

Elżbieta Dura

Automatisk extrahering av ordleder

A typical trait of Swedish word formation are words which are productive in word formation. They are not bound morphemes because they are either root morphemes or composites. They are not lexemes because they are used in a similar way as bound morphemes. These bound lexemes, as they are called here, constitute a crucial part of a compounding dictionary, particularly of a bilingual one. Potential bound lexemes are traced and automatically selected from a large text corpus of Swedish. Two parameters are established to make the extraction precise: compounding ability and compounding productivity. The first one provides measure of the relation between occurrences of a word as compound component and as an independent word. The second one is based on how many collocates it cooccurs with in compounds.

Inledning

Den nedan beskrivna korpusundersökningen av svenska texter har främst ett praktiskt syfte - att extrahera material till ett aktivt sammansättningslexikon. Det övergripande målet är att underlätta för även icke infödda talare av svenska att förstå, översätta och skapa nya sammansättningar. Arbetet ingår i projektet *Svenska sammansättningar och deras ekvivalenter i polska* som huvudsakligen behandlar verbala sammansättningar - dessa saknas nästan helt som produktiva mönster i slaviska språk. Projektet finansieras av HSFR och bedrivs på Institutionen för slaviska språk vid Göteborgs universitet. Prof. Roman Laskowski är projektledare, forskningsassistenter är fil.dr Elżbieta Dura och fil.dr Maria Toporowska-Gronostaj, samt under sista halvåret fil.mag. Morgan Nilsson.

Vid extraheringen användes texter från Språkbanken vid Göteborgs universitet (www.svenska.gu.se): 42 580 498 löpord, varav 33 245 519 från dagstidningar, och 9 337 979 prosa. Extraheringen genomfördes helt automatiskt med LexWare, ett korpusverktyg utvecklat av LexWare Labs i Göteborg (www.lexwarelabs.com). LexWare identifierar morfosyntaktiska enheter i den givna texten och väljer sedan ut de som efterfrågas. Dess textanalys bygger bl.a. på: över 70 000

uppslagsenheter¹, 6 tilläggsordlistor², 800 000 grammatiska former, 400 ordbildningsregler, och 500 regler för kontextbaserad disambiguering. Tester på en testkorpus (Stockholm-Umeå-Corpus) visar att man måste räkna med över 2% fel i LexWares morfosyntaktiska identifiering. Det innebär att beräkningarna inte skall betraktas som exakta, utan enbart som approximationer. Man bör även räkna med en viss övertolkning av sammansättningar p.g.a. felstavningar i de undersökta texterna, t.ex. analyseras felstavning av *betraktas* som *bet-rakats*.

I slutet av 1999 kommer de erhållna urvalen av ordleder att publiceras på projektets hemsida: bratislava.slav.gu.se/LexWare/sms. Här presenteras enbart korta utdrag.

1. Ordleder och bundna lexem

Huvuddelen i ett sammansättningslexikon är sammansättningar, huvuddelen i ett aktivt sammansättningslexikon är produktiva ordleder. De senare särskiljs som två ordledstyper i Nyord (1986): tidstypiska ordleder, t.ex. *tv*, både som för- och efterled, och de som är mera bestående, vars produktivitet förklaras med att "en viss bildning med ordet i fråga som förled eller efterled väckt sådan anklang att den kommit att stå som mönster för efterföljare. Ofta rör det sig om bildlig användning av ord". (1986:6). Den senare är en del av ett välkänt fenomen i svenskan: en operationalisering av sammansättningsleder. Detta kan illustreras t.ex. med: *riksdagsman* som avser inte bara *män* utan även kvinnor, *bilburen* som inte betyder att en bil bär något, *blixtfull* där det inte är någon blix inblandad, *barnvänlig* där ingen vänlighet förväntas av saker som är *barnvänliga*.

