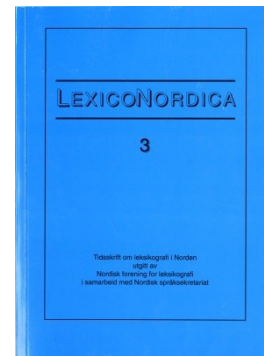


# LexicoNordica

Titel: Värdering av SAOB:s korpus och språkprovssamling  
Forfatter: Carl-Erik Lundbladh  
Kilde: LexicoNordica 3, 1996, s. 91-104  
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

## 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 LexicoNordica (1-16) 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.

*Carl-Erik Lundblad*

## **Värdering av SAOB:s korpus och språkprovssamling**

The Swedish Academy Dictionary bases its description of the Swedish language from 1521 till today on a collection of about 7.5 million slips containing a full sentence illustrating the use of a word or an expression. These are excerpts made from about 20,000 sources. Texts that have been systematically excerpted, thus contributing about three words per page, number about 11,000.

The collected material gives a good coverage of written standard Swedish of various styles. The quotations have mainly been made by selected and properly instructed persons. This is one reason for the reasonably high quality and usefulness of the material. The enormous amount of concerted work carried out on word selection and lemmatization is another. Concordances based on mainly 20th-century texts are used to fill in the material and to check the analysis based on the slips. Concordances of older texts are not yet available in great enough number to be of any importance.

Svenska Akademiens ordbok baseras på en korpus i bemärkelsen: textsamling sammanställd för visst ändamål. T.ex. Aitken (1978:30) använder också termen "korpus" så, men skulle lika gärna ha kunnat tala om sin ordboks källor. Det är inte den numera mest gångbara användningen av "korpus". De stora historiska ordböckerna bygger i allmänhet ordbeskrivningen på en korpus i denna äldre bemärkelse och inte i den moderna: textsamling för visst ändamål, där enskilda element kan (datamaskinellt) undersökas i hela samlingen. I denna uppsats har "korpus" mest den äldre betydelsen. SAOB:s korpus, dvs källor, och den därav gjorda materialsamlingen ska granskas.

SAOB:s källor kan med viss rätt men något överdrivet, vilket framgår nedan, uppges till ett antal av ca 20.000 från år 1521 och framåt (se Svensson 1991:372). Ur källorna har det excerperats ca 7,5 miljoner språkprov (se Hast 1983:160). Är detta material en tillräcklig bas vid ordboksredigeringen, och kan det brukas av andra än ordboksredaktionen? Genom att hänvisa till de redan utgivna 31 banden av SAOB vore det lätt att konstatera att materialet fungerar åtminstone någorlunda när man skriver en historisk ordbok över svenskan. Eftersom excerpten, utskrivna på lappar med stickord, kontext och källangivelse, ofta anlitas av forskare utanför redaktionen vore det lätt

att ge ett positivt svar även på frågan om materialet kan komma till användning av andra. Men svaren blir ointressanta utan precisering.

## Storlek och kvalitet

SAOB:s källor och material har utförligast beskrivits i Hast (1983: 159–198). Här ska endast erinras om vilka principer som gällt och gäller vid urval av källor och excerpt och därmed sätter sin prägel på ordbokens grundmaterial.

Korpusen har klar avgränsning bakåt i tiden. Ingen källa är äldre än från 1521. Åt andra hållet finns ingen gräns, utan vid redigerings-tillfället nyskriven litteratur kan användas. Härefter skiljer sig SAOB från den nederländska ordboken *Groot Woordenboek der Nederlandsche Taal* (WNT), som inte beaktar yngre källor än från 1921 (Sterkenburg 1984:86).

SAOB:s korpus är en expanderande sådan även i det avseendet att också äldre källor kan bli fler. Det kan alltså mycket väl tas språkprov t.ex. från en tidigare inte excerperad 1600-talskälla. Också i detta avviker SAOB numera från den nederländska ordboken, som har slutat att excerpera och alltså redan skulle kunna trycka en slutgiltig källförteckning.

