

Jonas Löfström
 Språkdata
 Göteborgs universitet

DOLDA ORDBILDNINGSMÖNSTER. NÅGRA PROBLEM INOM DATAMASKINELL LEXIKOLOGI

Med dolda ordbildningsmönster avses de regelbundenheter i vårt ordförråd som inte framträder i ordböcker så som dessa brukar vara uppbyggda. Det lingvistiska problem vi här har att lösa med datamaskinella metoder är hur ordbildningsstrukturerna i språket skall kunna registreras och presenteras i bearbetningar av lexikaliskt material. En utgångspunkt är att det vid lexikologiskt arbete ofta kommer fram information som inte är direkt avsedd för publicering respektive som publiceras men i ostrukturerad form. Med hjälp av datamaskinen är det möjligt att strukturera och lagra informationen så att den senare blir lätt åtkomlig. Man bygger upp ett välordnat lager av data. Detta kan ske på olika sätt rent tekniskt, men det går vi inte in närmare på här.

När vi talar om dolda ordbildningsmönster menar vi alltså att den alfabetiska förteckningen endast i begränsad utsträckning direkt synliggör ordbildningsmekanismen. En relativt vanlig och samtidigt enkel åtgärd för att råda bot på detta är att göra en finalalfabetisk sortering av det lemmatiserade ordmaterialet. Många sammansättningstyper och avledningstyper framträder då. Men redan det faktum att prefigurerade ord kan fungera som andra led i en sammansättning och att suffixavledningar kan följas av ett led i en sammansättning gör att den information vi får i en initial- respektive finalalfabetisk listning inte är fullständig. På motsvarande sätt uppträder vanliga grundmorfer på olika håll i orden. Det kan vara med praktiskt taget oförändrad betydelse eller med större eller mindre betydelseförskjutningar. För en komplett lexikonbeskrivning av ett språk behöver vi även denna information om grundmorfernas distribution. Detta har inte bara ett stort teoretiskt intresse. Den praktiska betydelsen av sådan kunskap torde vara uppenbar.

Den information som ligger dold eller osystematiserad i de flesta ordböcker kan man alltså lyfta fram. Det gäller då att veta vad som är relevanta upplysningar och att sedan bearbeta och sortera dem på ett konsekvent sätt. Om man lyckas lagra sin information väl kan man sedan få svar på sina frågor och till och med på oställda frågor. Det är det kanske mest lockande perspektivet – att genom ett konsekvent analysförfarande avslöja oväntade och okända egenskaper i ordbildningssys-

temet. Man står då inför uppgiften att modifiera sin modell.

Om man utarbetar en ordbok med annat huvudsyfte än att beskriva ordbildningsstrukturer ställs man inför valet att göra den morfologiska beskrivningen samtidigt med det övriga eller att göra en separat morfologisk omgång. Mycket talar för att en preliminär genomgång av materialet med avseende på segmentering (av orden), separering (av homonyma enheter) och klassificering av enheterna i morfem (d v s fastläggande av tillåten allomorfi) bör föregå det övriga arbetet.¹⁾ Vid arbete med datamaskinella metoder finns ju då all möjlighet att med utgångspunkt i regler och/eller en förberedande analys av ett begränsat material låta datorn presentera förslag som lexikologen har att acceptera, förkasta eller justera. Resultatet torde lättare bli enhetligt och arbetet gå snabbare med en sådan uppläggnig. Det innebär bl a att man har en ram att arbeta inom vilket gör att problem och egenheter lättare uppmärksammas. Annars blir det av förklarliga skäl lätt så att man bortser från det som man inte just för tillfället skall beskriva. Detta vill säga att frågor om kommutation, d v s vilken segmentering av orden man kan tillåta sig, morfotax (vilka morfemtyperna och deras funktion är) och allomorfi (vilken växling som ska accepteras) aktualiseras. Av det sagda framgår klart att ett fast format för den information man vill kunna ge är en stor fördel i arbetet. Till det som redan sagts kommer frågan om polysemi. Både vad det gäller grundorden och affixen är betydelsen viktig för att förstå ordbildningsmönstren.

