Språkdata i Göteborg håller på att upprätta över de tryckta banden av SAOB.

Svenska Akademiens ordbok (SAOB)¹ är en historisk-nuspråklig ordbok som behandlar svenskt skriftspråk från 1521 till nutiden. Hittills (1991) är 30 volymer utgivna. Band 1 publicerades 1898, band 30 1989. Jag skall nedan mycket översiktligt försöka ge en bild av SAOB:s material. Mina huvudkällor härvidlag har varit kapitel 1 och 2 i Ekbo-Loman, *Vägledning till Svenska Akademiens ordbok*, skrivna av förre ordbokschefen Sven Ekbo, och uppsatsen "Om SAOB:s material" (i *Nysvenska Studier* 1983), skriven av Sture Hast, som efter 30 års tjänstgöring som ordboksredaktör vid SAOB nyligen gått i pension. Jag kommer också att kort presentera SAOB:s *Källförteckning. Supplement II* (1990), som — i motsats till SAOB:s källförteckningar från 1939 och 1975 — är en bibliografi, inte bara en alfabetisk källförteckning.

7,5 miljoner egna excerpter

SAOB:s egen excerptsamling omfattar i dag ca 7,5 miljoner språkprov. Excerpterna, som är försedda med uppgift om källans namn (författare, boktitel o.d.), sida eller paragraf, datering o.d., förvaras i alfabetisk ordning efter de olika uppslagsorden i kapslar av papp. Varje excerptlapp omfattar ett enda språkprov. Excerpter som avser samma ord ordnas så att det äldsta språkprovet ligger först, det yngsta sist. Språkproven är skrivna för hand (det stora flertalet) eller på skrivmaskin eller utgörs av en fotograferad boksida o.d. eller urklipp av tidningstext som klistrats på excerptlapparna.

Material från andra samlingar

SAOB får också visst material från andra samlingar: språkforskaren Esaias Tegnér's språkprovssamling (flera tusen excerpter), arkivmannen och urkundsutgivaren Victor Granlunds ordspråkssamling, språkforskaren Olof Östergrens språkprovssamling (ca 250.000 excerpter) till *Nusvensk ordbok*, 10 band med riklig fraseologi, Tekniska Nomenklaturcentralen (TNC), ordförteckningar med hänvisningar till källor, Svenska Språknämnden (ca 1.000 excerpter per år), — Språknämnden får i gengäld visst, nytillkommet material från

SAOB —, Rolf Dunås klipp- och excerptsamling (flera tusen excerpter). — Dunås var lärare och svarade under flera år på språkfrågor i en dagstidning. Redaktionen får också sporadiska bidrag av enskilda forskare och ordälskare.

Här skall också nämnas en serie konkordanser på mikrofiche utarbetade av Språkdata vid Göteborgs universitet: press 76, omfattande 1, 3 miljoner löpande ord, romaner 76-77, 80-81, tillsammans omfattande 9,3 miljoner löpande ord ur 129 romaner, Svensk författningssamling 77-80, ca 1/2 miljon löpande ord. Språkdata-konkordanserna är ett viktigt, modernt material. De har icke excerperats men används för anskaffande av moderna belägg.

Räknar vi samman SAOB:s egen språkprovssamling och materialet från andra institutioner och samlingar får vi en imponerande och solid bas för redigeringen av ordboken. Följande siffror kan ge en antydning om materialets stora omfattning. Band 30 (*spår-stockna*) upptar 1.615 tättskrivna spalter, vilket motsvarar knappt 8 sidor i SAOL 1986, dvs. ca 202 spalter (eller ca 16.000 excerpter) i SAOB motsvarar ungefär 1 sida i SAOL. Det största parti som — hittills — delats ut är partiet "stor", dvs. substantivet och adjektivet "stor" jämte sammansättningar och avledningar. Det omfattade över 18.000 excerpter. Det tog ca 18 månader för redaktören att skriva första manus. — SAOB:s excerptmaterial står sig gott vid en jämförelse med utländska ordböcker av motsvarande typ: Ordbog over det danske sprog, ca 2,5 miljoner excerpter, Oxford-ordboken (*New English Dictionary*), ca 6 miljoner, *Thesaurus Linguae Latinae*, ca 10 miljoner, *Orðabók Háskólans*, ca 2,5 miljoner.

Säkerhetsfilmning av språkprovssamlingarna

Av säkerhetsskäl har den del av materialet som ännu icke redovisats i tryck (*ymb-ö*) mikrofilmats med ett datorsystem som medelst en mikrofilmläsare möjliggör återsökning av språkproven (ca 1,5 miljoner). En kopia av mikrofilmerna går att anskaffa från SAOB. I första hand borde universitetsbibliotek och språkliga institutioner vara intresserade av att ha tillgång till det utotryckta materialet. Observera att blott en ringa del (ca 10-15%) av det väldiga excerptmaterialet redovisas i tryck (för att ordbokens omfång skall begränsas). Så småningom kommer även materialet till den tryckta delen av ordboken att mikrofilmats.

Excerptering utanför och inom redaktionen

Excerpteringen — jag förbigår här ordbokens äldre historia — påbörjades i mitten av 1880-talet och pågår allttjämt. Den utförs i dag av 6-7 fasta excerptister utanför redaktionen, som excerpterar tidningar, tidskrifter, årsskrifter, skönlitteratur etc. Ordbokschefen och chefkontrollanten väljer ut den litteratur som skall excerpteras, ofta på förslag av redaktör, kontrollant eller annan medarbetare. Stundom föreslår också någon av SAOB:s facksakkunniga böcker för excerptering. Ofta får någon erfaren redaktör eller chefkontrollanten i uppdrag av ordbokschefen att göra en grov förexcerptering för att utvärdera om skriften i fråga kan tillföra något nytt utöver det som redan finns i språkprovssamlingarna. En *Hushållsbok*, t.ex., från slutet av 1700-talet kanske ratas som excerpteringstext. Belägg på ord som *öl*, *humle*, *salt*, *socker*, *gryn*, *ärter* osv. finns nämligen — med största sannolikhet — i redan excerpterade hushållsböcker från 1500- och 1600-talen.

Materialet kompletteras emellertid också av redaktionen, framför allt av assistenterna i den förberedande redigeringen, som genomgår bortemot 250 verk (ordböcker, ordlistor,

konkordanser, uppslagsböcker, florum, apotekartaxor, riksdagsprotokoll m.m.), men även av de assistenter som har till uppgift att kontrollera redaktörernas artiklar (beläggens datering, betydelseinnehåll och stavning, definitionernas hållbarhet i förhållande till språkproven osv.). Även redaktörerna själva kompletterar i viss utsträckning materialet under arbetets gång, ofta i samband med specialstudier av ett visst fackområde.

Excerptisterna

Under perioden 1894—1924 hade excerptisterna en egen chef, språkmannen Theodor Hjelmqvist. Nya excerptister fick under denna tid också en (alltför) utförlig och detaljerad instruktionsskrift på 13 sidor som hjälpmedel, "Anvisningar till insamlande af språkprof för Svenska Akademiens ordboksarbete". I dag får en ny excerptist instruktioner av chefkontrollanten, muntligen eller medelst brev. Viktig information om SAOB:s arbete och material meddelas också på de seminarier med föredrag och diskussioner, som arrangeras av SAOB-redaktionen med jämna mellanrum och till vilka excerptister och facksakkunniga särskilt inbjuds. — Sammantaget har SAOB hittills anlitat flera hundra, ofta handplockade, excerptister med fin språkkänsla och språkuppfattning, till yrket ofta skolmän och lärare, bibliotekarier, präster eller vetenskapsmän inom skilda fack och fält. Även pensionerade SAOB-medarbetare har anlitats och anlitas ännu som excerptister.

Noggrann kontroll av språkproven

Alla språkprov som medtages i trycket blir föremål för omsorgsfull kontroll. Excerpterna är i allmänhet tillförlitliga vad beträffar själva texten. Inte så sällan förekommer det dock korrupta textställen, utgivningsfel, skriv-, läs- och tryckfel, vilka kan vålla redaktör och kontrollant stort huvudbry. Felaktiga käll- och sidhänvisningar rättas till i manus av kontrollanterna, som också — om möjligt — flyttar belägg som icke excerperats i originalupplaga till en sådan. Ibland har excerptisten varit för snål med kontext, så att betydelsen av textstället inte klart framgår. I sådana fall måste man läsa en längre passage i källan och då är man tacksam för att redaktionen är inrymd i Lunds universitetsbibliotek, där redaktionspersonalen har tillstånd att röra sig fritt i magasinerna. — Före 1924 granskades excerptistens arbete mycket noga av excerptistchefen. Särskilt ordurvalet blev föremål för kritik eller beröm. Inlämnat excerptmaterial granskas också i dag men allvarlig kritik mot excerptisten är sällsynt.