Den grundläggande excerperingen gjordes under 1800-talets sista decennier, men excerperingen fortgår alltjämt. Att korpusen växer på detta sätt innebär naturligtvis att ord som hör till senare delen av alfabetet beskrivs med större material än andra. Tillväxten kan medföra den risk som ordbokschefen Esaias Tegnér d.y. hade i tankarna när han talade om att ordboken i stället för att närma sig slutet kom allt längre bort därifrån. Fullt så illa är det inte, men expansionen medför en ojämnhet i materialet och därmed en skillnad mellan de tidigare och senare banden. Allmänt sett är underlaget för beskrivningen magrare i de tidigare banden, och idag nya ord och ordanvändningar beskrivs bara i den mån de faller på den senare delen av alfabetet. Med tanke på ordbokens långa utgivningstid – efter drygt 100 år är 5/6 av ordförrådet avverkat – hade emellertid ett slutårstal för excerperingen varit ett sämre alternativ. Tack vare den kontinuerliga materialökningen har varje band skrivits med så fylligt material som möjligt. Senare betydelseutvecklingar beskrivs i sitt etymologiska sammanhang med äldre betydelser, och utvecklingen kan därmed förklaras tydligare. Det är inte otänkbart att sena utvecklingar ibland upptäcks eller att deras riktiga förklaring uppenbaras tack vare att sena språkprov behandlas till-

sammans med äldre. Att materialet är bättre i de senare delarna av alfabetet bör alltså ge ordboken ett extra värde.

Den kontinuerliga materialtillväxten har sannolikt inte spelat någon avgörande roll för ordbokens långsamma utgivning. Ordboken skulle ha kunnat skrivas betydligt snabbare trots att inget slutdatum för excerperingen har satts. Tidsutdräkten måste i första hand tillskrivas en självpåtagen och obefogad noggrannhetsiver och överdriven detalj-skräpa i redovisning av ordens ibland rent tillfälliga användningar. Dessutom beror förseningen på den ineffektivitet som med närmast naturlags förutsägbarhet drabbar vilken mänsklig organisation som helst där man nonchalerar riskerna och inte vidtar sådana motåtgärder som i konkurrensutsatta projekt lättare framstår som nödvändiga.

### **Korpusens täckning av svenskt skriftspråk**

Man har försökt att vid excerperingen täcka svenskt skriftspråk i olika genrer och ämnesområden. Svensson (1991:373) har kompletterat Hast (1983: 174) och gett en översikt av källornas fördelning på genrer. Svensson (1991:374) har vidare meddelat intryck i fråga om hur SAOB:s källor representerar svensk bokproduktion. Praktiskt taget allt 1500-talstryck har excerperats och det mesta av vikt från 1600-talet. Av den betydligt mer omfattande 1700-talslitteraturen har bara en del excerperats men med god spridning över genrer. Hur väl 1800- och 1900-talets svenska är representerad är svårt att ha en mening om, eftersom skrifterna från den perioden är många och därför inte har blivit lika bra översiktligt beskrivna som tidigare perioders vad avser relativ storlek och fördelning på genrer. Viss tillförsikt ges dock av den uppräknings i Svensson (1991: 376) av genrer eller ämnesområden utöver skönlitteratur som är väl representerade i SAOB:s material: biografi, botanik, ekonomi, etnografi, författning, förordning, geografi, historia, juridik, litteraturvetenskap, ordbok, periodica, samhällsvetenskap, sport, språkvetenskap, teknologi, teologi, tidningar, tidskrifter och undervisning.

Att exakt och med siffror ange hur väl svenskt skriftspråk är representerat i SAOB:s material är svårt dels av det skäl som redan nämnts, nämligen att vi inte har tillräcklig kunskap om de senaste seklernas svenska skriftspråk, dels därför att SAOB-källorna ofta har använts för bara något enstaka excerpt.

### **Antalet källor**

SAOB:s källförteckningar upptar 20.000 titlar. Siffran är imponerande och kan göras något högre. Man kan nämligen peka på det förhållandet att vissa källor omfattar flera volymer men har räknats såsom en enda, t.ex. Gustav I:s registratur, Publique Handlingar, Oxenstiernas skrifter och brevväxling. Men som mått på omfånget av SAOB:s korpus minskar siffran betydligt när man betänker att i källförteckningens supplement I resp. II alla citerade källor har medtagits, även om bara ett enda belägg skulle ha hämtats därur. Sådan komplettering med enstaka citat vid redigeringen är inte ovanlig. Enligt Hast (1983:173) torde bara 20 procent av källorna i Supplement I ha excerperats, dvs finnas representerade med åtminstone ett belägg per sida. Jfr Svensson (1991:372). Enligt stickprov som jag gjort i det senare supplementet (II) är det lika få källor där som har excerperats verket igenom. Fortsätter man beräkningarna i Hast (1983:174) och inkluderar källförteckningens supplement II torde antalet helt igenom excerperade källor inte uppgå till mer än ca 11.600.