Morfotax

Det är av grundläggande betydelse vilken morfotaktisk information man bestämmer sig för att ta hänsyn till och redovisa. I viss mån beror valet på vilket språk det gäller men även syftet kan spela in. För de nordiska språken är väl en rimlig indelning: prefix, grundmorf(er), fog, avledningssuffix samt böjningsändelse. Denna senare har en marginell betydelse för ordbildningen sett från en lexikalisk synpunkt. Till det nämnda kommer också infix som en enhet som kan visa sig vara en alternativ lösning när man stöter på problem i analysen av ordförrådet.

Kommutation

Fastställandet av ordledstyper ger oss en fastare ram för kommutationen. Det får i det enskilda fallet avgöras på vilka grunder kommutationen kan ske. I ett väl beskrivet språk har man ju ett försprång i det att en lång rad suffix och prefix är kända och väldefinierade. Med utgångspunkt i en förteckning över dessa förenklas kommutationsfasen. Det är dock

uppenbart att varje någorlunda stort material kommer att bjuda på svårigheter och att lösningen där till stor del får bero på syftet med undersökningen. Förutom hänsyn till språklig referensram (språkkänsla, kompetens), morfotaktiska mönster och semantisk affinitet har man att beakta variationen. Att *-are* och *-ing* är avledningssuffix i svenskan råder ingen tvekan om, men hur är det med *-ats*? Om man vid sidan av *segla* och *segla* har *segla* så torde man få notera *-ats* som ett substantivavledande suffix, även om det knappast är produktivt. Svårare är de fall där det som föregår ett eventuellt suffix inte finns i något annat ord, t ex *giljotin* och *magasin*. Det är först om man på andra grunder kunnat etablera *-in* som ett acceptabelt suffix som kommutationen här kan ske och ge restmorfemet *giljot-* respektive *magas-*.

Allomorfi

Det är inte ett helt trivialt problem att avgöra vilken allomorfi man skall räkna med. Skall man acceptera endast etymologiskt riktig växling eller även synkront sett rimlig variation? Frågan om semantikens roll vid bedömningen är avgörande. Syftet med den aktuella undersökningen får avgöra.

Den inomlemmatiska variationen som ju oftast finns förtecknad är en god utgångspunkt eftersom den ibland även fungerar mellan lemmen. En del ordbildning har ju historiskt sett skett med utgångspunkt just i någon av lemmats böjningsformer. I andra fall finns det liknande regulariteter mellan olika typer av ord i ordförrådet. Det ger oss då en sammanhållen typ av växling som det kan råda stor enighet om och som oftast speglar en tidigare produktivitet. I många fall har dessa regulariteter sannolikt också en psykologisk relevans. Även andra typer av regelbunden växling kan noteras. Det gäller t ex fonologisk (och grafonomisk) variation inom en morfotaktisk ram.

Polysemi

När det gäller betydelsen är den morfologiska nivån svårberäknad. En viss spännvidd i betydelsen inom ett morfem är nödvändig för att instanser från olika ord överhuvudtaget skall kunna föras samman i ett morfem. Men en gräns för hur abstrakt betydelsen får vara måste ändå sättas. Är t ex *sätt* i *sätta*, *bosättning*, *färgsättning*, *ifrågasätta*, *fortsätta*, *sättmaskin*, *besättning*, *undsättning* och *översättning* samma morfem? Det skall i detta sammanhang framhållas att olika delar av ordförrådet naturligen inte ligger på samma abstraktionsnivå. Speciellt det klassiska och romanska språkgodset bjuder på tydliga ordbildningsmönster men med grundmorfer vars betydelse för en nordisk språkkänsla är mycket vag.

Prefix och suffix företer en liknande skala av betydelser. Men redan en indelning av dem i typer (person- eller verktygsbeteckning, abstrakt substantiv o s v) ger en nyansering av betydelsestrukturen.