"Många av våra ordbildningsmedel, (...) intar en mellanställning mellan sammansättningsled och suffix." skriver Söderbergh (1967:30) "på detta sätt har vi fått ett suffix som *-lig*, vilket utvecklats ur den fornsvenska sammansättningsleden *-liker*". Även Liljestränd (1975:45) tar upp vad han betecknar som övergång från självständiga morfem till

¹Uppslagsord har hämtats mestadels från Språkbankens lexikaliska databas, affix från Liljestränd (1975) och Thorell (1981)

²Tilläggsordlistor innefattar bl.a. egennamn, datatermer, latinska eller engelska ord.

avledningsmorfem, ordleder som är övergångsformer mellan sammansättning och avledning. Enligt honom

"finns det ett stort antal produktiva för- och efterleder, vars klassificering som sammansättningsleder eller som avledningselement man kan vara tveksam om. Särskilt förefaller det vara fallet när det gäller efterleder. Det utmärkande för dessa efterleder är att de dels förekommer som självständiga ord, dels som delar av sammansättningar. Betydelsen i båda användningarna är densamma. Men efterhand som efterleden börjar uppträda i allt fler sammansättningar börjar den förlora sin ursprungliga, fixerade betydelse och får en allt allmänare innebörd."

Svårigheter med att behandla denna gråzon i svensk ordbildning gör den inte mindre viktig i språkbruket. Övergång från ord till ordleder "har skett oupphörligt under språkets historia" (Söderbergh 1967:30) och ord som är operationaliserade som ordleder har alltid funnits i det svenska lexikonet: de är inte affix och bundna morfem, utan ordleder och bundna lexem. Det krävs en separat artikel för avgränsning av bundna lexem, här karakteriseras de enbart utifrån kriterier som används i den aktuella korpusundersökningen.

Som lexem menas här lexikonenheter definierade i tre parametrar: form, betydelse och funktion i ord- och satsbildning. Definitionen följer Mel'čuks tredelad definition av en lingvistisk symbol: *signifiant*, *signifié*, *syntactics* (1982: 24-28). Det är alltså inte enbart lexikalisering av form och betydelse som kan skilja lexem åt, utan även lexikalisering av funktion. T.ex. är *ha* minst tre olika lexem: *ha* som perfektoperator, *ha* som stödverb och *ha* som innehållsord. Som bundna lexem räknas de ord som svårligen kan klassificeras som fria morfem därför att de används på liknande sätt som bundna morfem, och som inte heller kan klassificeras som bundna morfem därför att de är antingen rotmorfem eller flermorfemiga.

De enheter som eftersökes i extraheringen är alltså inte enbart de med en särskild ordledsform, som t.ex. *justitie-*, *general-* eller *-bent*. Även när de inte har någon särskild ordledsform betraktas de som bundna lexem när en särskild betydelse är knuten till användningen som

sammansättningsled, t.ex. är *vänlig* två lexem: det fria lexemet *vänlig* och det bundna lexemet *-vänlig*.³

Övergångsformer utgör ett problem för statiska ordböcker samtidigt som det är särskilt viktigt att de som lär sig svenska som andraspråk får veta att ett ord har operationaliserats som ordled. En infödd talare undviker intuitivt kompositionell sammansättning där ett ordmönster föreligger och det är just den typen av lexikala kunskaper hos en infödd talare som behöver återspeglas i en aktiv ordbok: t.ex. att *fågelfri* tolkas snarare som 'fri från fåglar' än 'fri som fågeln' (som annars är en helt vanlig fras). Det faktum att adjektivet *fri* även förekommer som huvudled i fraser med liknande betydelse som den som realiseras av ordleden *-fri* minskar inte dess roll som en produktiv ordbildningskomponent, och den aktiva ordboken behöver återspegla detta på något sätt. I synnerhet gäller detta de lexem som är produktiva i ordbildning och vars ordmönster kan beskrivas med selektionsrestriktioner, och inte enbart med kollokationsrestriktioner.

2. Ordleder i ordböcker

Någon transfer av sammansättningskunskaper kan inte väntas från slaviska språk och därför ställs speciella krav på att ordleder i en tvåspråkig ordbok med svenska som källspråk görs åtkomliga. Varken ordleder eller någon vägledning till ordmönster finns att hitta i svensk-polska ordböcker. Sammansättningar finns däremot representerade, ibland som uppslagsord och ibland som användningsexempel. Ett av syftena med projektet är att fylla den luckan. Den svenska delen i den största svensk-polska ordboken (Kubitsky 1988) följer mestadels *Svensk ordbok* (1986), utom just när det gäller ordleder. Om de överhuvudtaget får egna uppslag presenteras de som vanliga ord. T.ex. *-faldig*, *-årig*, *-makeri*, *ärke-* finns inte alls, *betonad*, *riktig*, *sinnad*, *tålig* finns enbart som vanliga adjektiv.