Siffran 11.600 kan nog även den vara i överkant, eftersom ingen modifiering av antalet verk i den grundläggande källförteckningen har gjorts. Visserligen är det allra mesta där regelrätt excerperat, men undantag finns i form av källor som bara har citerats någon enstaka gång. Förteckningen upptar även viss språkvetenskaplig referenslitteratur som inte alls har excerperats, t.ex. Axel Koch: Umlaut und Brechung im Altschwedischen. Men sådana verk är inte särskilt många.

## **Språkprovssamlingen**

Den språkprovssamling på uppskattningsvis 7,5 miljoner lappar (ett belägg i kontext på varje) som nu är SAOB:s material har växt fram under ett drygt sekel. Under redaktionens första decennier, närmare bestämt från 1884 till 1903, var tillväxten som intensivast. Den avtog markant från 1930-talet. Från 1966 till 1979 har den genomsnittliga årliga tillväxten varit knappt 15.000 excerpt. Se Hast (1983:169, 172). En överslagsberäkning som jag gjort visar att tillväxten efter 1979 är ännu lägre.

Ur språkprovssamlingen hämtas de språkprov som redaktörerna går igenom och beaktar och efter utförd analys väljer exempel ur för att illustrera betydelsebeskrivningen och belägga uppgifter om böjning och stavning. Alla belägg anges med noggrann uppgift om källa.

Ordböcker som likt SAOB baserar sin beskrivning på en excerpt-samling torde vara de enda som kan uppge en minimisumma för antalet

beaktade språkprov. Men det används alltid ännu fler. De språkprov som excerpister utanför redaktionen bidrar med kompletteras nämligen, främst med material från andra lexikon, samlingen till Östergrens Nusvenska ordbok, ordförteckningar från Tekniska nomenklaturcentralen, excerpt som ordboken får från Svenska språknämnden samt ord hämtade från floror, riksdagsprotokoll och några andra verk med alfabetiska ordförteckningar. Allt ger den stomme med vars hjälp redigeringen görs. Därtill kompletteras vid behov med datorlagrade konkordanser på Institutionen för svenska språket vid Göteborgs universitet och en finlandssvensk textkorpus (FISC) på Institutionen för nordiska språk vid Helsingfors universitet. Den excerpering som anlätade och per excerpt avlönade personer utanför redaktionen utför och har utfört ger emellertid den viktigaste delen av ordbokens material.

## Excerperingen

Materialet som ordbeskrivningen ska grundas på är redan från första stund valt med tanke på vad det ska användas till. Hur pass väl man lyckats uppnå representativitet i urvalet av källor är som nämnts ovan svårt att ha en bestämd uppfattning om, men mängden representerade genrer ger ett gott intryck och enligt Hast (1983:187) skulle ca 12.000 verk vara ett stort urval i jämförelse med andra historiska ordböcker.

Excerperingen innebär som regel att 3 ord per sida har valts. Detta val har ibland lite slarvigt och felaktigt beskrivits i termer av slumpmässighet (jfr Allén 1979:128 och Ekbo & Loman 1965:84). I stället för att vara slumpmässig innebär excerperingen den första av många absolut nödvändiga selektioner som måste göras efter källurvalet.

De av redaktionen anlätade excerpisterna är språkintresserade, läsvana personer, ibland och i synnerhet under redaktionens första årtionden språkvetenskapliga experter. Förr instruerades excerpisterna mycket noggrant och resultatet av deras arbete kontrollerades ingående (jfr Hast 1983:164ff, där sådana anvisningar återges i en bilaga). Numera inskränks anvisningarna till två skrivna A4-sidor och en kortare muntlig genomgång.

Excerpisternas allmänna uppgift är att välja sådana språkprov som ordboksredaktionen kan tänkas ha nytta av. Variation kan förekomma efter typ av litteratur. Excerpering av facklitteratur t.ex. är ofta av det slaget att man vill ha belägg på typiska, ämnesspecifika men ändå i allmänspråket påträffade ord och uttryck. Sådan excerpering görs inte sällan av sakkunniga inom det ämnesområde som boken behandlar.

Facksakkunniga excerptister har uppmanats att på lappen om nödvändigt och möjligt ge en kort förklaring av ordets betydelse om den skulle vara speciell och inte direkt begriplig. I övrigt ska alla språkprov alltid innehålla så mycket kontext att ordets konstruktion och användning blir belyst. Sådana språkprov som ger en definition eller förklaring eller på annat sätt är ägnade att kasta ljus över betydelsen är redaktionen tacksam för.