Ordförbindelser

Till ordbildning i strikt mening hör kanske inte förbindelser av ord, men viss fraseologi bör nog noteras för att fullständiga bilden av ordförrådets struktur. Att *hand* återfinns i *handtag*, *handgemäng*, *hantverk* (med allomorfen *hant*), *egenhändig* (*händ*) och *valhänt* (*hänt*) är klart, men det ligger nära till hands att redovisa även förekomsten i fraser som *i första hand*, *i sista hand*, *ta hand om* o s v. Jämför *omhändertagande* och *förstahandsinformation*.

Materialtyper

För lexikologiska syften torde man nästan helt kunna nöja sig med lemmatiserade ord som utgångsmaterial. En viss kontroll mot löpande text kan dock vara påkallad. En del ord förekommer endast i pluralis (t ex *vara i faggorna*) eller i bestämd form (*vara i görningen*) och de skall naturligtvis redovisas på något sätt. En typ av sammansättningar, *1900-talet*, förekommer knappast i obestämd form singularis, men man skulle ändå vilja registrera ordbildningstypen som ju har den egenheten att förledet är utbytbart men inte är av den typ som förtecknas i ordböcker. Detsamma gäller abbreviationer.

För att få fram språkets böjningsmorfologi behövs för de syften lexikologiskt arbete avser i ett väl beskrivet språk inte någon kommutation av det fullständiga olemmatiserade materialet. För vissa ord med mera fraseologisk användning samt för kontroll av polysemi är däremot den löpande texten värdefull som komplement. Dessutom innehåller böjningsformerna en hel del av den allomorfi som det är praktiskt att redovisa i ordbildningssammanhang.

Lagring

Lagringen av information torde kunna ske på olika sätt och får rättas efter syftet. En uppdelning av orden i morfem med möjlighet till märkning av dem är en första förutsättning. Vidare bör märkningen vara åtkomlig för sökning och sortering samt kunna avse olika nivåer eller typer av analys. Eventuellt kan information om alternativa morfemgränser vara aktuell. Det är då att notera att morfemens status och inbördes koppling noggrant bör kontrolleras så att resultatet av

maskinella bearbetningar av ett stort material inte blir felvisande. Ett kommutationstest kan ju inte utföras på båda analyserna av ett ord samtidigt. En annan möjlighet vid kommutationen är att laborera med olika starka morfemgränser och redovisa dem.

Några exempel

I det följande tar jag upp några exempel på de typer av problem som man ställs inför vid en morfologisk analys. De illustrerar också i viss mån de ovan nämnda aspekterna av morfologin.

Sammansättningar

Sammansättningarna kan vara två- eller flerledade. De tvåledade är hierarkiskt uppbyggda: *riksbanks-fond*, *arbetsmarknads-styrelse*, *krigs-skadestånd*, *turisttrafik-förening*. Bindestrecken här markerar en primär gräns. Men i leden *riksbanks*, *arbetsmarknads*, *skadestånd* och *turisttrafik* finns ytterligare, sekundära gränser. Dessa olika gränser kan markeras på önskat sätt för att möjliggöra en kartläggning av bruket av fogar (i det här fallet -s-) i olika sammansättningar. Kopulativa (flerledade) sammansättningar av typ *dansk-norsk-svensk* har inte den ovan beskrivna hierarkiska uppbyggnaden. De aktualiserar också ett ordfogningssätt som använder icke-alfabetiska tecken. Det kan vara värt att uppmärksamma eftersom en vacklan i bruket av böjningsändelser (jfr nedan) tycks göra sig gällande.

En märkning av ordleden med avseende på ordklass kan ge en kompletterande bild av ordbildningen. I många fall kan det vara svårt att från en synkron synpunkt avgöra ett sammansättningsledds ordklassstillhörighet. Är *spelhåla* sammansatt av *spel* (nomen) eller *spela* (verb) och *låda*? Möjligen kan etymologisk information vara intressant här.

