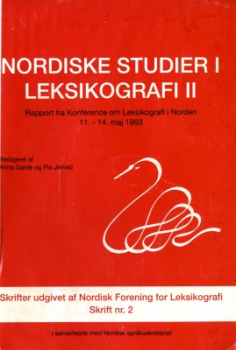


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Sammansättningsmorfologi och lexikografi	
Forfatter:	Sven-Göran Malmgren	
Kilde:	Nordiske Studier i Leksikografi 2, 1993, s. 179-184 Rapport fra Konference om leksikografi i Norden, 11.-14. maj 1993	
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.

Sammanställningsmorfologi och lexikografi

Sven-Göran Malmgren

I Illustrerad svensk ordbok, den svenska definitionsordbok som kom ut på 1950-talet, finner man bland sammansättningarna under **röd** bl.a. en kort artikel **rödbena** (en fågel) och strax därefter **rödbeneägg**, med förklaringen 'ägg av rödbena'. Vid första påseendet kan den sista artikeln verka totalt överflödig. Man kan omöjligt hävda att sammansättningen **rödbeneägg** är lexikaliserad, och definitionen utmanar nästan löjet. Men det finns faktiskt ett bra skäl till att ordet **rödbeneägg** är med i ordboken. Det är inte alls självklart vilket utseende ordet **rödbena** får som förled i sammansättningar. Vad artikeln upplyser om är i själva verket just att den rätta formen är **rödbene-** och inte t.ex. **rödbens-** (jfr *pepparkaka* – *pepparkaksdeg*). Den i ordboken upptagna sammansättningen ska alltså ses som representant för alla möjliga sammansättningar som utgår från **rödbena**. (Definitionen kunde dock med fördel ha utelämnats.)

Om alltså artikeln **rödbeneägg** faktiskt innehåller en värdefull upplysning, så kan man likväl fråga sig om ordbokens sätt att ge denna upplysning är det bästa. Man kan också fråga sig hur konsekvent uppgifter om sammansättningsfogar hos ord som själva är sammansatta ges i moderna svenska ordböcker (senare än Illustrerad svensk ordbok). I den här uppsatsen görs ett försök att besvara båda frågorna, och dessutom diskuteras en hittills knappast prövad möjlighet att ge information om sammansättningsmorfologin. Slutligen berörs möjligheterna att snabbt samla in nödvändiga data från korpusar och lexikaliska databaser.

Vi kan nöja oss med att se på elfte upplagan av Svenska Akademiens ordlista (SAOL, 1986), eftersom det är den enda moderna svenska ordbok som ger nämnvärd information av det aktuella slaget. I andra ordböcker, t.ex. Svensk ordbok (1986), tas många sammansättningar upp, men så gott som bara sammansättningar som är semantiskt intressanta. Sammansättningarna av typen **rödbeneägg** torde vara mycket lätt räknade i Svensk ordbok. Vi väljer vidare att koncentrera oss på substantiv av samma typ som **rödbena**, dvs. substantiviska sammansättningar som slutar på **-a**. Hur dessa ord förändras när de själva ingår som förled i dubbelt sammansatta ord av typen AB|C är synnerligen oförutsägbart, som Hellberg (1978) och framför allt Wennerberg (1962) visar. Jfr t.ex. *ficklampa* – *ficklampsbatteri*, *folkvisa* – *folkvisetradition*, *högmässa* – *högmässobesökare*, *vattenskada* – *vattenskidåkare*. Tilläggas kan, att problemen är likartade vid sammansättningar vars efterled slutar på **-e**. När efterleden slutar på konsonant är det betydligt lättare att förutsäga sammansättningsmorfologin; vi tar emellertid upp ett komplicerat fall i slutet av uppsatsen.