Många ordleder finns i svenska ordböcker men de kan vara svåra att komma åt. Den 12:e upplagan av *Svenska Akademiens ordlista* lägger ännu mer möda än de tidigare upplagorna på att återspegla hur orden är

³ Det faktum att vissa bundna och fria lexem associeras tydligt till en gemensam rot behöver också återspeglas i ordboken, det är dock alltid enklare att fälla samman enheter som har särskiljts än tvärtom.

uppbyggda, men den är huvudsakligen en lista av ord och inte ordleder. Förleder går att hitta i ordlistan men efterleder finns enbart utspridda som exempel: *-fri*, *-säker*, eller *-vänlig* hittas enbart under sammansättningar: *svavelfri*, *tjuvsäker*, *barnvänlig*, m.m. I *Svensk ordbok* och dess efterföljare - *Nationalencyklopedins ordbok* (1995) är ordleder rätt välrepresenterade som uppslagsord. När de är egna uppslagsord får de beteckning *förled* eller *slutled*, annars markeras funktionen med en kommentar. Hos t.ex. adjektivet *fri* i betydelsen 'fri från' finns kommentaren *ofta i sammansättningar*; adjektivet *special* är markerat med *nästan enbart i sammansättningar* men hos adjektivet *säker* markeras inte alls dess efterledsfunktion. Man kan inte vara säker på vad avsaknad av kommentar eller användningsexempel innebär, huruvida ordet antas vara frånvarande i ordbildning eller inte.

Det är alltså ingalunda en regel att ord som förekommer enbart eller företrädesvis i sammansättningar får egna uppslag som sådana, ens i svenska ordböcker. Många finns enbart som användningsexempel. Annorlunda är det i Nyord (1986), där just ordleder fokuseras. Ordboken upprättades som resultat av en undersökning av svenska texter från 1940 till 1980. Den har 7 500 enheter, däribland många produktiva ordleder som inte tas upp i de ovannämnda uppslagsverk, t.ex. *-bank* (*blodbank*), *mjuk-* (*mjukstart*), *när-* (*närdemokrati*), *snabb-* (*snabbval*), m.m. Men vissa ordled som är högfrekventa i den undersökta korpusen visar sig saknas i Nyord, t.ex. *-fylld* (*motsatsfylld*), *-laddad* (*ångestladdad*).

I ryskan är behovet av en speciell ordbildningsordbok inte enbart erkänt utan även tillfredsställt. En ordbok som ägnas uteslutande åt ordbildning kom ut 1996: *Tolkovyy slovar' slovoobrazovatel'nykh edinic russkogo jazyka* (Rysk ordbildningsordbok). Den har 1 875 välbeskrivna uppslagsenheter. Författarinnan T.F.Efremova skriver att idén är c:a 100 år gammal, förverkligandet har dock dröjt. Ett utkast till ett polskt ordbildningslexikon finns hos M. Majewska (1998): *O potrzebie rejestrowania hasel morfemowych* (Om behovet av att registrera morfemiska uppslagsenheter). Artikeln innehåller även en tillbakablick som visar att ordbildning brukade ta upp mycket plats i äldre polska uppslagsverk. Författarinnan argumenterar för behovet av att upprätta en speciell ordbok ägnad uteslutande åt ordbildning, eller åtminstone en lika systematisk och heltäckande förteckning av morfem som av lexem i

en gemensam ordbok. Egentligen borde det vara självklart, skriver Majewska. Om man finner uppslagsordet *auto* 'bil' i en ordbok, varför skulle man inte finna *auto-* 'själv-'? Båda kan förekomma även i ordbildning och behöver separeras i synnerhet då det föreligger homofoni, t.ex. *autoalarm* 'billarm' och *autobiografia* 'självbiografi'.