Excerptister med uppgift att gå igenom äldre källor är ofta vana läsare av historiska texter. De bör ha god förmåga både att välja ord och uttryck som avviker från senare eller tidigare språkbruk och att välja det idiomatiska och typiska.

De allmänna anvisningar som excerptister av moderna texter numera får handlar om att finna belägg på nya ord eller nya användningar av ord och uttryck. Excerptisterna uppmanas att undvika dels tillfälliga och självklara sammansättningar, dels uppenbara felaktigheter. De uppmanas också att undvika mycket vanliga och till betydelsen helt klara ord såsom *dörr* och *näsa*, om de inte ingår i speciella uttryck (*ha skinn på näsan*). I äldre, mer detaljerade anvisningar uppmanades emellertid excerptisten att ta med ett eller annat belysande språkprov med sådant vanligt ord. Se vidare Hast (1983:187 ff).

En selektiv, traditionell excerptering har sina nackdelar, men fördelarna torde överväga. Benägenheten att förbigå det typiska och lägga vikt vid det speciella kan ge för stor uppmärksamhet åt perifera ordanvändningar och en beklaglig försummelse av de vanligaste orden. Såväl högfrekvent, vardagsspråklig användning av betydelseord som typiskt bruk av formord riskerar att förbigås. Men medvetenheten därom och anvisningarna till excerptisterna liksom övervakningen av deras arbete under den intensivaste excerpteringsperioden motsvarar stränga krav på excerptering (jfr Reichmann 1990:1603 ff) och kan förmodas ha gett SAOB ett förhållandevis gott grundmaterial. För modernt språk har redaktionen numera god möjlighet att fylla luckor i det excerpterade materialet genom att anlita datorlagrade konkordanser när misstanke har väckts eller som rutinkontroll. För äldre språk erbjuds ännu inte samma möjlighet. Alltför få äldre texter från den av SAOB beskrivna perioden föreligger nämligen i datorlagrad form.

Det är en stor fördel vid redigering av en historisk ordbok att grundmaterialet samlas genom traditionell excerptering, eftersom excerptisterna på sitt stadium kan delta i den nödvändiga texttolkningen. De kan nämligen när de utför arbetet ta hänsyn till kontexten, tillkomsttiden, författaren, ev. översättning, genren, ordens frekvens, lexiseringsgrad, fraseologi, betydelse, speciell användning eller idiomatik. De kan vidare avgöra ordets grundform, normalisera stavningen av

stickordet och ange ordklass. Visserligen är det möjligt att använda dator både för lemmatisering och för sådan excerpering som modifieras t.ex. med hänsyn till högfrekventa ord, varigenom risken att drunkna i en uppsjö av de allra vanligaste orden elimineras, men de texter som en historisk ordbok med stor korpus har att beakta bjuder oftast så rik och oförutsedd variation i ordens böjning och stavning att automatiken blir omöjlig. Det arbete som därför måste utföras efter en automatisk excerpering blir tidsödande, dyrt och ineffektivt i jämförelse med arbetet under och efter en traditionell excerpering. En excerptist, som t. ex. kan beakta såväl den omedelbara kontexten som periodens språk i allmänhet, kan utföra en väsentlig begränsning snabbare och smidigare i samband med excerperingen än vad redaktören kan göra vid en genomgång av ett ordmaterial som har plockats maskinellt från många källor och flera århundraden. När en manuell selektiv excerpering ombesörjs av så många excerptister som fallet är i SAOB (enligt Hast 1983:161 har 300–400 hittills varit i verksamhet) minimeras dessutom risken för att individuell preferens präglar materialet.

Den manuell-cerebrala excerperingen är den enda som kan användas för handskrivna källor. SAOB har haft många skickliga handskriftsläsare som excerptister, främst Ludvig Larsson (se Hast 1983:186).

Ett material måste begränsas på något sätt för att effektivt kunna utnyttjas vid ordboksredigeringen. Den per år bearbetningsbara delen av en ordbokskorpus är enligt Aitken (1978:31) ca 15.000 språkprov, om arbetet ska bli grundligt gjort. Med hänsyn till avbrott och tillfälliga förseningar räknar Aitken med i runda tal 10.000 språkprov per år. SAOB har 11.000 språkprov som årspensum för redaktörerna, den nederländska ordboken (WNT) ca 8.000 (jfr Sterkenburg 1984: 77).<sup>1</sup>

### **Komplettering av excerpterat material**

Som nämnts ovan använder redaktionen datorbaserade konkordanser av 1900-talstexter, mest romaner och tidningar, men även t.ex. några Strindbergstexter från 1800-talet. Som komplement till det excerpterade materialet är dessa konkordanser av stort värde, t.ex. när man vet eller misstänker att en ordanvändning förekommer fastän den saknas i excerptmaterialet. På så sätt kan man ofta lätt hitta ett belägg på t.ex. *en*

---

<sup>1</sup> Skillnaden mellan ordböckerna säger förmodligen ingenting om kraven på redaktörerna eftersom det t.ex. kan variera hur mycket materialet behöver kompletteras, kontrolleras i källan och hur noggrant beskrivningarna utförs.