Vi ser först på några substantiv som liksom **rödbena** betecknar fåglar, närmare be-

stämt alla sammansättningar med *-kråka*, *-lärka*, *-skata* eller *-ärsla* som efterled som tas upp i SAOL. (Det finns en baklängessortering av ordboken.) Fördelen med att studera denna "tematiska" grupp är att den troligen inte genererar många lexikaliserade sammansättningar. Om alltså en dubbel sammansättning (t.ex. *sädesärleägg*) kommit med i SAOL, kan man räkna med att det är av samma skäl som att *rödbeneägg* togs upp i Illustrerad svensk ordbok, och det är det som är det intressanta i det här sammanhanget.

I tabell 1 redovisas resultatet av den lilla undersökningen:

Tabell 1. Information om sammansatta ord som förleder i SAOL

Ord	Sammansättning med ordet som förled finns/finns inte (+/-) i SAOL (1986)
blåkråka	-
nötkråka	-
spillkråka	+ (spillkråksbo)
piplärka	-
skärpiplärka	-
sånglärka	-
tofslärka	-
trädlärka	-
snöskata	-
strandskata	-
törnskata	-
gulärsla	-
sädesärsla	+ (sädesärleägg)

Vi ser att en språkbrukare som är osäker på hur man bildar en sammansättning med *sånglärka*, *tofslärka* etc. eller med *snöskata*, *strandskata* eller *törnskata* inte får någon vägledning av ordboken. Vill han bilda sammansättningar med *nötkråka* eller *gulärsla* får han inte heller någon **direkt** upplysning, men med lite tur och fantasi kan han hitta fram till *spillkråka* resp. *sädesärsla* och på grundval av sammansättningarna där gissa att det bör heta t.ex. *nötkråksbo* och *gulärleägg*. Men endast när det gäller *spillkråka* och *sädesärsla* får han explicit vägledning.

Nu finns det givetvis en gräns för hur många ord SAOL kan ta upp, och meningen är

ingalunda att hävda att alla de saknade dubbla sammansättningarna borde varit med. Men möjligen hade det varit bra om ett av de substantiv – och i så fall det man bedömer som det vanligaste – som slutar på *-lärka* och ett av dem som slutar på *-skata* hade försetts med en sammansättning, så att en användare åtminstone hade möjlighet att få indirekt vägledning beträffande de övriga orden med dessa efterleder. (En annan möjlighet att ge denna vägledning diskuteras i slutet av uppsatsen.)

Naturligtvis kan man säga att det inte är någon katastrof om en språkbrukare själv måste bestämma sig för om det ska heta *sånglärkebo*, *sånglärksbo*, *sånglärkbo* eller något annat. Och det ska medges att problemet inte hör till de mest centrala i språket. Men frågor föranledda av osäkerhet om förledens form i sammansättningar tycks höra till de vanligaste bland de frågor som ställs till Svenska språknämnden (Ulla Clausén, muntl. komm.). Det finns också en liten paradox i det här sammanhanget. Å ena sidan har man sällan anledning att bilda dubbla sammansättningar, och därför kan man säga att problemet med fogens utseende inte är så viktigt, men å andra sidan medför just de dubbla sammansättningarnas sällsynthet att språkkänslan är osäker, och därför kan behovet av vägledning vara särskilt stort när det någon gång finns behov av en sådan sammansättning.

För att illustrera ett par andra problem ser vi på ytterligare ett par grupper av ord i SAOL:s elfte upplaga. Först sammansättningarna med *-penna* som efterled. De är 28 till antalet. Endast vid *reservoarpenna* ges en dubbel sammansättning, *reservoarpenn(e)bläck*. Ser man nu efter i den närmast föregående upplagan av SAOL (1973), finner man utom *reservoarpenn(e)bläck* även det dubbelt sammansatta ordet *stålpennsfabrik*. Det fanns säkert mycket goda skäl att stryka detta ord i elfte upplagan av SAOL. Men strykningen har faktiskt en konsekvens som möjligen inte var avsedd: med den försvinner den enda indikationen i SAOL på att sammansättningar med efterleden *-penna* kan få fogelementet *-s-* när de ingår som förled i dubbelt sammansatta ord. En ambitiös ordboks-användare, eller en konstruktör av lemmatiseringsprogram eller program för stavningskontroll – beväpnad med baklängesordbok – som förlitar sig på SAOL måste dra slutsatsen att mönstret *reservoarpenn(e)bläck* numera, i motsats till för 20 år sedan, är genomgående. Kanske är det också så, kanske inte. Poängen här är bara att strykningen av ett helt perifert ord faktiskt kan innebära ett rätt kontroversiellt implicit påstående.