Avledda sammansättningar utgör ett annat problem. Det gäller att återge den annorlunda strukturen och eventuellt antyda den bakomliggande syntagmen. *Blåögd* ('med blå ögon') står ensamt utan något **ögd* att falla tillbaka på. Morfemet *öga* saknar verbavledning men har här i sammansättningen en participliknande form och adjektivisk funktion. En märkning med ordklassstillhörighet på *ög* och angivande av -d:s funktion ger en sammanställning som jämförd med andra liknande kan ge perspektiv på ordbildningsmönstren.

Det normala fallet är att förleden i en sammansättning är oböjd. Som vi nyss antydde kan den också utgöras av stammen så att ordklassen inte framgår. Dessutom finns fogalement: s

eller en vokal. Böjd form i egentlig mening förekommer också i några fall och är därför av intresse för förståelsen av (produktiv) ordbildning. Substantiv kan som förled stå i pluralis, *ländergrupp*, vilket kanske bäst noteras genom en markering på pluralmorfemet. Att nöja sig med att urskilja hela ordet *länder* eller bara *land* kan ställa till problem vid en mer generell behandling av allomorfi. Bestämd form i sammansättningar, *toppenbra*, *bottennapp*, tycks också kunna härledas från en syntagm. Vissa förled kan dock bli så produktiva att de snarare får betraktas som en ny enhet skild från simplex, t ex *toppen-* vid sidan om *topp*. Att detta får konsekvenser för analysen är uppenbart. Därför är en komplett kartläggning av sådana fall av bestämd form i förleden en hjälp när man kommuterar fram grundmorferna.

I vissa fall förekommer inre böjning av adjektiv, där båda leden är samordnade, *brittiskt-franskt*. Analysen bör lagras så att bruket av böjning inne i sammanställningen kan relateras till dels slutledets böjning, dels bruket av ickealfabetisk fog. Även komparationsformer kan ingå i ordbildningen: *mindretal* (jfr *mindervärdeskomplex* med en allomorf *minder*), och *förstfödd*. Att registrera dessa morfem och deras formella och semantiska egenskaper kan skapa förutsättningar för en detaljerad ordbildningslära.

De verb som har både s k lös och fast sammansättning utgör ett område som kan sprida ljus över intressanta ordbildningsprinciper. Med en genomtänkt analys och klassificering kunde man sammanställa olika fall av den här typen: *avbryta*, *bryta av*, *avbrott*, *avbruten* och eventuellt fler. Som synes kommer här även substantiv(avledning) in i bilden. Detta är ett bra exempel på behovet av ett effektivt grepp på den allomorfiska växlingen. Morfemet *bryt* har här allomorferna *bryt*, *brut* och *brot(t)*. Kan dessa presenteras samlat är mycket vunnet i åskådlighet. De semantiska egenskaper som finns här gör det fördelaktigt att även beakta lexem med mer än ett ingående ord. De sammansättningar som innehåller ickealfabetiska tecken, numeraler, abbreviationer, egennamn o s v saknas ofta i lexikon. Om man har tillgång till löpande text kan man få en bild av en mycket utbredd ordbildningstyp. Ex.: *1900-tal*, *50-kort*, *ABF-cirkel*, *Uddevallabo*. Det är inte helt orimligt att tänka sig en specialbehandling av dessa så att de åtminstone redovisas på efterleden.

Det finns en del sammansättningsled som fått en så allmän användning att de närmar sig prefix- eller suffixstatus. *Ur* i *urdum*, *uråldrig* och *vänlig* i *barnvänlig*, *riktig* i *sittriktig*. Skall dessa fall föras upp som egna enheter eller inte?

Avledning

I sammansättningar har man grovt taget att göra med hela ord som kombineras med varandra. I vissa fall faller ett ordbildningselement i slutet av förleden: *yxa* och *skافت* ger *yxskaft*. Låt oss kalla den del av ordet som återstår när ordbildningselementet avlägsnats för grundmorf.