3. Ordleder i texter

Två korpusundersökningar utfördes inom projektet: den första med fokus på nya verbala sammansättningar som helhet, den senare på leder i dessa; enbart den senare beskrivs här. Med verbala sammansättningar avses här inte enbart de med finita verb som huvudled utan även med particip, både perfekt och presens.⁴ Med nybildningar avses de ord som inte finns representerade som uppslagsord hos LexWare. LexWares uppslagsenheter överensstämmer i stort med de i *Nationalencyklopedins ordbok*, vilket innebär att de sammansättningar som väljs ut som nya inte finns bland uppslagsorden i *Nationalencyklopedins ordbok*, men de kan vara upptagna i t.ex. *Nyord* eller *Svenska Akademiens ordlista*. Undersökningen koncentrerar sig på nybildningar av två skäl: extraheringen väntas utgöra ett komplement till och inte en upprepning av det som redan finns i ordböcker, den skall förse oss med ordmönster som verkligen är produktiva.

89 664 grammatiska ord⁵: infinitiv, finita verb, presens och perfekt particip, finns belagda med 7 357 817 löpord i korpusen på över 42 miljoner löpord. 32 372 av dessa ord utgörs av nybildningar, 57 292 utgörs av redan kända ord. Antalet förekomster i löpord anges nedan parallellt med antalet förekomster i grammatiska ord. Ett urval av de mest frekventa ord i verbala sammansättningar presenteras nedan med ett exempel på en sammansättning som de funnits belagda i. Vad gäller efterleder omfattar antalet förekomster både finita verb, infinitiver och particip. I vissa fall är det inte particip utan endast participiell form, t.ex. motsvaras *färgad* som efterled i de flesta fall av *att ha färg*, inte av *att färga*.

⁴ Även deverbala substantiv undersöks i projektet.

⁵ *Grammatiskt ord och löpord* används som i Svenska Akademiens grammatik (1999:204).

förled	ord	löpord	exempel
<i>ny-</i>	854	4 613	<i>nyöversätter</i>
<i>väl-</i>	568	2 462	<i>välöverbägt</i>
<i>svår-</i>	304	925	<i>svåröverbakade</i>
<i>special-</i>	205	639	<i>specialvävda</i>
<i>färdig-</i>	270	631	<i>färdigåkt</i>
<i>små-</i>	209	451	<i>småöva</i>
<i>hög-</i>	107	451	<i>högväxte</i>
<i>stor-</i>	97	430	<i>störväxte</i>
<i>själv-</i>	196	413	<i>självöppnande</i>
<i>halv-</i>	192	396	<i>halvöppnade</i>

Tabell 1. Utdrag ur de mest frekventa förleder

verb i efterled	ord	löpord	exempel
<i>fylla</i>	354	1 236	<i>adrenalinfylld</i>
<i>bygga</i>	152	1 030	<i>amatörbyggda</i>
<i>spela</i>	197	879	<i>bollspelande</i>
<i>äga</i>	192	820	<i>vitägda</i>
<i>skada</i>	240	717	<i>ögonskadade</i>
<i>gå</i>	78	678	<i>östgående</i>
<i>likna</i>	439	667	<i>öliknande</i>
<i>färga</i>	305	600	<i>ölfärgad</i>
<i>betala</i>	76	575	<i>efterbetalt</i>
<i>köra</i>	133	548	<i>amatörkörd</i>

Tabell 2. Utdrag ur de mest frekventa verb i efterleder

Många sammansättningar skapas tillfälligt och därför finns det en risk att de återspeglar den dominerande texttypen (tidningstext) snarare än de dominerande ordmönstren i språket generellt. Frekvens behöver kompletteras med andra, mera lexikaliskt inriktade parametrar för att uppnå större träffsäkerhet. Två sådana har definierats och använts i extraheringen: sammansättningsbenägenhet och sammansättningsproduktivitet.

1. Sammansättningsbenägenhet

För alla lexikonverb som finns belagda i korpusen har alltså dessas sammansättningsbenägenhet beräknades enligt följande:

$$\frac{\text{förekomster av verbet i sammansättningar}}{\text{samtliga förekomster av verbet}}$$

Måttet anger hur ordets användning i ordbildning är relaterad till dess användning i satsbildning. Om textunderlaget antas vara tillräckligt stort, kan sammansättningsbenägenhet beräknas även i antal grammatiska ord och inte bara i löpord. Den kan beräknas för verb som lexikonord samt för olika verbformer separat.

