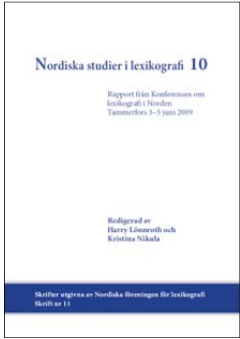


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Leksikaliserte ordsamband i norsk	
Forfatter:	Oddrun Grønvik	
Kilde:	Nordiska Studier i Lexikografi 10, 2010, s. 118-129 Rapport från Konferens om lexikografi i Norden, Tammerfors 3.-5. juni 2009	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© 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.

ODDRUN GRØNVIK

## Leksikaliserte ordsamband i norsk

What is the inventory of Norwegian phraseology, and how can one best capture and describe it? This task has been a challenge for the editors of *Norsk Ordbok* from 1946 onwards, both in terms of analysis and in terms of structural management. In 2003, when *Norsk Ordbok 2014* launched a new digital platform for integrated management of all materials and the whole editorial process, a separate unit was created for editing fixed, grammaticalised or lexicalised phrases, and a set of general guidelines for selection and treatment were drafted. The task of sifting texts and records and editing the result is the task of the *Norsk Ordbok* editorial staff. In this article, at the end of the eighth out of twelve volumes, the current inventory is analysed and discussed in the light of current approaches to handling phraseology in dictionaries.

*Nøkkelord:* fraseologi, ordsamband, leksikalisering, korpus, underoppslagfraseologi, ordsamband, leksikalisering, korpus, underoppslag

### Innleiing

Fraseologi er for tida eit internasjonalt vekstområde i lingvistikken. Det er lite utforska i mange språk, sjølv om mykje kartlegging er i gang. Litteraturen er heterogen både i teoretiske føresetnader og i grunnsyn.

Engelsk er det språket som er best utforska til no. Kartlegging av engelsk fraseologi dreiar seg mykje om å isolera ordsamband som i dei nordiske språka blir samanskrivne som samansetningar. Dermed blir isolering av ordsamband i norsk – og nordisk – noko anna enn i engelsk. Sidan mest alle samansetningar er substantiv, vil kanskje også norske ordsamband i mindre grad vera nominalfrasar og i større grad vera relaterte til andre ordklassar.

Det er grunn til å rekna med at typiske drag ved det fraseologiske inventaret i ulike språk kan vera språkspesifikke, og at dei mellom anna er knytte til dei ordlagingsmekanismane og dei (munnlege og ortografiske) konvensjonane som gjeld for kvart språk. Dei særeigne litterære tradisjonane må også reknast med.

Mål, teoriar, teknikkar og metodar har vidare bruk. Når det gjeld isolering og identifisering av faste ordsamband, er det jo slik at dei lingvistiske fagmiljøa internasjonalt nærmar seg emnet frå to kantar. Den eine tilnærminga er semantisk og språkfilosofisk, og går ut frå menneskeskapte definisjonar av kva tyding

er og korleis tyding kan kategoriserast. Det blir sett fram temmeleg detaljerte typeregister. (Jf. Svensén 2004: 220 ff., Fjeld 2009: 103.) Den andre tilnærminga tek utgangspunkt i ei grein av matematikken, sannsynsrekninga, og har som mål å finna operasjonelle metodar for å la dataprogram og sortera ut moglege leksikaliserte ordsamband frå store tekstmengder. Men skal denne tilnærminga få fram gode resultat, må det liggja føre hypotesar som kan testast, og teknikkar og metodar må vera skikkeleg forståtte.

Den første tilnærminga er då å definera krava til eit leksikalisert ordsamband på same måten som ein definerer krava til ei leksikalisert samansetning, eller kva tydingsberande ord som helst, etter ulike semantiske metodar. Den andre tilnærminga går ut på å sjå etter ytre kriterium (form, frekvens, samførekomst) som skal gjera det mogleg å sortera ut faste ordsamband frå ei stor tekstmengd med ulike metodar for sannsynsrekning (statistiske metodar).