Avledning blir då ordbildning med tillägg av prefix eller suffix till grundmorfen. När det gäller suffix är det vanligt med grupper av ord som hör ihop betydelsemässigt och skiljs åt genom olika suffix eftersom de har olika syntaktisk funktion: *vals*, *valsa*, *valsare*, *valsning*. För vidare analys och bearbetning är alltså uppgift om vilken ordklass suffixet bildar ounggänglig. Vissa suffix kan bilda olika ordklasser, så t ex *-a*: *kratta* (verb), *kratta* (nomen). Detta är ytterligare ett skäl att märka dem för att kunna göra relevanta bearbetningar.

Prefixen ger inte av egen kraft ordet (ny) ordklass. I viss mån skulle de därför kunna sägas ha en starkare semantisk innebörd. I fall som det negerande *o-* i *oordning* framträder detta drag. Det är emellertid vanligt att prefixet endast har en lätt nyanserande funktion, *fästa* - *befästa*, ofta på gränsen till rent formell funktion. Att så är fallet har bl a att göra med att många prefix kommit in via lån av hela ordet. Ett speciellt problem med avledningarna är att vi har ett germanskt och ett romanskt mönster. Som vi skall se ställer detta till problem bl a vid kommutationen.

Prefixen har ofta en mycket vag betydelse men kan ändå kommu-teras fram då de förekommer i många olika ord: *su-* i *suspekt* och *suterrängvåning*, *pro-* i *procent* och *producent*, *be-* i *beredskap* och *betyg*, *ge-* i *angelägen* och *gestalt*.

En grundmorf som *duk* i *introduktion*, *obduktion*, *produktion*, *konduktör* och *reduktion* framstår tydligt som en byggsten i vårt ordförråd. Simplex saknas (i den här aktuella betydelsen) och det är först vid analys som vi identifierar morfen i det enskilda ordet. Däremot accepterar språkkänslan lättare *produkt* i *produktion* och *produkt* som meningsfulla ordled. Om man tar med även *producera*, *producent* så framstår samhörigheten klart. Vi har då också en allomorfisk växling *produc/produkt*. Det finns anledning att acceptera olika nivåer i ordbildningen. De på olika kommutationsgrunder upprättade gränserna i orden kan med fördel hållas isär.²⁾ Om man utgår ifrån *duc/duk* kan man få följande lista:

<i>in duc era</i>	<i>konduk tör</i>
<i>in duk tion</i>	<i>obduc era</i>
<i>intro duc era</i>	<i>obduk tion</i>
<i>intro duk tion</i>	<i>produc era</i>
<i>via duk t</i>	<i>produkt tion</i>
<i>akve duk t</i>	<i>produc ent</i>
	<i>produkt t</i>

I den vänstra spalten identifieras en betydelse 'leda fram' hos *duc/duk* medan man i exemplen i högra spalten måst tillgripa en konkatenering av prefix och grundmorf för att få en psykologiskt godtagbar analys. På sitt sätt blir detta en kombination av synkrona och diakrona (etymologiska) aspekter.

Suffixen bildar flera räckor som *lära, lärare, (in)läring, lärarinna, lärning*, där suffixets modifiering av grundmorfen är ganska klar och förutsebar. Många analogibildningar förekommer. Lexikaliseringar förekommer givetvis. Även det klassiska eller romanska materialet innehåller den här typen av mönster. Ta t ex *informera, information, informator, informatör, informativ*. Ett påtagligt problem om man behandlar olika delar av ordbildningen separat utgör de inkonsekvenser som blir följden. Det är inte ovanligt att man i en svensk ordbildningslära hittar exempel som *arrang-era, arrange-mang; medic-in, medika-ment; arrendat-or* (trots *arrendera*); *reforma-tor* (trots *reformera*). Beskrivningen av ordförrådet och ordbildningsprinciperna kan inte bli annat än fragmentarisk om man inte beaktar denna typ av regelbundenhet.