*säker etta* i fråga om fotbollstips o.d. Konkordanserna är också smidiga när det gäller att få bekräftelse på eller stöd för en på excerperat material utförd betydelseanalys. Det är emellertid anslående att konkordanser över moderna texter förhållandevis sällan tillför det traditionellt excerperade SAOB-materialet något överraskande nytt och praktiskt taget aldrig uppvisar språkprov som tvingar till nya eller andra betydelsemoment än dem som baseras på språkprovslapparna. Det kan naturligtvis bero på redaktörens motvilja att ändra, men sannolikt är det traditionella materialet så grundligt att nytillkommet i stort sett bara kan bidra med tydligare exempel och sena belägg.

Det hade naturligtvis varit av stort värde med tillgång till även äldre texter i datorbaserade konkordanser, i synnerhet om de är lika lätthanterliga som eller ännu lätthanterligare än de som Institutionen för svenska språket i Göteborg tillhandahåller. De äldre texter som redan finns inlästa och sålunda kan fås i konkordansform är på grund av ringa antal och omfattning närmast försumbara vid ordboksredigeringen.

Lexikografiska institutionen vid Islands universitet har rapporterat om användning av en datorläsbar textbank för att fylla ut luckor i ett excerperat material (Pink et al. 1993:1–18). Något liknande vore önskvärt för SAOB. Ett antal av ordbokens excerperade källor från varje århundrade borde vara tillgängliga i datorläsbar form och som konkordanser. Förslagsvis borde 1 miljon ord från 1500-, 1600- resp. 1700-talet läsas eller skrivas in. Då skulle framförallt det excerperade materialets representativitet kunna kontrolleras, betydelseanalysen kunna bekräftas eller fördjupas och språkprovssamlingen kunna kompletteras med äldre eller bättre belägg. Om SAOB-projektet skulle ha börjat idag med förhållandevis lika mycket resurser som har använts hittills borde alla källor innan excerperingen sattes igång ha gjorts tillgängliga i en korpus i den moderna bemärkelsen. Att få denna korpus lemmatiserad hade naturligtvis gjort den oerhört värdefull. Det arbete som därvid krävs är egentligen detsamma som föregår normal ordboksredigering. Tillgång till konkordanser över ordförrådet i alla källor hade inneburit värdefulla möjligheter liksom detta att få uppgift om absoluta frekvenser och förekomst i vissa genrer o.d. samt att lätt få fram idiom och kollokationer genom att ordna materialet efter ordens högfrekventa samförekomster. Dessutom skulle hapax-ord och lexikonarv lättare kunna avslöjas.

### **Materialets tillgänglighet**

Så länge en ordboks språkprovssamling inte finns lagrad som datortext i sökbar form är förmodligen de traditionella lapparna med ett språkprov på varje den bästa lagringsformen och kanske är de också alltid att föredra som arbetsmaterial. De är behagliga att läsa och lätta att placera i grupper och ordning allt efter vad den språkliga analysen bjuder. Naturligtvis hade emellertid en inläsning för datorlagring av dem som nu finns inte bara inneburit en säkerhet mot brand eller annan skada utan också gett ökade sök- och bearbetningsmöjligheter. Om all text i språkproven hade varit inläst hade det varit allra bäst, men tidsåtgången för att skriva av alla de handskrivna lapparna skulle vara enorm även med SAOB-mått. Enbart en inskanning av språkproven som bild med komplettering på så sätt att stickordet och källans utgivningsår skrevs in och blev sökbart skulle kosta mycket tid och pengar. Säkerhetsvinsten vore emellertid fortfarande stor och även användbarheten, trots begränsningar på grund av att man inte kunde söka i textcitaten. Dessutom skulle man uppnå att materialet blev mer tillgängligt för forskare utanför ordboksredaktinen.