Selektiv excerptering

Det ordförråd som behandlas i SAOB är det svenska skriftspråkets från 1521 fram till vår egen tid. I källorna som excerperas väljer excerptisten ut 3-5 ord per sida. Vissa källor har emellertid blivit — eller blir — föremål för en betydligt utförligare excerptering (t.ex. Bibeln, Olaus Petris skrifter, Linnés skrifter), dock aldrig så, att alla löpande ord i en text blivit excerperade. Det finns naturligtvis en risk med selektiv excerptering: de mest frekventa orden och uttrycken kan bli dåligt representerade. Excerptisten går lätt förbi de vanliga orden och noterar i första hand språkliga "egendomligheter". Å andra sidan är det betydligt lättare för redigerande redaktör och kontrollant att komplettera bristen på "vanliga" ord och uttryck än

på "ovanliga", om så skulle behövas. SAOB iakttar och måste iaktta restriktivitet i fråga om utökande av materialet. Ett alltför stort material hindrar starkt utgivningstakten. Ett franskt ordboks företag, som med dator hade anskaffat 90 miljoner excerpter, använde i realiteten bara 6 miljoner vid redigeringen. SAOB:s språkprovssamling utökas sedan ca 30 år tillbaka årligen med ca 10.000-12.000 excerpter. Excerpteringen sker sedan 1959 likformigt över hela alfabetet. En del av det nytillkomna materialet kommer således att behandlas i ett kommande *Supplement* (eller i en ny upplaga av ordboken).

Urvalsprinciper

I SAOB behandlas svenskt skriftspråk från nästan 5 sekel men icke utan inskränkning. Gränserna för vad som kommer med eller utesluts blir ibland flytande. Följande kategorier av ord medtas i regel icke i SAOB:

- a) namn utan användning såsom appellativ, dvs. personnamn och ortnamn,
- b) tillfälliga ord (särskilt sammansättningar) och bildningar,
- c) extremt fackspråk, särskilt sådant som ej nått utanför den trängre fackkretsen,
- d) dialektord i dialektordböcker eller dialektundersökningar eller i text som är skriven helt på dialekt,
- e) främmande ord som ej har svensk böjning och ej vunnit hävd i svenskan,
- f) rena citatord från främmande språk.

Kommentar

Viktiga litteraturverk och kända litteraturställen prioriteras. Man är i allmänhet också generösare mot belägg från 1500- och 1600-talen än mot senare belägg. Kryptogamnamn belagda endast i 1900-talsfloror förbigås liksom i regel termer som utgör översättning från främmande språk och endast är belagda i referat av utländska undersökningar eller förhållanden i källor som *ÅrsbVeta*, *KrigVAH* och *KrigVAT*. Individuella ordskapelser (t.ex. hos Olof Högberg eller Carl August Hagberg, i hans Shakespeare-översättningar) förbigås i allmänhet. Sammansättningar till sammansättningar utgallras strängt. Använder en författare dialektord i berättande framställning tas de i regel med.

Källorna

Medeltidens svenska skriftspråk är täckt genom Söderwalls och Schlyters ordböcker. I svensk språkhistoria brukar man räkna den äldre nysvenskan från 1526, *Nya Testamentets* utgivningsår, men startåret för SAOB:s excerptering har satts till 1521, för att *GIR* skall komma med i sin helhet — de äldsta akterna där är daterade 1521. SAOB:s källförteckningar — 1898 (i första bandet), 1939, 1975, 1990 (den sista är ännu icke publicerad) — upptar ca 20.000 titlar men antalet texter är betydligt större. En källa som omfattar flera band eller volymer uppföres nämligen i *Källförteckningarna* som en. Så omfattar t.ex. källan *GIR* 29 band, *PH* 15 band, *OxBr.* 12 band. Likaså anges en tidning eller tidskrift eller årsbok o.d. som en enda källa, även om flera årgångar har excerpterats. Omvänt förekommer naturligtvis också att åtskilliga texter, t.ex. almanackorna under 1600-talet, bara består av ett 10-tal sidor. I

Källförteckningen 1939 har arbeten, som endast citeras någon enstaka gång, ej medtagits. I *Supplement I* och *II* har däremot alla citerade källor medtagits, även sådana, varifrån endast enstaka belägg hämtats. Sture Hast (*Nysvenska Studier* 1983, s.173 f.) har efter en stickprovskontroll kommit fram till att endast 1/5 av de citerade texterna i *Supplement I* verkligen är excerperade, dvs. att excerptisten har skrivit av 3-5 belägg per sida. I flertalet källor i *Supplement I* har alltså blott enstaka belägg excerperats.

Med excerpteringen har man — samtidigt som man eftersträvat kronologisk spridning av materialet — velat täcka så mycket som möjligt av svenskt språk, t.ex. olika stilsfärer och olika språk i språket (bygdemålsfärgat språk, vardagligt språk, högtidligt språk, religiöst språk, ålderdomligt eller arkaiserande språk, slang, *arkeologi, botanik, ekonomi* etc.). *Källförteckningarna* omfattar ca 150 genrer och ämnesområden: vitterhet, teologi, språkvetenskap, förordningar, kanslihandlingar, astronomi, kemi, militärväsen, sport, brev, dagböcker, domböcker, tänkeböcker (dvs. stadsböcker innehållande anteckningar av judiciell, kommunal och kameral art), biografi, botanik, zoologi, reseskildringar, tidningar, tidskrifter. Osv. Även barn- och ungdomsböcker, t.o.m. serietidningar, är representerade. Excerptsamlingen innehåller inte bara tryckt material utan också ett mycket omfattande handskriftsmaterial (räkenskapshandlingar, bouppteckningar o.d.). Detta är ett mycket viktigt material. Kulturord, vanligen lånord, uppträder ofta först i tullhandlingar och slottsräkenskaper och får sedan vidare spridning. I bouppteckningarna anträffas äldre benämningar på redskap, verktyg, husgeråd o.d.

I detta sammanhang kan nämnas att Bibliotekstjänst i Lund, dit praktiskt taget allt svenskt tryck skickas för recension, kontinuerligt (2 gånger i månaden) sänder oss sambindningsrecensionerna (som utgör underlag för bibliotekens inköp av litteratur), varigenom vi fortlöpande underrättas om allt nytryck i Sverige.

Fig. 1 Några viktigare, utförligt excerperade källor.

1. Bibeln och psalmboken.
2. GIR (Konung Gustaf den förstes registratur)
4. PH (Utdrag utur alle ifrån .. 1718 utkomne publique handlingar)
5. Tänkeböcker
6. Encyklopedier
7. Protokoll från ständsriksdagarna
8. Axel Oxenstiernas skrifter och brevvetling
9. Svenska Akademiens handlingar
10. Vetenskapsakademiens handlingar
11. Tidningar och tidskrifter
12. Handskrifter.

(Efter Sture Hast, *Nysvenska Studier*, 1983, s.180)

Källornas fördelning på genrer

Sture Hast har i *Nysvenska Studier* 1983, s.174 f. gett en översikt av källornas fördelning på olika genrer eller fackområden. Han har granskat *Källförteckningen* i sin helhet från 1898 (24 sidor i band 1), 50 spalter i *Källförteckningen* tryckt 1939 och *Källförteckning. Supplement I* tryckt 1975. I tabellen nedan har jag lagt till siffror ur *Källförteckning. Supplement II* från 1990. Jag har också ändrat uppställningen. Den största genren (i *Supplement II*) kommer först.

Fig. 2 Jämförelse mellan källförteckningarna med avseende på genrefördelning (i procent).

Genrer	1898	1939	1975	1990	antal excerpt. texter 1990
vitterhet	29,0	21,8	42,0	18,1	329
handskrifter	0,1	1,2	1,7	7,5	136
tidskrifter	2,3	1,8	1,1	6,6	121
teknik	0,6	1,4	2,5	6,3	114
juridik	1,6	2,1	0,7	4,8	88
språkvetenskap	4,0	3,8	2,4	5,2	95
biografi	2,7	8,3	4,2	4,3	78
ekonomi	2,5	2,5	4,3		79
sport	1,6	0,8	2,0	3,9	71
historia	6,0	5,5	5,1	3,7	68
naturvetenskap	1,5	1,1	1,2	2,9	53
officiellt tryck	1,7	3,3	0,5	2,7	54
teologi	12,5	6,5	3,1	2,6	47
tidningar	1,6	1,3	0,7	2,2	40
geografi	1,4	3,0	1,8	2,2	40
medicin	2,2	3,3	2,0	2,1	39
matematik, fysik, kemi	1,1	3,4	1,3	1,6	29
litteraturvetenskap	1,1	2,2	2,5	1,5	27
ordböcker	4,9	1,1	2,0	1,2	21
botanik och trädgård	1,2	3,2	0,8	1,2	23
resor	1,6	2,1	3,2	0,9	17
militärväsen	4,1	2,1	2,3	0,6	11
sjöväsen	1,0	0,3	0,4	0,5	9
åkerbruk	1,0	2,4	1,3	0,3	5
pedagogik	1,4	2,4	0,8	0,2	3

Kommentar

Som framgår är *vitterhet* fortfarande den största genren. Att siffrorna är relativt små 1898 och 1939 beror säkerligen på att åtskilliga "samlade verk" tagits upp som en enda källa. *Handskrifterna* i *Supplement II* (1990) är påfallande många. Detta torde bero på att samtliga källskrifter upptas här, även sådana som endast excerperats för något enstaka belägg eller endast citerats någon enstaka gång. I *Supplement II* har siffrorna för *tidskrifter*, *teknik*, *juridik*, *språkvetenskap* (i mina siffror ingår även språkvetenskapliga *ordböcker*), *biografi*, *ekonomi*, *sport*, *naturvetenskap*, *officiellt tryck* och *tidningar* ökat. Siffrorna för *teologi*, *militärväsen* och *pedagogik* fortsätter att rasa. Även *reseskildringarna* har avtagit.