Det er ingen grunn til å hevda at den eine tilnærminga gjer den andre uaktuell (jf. Farø & Lorentzen 2009: 75 f.). Definisjonar og typekategorisering av faste ordsamband kan gjera det mogleg å ta stilling til konkrete døme. Men slikt førebuaende teoretisk arbeid skaffar ikkje fram tilfanget som finst i eit gjevi språk, og om definisjonane og kategoriane dekkjer tilfanget, veit ein ikkje før ein har oversyn over heilskapen. Automatisert sorteringsarbeid for å få fram faste ordsamband, har heller ikkje synt seg 100 prosent treffsikkert, og listene som kjem ut, syner tydeleg at det framleis trengst kompetent lingvistisk vurdering av resultatet.<sup>1</sup>

Frå empirisk ståstad vil ein rekna det som ein føresetnad for god vurdering at ein ikkje berre har definisjonar og (prosessert) råmateriale, men også eit brukande oversyn over feltet, i dette tilfellet over inventaret av faste ordsamband i norsk. Det næraste ein kan koma til å få oversyn over feltet innanfor eit einskild språk, er å ta for seg ei stor mengd med tekst og metaopplysningar om språk (samlingar, korpora) og handsama eigenskapane til kvart lemma, slik ein lingvist gjer når ho eller han redigerer artiklar i eit stort ordboksverk. Redigeringa av *Norsk Ordbok 2014* gjev såleis eit truleg eineståande høve til å dokumentera typegalleriet, rikdomen og variasjonen i norsk leksikalisert fraseologi. Dette høvet blir no utnytta.

I den vidare drøftinga går eg, i tillegg til det som er sagt ovanfor, utfrå føresetnadene for og føremålet med *Norsk Ordbok 2014*. Dei er 1) at det finst eit behov for å skilja ut faste ordsamband i papirutgåva av ordboka, og å gjera dei søkbare i den komande elektroniske utgåva. Dette både på grunn av eigenska-

1 Jf. innlegg MONS 13 ved Gunn Inger Lyse og Gisle Andersen om fleirordsuttrykk i *Norsk aviskorpus*.

par ved tilfanget (faste ordsamband finst) og av di dei som bruker ordbøker, er interesserte i enkelt tilgjenge til faste ordsamband. Vidare er det 2) ikkje nødvendig frå brukarsynsstad å skilja nøye mellom ulike typar faste ordsamband. Derimot er det 3) nødvendig at retningslinjene til redaksjonen for å skilja ut faste ordsamband er klåre nok til å vera operasjonelle, både i å skilja ut det vi har sikre rutinar for og i vurderinga av gråsonane i tilfanget.

## Korleis finna og lagra leksikaliserte ordsamband?

*Norsk Ordbok* har valt ein tofrontsstrategi, som vonleg gjev eit stort tilfang av faste ordsamband både frå skrift og tale. Det er 1) utforma formelle krav til faste og leksikaliserte ordsamband, som finst (i ein førebels versjon) i redigeringshandboka for *Norsk Ordbok 2014*. Vidare vart det 2) i samband med omlegginga av redigeringsformatet for *Norsk Ordbok* laga ein eigen underkategori for faste ordsamband i redigeringsformatet for ordartikkelen. Underkategorien heiter *underoppslag*, og har form som ein redusert ordartikkel.

Sidan tyding er ein mental storleik, som byggjer på dokumentert konsensus mellom språkbrukarane, må det vera ei redaktør oppgåve å avgjera om eit ordsamband er ei konvensjonell samanstilling med føreseieleg tyding, og dermed skal stå som redaksjonelt døme, eller om det blir nytta biletleg eller overført på ein slik måte at tydinga ikkje framgår av einskildorda, slik at det må reknast som eit leksikalisert eller grammatikalisert ordsamband. Redaktørane har redigeringsreglane for utval av underoppslag, og siktar tilfanget for kvart ord dei redigerer. Dei har dermed dagleg høve til å testa retningslinjene mot den språklege røyndomen.