### **Förväntningar på språkprovssamling och ordbok**

Alltsomoftast ställs orimliga krav eller förväntningar på SAOB:s språkprovssamling, även från ordbokens egna led. T.ex. ville en av dess första redaktörer, den junggrammatiskt skolade Gustaf Cederschiöld, att ordboken skulle beskriva orden i alla dess tänkbara sammanhang. Sålunda skulle ett ord inte bara beskrivas med avseende på dess stilvärde och varierande bruk i olika samhällsgrupper (något som ordboken naturligtvis försöker informera om ifall det är möjligt) utan orden skulle helst också beskrivas i fråga om variation med andra ord. Även sådan information ger ordboken, men i blygsam utsträckning. Över huvud taget ger ordboken förhållandevis sällan direkt uppgift om synonymi- och antonymirelationer eller annan opposition och (hierarkisk) ordning inom ordförrådet. Cederschiöld ville emellertid ha upplysningar om t. ex. alla synonymer som ordet ersatts av eller ersatt, när det skett och hur förhållandet mellan synonymer varit under olika tider (se Cederschiöld 1894). Sådant kräver ingående undersökningar som inte utan besvär kan utföras ens med hjälp av en fullbordad SAOB. Då behövs en väldefinierad korpus med framför allt i fråga om genre representativa texter. Dessutom borde SAOB inte bara vara klar utan så beskaffad att ordens betydelsebeskrivning enkelt kan knytas till förekomst i korpus. Korpusen skulle alltså vara helt betydelseanalyserad. Det torde vara oerhört svårt att göra en så noggrann ordbeskrivning som Cederschiöld efterlyste, och den blir inte mycket lättare att utföra med hjälp av en klar SAOB.

Ordboken beskriver orden vart och ett. Den kan knappast läsas på annat sätt än så som den är skriven, ord för ord. Därmed inte sagt att ordboken med sitt material inte skulle vara en värdefull om inte oumbärlig källa vid överlexematiska studier av svenskans ordförråd. Men om man t.ex. vill ta reda på hur *snille*, *geni* och *begåvning* förhåller sig till varandra eller jämföra användningen av *civilisation*, *samhälle*, *stat*, *rike* och *nation*, så krävs eget forskningsarbete.

Den historiska ordbokens uppgift är att beskriva ordens historia, framför allt deras betydelse och betydelseutveckling. Det är inte ordbokens uppgift att svara på frågor om hela eller ens delar av ordförrådet. Vid sådan forskning är en historisk ordbok visserligen en hjälp, men den utgör en begränsad faktabank med upplysningar om orden var för sig och bara lite om deras förhållande till varandra. Det är i stort sett bara indirekt genom de distinktiva definitionerna som SAOB ger information om ordens (lexemens) betydelseförhållande till ett överordnat begrepp och talar om vilka specifika särdrag som utmärker orden och därmed skiljer dem från andra.

En av SAOB:s brister är dess långa tillkomsttid, som medför att t.ex. upplysning om ordens bruklighet måste tas emot med beaktande av när den aktuella ordboksartikeln skrevs. Ett "numera föga br." kan vara ett omdöme om ordet i en av dess betydelser för 100 år sedan. Nu kanske ordet inte längre ingår i vårt ordförråd och borde ha dödförklarats. Men det kunde ha varit ännu sämre i detta avseende, nämligen om ordboken försetts med ett eller flera integrerade supplement på samma sätt som t.ex. OED (Oxford English Dictionary). Där kan t.ex. bruklighetsbeteckningen "now rare" referera antingen till tidpunkten för redigeringen av det ursprungliga verket eller till endera av de båda supplementens tillkomsttid. Se Duekilde (1995).

Möjligheter som ordboken och dess samlade material kan erbjuda upptäcks ibland under inflytande av sådan entusiasm som nya hjälpmedel väcker. Den datorlagrade bibliografi över SAOB:s källor som beskrivs i Svensson (1993:376 f) är mycket värdefull, framför allt som lättanvänd källförteckning. Om man dessutom kan göra undersökningar på tvärs i ordboken med dess hjälp återstår att se. Det beror bl.a. på hur representativt och vältäckande inte bara excerperat material utan även det därav valda och i ordboken citerade materialet är för en bestämd tids ordförråd. Är det möjligt att t.ex. via bibliografin spåra texter från ett visst fackområde, t.ex. juridik, och sedan söka källorna i databasen OSA (en inläsning av SAOB:s tryckta delar) för att gå vidare till uppslagsorden och därefter upprätta en förteckning över ordförrådet i juridiskt fackspråk från t.ex. 1700-talet (jfr Svensson 1991:377)? Sannolikt är just det inte möjligt, bland annat eftersom andra ord än

fackord kan ha excerperats ur fackspråkliga källor. Vi vet inte heller hur mycket av det totala (fack)ordförrådet som uppmärksammas i ordboken, men det är i alla fall bara ett urval, om än stort. Dessutom återstår det att fullborda kodningen och göra OSA ännu mer användbar än nu. Fulltextsökning är än så länge visserligen inte den enda men den viktigaste funktionen. Men prova gärna! Det väcks ofta nya förväntningar ur misslyckade försök och alla kommer inte på skam.