Hur täcker materialet svensk bokproduktion?

1500- och 1600-talen

En snabb genomgång i Collijns 1500- och 1600-tals-bibliografier visar att praktiskt taget allt svenskt 1500-talstryck excerperats och det mesta från 1600-talet (i varje fall det som är av vikt), vilket stämmer bra överens med min uppfattning av perioden från egen forskning. Det gäller biblarna, reformatorens skrifter, katekeser, psalmböcker, bönböcker, evangelier och epistlar, läkeböcker, pestböcker, mandat, stadgor, bondepraktikan och ordböcker. Därtill kommer skrifter som ej trycktes i samtiden utan senare (tänkeböcker, skotteböcker, G1R, visböcker osv.). Likaså har samlade verk av periodens förnämsta författare excerperats: Messenius, Bureus, Stiernhielm, Lucidor, Columbus, Lindschöld, Werving, Israel Holmström, Dahlstierna, fru Brenner, Runius, Frese, Triewald, Urban Hiärne, Spiegel, Swedberg, Rudbeck,

Verelius översättningar av isländska sagor, Johann Arndts skrifter. Osv. Av de 41 skrifter som Collijn förtecknar av Ericus Schroderus, som var kunglig korrektör och translator och bedrev en imponerande översättarverksamhet åren 1610-1642, har SAOB excerperat 25. Osv. Vad som ej excerperats — det gäller 1600-talet — är bl.a. ABC-böcker, enstaka visor och skrifter av mindre kända eller anonyma författare. Vad gäller ABC-böckerna är i allmänhet innehållet hämtat ur katekesen eller psalmboken, två källor som är väl excerperade och följer Bibelns språk, varför dessa källor knappast skulle tillföra ordbokens samlingar något nytt i fråga om ordförråd (eller böjnings- eller stavningsformer).

1700-talet

1700-talets samlade svenska bokproduktion är enorm. Den är i sin helhet ungefär 10 gånger så stor som Collijns 1600-talsbibliografi, som upptar ca 4600 titlar. Den nyligen färdiga bibliografiska databasen *Svensk Bibliografi 1700-1829*, KB, Stockholm, som grundar sig på KB:s samling av katalogiserat tryck och som är tillgänglig för alla som är anslutna till LIBRIS-systemet, uppgår till ca 40.000 tryck. I denna siffra är vissa tryckalster icke medtagna såsom årstrycket (dvs. författningstrycket, från lagar och förordningar ned till ordningsstadgor och taxor), periodica, dissertationer och program, tillfällestryck (dvs. bröllopsdikter, sorgeskväden, gravtal, namnsdagshyllningar, välgångsönskningar o.d.), sagor, skillingtryck och det kommunala förvaltningstrycket. Inbegriper man dessa tryck blir siffran 4 gånger så stor. För att få en uppfattning om hur SAOB täcker 1700-talet har jag använt *Swedish Imprints 1731-1833 (SWIM)*, ett engelskspråkigt, häftesvis utkommande arbete av nationalbibliografisk karaktär, som ges ut i Uppsala av Rolf du Rietz, Sten Lindroths *Svensk Lärdomshistoria*, Lars Svenssons *Förteckning över bibliografier rörande 1700-talets bokproduktion* och — delvis — också databasen *Svensk Bibliografi 1700-1829*. Självklart har SAOB bara excerperat en del av 1700-talets litteratur men mitt huvudintryck är att excerperingen täcker väl de viktiga och stora genrererna: vitterhet, teologi, språkvetenskap, historia, juridik, resor, geografi och topografi, juridik, teknologi, ekonomi, naturvetenskap och medicin. T.ex. har SAOB excerperat första och femte upplagan av vår första svenska nationalgeografi på modersmålet, *Inledning til geographien öfver Sverige*, författad av geografen och kanslisten Eric Tuneld 1741. De texter som Lindroth behandlar av Anders Berch, Torben Bergman, Carl Carleson, Celsius, Gadd, Klingensstierna, Linné, Nordencrantz, Rosén von Rosenstein m.fl. har excerperats av SAOB. Detsamma gäller för den topografiska litteraturen företrädd bl.a. av Broman, Broocman, Faggot, Fernow, Högström, Lars Salvius, Eric Tuneld och Oedman.

Jag har ofta förvånat mig över hur väl 1700-talet är excerperat. De äldsta SAOB-redaktionerna har haft stor litteraturkännedom men naturligtvis också använt sig av bibliografisk litteratur vid litteraturvalet. Särskilt viktiga hjälpmedel torde ha varit Lénströms *Sveriges litteratur- och konsthistoria i utkast* (1841), som innehåller över 100 sidor utförliga bibliografiska förteckningar ordnade under huvudrubriker: teologi, estetik, filosofi, botanik, zoologi, kemi, bergsvetenskap etc. Andra viktiga bibliografiska hjälpmedel torde ha varit Elmgrens *Öfversigt af Finlands litteratur* (1861-65), Lüdekes *Allgemeines schwedisches Gelehrsamkeits-Archiv unter Gustafs des dritten Regierung* (1781-1796), och Warmholtz *Bibliotheca historica Sueo-gothica* (1782-1817).

1800- och 1900-talen

1800- och 1900-talen är också väl representerade i SAOB:s språkprovssamling. Alla klassiska författare har excerperats från Tegnér, Geijer, Atterbom, Stagnelius, Almquist, Strindberg och Selma Lagerlöf till Hjalmar Gullberg, Edfelt, Ekelund, Ekelöf, Evert Taube m.fl. Några författare från vår egen tid som excerperats under den senaste 15-årsperioden är Stig Claesson, Sven Delblanc, Kerstin Ekman, Per Gunnar Evander, Maria Gripe, Lars Huldén, Gösta Knutsson, Ulf Lundell, Göran Palm och Bo Setterlind. Andra väl representerade genrer eller ämnesområden o.d. är biografi, botanik, ekonomi, etnografi, författning, förordning, geografi, historia, juridik, litteraturvetenskap, ordbok, periodica, samhällsvetenskap, sport, språkvetenskap, teknologi, teologi, tidningar, tidskrifter och undervisning.

Källförteckning. Supplement II

SAOB:s alfabetiskt uppställda tryckta källförteckning från 1939 upprättad av Sigrid von Greyerz och dess *Supplement* från 1975 upprättat av Gulli Starfelt-Brown ger inte upplysningar om vilka ämnesområden som excerperats eller vad som excerperats från en viss period utan är avsedda att förklara titelförkortningarna i de tryckta ordboksbanden. SAOB:s *Källförteckning. Supplement II*, som nyligen färdigställts, är en bibliografi över ca 1.800 texter excerperade perioden 1975-1990. Bibliografien, som lästs in i referenshanterings- eller databasprogrammet Reference Manager föreligger i datorversion (Macintosh) och är i första hand avsedd för SAOB-redaktionen men kan givetvis också vara av stort intresse för andra institutioner och forskare. Den kommer att inom kort tryckas i följande 4 delar:

- 1 en strikt alfabetisk del,
- 2 en kronologisk del,
- 3 en ämnesinriktad del och
- 4 ett ämnesordsregister.

I sökregistret kan man söka på ca 550 olika ämnen. Som sökord använder jag alla de ämnen och fackområden som finns i SAOB. Jag har dessutom lagt till ett antal sökord. Man kan också söka på författare, titel, kommentar, tryckort, tryckår, personnamn och ortnamn. Varje post eller referens har försetts med 5-7 sökord jämte underavdelningar. Den vittra litteraturen har jag försett med underavdelningarna poesi, prosa, drama, tal, svenska, översättning (från vilket språk). Söker man på "vitterhet:svenska:poesi" får man således fram all svensk originallitteratur som klassats som poesi. Söker man på "Shakespeare" så får man fram alla excerperade Shakespeare-översättningar. Söker man på "språkvetenskap:ordbok:-svensk-engelska" så får man fram svensk-engelska ordböcker. Är man intresserad av att veta var t.ex. 1700-talets litteratur trycktes utanför Stockholm, som givetvis var den dominerande tryckorten (liksom i våra dagar), så går det lätt att få fram denna uppgift. Osv. Jag har även noterat om författaren är manlig eller kvinnlig. Man kan alltså få svar på frågan hur många kvinnliga författare det fanns t.ex. under 1600-talet och vad de skrev. Sökmöjligheterna är legio. — Materialet kan sorteras med avseende på författare, titel, tryckort, utgivningsår och löpnummer.