Nedan presenteras den genomsnittliga sammansättningsbenägenheten hos verb. Beräknad i förekomster blir den genomsnittliga sammansättningsbenägenheten låg p.g.a. högfrekventa hjälpverb. Beräknad i grammatiska ord är den däremot mycket hög, inte bara hos particip utan även hos finita verb former och infinitiv. Om de verb som inte alls belagts i sammansättningar inte tas med i beräkningen av den genomsnittliga sammansättningsbenägenheten, blir den mycket högre för de verb som överhuvudtaget förekommer i ordbildning. Även om ett par procent räknas bort som fel, visar resultaten ändå tydligt att sammansättning utgör en betydande andel i nybildning även när det gäller verb.

	samtliga verb dvs sammansättnings- benägenhet ≥ 0		enbart verb med sammansättnings- benägenhet > 0	
	löpord	ord	löpord	ord
samtliga former	1.08	36.10	2.25	55.35
finita och infinitiv	0.24	22.92	0.62	45.31
presens particip	15.89	58.31	24.00	78.81
perfekt particip	3.74	40.47	4.71	56.03

Tabell 3. Genomsnittlig sammansättningsbenägenhet hos verb

Verb som visar sig helt saknas i ordbildning är som väntat hjälpverb: *bli, bör, lär, kunna, måste, skall, vilja*. Vissa verb förekommer inte i sammansättningar som finita former men gärna som presens particip, t.ex. *behöva*. Det är en intressant och öppen fråga varför vissa innehållsverb med mycket hög frekvens i de undersökta texterna inte alls förekommer i sammansättningar. Nedan presenteras en rätt brokig lista av verb som har hög frekvens men är helt sammansättningsobenägna, åtminstone som finita former eller perfekt particip.

<i>befara</i>	<i>hävda</i>	<i>medverka</i>	<i>sluta</i>	<i>uttrycka</i>
<i>betyda</i>	<i>inleda</i>	<i>minska</i>	<i>tillhöra</i>	<i>veta</i>
<i>fortsätta</i>	<i>innebära</i>	<i>påpeka</i>	<i>undvika</i>	<i>våga</i>
<i>föredra</i>	<i>innehålla</i>	<i>riskera</i>	<i>understryka</i>	<i>värna</i>
<i>förefalla</i>	<i>inse</i>	<i>räcka</i>	<i>upphöra</i>	<i>ytra</i>
<i>försöka</i>	<i>instämna</i>	<i>sakna</i>	<i>uppmåna</i>	<i>ägna</i>
<i>gilla</i>	<i>kosta</i>	<i>ske</i>	<i>uppmuntra</i>	<i>övertaska</i>
<i>gälla</i>	<i>meddela</i>	<i>skylla</i>	<i>utesluta</i>	<i>övertyga</i>
<i>hänvisa</i>	<i>medföra</i>	<i>slippa</i>	<i>utgöra</i>	

Det är dock inte verb med låg, utan med hög sammansättningsbenägenhet som är mest relevanta för ett sammansättningslexikon. Nedan presenteras de ordnade enligt frekvens och sammansättningsbenägenhet. Sammansättningsbenägenheten har räknats fram på samma sätt som genomsnittet för alla verb.

efterled	benägenhet	löpord
<i>fylla</i>	17.42	1 236
<i>bygga</i>	5.58	1 030
<i>spela</i>	2.82	879
<i>äga</i>	11.48	820
<i>gå</i>	0.60	678
<i>skada</i>	11.65	717
<i>betala</i>	3.50	575
<i>likna</i>	19.87	667
<i>köra</i>	29.47	529
<i>skriva</i>	2.12	534

Tabell 4. Verb med hög frekvens och hög sammansättningsbenägenhet

1. Sammansättningsproduktivitet

Sammansättningsproduktivitet är en parameter som skapats för att ytterligare höja träffsäkerheten i urvalet av ordleder. Den är särskilt inriktad på att välja de ordleder som allra troligast kan utgöra bundna lexem. Parametern fungerar som ett komplement till både frekvens och sammansättningsbenägenhet och tillför den dimension som lexikaliskt sett verkar viktigast.

