

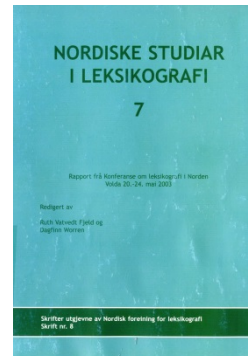
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Ord som går i pension - Om arbejdet att rensa bort gamla ord i SAOL

Forfatter: Martin Gellerstam

Kilde: Nordiske Studier i Leksikografi 7, 2005, s. 133-139
Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24. maj 2003

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.

Martin Gellerstam

Ord som går i pension

Om arbetet att rensa bort gamla ord i SAOL

The paper tackles the problem of identification and deletion of obsolete words in the present edition of the Swedish Academy Glossary (the Swedish abbreviation is SAOL). The first edition was published in 1874 and the different editors have put more effort into the introduction of new words than into the deletion of old ones. The vocabulary of the 12th edition was compared with a modern corpus of some 90 million words and surprisingly many of the words in SAOL did not appear in the corpus. A substantial number of these words will have to be deleted. The paper discusses the thoughts behind and the methods for such a deletion.

1 Om gamla ord i moderna ordböcker

Ord föds och dör. Nya ord är i allmänhet lätta att registrera, de förväntas dyka upp i nya upplagor av ordböcker, de är intressanta för språkbrukaren och tacksamma för förlagen att publicera böcker om. Ord som börjar komma ur bruk är svårare att registrera eftersom de sällan används, de är på sin höjd intressanta som kuriositeter och svåra att hantera för lexikografen. Det är i allmänhet bättre att i en ordbok fria än fälla och därför får utjänta ord stå kvar i ordböcker som stumma vittnen från en annan tid och eventuellt med en markering som «mindre brukligt». Så växer ordboken och problemet blir allt större med åren.

Svenska Akademiens ordlista (SAOL) har varit i bruk sedan 1874 och benägenheten att rensa ut gamla ord har varit mindre än önskan att introducera nya ord. Ändå är omsättningen av ord större än man kanske tror när man vill spegla 125 år av snabb samhällsutveckling.

För att göra saken svårare är ordlistans uppgift inte bara att spegla dagens ordförråd utan också gårdagens i viktiga avseenden. Ord som *armborst*, *hillebard* och *örtug* har inga moderna motsvarigheter men försvarar sin plats i SAOL (visserligen med beteckningen «historiskt») även om Adolf Noreen kallar dem «stendöda». Vidare kan en sak vara oförändrad medan beteckningen byts ut: *räddningstjänsten* ersätter *brandkåren*, *försäkringskassan* ersätter *sjukkassan*. Alla är dock inte överens om att det bara är en verbal förändring och allmänheten fortsätter ofta att använda den invanda termen. Ett ord som *kaffesurr* «kaffesurrogat» kan man undvara eftersom saken inte längre finns i vår välfärdsstat och ordet inte har någon bildlig betydelse som kan få det att överleva. Samma väg går *negerby* som visserligen förekommer i en känd sång men avspeglar en oönskad etnisk syn. Av andra men goda skäl faller det vackra ordet *själaspis* från äldre andlig litteratur. Också funktionsord som *evem* «vem som helst» liksom *medelbarligen* «indirekt» och *ogensägligen* faller för åldersstreck- et när andra funktionsord som betyder samma sak tar över.

Hur görs då denna utsällning? Vilka objektiva verktyg har man till sitt förfogande förutom den subjektiva språkkänslan? Det ska denna uppsats handla om.

2 Underlag för strykning av ord

2.1 SMDB och jämförelsetexter

SAOL föreligger dels i en databas som innehåller all information som finns i den tryckta upplagan (för närvarande tolfte upplagan), dels i en renodlat morfologisk databas, Svensk morfologisk databas, SMDB, med uppgifter om framför allt ordens böjning.

SMDB är resultatet av ett omfattande arbete med att utifrån SAOL:s ordförråd göra en noggrann beskrivning av svensk böjningsmorfologi. Arbetet påbörjades i liten skala efter publiceringen av SAOL11 men har getts slutgiltig form efter publiceringen av SAOL12 (1998). Den språkliga beskrivningen och klassifikationen (över 300 klasser) har gjorts av Sture Berg och databasuppbyggnaden av Yvonne Cederholm och på senare tid också Susanne Mankner. Dataansvarig för SMDB är för närvarande Christian Sjögreen.