I kommentarfältet har jag skrivit in uppgifter om eventuella original, förlagor, upplagor, översättningar, tillförlitlighet m.m., ibland också biografiska upplysningar om författaren (yrke, födelseår, födelseort o.d.). — I samband härmed vill jag nämna att docent

Sven Lange, Stockholm, använt SAOB:s källförteckning för upprättande av en databas för undersökning av svenskt skriftspråk under perioden 1526-1799. Man kan söka på författarnamn och ämnesområde men också på födelseår, födelseområde och generation, vilket ger möjlighet att undersöka om det finns någon regional fördelning när det gäller vissa ämnesområden. De biografiska uppgifterna är mycket värdefulla för SAOB och kommer — med benäget tillmötesgående av Sven Lange — att arbetas in i SAOB:s bibliografi. De tryckta källförteckningarna har nyligen optiskt lästs in i dataprogrammet och kommer inom kort att bearbetas. Bibliografien kommer att bli en ovärderlig hjälp för ordboksredaktionen när det gäller att snabbt upptäcka luckor i excerperingen. Det blir också lätt att uppdatera databasen.

Fig. 3 Sökning på några slumpvis valda sökord i databasen Källförteckning. Supplement II.

sökord	antal källor	tryckår
bin	1	1969
dator	4	1973, 1979, 1980, 1982
fiske	7	1754, 1914, 1935, 1941, 1947, 1952, 1954
schack	6	1839, 1869, 1902, 1908, 1972, 1979

De undersökta ämnesområdena är förutom "bin" väl representerade, även kronologiskt.

Optisk inläsning av SAOB

Artiklarna i SAOB innehåller en mångfald uppgifter av skilda slag. Läsaren får information om uttal, ordklass, böjning, sido- eller variantformer, stavning, etymologi, bruklighet, betydelse, äldsta (och för ord med inskränkt bruklighet också yngsta) belägg, sammansättningar, avledningar, och i fråga om verb särskilda förbindelser osv. Vid Språkdata i Göteborg pågår sedan ca 10 år tillbaka ett projekt med syfte att upprätta en databas över de tryckta banden av SAOB. I dag är artiklarna i banden 1-28 inlästa och kontrollarbete pågår. Databasen, som kommer att ge en mängd sökmöjligheter — kronologiska, morfologiska, stilistiska, etymologiska osv. — kommer att bli ett värdefullt hjälpmedel för ordboksredaktionen och annan lingvistisk forskning. Ordboksredaktionen kommer inom kort att kunna utnyttja databasen — i varje fall delar av den. Det kommer också att vara fullt möjligt att koppla databasen till den *Bibliografi över SAOB:s källmaterial* som utarbetas vid redaktionen i Lund. Genom att med Bibliografins hjälp spåra texter från ett visst ämnesområde — t.ex. juridik — kan man med databasens hjälp komma åt källhänvisningarna och därmed uppslagsorden och på så sätt upprätta en förteckning över ordförrådet i juridiskt fackspråk från t.ex. 1700-talet. Tanken svindlar inför de möjligheter som yppar sig.

SAOB:s språkprovssamling tillgänglig för forskning

Ett stort antal forskare från Norden och andra länder har under årens lopp besökt ordboksredaktionen i Lund för att undersöka och excerpera SAOB:s väldiga material, varav endast ca 10-15 % redovisas i tryck. Excerptsamlingen utnyttjas inte bara av språkforskare utan

också av litteraturhistoriker, teologer, historiker, folklivsforskare, jurister osv. Särskilt för ordforskare erbjuder språkprovssamlingen ett oskattbart och oundgängligt material. Erik Terner, t.ex., utvidgade sin ordboksartikel om räkneordet *En* till en doktorsavhandling (1922), och Emil Olson gav ut sin ordboksartikel *Den* som en egen skrift (1913). Elof Hellquist fick — har det berättats — idén till sin monumentala *Svensk etymologisk ordbok* (1920-22) genom sitt arbete som redaktör vid SAOB, där han var anställd 1893-1904. Och — för att avsluta i vår egen tid — för något år sedan har Elsie Wijk-Anderssons doktorsavhandling utkommit, *Bara i fokus*, som är en semantisk-syntaktisk studie av adverbet *bara* och dess ekvivalenter (*allena, allenast, blott, enbart, endast, uteslutande*) i nysvenskt skriftspråk, ett ämne som knappast skulle kunna ha behandlats utan SAOB och dess materialsamling.

Noter

1. Boktitlar o.d. förkortas enligt SAOB:s principer. Se Litteratur och förkortningar sid. ??

Litteratur och förkortningar

- Allén, Sture. 1980. Hur djupt går SAOB? I: *Nysvenska Studier* 60: 119-129. Upps.
- Allén, Sture. 1986. Inför tredje seklet. I: Allén, Sture, Loman, Bengt, Sigurd, Bengt, *Svenska Akademien och svenska språket. Tre studier.* 233-273. Sthm.
- Collijn, Isak. 1927-38. *Sveriges bibliografi intill år 1600.* 1-3. Upps.
- Collijn, Isak. 1942-44. *Sveriges bibliografi. 1600-talet.* 1-2. Upps.
- Ekbo, Sven. 1965. Historik, organisation och arbetsmetod. I: Ekbo, Sven, och Loman, Bengt, *Vägledning till Svenska Akademiens ordbok.* NSS 32. Sthm.
- Ekbo, Sven. 1965. Ordboksartikelns uppbyggnad. I: Ekbo, Sven, och Loman, Bengt, *Vägledning till Svenska Akademiens ordbok.* NSS 32. Sthm.
- Elmgren, Sven Gabriel. 1861-65. *Öfversigt af Finlands litteratur* [1-2]. Hfors.
- Falck-Kjällqvist, Birgit. 1987. Om Svenska Akademiens ordbok. I: *Språkbruk* 1. 16-20. Hfors.
- GIR = Konung Gustaf den förstes registratur.* 1-29. 1861-1916. Sthm.
- Hast, Sture. 1983. Om SAOB:s material. I: *Nysvenska Studier* 63. 159-192. Upps.
- Jonsson, Hans. 1981. Ekeblad och SAOB. I: *Nysvenska Studier* 61. 134-138. Upps.
- Jonsson, Hans. 1983. Om historisk lexikografi. I: *Nysvenska Studier* 63. 128-158. Upps.
- Jonsson, Hans. 1990. Svenska Akademiens Ordbok (SAOB). I: *Wörterbücher .. Ein internationales Handbuch zur Lexikographie...* Edited by .. Franz Josef Hausmann et.al. Second Volume. 1940-1943. Berlin • New York.
- KrigVAH = Kongl. krigs vetenskaps akademiens handlingar.* 1805 ff. Sthm.
- Krig VAT = Kongl. krigs-vetenskaps akademiens tidskrift.* 1833 ff. Sthm.
- Källförteckning* [till SAOB]. [1898]. Lund.
- Källförteckning* [till SAOB]. [1939]. Lund.
- Källförteckning. Supplement* [till SAOB]. [1975]. Lund.
- Källförteckning. Supplement II* [till SAOB]. 1990. Lund.
- Lange, Sven. 1986. Databas för undersökning av svenskt skriftspråk perioden 1526 till 1799
I: *Datorstödd språkforskning. Fyra uppsatser om lånord, växtnamn, språk- och stilhistoria samt ordgeografi.* MINS 22. 25-32. Sthm.
- Lénström, Carl Julius. 1841. *Sveriges litteratur- och konsthistoria i utkast.* Ups.

- Lindroth, Sten. 1975. *Svensk lärdomshistoria. Frihetstiden*. Sthm.
- Lindroth, Sten. 1975. *Svensk lärdomshistoria. Gustavianska tiden*. Sthm.
- Lüdeke, Christoph Wilhelm. 1781-1796. *Allgemeines schwedisches Gelehrsamkeits-Archiv unter Gustafs des Dritten Regierung*. 1-17. Leipzig.
- Olson, Emil. 1913. *Studier över pronominet den i nysvenskan*. I: Lunds universitets årsskrift. NF. Avd. 1. Band 9. Nr 3. Lund.
- OxBr. = *Rikskansleren Axel Oxenstiernas skrifter och brevexling*. Afd. 2: 1 ff. 1888 ff. Sthm.
- PH = *Utdrag utur alle ifrån .. 1718 utkomne publique handlingar*. 1-15. 1742-1829. Sthm.
- Rietz, Rolf E. du. 1977 ff. *Swedish Imprints 1731-1833 [SWIM]. A Retrospective National Bibliography ..* Upps.
- Rydstedt, Rudolf. 1988. Creating a Lexical Database from a Dictionary. I: *Studies in Computer-Aided Lexicology*. 228-267. Gbg.
- SAOB = *Ordbok öfver svenska språket*. 1898 ff. Lund.
- SAOL = *Svenska Akademiens ordlista över svenska språket*. 11 uppl. 1986. Sthm.
- Starfelt, Gulli. 1975. Något om citatkontrollen vid Svenska akademiens ordboksredaktion. I: *Festskrift till Sven Ekbo*. 122-128. Sthm.
- Svensk Bibliografi 1700-1829*. [Databas upprättad av Sune Lindqvist, Gunilla Jonsson och Ylva Tjeder. Kungliga Biblioteket, Sthm.]
- Svensson, Lars. 1988. En bibliografi över SAOB.s källmaterial på persondator. I: *Sagt och skrivet. Festskrift till David Kornhall*. 155-157. Lund.
- Svensson, Lars. 1985. *Förteckning över bibliografier rörande 1700-talets bokproduktion. Ett urval*. Nordlund. 8 [stencilupplaga]. Lund.
- Svensson, Lars. 1985. Om 1700-talets bokproduktion. I: Ulf Teleman, red. *Det offentliga språkbruket och dess villkor i Sverige under 1700-talet*. Nordlund 7 [stencilupplaga]. 61-78. Lund.
- Terner, Erik. 1922. *Studier över räkneordet en och dess .. användning .. i nysvenskan*. Upps.
- Warmholtz, Carl Gustaf. 1782-1817. *Bibliotheca historica Sueo-gothica; eller förteckning uppå så väl trykte, som handskrifne böcker, ..* 1-15. Sthm och Ups.
- Vifot, Eva. 1975. Hur ordboksartiklar förbereds. I: *Kring en ordbok. Festskrift till Sven Ekbo*. 195-202. Sthm.
- Wijk-Andersson, Elsie. 1991. *Bara i fokus. En semantisk-syntaktisk studie av bara och dess ekvivalenter i nysvenskt skriftspråk*. Upps.
- ÅrsbVetA = *Årsberättelse om vetenskapernas framsteg, afgifne af kongl. vetenskaps-- akademien embetsmän*. 1822-59. Sthm.