För de lexikonenheter som identifierats i sammansättningar har dessas sammansättningsproduktivitet beräknats enligt följande:

lexikonenheter förekommande tillsammans med den valda enheten/
samtliga sammansättningar med den valda enheten

Produktivitet kombinerad med frekvens och sammansättningsbenägenhet ger en mycket värdefull sortering. De listor av för- och efterleder som presenteras nedan visar detta tydligt.⁶

förled	produktivitet (%)	antal ord	antal kollokater	exempel
<i>ny-</i>	42	854	358	<i>nybildade</i>
<i>väl-</i>	45	568	258	<i>välskötta</i>
<i>färdig-</i>	61	270	166	<i>färdigbehandlad</i>
<i>svår-</i>	53	304	162	<i>svårtolkat</i>
<i>små-</i>	71	209	149	<i>smådjävlat</i>
<i>lätt-</i>	65	214	139	<i>lättanvända</i>
<i>själv-</i>	70	196	138	<i>självkrönt</i>
<i>halv-</i>	62	192	119	<i>halvciviliserade</i>
<i>snabb-</i>	69	133	92	<i>snabbinkallad</i>
<i>fel-</i>	53	172	91	<i>felklädd</i>
<i>hård-</i>	61	143	88	<i>hårdbantad</i>

Tabell 5. Högproduktiva förleder

⁶ Med *kollokat* avses endast tillsammansförekommande lexikonenheter och inte skillnaden mellan *bas* och *kollokat*.

efterled	produktivitet (%)	antal ord	antal kollokater	exempel
<i>-liknande</i>	99	439	433	<i>dvalliknande</i>
<i>-fylld</i>	58	354	207	<i>adrenalinfylld</i>
<i>-baserad</i>	68	250	169	<i>bildbaserad</i>
<i>-färgade</i>	55	305	168	<i>blodfärgade</i>
<i>-klädda</i>	66	255	168	<i>affärsklädda</i>
<i>-inriktad</i>	59	269	160	<i>barinriktad</i>
<i>-styrd</i>	65	217	142	<i>bildstyrd</i>
<i>-formad</i>	59	236	140	<i>bananformade</i>
<i>-intresserad</i>	55	210	116	<i>fotbollsintresserad</i>
<i>-relaterade</i>	68	161	109	<i>inkomstrelaterade</i>

Tabell 6. Högproduktiva efterleder

1. Lexikologisk relevans hos urvalen

En intelligent korpusundersökning kan vara till hjälp i lexikologiskt arbete på många olika sätt.

Den mest uppenbara nyttan med att extrahera nybildningar är att ens lexikon kan kompletteras med nytillkomna lexikaliseringar. Vissa nya sammansättningar verkar förtjäna egna uppslag i ordboken, t.ex. *intelligensbefriad*, *kritikerrosad*, *långtgående*, *pådrivande*, *räntebärande*, *synliggöra*, *återuppstå*, m.m. Även bland icke-sammansatta nya verb finner vi möjliga uppslagsverb, som t.ex. *destabilisera*, *desinformera*.⁷

För en aktiv ordbok är dock de nya sammansättningar som följer ett produktivt ordmönster mest intressanta. Om korpusen t.ex. visar att de högfrekventa efterlederna *-fylld* och *-laddad* utgör ett produktivt ordmönster bör dessa tas upp som uppslagsenheter och förses med både betydelse och form som är specifik för användningen som efterled. Korpusundersökningen visar en påtaglig homogenitet i användning av *-fylld* och *-laddad*. I få fall är det fråga om att något verkligen 'fylls' eller 'laddas' bokstavligen, även om sådana fall förekommer, t.ex. *Liberoladdad*. De flesta belägg avser en bildlig användning, som även

⁷ I Nationalencyklopedins ordbok finns t.ex. endast *destabiliserande*, *desinformation*.

finns i sådana konventionaliserade sammansättningar som *farofylld* och *känsloladdad*; t.ex. *prestigefylld*, *lustfylld*, *skräckfylld*, *kampfylld*, *motsatsfylld*, *prestigeladdad*, *ångestladdad*.

Bland leder som utgör ordmönster utan att någon metafor är inblandad men som ändå är specialiserade som ordled återfinns *hand-*, *ny-*, *special-*, vilka egentligen betyder 'manuellt', 'nyligen', och 'särskilt'. Att *hand-* som en verbal förled kommer av frasen *för hand* inte substantivet *hand*, visas i urvalet av dess sammansättningar.