Orden är indelade i böjningsklasser och alla böjningsformer är genererade. Varje enhet har en grammatisk böjningsspecifikation (en morfosyntaktisk «tagg»). SMDB är kopplad till korpusdata från Språkbanken (närmare 90 miljoner ord), vilket betyder att varje ordforms förekomst i texter kompletteras med en frekvensuppgift. Basen omfattar totalt ca 1,2 miljoner ordformer (varav närmare 900 000 olika ordformer).

Databasen är framtagen av flera olika skäl: för att underlätta arbetet med uppdatering av svensk ordböjning och för att göra vägen mellan text och lexikon möjlig med allt vad detta innebär, bl.a. möjligheten att se hur lexikonet utnyttjas i bruket. För det redaktionella arbetet med SAOL betyder detta bland annat två viktiga fördelar: man kan samköra SAOL med stora textmängder och därigenom få fram ord som finns i texter men inte i ordlistan (nyord) och omvänt få fram ord som finns i ordlistan men inte i texter (ord som kan strykas). Till fördelarna med SMDB hör också att databasen kan användas för ordklassmärkning («taggare») av texter.

För att få fram förslag till strykningar av ord i SAOL kördes SMDB mot närmare 90 miljoner ord jämförelsetext, till stor del tidningstext (inklusive telegrambyråmaterial) men också skönlitteratur och offentligt språk. Texterna är framtagna inom ramen för Språkbanken, institutionens organ för bl.a. textinsamling, konkordansuppbyggnad och taggning av texter.

2.2 Andel ord som saknar belägg

Vid en första köring visade det sig att 16.8% (ca 20 000) av orden i SAOL (ca 120 000 ord) inte fanns i jämförelsetexterna. När jämförelsetexterna utvidgades från 90 miljoner till databasen PressText som innehåller en mångdubbelt större mängd text från tidningar och tidskrifter, reducerades denna siffra till ca 10%. Men fortfarande innebär detta att ca 12 000 ord i SAOL inte får några träffar i en stor mängd moderna texter.

Nu behöver detta inte betyda att en så stor ordmängd faktiskt skall tas bort från ordlistan. En hel del ord kan få träffar genom att andra typer av källor utnyttjas. En sökmotor som Google kan utnyttjas, både på grund av att den innehåller andra genrer och stilar än tidningstext och genom sin blotta storlek (kanske tio gånger större än PressText). En annan viktig källa är Nationalencyklopedins databas (på nätet) som innehåller en mycket stor mängd fackord.

3 Vilka typer av ord saknar belägg?

3.1 Typologisk genomgång av utgående ord

De ord som stryks i SAOL är föråldrade på olika sätt. Kriteriet på föråldrade ord är i brist på annat att orden förekommer sparsamt eller inte alls i moderna texter. Nu blir detta kriterium naturligtvis beroende av vilka texter som används som underlag. Någon strikt balanserad korpus rör det sig inte om utan till stor del tidningstext och pressbyråtexter jämte något tjugotal miljoner löpord litterär text och officiell text. Texterna är i huvudsak från 1990-talet (och början på 2000-talet) men några delkorpusar går tillbaka till slutet av 1970-talet. Totalt rör sig korpusen om ca 90 miljoner löpord.

I det följande ges exempel på typer av föråldrade ord, hämtade från en tidigare utrensning av ovanliga ord inför tolfte upplagan av SAOL (1998). Exemplifieringen i Tabell 1 hämtas alltså från arbetet inför förra upplagan.

Det bör nämnas att ovanliga ord kan leva kvar i språket (och därmed i SAOL) eftersom de har blivit termer för historiska begrepp. Sådana ord kan exempelvis syfta på äldre folkslag (*goter, meder, kelter*), organisationer (*andrakammarledamot, indelningsverket*), äldre titlar (*länsman, fjärdingsman*), äldre föremål som spelat en viktig historisk roll (*musköter, armborst, årder* «en primitiv plog»), äldre fortskaffningsmedel (*knarr, landå*) osv. Dessa ord är normalt inte frekventa i det svenska allmänspråket men är fullt levande i historiska framställningar. De faller alltså utanför indelningen av utgående ord. Detta betyder inte att historiska ord genomgående bevaras utan bara sådana som fortfarande har hävd i språket (i Tabell 2 anges t.ex. att ord som *laddstock, ladufogde* och *landsbibliotekarie* föreslås utgå ur SAOL).