En del avledningar innehåller svårplacerade infix: *beskaffenh*et kopplas naturligen ihop med *beskaffad* och avledningssuffixet *-het* står för abstrakt substantiv. Men *-en-* låter sig inte så lätt föras till någondera av de ovan nämnda ordleden. Vi får helt enkelt räkna med ett extra infix som inte tycks bära någon egen betydelse eller ens funktion. Jämför *rosenknopp*. Även det inlånade klassiska eller romanska ordförrådet har – åtminstone sett ur svenskans perspektiv – sådana infix: jämför *human, humanitär* med *revolution, revolutionär*. Mellan grundordet *human* och den etablerade adjektivändelsen *-är* finns det ett "överflödigt" infix *-it-*. En rimlig lösning vore kanske att behandla *-itär* som en variant av *-är*. Alternativen är att behandla det som en fristående enhet i ordbildningen eller att slå ihop *-it-* med det föregående grundledet *human* till *humanit-*. Detta senare framstår inte som någon lockande lösning; den kan bli aktuell först när grundmorfen inte kan fungera som simplex eller har någon klar betydelse. Vi står således med ett infix *-it-*. Andra sådana är det ovan nämnda *-en-*, *-nd-* som i *uppståndelse* där *-else* och (*upp*)*stå* kommuteras ut. Oftast ger väl infixen en bild av de rester som finns kvar i ordförrådet från äldre språkskeden eller från inlånat material.

Ett specialfall av ordbildning är att ordet byter syntaktisk funktion, d v s byter ordklass. Hur vanligt är det och vilka regler följer denna utveckling? Kortord som *bil* ur *automobil* (jämför *mobíl*) ställer frågan om orden ska relateras till varandra på ett sådant sätt att *mo-* och *-bil* kommuteras ut eller inte. Det saknas alltså inte problem att ta ställning till.

Användningsområden

En detaljerad och genomtänkt märkning av ordleden ger möjlighet till sammanställningar med fyllig information. För internt bruk vid ordboksframställning och för forskningsändamål är det både möjligt och lämpligt att sortera sitt material på skiftande sätt alltefter de analyser som gjorts. Men även för en ordbok av ordinär typ kan man tänka sig en komprimerad morfotaktisk information. Sådan förekommer redan nu. Vissa avledningar och sammansättningar anges gärna under ett uppslagsord. Med det här skissade sättet att utmärka ordledstyper o s v, öppnar sig möjligheten att ge informationen i en mer systematisk form, t ex med angivande av vilken typ av ordbildning ett visst ord ingår i. Det skulle mer explicit än den alltid ofullständiga exemplifieringen kunna ange möjligheterna och göra oss mer medvetna om vilka faktorer som påverkar ordbildningen.

En frestande möjlighet, när man har en någorlunda komplett beskrivning av ordbildningsprinciperna, är att på grundval av explicita regler generera ord för att så testa reglernas hållbarhet.

Noter

- 1) Framställningen präglas säkerligen en hel del av erfarenheter från arbetet med Nusvensk frekvensordbok 4 (observera särskilt inledningen) och de många och långa diskussionerna med mina arbetskamrater och medförfattare till ordboken: Sture Allén, Sture Berg, Jerker Järborg, Bo Ralph och Christian Sjögreen. Jag tackar dem samt Åsa Abelin och Erland Gadelii som givit mig synpunkter på manus till detta föredrag.
- 2) Se t ex inledningen till Nusvensk frekvensordbok 4 och Alhaugs uppsats från Nordiska datalingvistikdagarna 1977.

Några referenser

- Alhaug, G., Noen problemer ved opbygging av et datamaskinelt morfem-lexikon. Göteborgs universitet. Rapporter från Språkdata. 3. Nordiska datalingvistikdagar 1977, s. 5-12.
- Nusvensk frekvensordbok baserad på tidningstext. 4. Ordled, betydelser. Data linguistica. 14. Almqvist & Wiksell International. Stockholm. 1980.
- Liljestränd, B., Så bildas orden. Studentlitteratur. Lund 1975.
- Söderbergh, R., Svensk ordbildning. Skrifter utgivna av Nämnden för svensk språkvård. 34. 2. uppl. Läromedelsförlagen. Stockholm 1971.