<i>handblåst</i>	<i>handfärgade</i>	<i>handrullad</i>	<i>handsätta</i>
<i>handbroderat</i>	<i>handknutna</i>	<i>handsmidda</i>	<i>handtextade</i>
<i>handbyggd</i>	<i>handknypplade</i>	<i>handsnidad</i>	<i>handtryckta</i>
<i>handdrejade</i>	<i>handkolorerade</i>	<i>handstickad</i>	<i>handvevad</i>
<i>handdrivna</i>	<i>handritad</i>	<i>handsydd</i>	<i>handvirkad</i>

En rangordning av verb enligt produktivitet fokuserar de verb som är produktiva efterleder och vars kollokationsmönster i ordbildning är välrepresenterade i texter. Extrahering av kollokater är särskilt viktig i det aktuella projektet då det skall kartlägga hur verbets valensram realiseras i sammansättningar. Ett exempel på ett urval av olika förleder till ett specifikt verb presenteras nedan.

<i>behovsstyrda</i>	<i>lagstyrda</i>	<i>målstyrda</i>	<i>röststyrda</i>
<i>centralstyrda</i>	<i>lättstyrda</i>	<i>planstyrda</i>	<i>självstyrda</i>
<i>datorstyrda</i>	<i>maffiastyrda</i>	<i>radarstyrda</i>	<i>statsstyrda</i>
<i>detaljstyrda</i>	<i>maktstyrda</i>	<i>regelstyrda</i>	<i>svårstyrda</i>
<i>föräldrastyrda</i>	<i>militärstyrda</i>	<i>resultatstyrda</i>	<i>terminsstyrda</i>
<i>högerstyrda</i>	<i>miljöstyrda</i>	<i>rollstyrda</i>	<i>toppstyrda</i>
<i>indiskstyrda</i>	<i>moderatstyrda</i>	<i>räntestyrda</i>	<i>trådstyrda</i>

En värdefull upplysning om egenheter hos ordleder får man genom att extrahera och jämföra närsynonyma ordleder som förekommer med en och samma ordled. T.ex. används endast *väl* i förled som en allmän positiv bestämning vid particip, det närbesläktade *bra* förekommer inte alls i den rollen. Jämför man användning av *färdig-* och *slut-*, samt *hel-* och *total-* visar det sig att *färdig-* och *hel-* förekommer mest med perfekt particip medan *slut-* och *total-* även förekommer med finita verb.

färdig-	slut-	verb i efterleder
<i>färdigbehandlad</i>	<i>slutbehandlad</i>	<i>behandla</i>
<i>färdigbesiktigad</i>	<i>slutbesiktigas</i>	<i>besiktiga</i>
<i>färdigförhandlad</i>	<i>slutförhandla</i>	<i>förhandla</i>
<i>färdigräknade</i>	<i>sluträknad</i>	<i>räkna</i>
<i>färdigspelad</i>	<i>slutspelad</i>	<i>spela</i>

Tabell 7. *Färdig-* och *slut-* med samma verb i efterleder

hel-	total-	verb i efterleder
<i>helstoppad</i>	<i>totalstoppa</i>	<i>stoppa</i>
<i>helmissade</i>	<i>totalmissa</i>	<i>missa</i>
<i>helrenoverade</i>	<i>totalrenovera</i>	<i>renovera</i>

Tabell 8. *hel-* och *total-* med samma verb i efterleder

På liknande sätt kan man jämföra antonyma ordleder som förekommer med en och samma ordled. Att förse dessa med ekvivalenter i ett annat språk hjälper även till att avgöra vilka som är viktigast att ta upp som uppslagsord. T.ex. kräver sammansättningar med *stor-* och *små-* egna ekvivalenter, medan *lätt* och *svår* verkar vara kompositionella även i engelskan.

stor-		små-	
<i>storförtjust</i>	mighty delighted	<i>småförtjusta</i>	rather delighted
<i>storskrattade</i>	laugh loudly	<i>småskrattade</i>	chuckle
<i>störväxte</i>	tall of stature	<i>småväxte</i>	short of stature