En första typ av utgående ord enligt Tabell 1 avser ordbildning och gäller fast komponerade partikelverb (*avbita, insläppa* etc.) där modern svenska genomgående väljer en lös ordbildning (*bita av, släppa in* osv.). Den fasta partikelbildningen har länge bibehållits i lexikaliska sammanhang av det enkla skälet att löst sammansatta partikelverb av tradition inte har förekommit som uppslagsord. Resultatet har blivit ålderdomliga eller helt utdöda infinitiver som nu successivt rensas ut ur ordlistan. Ofta har dessa verb haft levande participformer (*avbiten, insläppt*), som är gemensamma för fast och lös partikelbildning och därför har kunnat få en självständig plats i lexikon, i form av participiella adjektiv.

Vidare kan ordförrådet ändras beroende på en förändrad syn på manligt och kvinnligt. Så har t.ex. rätt många kvinnliga avledningar försvunnit och ersatts av den manliga avledningen: detta gäller inte bara många gamla ord som *angiverska, befrierska* etc. utan också en mängd yrkesbeteckningar i dagens samhälle: *dansare, frisör, lärare* osv. Denna utveckling är inte självklar – i andra länder är det tvärtom viktigt att markera bägge könen (tyska *Student – Studentin* osv.)

Ord försvinner också genom att enskilda avledningar eller sammansättningar blir ovanliga. Man känner igen grundordet men repertoaren av avledningar minskar. Så kan något/någon *vara i antågande* men substantivet *antåg* och verbet *antåga* har försvunnit. En sjukdom kan vara *obotlig* men motsatsen *botlig* tillhör en äldre tid.

De ord som stryks kan vara isolerade morfem som bara finns i en enda form. Ofta är det fråga om lånord som inte har annat stöd i morfologin, men det kan också handla om gamla svenska ord som på samma sätt förlorat stödet av besläktade ord (*abderit, amendemang blouson; farnöte*).

Språklig typ	Beskrivning av förändring	Exempel
Ordbildning	Fast sms ersätts av lös sms	<i>aktge, avbedja, avbita, avborsta, insläppa, påbreda, påvrda</i>
	Kvinnlig avledningsändelse ersätts av manlig	<i>angiverska, bedjerska, befrierska, budbärrinna; dansare, frisör, lärare</i>
	Ovanlig avledning till känt grundord	<i>antåg, antåga, botligt, dristeligen, förlångsamma, försnabba</i>
Morfem	Främmande eller isolerat semantiskt morfem	<i>abderit, amendemang, blouson, brouillerad, faute, farnöte, fåvitsko</i>
Stavning	Ovanlig alternativ stavning	<i>angelsachsare, jos</i>
Form & innehåll	Föråldrad form – aktuellt innehåll	<i>galanteriaffär, hökare</i> "matvaruhandlare"; <i>annanstädes, evar</i>
	Föråldrad form – föråldrat innehåll	<i>bykdag, datamaskintid, enbetsvagn</i>
Konnotationer	Eufemistisk bildning	<i>alkoholproblematiker, evakostym</i>
	Socialt olämpligt eller fördomsfullt ord	<i>negerbarn, negerby</i> (jfr dock <i>svartskalle</i> som kvarstår), <i>hednamission</i>

Tabell1. Typer av strukna ord i SAOL12

Stavningsvariation leder ofta till att en av varianterna utgår. Detta innebär inte att ordet som sådant försvinner, bara enskilda former som blir mindre och mindre utnyttjade (eller kanske aldrig blir accepterade, t.ex. *jos* «juice»).

Ibland ändras ett ords form även om innehållet är levande – man går exempelvis inte längre till en *hökare* för att handla. Tydligt syns detta på formord som av olika skäl försvinner och ersätts av nya (*evar* heter idag *var helst*, *annanstädes* heter snarare *annorstädes*). Ofta kan såväl formen som innehållet vara döda (man talar inte längre om *forwardskedja* i vår tid när spelstrategin i fotboll kan se ut på mångahanda sätt).

Synen på vad som är en eufemistisk bildning växlar med tiden och att använda *evakostym* uppfattas idag som preciöst av en generation som inte är chockerad av nakenhet. Svårare att hantera är ord som av någon anledning är socialt olämpliga och därmed blir ovanliga. Hit hör en rad sammansättningar med *neger*, ett ord som i sig självt är svårhanterligt och i ordlistan betecknas som «något nedsättande». Ett annat exempel på förlegad samhällssyn (i detta fallet inom kyrkan) som utgått är sammansättningar med *hedna-*, t.ex. *hednamission*, som avsåg mission bland ickekristna (företrädesvis i Afrika).