En ordbog og dens informanter

På Orðabók Háskólans i Reykjavík, Islands leksikografiske institut, har man i næsten femti år indsamlet materiale til en historisk ordbog¹. Jeg vil fortælle om ét led i dette indsamlingsarbejde, det vil sige hvordan Ordbogen har draget nytte af den almene sproglige interesse i landet ved at rette spørgsmål til folk angående ord der ikke er fundet i skriftlige kilder eller der mangler nærmere oplysninger om. Jeg vil dog først give et kort overblik over den virksomhed der er foregået på Ordbogen i de forløbne årtier.

Planer om en historisk islandsk ordbog blev først fremsat i begyndelsen af dette århundrede, og da ville man stile mod en ordbog som dækkede islandsk sprog fra de ældste kilder og op til sidste århundredskifte. Man startede arbejdet, men på grund af svigtende bevillinger måtte man snart give op. Idéen blev dog holdt i live, og i begyndelsen af firtiårene gik man på ny i gang med arbejdet med en historisk ordbog, som ifølge nye planer skulle omfatte perioden fra 1540 til nutiden. Sproghistorisk set er den nye tidsgrænse bagud tilfældig, men 1540 er årstallet for udgivelsen af den første bevarede trykte bog på islandsk, Det ny testamente, og en forklaring på at man nu satte denne grænse var den at man da i København var begyndt at forberede et andet historisk ordbogsværk, som skulle omfatte norrøn prosa, og heri indgik islandsk op til 1540.

I de forløbne år har indsamlingsarbejdet bestået i excerpering af skrifter fra denne periode, især trykte bøger, og man kan sige at man allerede har excerperet de vigtigste trykte kilder op til slutningen af forrige århundrede foruden et betydeligt materiale fra det tyvende århundrede.

Imidlertid forholder det sig sådan, at de trykte islandske bøger fra det 16., 17. og op i det 18. århundrede først og fremmest indeholder kirkelig og anden opbyggelig litteratur, og derfor har det været nødvendigt at foretage en betydelig excerpering på grundlag af et bredere tekstudvalg, det vil sige fra håndskrifter, for at udfylde lakuner. Blandt andet har man søgt at inkorporere alle håndskrevne ordbøger og ordsamlinger. Samlingen fra trykte kilder og håndskrifter består nu af over 2 millioner ordbogssedler, og i de senere år har man arbejdet med at indtaste dette materiale, således at det bliver maskinelt tilgængeligt.

Men en historisk ordbog må give et så fyldigt og detaljeret billede som muligt af sproget i den periode den omfatter, og skriftsproget alene dækker aldrig det levende sprog til fulde. Til trods for omfattende bogproduktion inden for mange forskellige emner i løbet af de sidste to århundreder, dækker disse kilder aldrig hele det ordforråd som er eller har været forbundet med forskellige arbejdsmetoder og kulturelle foreteelser. Mange slags udtryksmåder bliver således sjældent eller aldrig skriftfæstet, i hvert fald i trykte bøger, og et stort antal ord og ordforbindelser, særlige betydninger og talemåder, som hører hjemme i en udførlig

¹ Senere omtalt som Ordbogen.

historisk ordbog, er således i fare for at gå tabt. Derfor er det nødvendigt for leksikografer også at søge materiale i det talte sprog, hvis de skal være i stand til at foretage en historisk beskrivelse af ordforrådet, sprogets udvikling og betydningsændringer.

Det er en lang islandsk tradition for at interessere sig for talesproget og optegne dialektale ejendommeligheder. I begyndelsen af det 18. århundrede indsamlede håndskriftsamleren og forskeren Árni Magnússon fra forskellige landsdele ord og talemåder, som var ham selv fremmede og som han mente var karakteristiske for de pågældende landsdele. Hans amanuensis, Jón Ólafsson fra Grunnavík, som gennem et langt liv arbejdede med et stort anlagt ordbogsværk, optegnede meget fra sin samtids talesprog og bestræbte sig på at oplyse om kilderne til de optegnede ord og ordforbindelser.

Den danske sprogforsker Rasmus Kristian Rask foretog en rejse i det østlige Island tidligt i det 19. århundrede og optegnede på sin rejse ord som han anså for kendetegnende for den østlandske dialekt.

Den islandske filolog Björn M. Ólsen, som senere blev professor og den første rektor ved Islands universitet, viste stor interesse for talesproget og foretog optegningsrejser i landet i årene 1884-93, det meste af tiden med støtte fra Carlsbergfondet.

Man kan sige at det første forsøg på en organiseret indsamling fra talesproget begyndte i 1917 da forfatteren Þórbergur Þórðarson fik en bevilling fra Altinget til indsamlingsarbejde. Han fik årlige bevillinger indtil 1924, da de blev sløjfet og dette arbejde ophørte. Þórbergur lagde stor vægt på at få kontakt med meddelere så mange steder som muligt over hele landet og opbygge et samarbejde mellem frivillige. Hans formål var en systematisk indsamling i hver syssel af ordforråd der var knyttet til bestemte arbejdsområder og -metoder fremfor tilfældige optegnelser af enkelte ord. I en periode forsøgte han også at udvide indsamlingsarbejdet gennem en fast avisspalte, og det lader til at en del folk har reageret herpå. Da bevillingerne til ham ophørte, opgav han at have omkring 300 informanter.

Alle de samlinger som jeg her har omtalt er nu helt eller delvis i Ordbogens varetægt. Dog ser det ud til at noget af Þórbergurs materiale er gået tabt. På Ordbogen opbevares endvidere en del mindre samlinger som jeg ikke har nævnt. Dette materiale er kun delvis blevet behandlet endnu, men det er tilgængeligt for alle som vil undersøge ordforrådets historie eller dialektale udtryk i enkelte landsdele. Samtidigt er disse kilder et uvurderligt bidrag til det indsamlingsarbejde som foregår på Ordbogen. Uden dem ville en del udtryk som kendes fra skriftsproget være svære at forklare, og en del oplysninger som disse samlinger har bevaret ville ellers være gået tabt.

Ordbogens stab har uden tvivl fra starten gjort sig klart hvor vigtigt det var at indsamle materiale fra talesproget, men det første årti har man dog ikke taget noget initiativ på dette felt, nok især som følge af hvor fåtallige man var og hvor meget andet arbejde der skulle varetages.

Midt i halvtredserne blev Ordbogens redaktører opfordret til at tage sig af regelmæssige radioudsendelser om islandsk sprog, indslag som radioen flere år forud havde haft med forskellige løse medarbejdere. Der var her tale om oplysende programmer med vejledning med hensyn til sprogbrug i skrift og tale. Under ordbogsfolkenes redaktion udviklede disse programmer sig efterhånden til at blive en blandning af oplysning og indsamling. Redaktørerne havde da siddet med excerperingsarbejde i flere år og kunne formidle mange interessante oplysninger til lytterne, blandt andet om ord og udtryk som tilsyneladende var begrænset til bestemte landsdele eller mindre områder. Samtidig kunne de drage nytte af lytternes reaktioner ved deres indsamlingsarbejde, idet lytterne begyndte at formidle oplysninger til Ordbogen. Disse udsendelser blev meget aflyttet, og Ordbogen opnåede hurtigt kontakt med mange gode informanter, og nogle af disse har indsendt

oplysninger i årtier. Enkelte af de første informanter havde i sin tid bidraget til Þórbergur Þórðarsons indsamlingsarbejde.