Tabell 9. *stor-* och *små-* med samma efterleder

lätt-		svår-	
<i>lättavlästa</i>	easy to read from	<i>svåravläst</i>	difficult to read from
<i>lättadministrerad</i>	easy to administrate	<i>svåradministrerad</i>	difficult to administrate
<i>lättanalyserat</i>	easy to analyse	<i>svåranalyserat</i>	difficult to analyse

Tabell 10. *lätt-* och *svår-* med samma efterleder

1. Slutord

Bundna lexem är både svårast och viktigast att ta upp i tvåspråkiga ordböcker. En klar vägledning behövs för en icke-svensk både vad gäller lexikaliserade sammansättningar och sammansättningsmönster med bundna lexem. Bundna lexem är viktiga även i svenska ordböcker. Korpusundersökningens resultat tyder på att sammansättning är så pass produktiv i författande av skriven svenska att bara detta motiverar ett modernt aktivt lexikon ägnat enbart åt ordbildning. Alla produktiva för- och efterleder skulle få egna uppslag i en sådan ordbok. Man skulle ha möjlighet att slå upp en enskild ordled på samma sätt som man slår upp ett ord i en ordbok, och finna alla relevanta uppgifter om dess form, betydelse och användning samlade på en plats, som t.ex. i den ovannämnda ryska ordboken.

Den typ av intelligent extrahering som presenteras ovan är värdefull i lexikologiskt arbete. Med ett tillräckligt brett textunderlag kan huvudfunktionen hos en enhet spåras i texter. Med tanke på att övergångsformer inte bara är enheter utan även en process, behövs den typen av extrahering från nya texter nästan jämt. Med denna som hjälp kan en statisk ordbok förses med ett dynamiskt komplement, där uppdateringar automatiskt skulle följa nytillkomna texter.

Litteratur

- Efremova, T.F. 1996. *Tolkovyj slovar' slovo-obrazovatel'nych edinic russkogo jazyka*. Moskva: Russkij jazyk.
- Hellberg, Staffan. 1978. *The Morphology of Present-Day Swedish*. Stockholm: Almqvist och Wiksell.
- Johannisson, Ture. 1964. Tendenser i nutida svensk ordbildning. I: *Nutidssvenska*. Stockholm: Almqvist & Wiksell.
- Josefsson, Gunlög. 1997. *On the principles of word formation in Swedish*. Lund: Lund University Press.
- Kubitsky, Jacek. 1988. *Słownik szwedzko-polski. Svensk-polsk ordbok*. Stockholm: Natur och Kultur.
- Liljestränd, Birger. 1975. *Så bildas orden*. Lund: Studentlitteratur.

- Majewska, Małgorzata. 1998. O potrzebie rejestrowania haseł morfemowych w słownikach języka polskiego. I: *Poradnik językowy* 8-9. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Malmgren, Sven-Göran. 1993. Sammansättningsmorfologi och lexikografi. I: *Nordiske studier i leksikografi* II. Oslo: Nordisk Forening for Leksikografi.
- Mel'čuk, I.A. 1982. *Towards a Language of Linguistics*. München: Wilhelm Fink Verlag.
- Noreen, Adolf. 1906. *Vårt språk*, del VII. Lund: CWK Gleerups Förlag.
- Nationalencyklopedins ordbok*. 1995. Höganäs: Bra Böcker.
- Nyord i svenskan från 40-tal till 80-tal*. 1986. Stockholm: Esselte Studium.
- Språkbanken. 1999. www.svenska.gu.se
- Stockholm-Umeå-Corpus. 1999. www.ling.su.se/DaLi/Projects/SUC/
- Svensk ordbok*. 1986. Stockholm: Esselte Studium.
- Svenska Akademiens grammatik*. 1999. Stockholm: Norstedts ordbok.
- Svenska Akademiens ordlista över svenska språket*. 1998. Stockholm: Norstedts Ordbok.
- Söderbergh, Ragnhild. 1968. *Svensk ordbildning*. Stockholm: Nämnden för svensk språkvård.
- Teleman, Ulf. 1972. *Om svenska ord*. Lund: Gleerups.
- Thorell, Olof. 1981. *Svensk ordbildningslära*, Stockholm: Esselte Studium.