3.1 Motsättning mellan bruk och norm

I en delvis normativ ordlista som SAOL förväntas läsaren inte bara kunna inhämta kunskap om enskilda ord utan också om normal ordbildning i ord som råkar saknas i det urval som förekommer. Om vi här inskränker oss till hur fogen ser ut i sammansatta ord, något som kan variera (t.ex. *bergstopp* eller *bergtopp*, *gatuskylt* eller *gatskylt*), så anges detta normalt genom att relativt många sammansättningar ger explicita upplysningar om fogen. Men det finns ovanliga ord som har få eller inga sammansättningar och därför ger otillräckligt besked om

hur en sammansättning skulle se ut om den bildades. Här har SAOL sedan länge tillämpat modellen att skapa en sammansättning även om den är mycket ovanligt, bara för att ge en upplysning om hur bildningen skulle se ut om man skapade en sammansättning (exempel på denna praxis är sammansättningar som *langustfiske* och *lavethjul* som saknas i jämförelsematerialet). Här kommer alltså bruket och behovet av norm i konflikt med varandra. Detta skulle kunna lösas genom att varje substantiv innehöll en anvisning om hur fogen ser ut i sammansättningar. Så har inte skett tidigare men det kommer att övervägas i nästa upplaga.

Andra exempel på samma sak är det väl företrädda adjektivet *ömhjärtad* som vid sin sida har substantivet *ömhjärtenhet*, icke förekommande i våra texter. Men substantivet ger besked om hur substantivet skulle se ut om det skulle behöva användas, något som inte är självklart för språkbrukaren. Andra exempel är det ovanliga adjektivet *avvisbar* till verbet *avvisa* och det ovanliga *regnklädesfabrik* till det vanliga *regnkläder*. Vidare kan man undra vad den manliga motsvarigheten till *järphöna* är – jo *järptupp*, även om det är sällan förekommande.

I alla dessa fall ser vi hur en ren avläsning av bruket inte skulle ge tillräckliga fingervisningar om språkstrukturen vid skapandet av ord som inte finns i ordlistan. Detta behöver kanske i många fall inte ske genom att man skapar ovanliga ord utan kan ges i anslutning till det normala ordet genom en kortfattad regel, t.ex. att sammansättningar med ett visst ord normalt har -s- i fogen.

4 Strategi för utrensning av ord

Arbetet på att rensa ut ord inför den nya (trettonde) upplagan av SAOL baseras på det underlag av ca 12 000 ord som beskrivits ovan. Den första sällningen innebär att ordet har fallit ut som «nollord» i såväl Språkbankens texter som i Presstext. Arbetet innebär därefter följande steg.

Steg 1 Leta på Internet med sökmotorn Google. Finns flera exempel på ordet i naturliga och moderna kontexter skall ordet kvarstå i SAOL. Saknas ordet går man vidare till Steg 2.

Steg 2 Om ordet är ett möjligt fackord kontrolleras med Nationalencyklopedin (NE). Finns ordet där behålls det i SAOL. Saknas det går man vidare till Steg 3.

Steg 3 Om ordet saknas i Google och NE kontrolleras först om det är medtaget i SAOL av strukturella skäl, t.ex. för att ange fog vid sammansättning. Om ordet kan tänkas har den funktionen läggs denna information in explicit i första delen av det sammansatta ordet och den icke brukade sammansättningen utgår.

I steg 1 hänvisas till förekomst i ordlistor, lexikon och andra ordsamlingar, alltså metaspråkliga kontexter. Sådana exempel är inte ovanliga men de räknas alltså inte till «naturliga». Med «moderna» kontexter avses användning under senare delen av 1900-talet. I Google förekommer relativt många historiska texter och litterära texter men sådana kontexter accepteras alltså inte.