Ordbogsfolkene tog ikke bare imod svar på deres forespørgsler. De meddelte lytterne oplysninger som de modtog, i det omfang svarene gav anledning dertil, og det har man holdt fast ved siden. Og lytterne svarer ikke bare spørgsmål. De stiller ofte selv spørgsmål om ord og udtryk, om oprindelse, brug og betydning, og i det omfang tiden tillader det benyttes radioudsendelserne også til at give svar på sådanne spørgsmål på grundlag af de oplysninger man allerede sidder inde med. I andre tilfælde må spørgsmålet sendes videre til andre lyttere i håb om reaktioner fra dem. Selv om Ordbogens hovedformål med udsendelserne naturligvis er at indsamle materiale, er det på denne måde lykkedes at opholde en tosidet kommunikation med folk fra forskellige samfundslag rundt omkring i landet. Det er nu tale om et nyt indslag hver uge i vintermånederne på program 1, og hver udsendelse gentages et par dage senere. Længden har været noget varierende, som regel tyve minutter, men i nogle perioder ti eller femten minutter. I den første tid modtog Ordbogen næsten altid lytternes svar i skriftlig form, men efterhånden som telefonsystemet blev udbygget og det blev nemmere og billigere at telefonere fra landet, blev det mere og mere almindeligt at lytterne i fortsættelse af en udsendelse ringede den pågældende ordbogsredaktør op - enten på Ordbogen eller privat. Selv om telefonsamtalerne unægtelig tager tid - bl.a fordi nogle informanter også trænger til en personlig sludder - så er telefonsamtalerne i visse henseender mere givende end brevene, idet de giver lejlighed til at spørge nærmere om de behandlede ords betydning og samtidig eventuelt om andre ord inden for det samme område. Man kan også benytte lejligheden til at spørge om ord som man har modtaget oplysninger om fra andre personer i nabolaget.

Foruden svar på direkte forespørgsler har flere informanter sendt Ordbogen kortere eller længere lister over sjældne ord som de har hørt og noteret hos sig, måske gennem en længere periode, også lister over ord som de har stødt på under læsning og som har været dem fremmede. Dagbøger og andre skrevne kilder har vi også modtaget til excerpering. Alt dette materiale bliver gennemgået på Ordbogen, og vi får på denne måde tit oplysninger om ord som i forvejen ikke eller kun i ringe grad er belagt i vores seddelsamlinger. Disse oplysninger kan i visse tilfælde give anledning til omtale i radioudsendelserne, og ofte har det vist sig at ord eller udtryk, som man tidligere havde troet var meget isolerede, faktisk hørte til det levende sprog i et større område.

De oplysninger som Ordbogen modtager fra lytterne bliver omhyggeligt registreret på sedler. Foruden selve opslagsordet anføres informantens definition af ordets betydning så præcis som muligt. På sedlen noteres informantens navn og bopæl og hans eller hendes kilde. Til slut oplyses om informationen er givet per brev eller i en samtale. Alle indsendte breve arkiveres, og de udgør efterhånden en stor samling.

Sedler som indeholder oplysninger fra det talte sprog holdes for sig på Ordbogen. Det vil sige at de ikke ordnes ind imellem sedler der bygger på excerpering af skriftlige kilder, idet disse to samlinger er væsensforskellige. Eksemplerne fra de skriftlige kilder viser ordenes brug på en bestemt tid i en bestemt kontekst, medens eksemplerne fra det talte sprog snarere ledsages af informanternes oplysninger om betydning og udbredelse. I Ordbogens talesprogs-samling er der nu over 190 tusinde sedler, men man har ikke optalt hvor mange opslagsord de dækker.

I løbet af af de tre-fire årtier som Ordbogens redaktører har stået for disse sprogindslag har Ordbogen, som før nævnt, fået kontakt med et stort antal informanter i alle aldersgrupper. For fem år siden lavede man en statistik over Ordbogens informanter før og nu, som indeholdt alle de oplysninger om vedkommende, som vi var i besiddelse af. Antallet viste sig til at være næsten 2400, heraf knap 1500 mandlige, og den geografiske spredning var meget god. Nogle

af disse informanter har regelmæssigt eller med mellemrum formidlet Ordbogen oplysninger gennem det meste af denne periode, andre har optrådt mere sporadisk og en del er naturligvis i tidens løb afgået ved døden. Hvert år dukker der imidlertid nye informanter op, således at det samlede antal er vokset.

Måske har det spørgsmål meldt sig om oplysninger indhentet på denne måde er pålidelige. Det kan man i de fleste tilfælde svare ja til. Den nære kontakt, som der normalt er mellem informanterne og Ordbogen, gør at de ordbogsredaktører som tager sig af udsendelserne og modtager svarene lærer de enkelte informanter personligt at kende og er derved i stand til at evaluere de oplysninger de afgiver. Ved en gennemgang af talesprogs-samlingen for nogle år siden viste det sig, som det var at forvente, at den indeholdt en del hapax legomena som krævede nærmere undersøgelse. Man spurgte i radioen om hovedparten af disse ord, men kun i få tilfælde modtog vi en bekræftelse af deres eksistens. Man kan således regne med at nogle af disse ord er opstået ved at informanten har hørt forkert eller misforstået. Men dette er et problem som ikke er specielt for disse oplysninger, idet tilsvarende problemer opstår når skriftlige kilder excerperes, og sådanne isolerede former, som ikke får nogen støtte, bliver lagt til side.

Som det er fremgået af det der allerede er sagt, består udsendelserne for det meste af spørgsmål og svar angående enkelte ord og ordforbindelser. Spørgsmålene giver lytterne imidlertid anledning til også at nævne andre ord inden for det samme område, eksempelvis husdyrfarver, høhøst og forskellige slags redskaber, således at betydelige oplysninger foreligger nu angående enkelte fænomener. Det er lykkedes at indsamle et betydeligt ordforråd der er forbundet med dele af dagliglivet som der ikke er skrevet meget om. Dette gælder for eksempel vejrforhold, forskellig slags håndarbejde som dels blev udført af mænd og dels af kvinder, tøj, madlavning og det daglige arbejde i det hele taget både til sø og til lands. Det drejer sig her naturligvis ikke mindst om ordforråd der er forbundet med aflagte arbejds-metoder, og radioudsendelserne har bidraget væsentlig til at en betydelig del af dette ordforråd ikke er gået i glemmebogen.

Det er dog ikke gennem radioen foretaget en systematisk indsamling af ordforråd der er forbundet med bestemte arbejdsområder eller -metoder, og det materiale der er blevet indsamlet er ikke blevet klassificeret emnemæssigt. De enkelte redaktører har ganske vist ofte i en periode valgt først og fremmest at behandle specielle områder og har høstet meget materiale, men lytternes reaktioner har været meget forskellige af omfang. Ændrede arbejdsmetoder medfører at ordforråd, der har været forbundet med det gamle, hurtigt bliver fjernet og kun huskes af få, og i de tilfælde, hvor det drejer sig om arbejde der kun er blevet udført af et fåtal mennesker, er lytternes reaktioner ofte minimale. Derfor er det vigtigt at sammenflette det almene og det mere specifikke i udsendelserne. Så snart man kommer ind på de mere almindelige områder, får vi som regel mange svar og der dukker gerne nye hjemmelmænd og -kvinder op, som tidligere havde ment at de ikke havde så meget at formidle, men som kan vise sig at være meget vigtige informanter inden for forskelligartede emner.

Jeg har nu givet en kort oversigt over tidligere indsamling fra det talte sprog på Island og beskrevet den metode som Islands leksikografiske institut har benyttet sig af for at forøge sine samlinger. Udsendelserne har haft en trofast lytterskare, og aflytningen og interessen forekommer ikke at være blevet væsentlig formindsket til trods for at Statsradiofonis program 1, som i starten var landets eneste ærtermedie, efterhånden har fået konkurrence af flere radioprogrammer og to fjernsynskanaler. Men hvilket udbytte har Ordbogen selv haft af sine udsendelser i årenes løb?

Der har været en betydelig forbindelse mellem excerpering fra skriftlige kilder og indsamling fra det talte sprog. Ved excerpering fra bøger dukker der ord og ordforbindelser op, hvis nøjagtige betydning ikke fremgår af konteksten og som man ikke i forvejen har belæg på. Man kan slå forgæves op i eksisterende ordbøger, og da er det oplagt at søge hjælp hos Ordbogens informanter. I mange tilfælde er det at den vej lykkedes at indhente gode beskrivelser af vedkommende udtryks betydning og anvendelse.

Talesprogssamlingen er i høj grad blevet brugt ved indsamling af materiale til forskellige undersøgelser inden for etymologi, leksikografi, dialektgeografi og morfologi. Det har vist sig at samlingen rummer betydelige oplysninger om udtryk som ikke kendes fra trykte bøger. Den indeholder mange eksempler på talemåder og ordsprog som ikke findes i de bøger der er blevet skrevet om disse emner. Der kan man også finde forskellige bøjningsformer og udtaleformer som i visse tilfælde er lokalt begrænsede og kan bidrage til at kaste lys over den sproglige udvikling i bestemte dialekter.