De sista exemplen vi tar upp är sammansättningarna med efterleden *-stuga*. De är 47 till antalet i SAOL (1986), och vid 6 av dem ges dubbelt sammansatta ord. Dessa är *backstugusittare*, *förstu(gu)trappa* (m.fl.), *rådstu(gu)rätt*, *småstugebebyggelse*, *sommarstugebebyggelse* och *sportstugeområde*. Om man bortser från variationen vid *förstuga* och *råd-stuga* verkar alltså *-stuge-* och *-stugu-* vara de två möjligheter som finns. Nu kan man emellertid observera att *-stugu-* i SAOL endast förekommer i mer eller mindre ålderdomliga ord, och troligen används denna form sällan i nybildade sammansättningar. Kanske vore en explicit upplysning om detta av värde, även om det redan nu antyds av SAOL:s urval av dubbla sammansättningar.

På vilket sätt skulle dessa upplysningar om sammansättningsmorfologi kunna ges? Om

vi stannar vid det sista exemplet, kunde man kanske tänka sig något i stil med följande (t.o.m. **-värme** är det den autentiska artikeln i SAOL):

stug|a -an -or s. stug|by -dörr [...] -värme; -stuge-, (förr) -stugu-

Informationen om hur *tvättstuga*, *fritidsstuga* etc. bör se ut som förled i dubbla sammansättningar skulle alltså samlas under det enkla ordet *stuga*. En användare, som söker råd när han ska bilda en sammansättning av *tvättstuga* och *tid*, behöver veta att han ska gå från artikeln **tvättstuga** till artikeln **stuga**, när han inte får vägledning på det första stället. Den upplysningen skulle han givetvis få i ordbokens inledning. När han väl hittat fram till artikeln **stuga** får han en direkt rekommendation att använda formen *tvättstugetid*.

På motsvarande sätt skulle man kunna ge den information som nu saknas vid de övriga orden vi diskuterat. Man skulle då också tvingas ta ställning till om möjligheten *-penns-* eventuellt bör anges vid sidan av *-penn-* och *-penne-* i artikeln **penna**.

Den föreslagna modellen fungerar emellertid tillfredsställande bara under förutsättning att alla sammansättningar med en viss efterled följer samma mönster som förleder. När detta inte är fallet, vilket är långtifrån ovanligt, får det anges på något sätt. Vi kan ta sammansättningarna med *-vatten* som efterled som exempel. SAOL ger enbart *regnvattens-* (*regnvattenstunna*) som förledsform av *regnvatten*, enbart *grundvatten-* av *grundvatten* (*grundvattennivå*), men både *lågvattnens-* och *lågvattnen* (*lågvattnen(s)märke*) av *lågvattnen*. Av artikeln **vatten**, där möjligheterna *-vatten-* och *-vattens-* enligt den föreslagna modellen skulle anges, skulle det då behöva framgå att valet inte alltid är fritt utan att användaren bör gå vidare till de enskilda sammansättningarna med *-vatten* som efterled. Men finner han vid en sådan sammansättning inget exempel på ett dubbelt sammansatt ord, som t.ex. vid *badvatten* i den aktuella upplagan av SAOL, bör han kunna dra slutsatsen att båda varianterna är möjliga just vid detta ord.

Den information som här har föreslagits kan mycket väl ges också i definitionsordböcker. Kanske passar den t.o.m. bättre där, om inte annat så av typografiska skäl. (Definitionsordböcker brukar ha något mindre kompakt typografi än SAOL.)