Det seier seg sjølv at redaksjonelle retningsliner kan bli ulikt oppfatta av ulike redaktørar, og særleg når ein vågar seg til å bryta ny mark. Redaksjonen har no arbeidd i seks år og har redigert tre og eit halvt band. Dei to første banda er detaljinnpassa i ordboksdatabasen (noko som medførde ein byrjande revisjon), to og eit halv band er delvis innpassa. Med to tredelar av verket på plass, og halvparten i detalj, er det grunnlag for å sjå på kva innhaldet i kategorien underoppslag fortel om norsk fraseologi.

## Definisjonar

Definisjonane som blir nytta i det redaksjonelle arbeidet i *Norsk Ordbok* er som følger: 1) Eit fast ordsamband er lik ei samanstilling av minst to ord som re-

gelmessig opptrer saman. 2) At eit slikt ordsamband er grammatikalisert eller leksikalisert<sup>2</sup> vil seia at det har minst ei tyding eller ein bruk som ikkje framgår av tyding og bruk for kvart av orda i ordsambandet. Det er ikkje skilt mellom ulike typar ordsamband når det gjeld ordtypar, opphav eller anna.

I grunnlagstilfanget er det lett å finna døme på faste ordsamband med høg frekvens: til dømes *har vori, i dag, til å, av og til, pro rata, kvit dverg, brun kafé, leggja seg ut med*. Men kva av dette skal reknast som moglege underoppslag i *Norsk Ordbok*?

Det viktigaste kriteriet er at ordsambandet har ei eller fleire tydingar som ikkje er føreseieleg eller openberr ut frå einskildorda, med vanleg språk- og kulturforståing hos ein vaksen morsmålsbrukar som utgangspunkt. Støttekriterium er at ordsambandet fungerer som setningsledd og dermed kan ha funksjonell ordklasse, at det er rimeleg fast i form og ordfylgje, og at det er etter måten frekvent.

Frekvenskriteriet fortener ein ekstra kommentar. *Norsk Ordbok* dekkjer norske dialekter frå landsende til landsende, og underlagstilfanget har fylgjeleg ein del døme på faste ordsamband som har høg frekvens lokalt, men som er udokumenterte i standardspråket. I slike tilfelle må vi lita på gode informantar. Det at eit fast ordsamband ikkje er dokumentert i skriftspråket eller allment kjend i talemålet, er dermed ikkje avvisingsgrunn.

## Eigen ordartikkel eller underoppslag

Eit ordsamband kan plasserast i ordboka som 1) lemma med eigen ordartikkel, 2) som underoppslag med redusert artikkelformat innordna i ein større ordartikkel eller 3) som redaksjonelt døme med forklaring. Fordelinga på dei tre føringsmåtene er dels styrt av tidlegare redaksjonell praksis, dels av eigenskapar ved sjølve ordsambandet og det underliggjande tilfanget.

Eit viktig ytre rammevilkår er at eit ordsamband skal ha éi oppslagsform, og skal handsamast redaksjonelt berre på ein stad. Same ordsambandet kan dermed ikkje både ha eigen artikkel og finnast i eit underoppslag, eller vera handsama i fleire underoppslag. Dette er ein banal, men nødvendig regel.<sup>3</sup>

2 Eg skil her ikkje mellom leksikaliserte og grammatikaliserte ordsamband.

3 Homografar er ulike lemma og kan sjølvstøtt knyta til seg dei same tilleggsorda i faste ordsamband som i sin tur blir førde som underoppslag. Ein får då underoppslag som er homografar. Slike døme finst, men dei er sjeldsynte.