I ældre ordbøger findes der tit oplysninger om at bestemte ord og ordforbindelser er lokale, men ofte har forespørgsler til lytterne afsløret at de er kendt over et langt større område, i visse tilfælde over hele landet. Forklaringen er at tidligere ordbogsforfattere har holdt sig til de begrænsede oplysninger de havde, og de kunne stamme fra én bestemt egn. De havde ingen mulighed for at undersøge den faktiske udbredelse på den måde som vi nu kan.

Der findes eksempler på at udtryk i talesprogssamlingen har kunnet bidrage til en sikrere forståelse af tvivlsomme steder i det gamle sprog, og det viser sig også nogle gange at ord, som man kun havde belæg på i middelaldersproget, stadig hører til det levende ordforråd visse steder i landet.

Af det foregående skulle det være klart at indsamling af materiale fra det talte sprog må gå hånd i hånd med anden indsamling til en historisk ordbog. En historisk ordbogs seddelsamlinger tjener ikke bare det formål at være ordbogsmateriale, idet de også undersøger en uundværlig kilde ved de fleste undersøgelser af sproget og dets udvikling. Det institut, som har til formål at udarbejde en historisk ordbog, er således samtidig et alsidigt leksikografisk institut.

Det er nødvendigt at gøre talesprogssamlingen maskinelt tilgængeligt i en database som ville gøre den lettere at anvende til forskelligartede formål. Man burde således for eksempel anføre nøjagtigt den geografiske placering, således at man kunne kalde frem ord fra bestemte landsdele. Man burde endvidere markere talemåder, ordsprog, bøjningsformer og udtaleejendommeligheder. Det ville også være ønskeligt at klassificere materialet, således at man kunne kalde frem ordforråd inden for bestemte områder såsom folketro, folkekultur, arbejdsmetoder og værktøj.

Det står klart at radioprogrammerne alene ikke er tilstrækkelige til at opsamle alt det ordforråd som efterhånden går tabt. Det ville være nødvendigt at indsamle systematisk gennem udarbejdelse og udsendelse af ordlister og ved besøg hos informanterne, men en sådan indsamling kræver grundig forberedelse, og den vil være tidskrævende og bekostelig. I den forbindelse skal det dog nævnes, at Nationalmuseets afdeling for folkekultur i de sidste årtier har gennem spørgelister indsamlet stort materiale på forskellige områder, og det materiale gøres nu efterhånden maskinelt tilgængeligt og dermed lettere anvendeligt for Ordbogen til excerpering; men denne indsamling er naturligvis ikke primært foretaget med et leksikografisk sigte. Vores radioprogrammer har til trods for deres begrænsninger gjort stor nytte, men denne metode er muligvis bedst egnet i et lille og forholdsvis homogent sprogsamfund. Metoden er også blevet anvendt på Færøerne, hvor den også har vundet popularitet. Og den kunne måske også tilskynde andre ordbogsredaktører til et øget samarbejde med almindelige sprogbrugere.

Helle Degnbol, Guðrún Ása Grímsdóttir, Bent Chr. Jacobsen, Jette Knudsen, Eva Rode & Christopher Sanders

En ordbog er en database

Under produktionen af *Ordbog over det norrøne prosasprog: Registre* (første bind af Den arnamagnæanske kommissions ordbog i 12 bind, udkommet i slutningen af 1989) lagdes alle oplysninger om ordbogens kildemateriale (håndskrifter, udgaver, norrøne og udenlandske) i database. Fra databaserne udvalgte redaktionen, hvilke data pr. automatik skulle bringes i registerbindet og under hvilken præsentationsform (typografi, layout). Det traditionelle stadium i en bogs tilblivelsesproces med udfærdigelse af "tekstbehandlet" manuskript og efterfølgende korrekturlæsning kunne således overspringes, og arbejdet kunne koncentreres om bogens indhold.

For halvandet år siden udkom det første af tolv ordbogsbind: *Ordbog over det norrøne prosasprog: Registre / A Dictionary of Old Norse Prose: Indices*. Bindet præsenterer ordbogens korpus, det norrøne (dvs. det middelalderlige islandsk-norske) prosasprog, i form af et signaturregister med oversigt over benyttede udgaver og håndskrifter (selve "nøglen"), et register over benyttede middelalderhåndskrifter med dateringer og litteraturhenvisninger, samt et register over benyttet litteratur, første del af ordbogens akkumulerende bibliografi.

For en kort beskrivelse af ordbogsprojektet henvises til *Lexikonord: Leksikografi i Norden*, Nordisk språksekretariats rapporter 14 (Oslo 1991:67-71).

Specielt om baggrunden for registerbindet og dets produktion kan man læse i The Arnamagnæan Institute & Dictionary: *Bulletin 17: 1988-1989* (København 1990:23-27).

Prøver på ordbogsartikler ses i *Ordbog over det norrøne prosasprog: Prøvehæfte / A Dictionary of Old Norse Prose: Prospectus*, Kbh. 1983.

Al virksomhed er nu koncentreret om produktion af artikler til det næste ordbogsbind. I det følgende præsenteres glimtvis hvordan vi rent teknisk arbejder med at lave "egentlig" ordbog: i princippet som ved udarbejdelsen af registerbindet, dvs. i database fra første færd og så nær hen til den endelige bogproduktion som muligt, i et system, vi selv har opbygget og til stadighed udvikler.

Database

Hvad vi forstår ved en database, er en samling oplysninger indlagt i en fast struktur. Ordbogens database er emnemæssigt opdelt i enheder kaldet filer.

Centralt står ord-/hovedfilen med lemma, ordtypebestemmelse (til udskillelse af fx. person- og stednavne og poetisk ordforråd), grammatisk bestemmelse og oplysning om ordenes bøjning og formvariation (af denne fil dannes *artikelens hoved*), citatfilen og definitionsfilen (af hvilke dannes *artikelens krop*).

Perifert står filer med komposita-forled og med oplysninger om ordenes behandling i anden litteratur, herunder ordbøger (af hvilke filer dannes *artikelens hale*, afsnittene Comp., Litt. og Gloss.).

Ligeledes perifert står en fil med oplysning om de til grund for signaturerne liggende udgaver og håndskrifter, en værkfil, samt en håndskriftfil med oplysning om datering, proveniens mm. (Disse filer danner grundstammen i *registerbindet*.)

Endelig hører der til hele systemet en bibliografisk fil (efter hvilken den akkumulerende *bibliografi* trykkes).

Hver af disse filer er opdelt i mange, mindre felter, hvert felt rummende sin information. Således har alle oplysninger, store som små, deres plads i en ordnet struktur. Med sin konsistente arkitektur er systemet fleksibelt og giver mulighed for datavalidering på højt niveau, fx. kontrol af signaturer; udnyttelse af allerede lagrede informationer, fx. datering af citatmaterialet; beskuelse af materialet under forskellige synsvinkler; rettelser og omflytninger; vilkårligt bestemte udskrifter; automatisk registrering af hvert ords status i redaktionsprocessen. Alt bestemmes og udføres direkte af redaktørerne.

Citatmateriale

Kernen i ordbogsartikelen og dens hele udgangspunkt er citaterne, og den mest fundamentale enhed i databasen er den fil, der indeholder ordbogens citatmateriale.

ONP's seddelsamling rummer mange forskellige citattyper: enkle citater med kun én tekst (hentet enten i en udgave eller direkte i et håndskrift); citater med en eller flere varianter, repræsenterende forskellige versioner af samme tekst; citater med udenlandsk parallel; citater med udgivers rettelse; citater med udgivers kommentar; citater med ordbogsredaktørs kommentar, evt. i form af norrøn hjælpetekst eller i form af oversættelse til dansk/engelsk. Ofte er disse citattyper kombineret, og der er derfor i databasen brug for mange felter svarende til alle de forskellige lag og typer af oplysninger. Citatfilen har således henvend 100 felter.

Databasen hjælper på forskellig vis til rationel behandling af materialet, og der skal her gives et enkelt eksempel (et mindre pilotprojekt): En type sedler, der traditionelt vil koste en del arbejde, er dem, der rummer citater afskrevet direkte fra håndskrifterne, idet de pågældende tekster endnu ikke er udgivet. Ordsedlerne indtastes, som de optræder i skufferne, dvs. med citater fra mange forskellige håndskrifter i tilfældig rækkefølge. Herpå sorteres de efter håndskrift og håndskriftside. Man har nu et praktisk overblik over materialet og kan sammenligne citat og originaltekst håndskrift for håndskrift og side for side, man kan nemt indsætte linjenummer, som er det der oftest mangler, kontrollere ortografien, tilføje mere kontekst osv. På denne måde er det kun nødvendigt at konsultere det samme håndskrift én gang.

På samme måde kan databasen sikre rationel korrekturlæsning af ordbogens mange citater imod udgaverne.

Når redaktøren modtager citaterne med henblik på redigering, er de blevet indtastet af assistenter. I samme proces er signaturerne kontrolleret og revideret og visse systematiske fejl påvist og udryddet.