Återstår frågan hur man bäst får fram det empiriska underlaget för de uppgifter om sammansättningsmorfologi som här diskuterats. För det första bör man kunna söka på strängar inne i ord i en lexikalisk databas. Den möjligheten finns redan nu vid Institutionen för svenska språket vid Göteborgs universitet (där arbetet på tolfte upplagan av SAOL och på en utvidgning av Svensk ordbok pågår). Den maskinläsbara ordlista man kan söka på detta sätt i innehåller mer än 150 000 ord, huvudsakligen från SAOL. Om man t.ex. beställer fram alla ord som innehåller strängen *stug* föregående och följd av minst två bokstäver får man efter några sekunder följande lista på 15 ord på terminalen:

Fig. 1. Ord med strängen *stug* föregången och följd av minst två bokstäver i en lexikalisk databas på 150 000 ord

Söksträng:

Maxant

%__stug__%

500

Hämta

Lemman i L

Grundform	Neo	Hv	Lx	Ordklass
backstugusittare	1		1	subst.
förstugubro	1	#		subst.
förstugukvist	1	#		subst.
förstugutrappa	–			subst.
ristuggare	–			subst.
rådstugurätt	1	a	1	subst.
rådstugurättspredikan	–			subst.
småstugebebyggelse	–			subst.
snustuggande	–			adj.
snustuggande	–			subst.
snustuggning	–			subst.
sportstugeområde	–			subst.
testuggare	1		1	subst.
tobakstuggning	–			subst.

Vi får alltså fram de ord vi tidigare diskuterat och dessutom en del ord som vi inte är intresserade av. Det är uppenbart att man vill kunna söka på **morfemet** *stug*. Detta har ännu inte utförts i praktiken men det är i princip möjligt – morfemgränserna är markerade i de flesta av de 150000 orden.

Men för att kunna ge tillförlitliga uppgifter om sammansättningsmorfologi behöver man i många fall ha tillgång till data från en **korpus**. Idealet vore att från en korpus på exempelvis 10 miljoner ord snabbt kunna få fram grundformerna av alla ord som innehåller ett visst morfem, sorterade i första hand efter varianterna av morfemet (eller av morfemet plus den tillhörande sammansättningsfogen) och med frekvensuppgifter. För detta behöver man, utom markeringar av morfemgränser i orden i den lexikaliska databasen, ett system för automatisk lemmatisering. Vi vet att det bästa svenska lemmatiseringssystemet ger rätt analys av mer än 99% av orden i en allmänspråklig korpus (Karlsson 1992, om det s.k. SWETWOL-systemet). Många fall av homografi som normalt väl-

lar problem är harmlösa om man i första hand söker efter morfem **inne** i ord. Att från en stor korpus snabbt få fram exempelvis alla dubbelt sammansatta ord där förleden slutar på *-stugu-* eller *-stuge-* ligger därför inom möjligheternas gräns.

Ett litet problem återstår dock. Ett lemmatiseringssystem måste bygga på existerande ordböcker, t.ex. SAOL. Antag att lemmatiseringssystemets lexikon i likhet med SAOL inte räknar med möjligheten *-penns-*. Om ett ord som *kulpennsfodral* förekommer i korpusen kommer det då att höra till den knappa procent ord som lemmatiseringssystemet inte klarar att analysera. Den givna slutsatsen är att dessa ord är mycket viktiga att gå igenom eftersom de kan indikera förändringar i sammansättningsmorfologin.

Litteratur

Illustrerad svensk ordbok. 1956. Stockholm.

Hellberg, Staffan 1978. *The Morphology of Present-Day Swedish*. Stockholm.

Karlsson, Fred 1992. SWETWOL: *A comprehensive Morphological Analyser for Swedish*. *Nordic Journal of Linguistics*, 15, s. 1-45.

SAOL=*Svenska Akademiens ordlista*. 10, 11. uppl. 1973, 1986 Stockholm.

Svensk ordbok. 1986. Stockholm.

Wennerberg, John. 1962. *De dubbelt sammansatta orden i svenskan*. *Nysvenska studier*, 41, s. 5-54.