Hovudregelen for plassering i eigen ordartikkel, i motsetning til underopp- slag, er at faste ordsamband med einskildord som ikkje er egne lemma i norsk, får eigen ordartikkel. Mange er importerte: døme er *a capella*, *ex libris*, *rock and roll*. Men også ein del fleirordige interjeksjonar, og funksjonsord har fått egne ordartiklar, trass i at komponentane er egne lemma: døme er *ja så*, *nei visst*, *i det*, *etter kvart*, *nokon kvar*, *opp ned*, *pø om pø*. Sams for desse ordsambanda er at dei ofte har vori samanskrivne, særleg i eldre litteratur, og dei er til dels hand- sama som egne lemma i rettskrivingsordbøker. Det er dei ortografiske konven- sjonane i norsk som har gjort dei til ordsamband.

I gråsona mellom plassering i underoppslag eller eigen ordartikkel står ord- samband der ei av ordformene er ei stivna kasusform, til dømes *av garde*, *av dage*, *i live*. Argumentet for plassering i eigen artikkel er at desse ordformene ikkje lenger er del av bøyingsparadigmet til substantiva *gard*, *dag* eller *liv* og derfor bør sjåast som avleidde men sjølvstendige ordformer. På alfabetisk plass blir dei også lettare å finna – i papirutgåva. Argumentet for plassering som un- deroppslag er alfabetisk og etymologisk koherens, og redigeringstradisjonen i *Norsk Ordbok*. Førabels ser det ut til at tradisjonen veg tyngst for redaktørane.

## Underoppslag eller redaksjonelt døme

Faste ordsamband som har ei eller fleire leksikaliserte tydingar eller uføresei- elege bruksmåtar skal førast som underoppslag, medan faste ordsamband med føreseieleg tyding skal handsamast som redaksjonelle døme (døme utan kjelde). Dette er i praksis det vanskelegaste skiljet å dra, og det er fleire grunnar til det. Ein viktig grunn er heterogeniteten i tilfanget. Eit underoppslag skal vera fast i forma, men framtrer ofte i ein overflod av nesten like former i belegg frå ulike dialekter eller skriftstykkje frå ulike tider. I slike tilfelle inneber det å velja ut éi form til underoppslag ei språkleg standardisering. I redaksjonelle døme har ein derimot høve til å syna alternative former. Men om tyding og bruk syner klar leksikalisering, skal ein velja underoppslag, og har ordsambandet fleire tydingar er underoppslag einaste føringsmåten. Problemet med variabel form kan då til dømes løysast slik:

”**koma under prestehanda** (òg med subst i ubf). **1** bli konfirmert (..).  
**2** bli gift (..).”

Ein annan grunn til at det kan vera vanskeleg å velja ut faste ordsamband til underoppslag, er at dei frekvente grunnorda genererer ein svært rik metaforikk,

ofte i mange vendingar med nærliggjande tydingar. Då kan det vera vanskeleg å dra skiljet mellom dei fastaste og mest frekvente, og resten.<sup>4</sup> Desse flokane blir handtert noko ulikt frå redaktør til redaktør. Tendensen i *Norsk Ordbok* er likevel at redaktøren tek utgangspunkt i regularitet og semantisk konsistens, og vel éi normalform for underoppslaget på grunnlag av mellom anna frekvens og sentralitet.

Ei anna gruppe faste ordsamband som ser ut til å valda redaksjonell uvisse, er reduserte ordtak. Hovudregelen for *Norsk Ordbok 2014* er at ordtak skal ha setningsform, og først som sitat (gjerne med kjelde), medan underoppslag har form som hypotagme. Ein gjennomgang av databasen syner likevel at somme verbfrasar er blitt førde som ordtak.<sup>5</sup>

For å oppsummera:

Underoppslaget i *Norsk Ordbok 2014* er skapt for å gje eit lagringsformat som er framheva (i papirversjonen) og søkbart (i den elektroniske versjonen) til faste ordsamband som

- inneheld minst to ord, skilde med mellomrom
- er setningsledd, men ikkje full setning
- har tyding og bruk som ikkje framgår av einskildorda, og derfor krev definisjon
- kan vera polysem

Resten av artikkelen skal brukast til å drøfta bruken av underoppslaget i *Norsk Ordbok* så langt.