For såkaldt mindre ords vedkommende indtastes alle ordsedler. Inden for bogstavet a er der ca. 40.000 citater fordelt på ca. 4.000 ord; 83 % af ordene er belagt med mindre end 9 eksempler, og hele 96,5 % af ordene er på under 50 belæg. De indtastes rub og stub. Inden for de såkaldt større ord, de 3,5 % af ordene, der er belagt med mere end 50 (enkelte med

over 1000) eksempler, sorteres der fra inden indtastning og redigering. Selv om det er ærgerligt at måtte undvære en del af citaterne i databasen, ses det klart, at frasorteringen betaler sig, da disse få større ord udgør hele 55 % af seddelmaterialet.

Citaterne kan redaktøren få i hænde (dvs. op på sin skærm eller printet ud i et hvilket som helst ønsket format, fx. seddelformat) sorteret efter eget valg: altid naturligvis efter ord (lemma), men dernæst fx. alfabetisk efter navnene på de værker, der er citeret (det kan være praktisk ved meget store ord) eller kronologisk efter de tilgrundliggende håndskrifers alder. Håndskriftdatering er automatisk indført i filen og påført udskriften, og ordets ældste citat er automatisk markeret, med et mærke, der følger citatet hele vejen ud i den trykte ordbog. Det vil altid være værdifuldt for redaktøren at få dette kronologiske overblik over ordet. Det kan også være frugtbart ved databasemæssig sortering at få samlet alle citater, til hvilke der er udenlandske paralleller, ligesom citaterne kan vælges sorteret efter værkgenre.

Redigering

Redigeringen er den mest udfordrende del af ordbogsproduktionen, både i sig selv og i forhold til databasesystemet.

Redigeringsarbejdet i alle dets faser foregår (eller kan foregå) i database. Mest aktivt redigeres der på de centrale filer: ord-/hoved-, definitions- og citatfilerne, som tilsammen skal udgøre artikelens hoved og krop.

Redaktøren kan i princippet begynde redigeringen hvorsomhelst i systemet, men han begynder gerne i *citatfilen*, ligesom man ved traditionel redigering oftest tager fat på citatmaterialet først og sorterer sedler. Ved hjælp af særlige redigeringsfelter foretager redaktøren den første analyse af citaterne og tager for hvert citats vedkommende stilling til ordets/citatets betydning og brug, indsætter evt. kommentar til eller særoversættelse i citatet, ser på ordets syntaktiske stilling og eventuelle indgåen i en ordforbindelse af løsere eller fastere karakter, samt danner sig et overblik over norrøne varianter og udenlandske paralleller.

Det er på dette stadium naturligt at formulere definitioner og beslutte sig for artikelens inddeling og opstilling, kort sagt lægge data i *definitionsfilen*. I citatfilens særlige indekseringsfelter markeres det, hvilke citater hører til hvilken betydningsafdeling (evt. syntaktisk afdeling eller afdeling med ordforbindelse), hvilke belæg skal citeres og i hvilken rækkefølge, hvilke blot bringes som 'nøgen henvisning' (signatur uden citat), og hvilke ikke skal med overhovedet. I det integrerede databasesystem ruller definitioner og citater koordineret hen over skærmen. Indekseringsfelterne gør det let og smidigt at ændre i artikelen, idet den så at sige sorterer sig selv. Fx. vil et citat ved ændring af dets betydningsnummer automatisk flytte plads, og det er tilsvarende nemt at ordne citaternes rækkefølge inden for de forskellige afdelinger. I hver afdeling følger efter citaterne et udvalg af nøgne henvisninger i kronologisk orden.

I databasesystemet får man et klart overblik over ordets ortografiske variation og dets bøjningsformer og kan dermed færdigudfylde *ord-/hovedfilen*.

Når centralfilerne er færdigbearbejdet, er artikelens hoved og krop formet. De afsnit, der udgør artikelens hale med stof fra både de centrale og de perifære filer, genereres næsten automatisk, dvs. med meget lidt manuel indblanding fra redaktørhold. Latinske, franske, tyske, engelske og danske paralleller indhentes fra citatfilen, sammensætningsforled fra komposita-filen og ordbøger/glossarer, der behandler det pågældende ord, fra glossarfilen. Hermed er artikelen klar til at forlade førsteredaktørens bord.

Redaktionelt samarbejde

Databasen byder på særlige muligheder, når flere redaktører vil forsøge sig med redigering af det samme intakte citatmateriale. Når førsteredaktøren af en artikel har lavet sit udkast, gennemgås det grundigt af andenredaktøren. I databasen kan andenredaktøren komme med egne forslag til beskrivelse af ordet og til strukturering og anden tilskæring af artikelen, og han kan med få tryk på tastaturet ved hjælp af egne indekseringsfelter afprøve sine ideer. Oftest, i de mange mindre artikler, viser han dog sine forslag blot ved at gøre antegninger i en arbejdsudskrift.

Arbejdsudskriften kan bestå ene og alene af artikelens skelet, dvs. dens afdelinger og definitioner; eller den kan have den længst mulige form, vise alle citater, også dem, der foreslås bragt som nøgne henvisninger, og dem, der foreslås udeladt, samt alle interne kommentarer; eller udskriften kan være udformet præcis som en artikel i den trykte bog.

Samme mulighed som andenredaktøren har den engelske redaktør, der gennemser alle artikler, først og fremmest med henblik på de danske/engelske definitioner. Den engelske redaktør kan vælge at arbejde i et særligt skærmformat, hvor engelsk tekst praktisk kan indsættes og kommentarer tilføjes. Undertiden gives der danske/engelske kommentarer eller oversættelser nede i selve citaterne, også til det formål er der udviklet skærmformater, og et lille program finder vej til de få steder, det drejer sig om.

Fra database til trykt bog

Takket være databasesystemet kan ordbogens medarbejdere indsamle og indskrive materiale, automatisk indhente oplysninger fra tilkoblede registre, analysere på materialet, redigere det, samle baggrundoplysninger, diskutere med hinanden, gøre interne notater, alt sammen på en særdeles praktisk måde.

Arbejdet foregår inden for et par simple business-databasesystemer: dBASE III og SmartWareII. Vi vil, på vej til fotosætningen, benytte tekstformatteringssystemet TEX, som også blev benyttet ved fremstillingen af registerbindet (interface: Postscript). Helt endelig beslutning om typografi behøver man ikke træffe før sent i forløbet. Hvis bare man har stoffet fornuftigt struktureret og fordelt på diskrete kategorier, kan alt lade sig gøre typografisk. Tegnproblemer og alfabetiseringsproblemer er løst. Systemet er helt åbent. Det kan endnu ikke siges at fungere smidigt, men det vil blive smidigere efterhånden, og snart har vi, hvis vi skulle ønske det, et godt grundlag for at kunne definere et evt. skræddersyet ordbogssystem. Det, der ligger fast, er, at vi hele tiden grundigt kender og kan overskue og håndtere både system og materiale, og at vi hele tiden arbejder henimod nøjagtigt det samme produkt. Som nævnt kan vi trække alskens arbejdsudskrifter ud af databasen undervejs, og vi kan definere et hvilket som helst skærmformat, men vigtigst af alt kan vi til sidst direkte ud af én og samme base trække det destillerede produkt: ordbogsartikelen og dermed ordbogen.

Som vi fik det afprøvet under arbejdet med registerbindet, hvor vi gik så direkte som overhovedet muligt fra database til trykt bog, indebærer en sådan bogproduktion indlysende praktiske fordele: ingen ventetider og mellemstationer, ingen traditionel tekstbehandling (det vil fx. sige ingen tegnsætning mellem de forskellige slags oplysninger) og ingen korrekturlæsning i traditionel forstand. Redaktørerne kan i stedet og indtil sidste øjeblik koncentrere sig om den vigtigste indholdskorrektur i selve databasen.

Vi har bevidst udskudt at arbejde med de allermost komplicerede artikeltyper, det gælder verberne med deres partikler og konstruktionseksempler. Men systemet er gjort klar til modtagelse og manipulering af også komplekse strukturer, og ikke mindst, når vi skal til

at håndtere disse mere tunge sager, har vi gode forhåbninger til databasesystemet både som opbevaringskasse og som redigeringsværktøj.

Når vi bliver spurgt, hvad vi vil stille op med de helt store ord, fx. de store præpositioner på over 1000 sedler, om vores databasesystem virkelig kan håndtere dem, så må vores svar være: I krig og kærlighed gælder alle kneb, og stillet over for særlig store leksikografiske problemer bør overhovedet ingen metode foragtes. Problemerne skal søges løst og artikelen præsenteres på den bedst mulige måde. Men grundtanken må fastholdes. Ikke nok med at databasesystemet er et fint stykke værktøj for mangen en videnskabsmand. En ordbog er en database, og i bund og grund betaler det sig at lade den være det fra først til sidst.

Litteratur

- Nordisk sprogsekretariats rapporter 14. 1991. *Lexikonord*. Oslo
Ordbog over det norrøne prosasprog. 1983. *Prøvehæfte*. København
Ordbog over det norrøne prosasprog. 1989. *Registre*. København
The Arnamagnæan Institute & Dictionary. 1990. *Bulletin 17*. København