I det följande ges exempel på vad som kan återstå (på en bit av bokstaven L) när dessa kontroller har skett. För varje ord ges en antydning om varför ordet har utgått.

labiologi	äldre ord för "läppavläsning"
labiologisk	se ovan!
labyrintfönster	anatomi (ej i NE)
lackerpensel	ev. tilläggs <i>lackpensel</i>
laddstock	historiskt; till mynningsladdare
ladufogde	historiskt "befallningsman i södra Sverige"
ladugårdskarl	historiskt
lagfarenhet	äldre ord för "juridik"
lakejsjäl	senaste belägg 1910 enl. SAOB
laksaltad (fisk)	"(fisk) i saltlake"
lallning	till lalla
lamellera	utgår som infinitiv, ersätts av <i>lamellerad</i>
lamentering	vanligare: lamentation
landsbibliotekarie	historisk titel
landsmålsarkiv	tidigare beteckning på dialekt- och ortnamnsarkiv
landssamtal	tidigare typ av teletjänst?
lanfiske	(exempel på sms-fog)"fiske med <i>lana</i> "
langamsrulle	med langarn (exempel på sms-fog)
langustfiske	(exempel på sms-fog)
lansskaft	
lantanglas	(<i>lantan</i> grundämne + glas) (exempel på sms-fog)
lantmätarinstrument	
lantpolis	mest i Grönköpings Veckoblad
lantvistelse	
lapisstift	medicinhistoriskt
lappjäxa	"samepjäxa"?
lappmarksförsamling	historiskt
lappsalvning	sjöterm (<i>lappsalva</i> behålls)
lapsktalande	tidigare beteckning på "samisktalande"
larsmässa	historiskt
lasarettbyggnad	<i>lasarett</i> vanl. ersatt av <i>sjukhus</i>
lasarettsvård	se ovan!
lastfull	äldre bildning av ordet <i>lastbar</i>
lastfullhet	se ovan!
lasurblått	skrivs vanl. <i>lasyr-</i>
lavetthjul	(exempel på sms-fog)
laxkar	historiskt
laxkista	historiskt
lazzaronliv	(ev. utgår även <i>lazzaron</i>)
ledartrissa	tekn. "glidande trissa vars rörelse överförs till ett arbetsverktyg" (ej i NE)
ledera	"skada, sára"
ledering	se ovan!
ledhål	"grindöppning"
ledren	"ren som går i täten"
ledvatten	den vanliga termen är <i>ledvätska</i>
legaliter	äldre ord för "enligt lag"
legitimitetsprincip	för lagligen arvsberättigad furste
lejdra	"redskap som användes vid omröring o. avjämnande av vaskgods på en planhård" (ej i NE)

Tabell 2. Exempel på utrensning av ord (del av bokstaven L)

5 Slutord

Arbetet på att plocka bort föråldrade ord i SAOL har visat sig mer arbetskrävande än vad vi från början tänkte, trots att vi har haft utmärkt hjälp av samkörning mellan SAOL och stora textmängder. Det framstår som alltmer begripligt att tidigare redaktörer har ägnat större flit åt att lägga in nya ord än åt att rensa i det redan etablerade ordförrådet. Detta desto mer som tidigare redaktörer inte haft de verktyg till förfogande för jämförelse mellan SAOL och texter som vi har.

Det största problemet är kanske att vi inte har en enda stor balanserad korpus som vi kan köra mot SAOL – och därmed nerbringa antalet kandidater för uteslutning ur ordlistan. Vi har av naturliga skäl inte heller kunnat köra SAOL direkt mot Google eller ens mot PressText eller Nationalencyklopedin utan «bara» mot de ca 90 miljoner orden i Språkbankens förråd av moderna texter som vi själva förfogar över. Å andra sidan är det otänkbart att enbart låta frekvensen fälla utslaget – en genomgång på grundval av datorns resultat är under alla omständigheter nödvändig.

Men är det så viktigt att rensa bort ord som är på väg ut ur ordförrådet? Särskilt som konstruktörer av korsord och olika språkliga tävlingar särskilt omhuldar de lite halvkonstiga ord som vi tar bort och verkligen inte tackar oss för det arbetet. Ja, det är viktigt av flera skäl. I en källa som SAOL kan ord dröja sig kvar i långa tider eftersom det är sparsamt med ordförklaringar som kan sätta ordens existens ifråga. Dessutom tar de plats, det rör sig dock om åtskilliga tusen ord som inte bör vara med och det finns ett stort ordförråd som väntar på att bli upptaget i SAOL, framtaget med liknande datamaskinella metoder som de utgående orden. Därför fortsätter vi att stryka.

Källor

Nationalencyklopedin. 2003. Nättupplagan.

Svenska Akademiens ordbok (SAOB) (A- Talkumera). 2003. Nättupplagan.

Svenska Akademiens ordlista (SAOL). 12 upplagan. 1998. Norstedts.

Sökmotorn Google.