Varken ordboken eller dess språkprovssamling kan alltså ge direkta svar på frågor om ordförrådet i dess helhet. Även om språkprovssamlingen hade förelegat i lättillgänglig form skulle den ha svikit. Materialinsamlingen har gjorts med tanke på att ordboken ska beskriva enskilda ord, och excerperingen är oerhört gles; antalet språkprov är väldigt litet i förhållande till det enorma totala ordantal som korpusen omfattar. Vi erbjuds ingen säker information om ordens förhållande till andra ord. Tyvärr räcker alltså varken SAOB:s materialsamling eller den skrivna ordboken till om man vill ha direkta upplysningar om relativa ordfrekvenser, ordförrådets växlingar under årtionden eller sekler, växlingar i olika genrer o.d. Det kan bara ges antydningar om sådant. Däremot tillhandahåller ordboken mycket upplysning om enskilda ords betydelse och användning under olika tider och med olika stilvärde. Ordbokens korpus och materialsamling är en nödvändig och bra grund att stå på vid redigeringen av denna historiska faktabank. Men det är rättmätigt att önska mer, och för den uppfinningsrike och tålmodige finns alltid mer information att få. Som hjälp står språkprovssamlingarna till allmänt förfogande. Än så länge krävs det emellertid att man tar sig till redaktionen i Lund.

### **Traditionell språkprovssamling kontra modern lexikografisk korpus**

Det har ovan pekats på några fördelar hos den traditionella typ av excerptsamling som utgör det förnämsta materialet för de historiska ordböckerna. Det har förhoppningsvis framgått att datorbearbetad och bearbetningsbar korpus är till stor fördel som komplettering av en excerptsamling. Att få till stånd en korpus i den moderna bemärkelsen med vilken man som enda hjälp skulle kunna skriva en historisk ordbok torde emellertid vara oerhört svårt och tidskrävande. Eiríkur Rögnvaldsson (1995) presenterar en värdefull konkordans över de isländska sagorna och nämner fördelar med korpus till skillnad från excerptsamling vid redigering av ordbok över sagalitteraturen. Huruvida hans bedömningar i fråga om en sagaordbok är korrekta får andra

avgöra. Men det som gäller de isländska sagorna behöver naturligtvis inte vara direkt allmänt tillämpligt. Det ska t.ex. betänkas att sagalitteraturen är en tämligen enhetlig genre med litet ordförråd jämfört med t.ex. den skrivna svenskan från de senaste seklerna. I nysvenskan är såväl orden, böjningarna, formerna som betydelseerna många fler. Dessutom ska man ha i minne att de historiska ordboksprojekten gör en etymologiskt baserad betydelsebeskrivning, vilket en periodordbok över sagalitteraturen också kan välja att göra. För den uppgiften är emellertid frekvensundersökningar inte fundamentala, varför t.ex. en beskrivning av *heimsókn* hos Fritzner (1954) inte behöver vara mindre lyckad av det skäl allena att en korpusundersökning eller flera skulle visa att den betydelse som Fritzner angett först som nummer 3 är den vanligaste (jfr Eiríkur Rögnvaldsson 1995:130f).

Den höga status som korpusbaserad lexikografi numera har kan leda till ett ointresse eller i vissa fall förakt för områden där den inte är fullt eller ensamt tillämplig, t.ex. historisk lexikografi. De historiska ordboksprojekten förtjänar emellertid att bedömas för vad de faktiskt åstadkommer, inte förutfattat för det sätt på vilket de har samlat och organiserat sitt material. Det vore värt en ingående studie att få veta hur de historiska ordböckerna bäst når målet att göra en etymologisk betydelsebeskrivning. Det är också intressant att få de korpusbaserade ordböckerna noggrant granskade med tanke på vilken effekt korpushantering har, framför allt på betydelseanalysen och -beskrivningen. Beskrivs orden verkligen med beaktande av betydelseernas frekvens eller görs andra överväganden? Utnyttjas möjligheten till automatiska, statistiska undersökningar av ordens samförekomst för att få exempel på vanliga kollokationer (*ha ont i magen, kulmage, ha mage att, mat i magen, ligga på magen; borsta tänderna, borsta bort dammet, borsta håret*) eller används detta för att spåra grundläggande betydelser? Det senare vore en tvivelaktig metod. En kritisk granskning av såväl historisk som nuspråklig lexikografi efterlyses alltså, men tyvärr finns för närvarande ett ointresse för historisk, etymologiskt baserad lexikografi.

## Litteratur

- Aitken, A. J. 1980. On Some Deficiencies in our Scottish Dictionaries. I: W. Pijnenburg & F. de Tollenaere (red.): *Proceedings of the Second International Round Table Conference on Historical Lexicography*, 33–56. Foris Publications, Dordrecht.

- Aitken, A.J. 1978. Historical Dictionaries, Word Frequency, Distributions and the Computer. I: *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*. Vol. XXXII nr 1, 28–47. Besançon.
- Allén, Sture 1979. Hur djupt går SAOB? I: *Nysvenska studier* 59, 119–129.
- Cederschiöld, Gustaf 1894. *Om de senast framställda fordringarna på en historisk ordbok*. Göteborg. Omtryckt i: *Om ordlekar och andra uppsatser i språkliga och historiska ämnen*, 83–120. Stockholm 1910.
- Duekilde, Anne 1995. Hovedværk og supplement. I: *Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden 7.–10. juni 1995. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr 3*, sid 83–92. Reykjavík.
- Eiríkur Rögnvaldsson 1995. A Concordance to Old Icelandic Texts and its Lexicographic Value. I: *Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden 7.–10. juni 1995. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr 3*, 123–135. Reykjavík.
- Ekbo, Sven & Loman, Bengt 1965. *Vägledning till Svenska Akademiens ordbok*. Skrifter utgivna av Nämnden för svensk språkvård 2. Läromedelsförlagen, Stockholm.
- Fritzner, Johan 1954. *Ordbog over Det gamle norske Sprog*. Nyt uforandret opptrykk av 2. utgave (1883–1896) med et bind tillegg og rettelser redigert av Didrik Arup Seip og Trygve Knudsen. Oslo: Tryggve Juul Møller.
- Hast, Sture 1983. Om SAOB:s material. I: *Nysvenska studier* 63, 159–192.
- Ordbok över svenska språket, utgiven av Svenska Akademien. Källförteckning*. 1939. Lund.
- Supplement I. 1975. Lund.
  - Supplement II. 1990. Opublicerad. Svenska Akademiens ordboksredaktion. Lund.
- Pind, Jörgen & Bjarnadóttir, Kristín & Jónsson, Jón Hilmar & Kvaran, Gudrun & Magnússon, Fridrik & Svavarsdóttir, Ásta 1993. Using a Computer Corpus to Supplement a Citation Collection for a Historical Dictionary. I: *International Journal of Lexicography*. Vol. 6 Nr 1. Oxford.
- Reichmann, Oskar 1990. Formen und Probleme der Datenerhebung I. Synchronische und diachronische Wörterbücher. I: *Wörterbücher*.

- Ein Internationales Handbuch zur Lexikographie.* Utgiven av Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta. (Handbücher zur Sprach- und Kommunikationswissenschaft. Band 5.2) Walter de Gruyter. Berlin, New York.
- Sterkenburg, P. G. J. van 1984. *Van woordenlijst tot woordenboek. Inleiding tot de geschiedenis van woordenboeken van het Nederlands.* E. J. Brill, Leiden.
- Svensson, Lars 1991. Om SAOB:s material och en bibliografisk databas. I: *Nordiske studier i leksikografi I. Rapport fra Konference om leksikografi i Norden 28.–31. mai 1991. Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2.* Oslo.

### **Förkortningar**

- DOST = *Dictionary of the Older Scottish Tongue from the Twelfth Century to the End of the Seventeenth.* 1931–. Chicago.
- FISC = Finland Swedish Corpus. (Finlandssvensk textkorpus. Se <http://www.ling.helsinki.fi/dept/nord/pro.html>)
- OED = *Oxford English Dictionary.* Second edition. 1989. Clarendon Press, Oxford.
- OSA = Om svar anhålles. (SAOB som databank. Se <http://logos.svenska.gu.se/saob/osa.html>)
- SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien.* (Svenska Akademiens ordbok). 1898–. Lund.
- WNT = *Groot Woordenboek der Nederlandsche taal.* 1882–. 's Gravenhaage och Leiden.