## Litt historikk

*Norsk Ordbok* gjekk over til digitalt redigeringsformat i 2003. Då var fire band utgjevne, og alfabetet var dekt til midt i bokstaven *h*. Ein føresetnad for omlegginga var at ein skulle unngå unaudsynte brot med tidlegare redigeringspraksis. Ein kunne endra for å betra kvalitet og konsistens, men ikkje elles. Redigeringsformatet som vart utvikla for dei siste åtte banda byggjer derfor på det som vart gjort i dei første fire, og er ei innstramming og systematisering av tidlegare praksis (1966–2002).

4 Sjø t.d. *I liv 1 b* og *II lag 16* som døme på to ulike metodar for handtering av metaforikk.

5 Såleis er t.d. *ana fred* og *ingen fare* ført som ordtak.

Det er mange døme i band 1–4 på at redaktørane sleit med plassering, definering og eksemplifisering av faste ordsamband. I det gamle formatet skilde ein mellom definisjon og bruksdøme, men det mangla eit fast system for handtering av fraseologi innanfor ordartikkelen. Ordsamband vart derfor førde som sitat. Dette og utløyste ein redaksjonell praksis med definering av sitat, ofte med fleire, nummererte tydingar som i sin tur skulle ha bruksdøme med kjelder og kommentarar. Slik praksis skaper strukturelle spiralar og sløyfer, og resultatet vart uoversiktlege ordartiklar. Underoppslaget er skapt som svar på behovet for å definera og eksemplifisera ordsamband etter same prinsipp som for lemma, men likevel kunna handsama dei innanfor ordartikkelen.

## Underoppslag i Norsk Ordbok – mengd, fordeling og lengd

Underoppslagsformatet er no brukt i seks år og fire band av *Norsk Ordbok*. I tillegg er revisjonen av band 1 og 2 påbyrja, slik at definerte ordsamband frå *a* til *flusken* ligg som underoppslag.

I databasen for *Norsk Ordbok* ligg det no opp mot 8 000 underoppslag. Utlikna på seks band blir dette 1 200–1 500 underoppslag per band. Bokstaven *k*, som fyller eit band, er såleis typisk for fordelinga. Bruken er kanskje litt aukande i dei siste banda, men ikkje mykje. Fordelinga av underoppslag gjennom alfabetet er også temmeleg jamn.<sup>6</sup>

Kor mange ord har dei faste ordsambanda, og korleis er fordelinga på lengd? Nedre grense går ved to. Dei lengste underoppslaga er på 8–9 ord. Slik det ser ut no, er dei aller fleste underoppslag fåordige, 56 prosent er på to ord og 32 på tre. Underoppslag på fire eller fleire ord er snau 12 prosent av heilskapen.

Mest alle dei lange ordsambanda som har fått underoppslag er verbfrasar med fleire kollokat. Døme kan vera *stå med eine foten i grava* eller *ta mannen i staden for ballen*. Den lengste er også eit paratagme: *stikka fingeren i jorda og lukta kvar ein er*.

6 Tilhøvet mellom den prenta ordboka og ordboksdata-basen er per dato som fylger: *a–e* er fullstendig innlasta og under revisjon, *f–h* er ufullstendig innlasta, og underoppslag stort sett ikkje skilde ut (tala dermed ikkje representative og ikkje rekna med i gjennomsnitt per band), *i–q* er ferdigredigerte og utgjevne og *r* så vidt påbyrja.



Tabell 1. Fordeling av underoppslag på bokstavar og band

Bokstav	Band	Tal UO	Bokstav	Tal UO	Band
A	1 a-doktrinær	247	J	233	band 5
B		824	K	1 419	
C		0	L	1 271	band 6 k-kåvå
D		509	M	659	
E	2 dokument-flusken	354	N	331	band 7 l-mugetuft
F		860	O	175	
G		43	P	336	band 8 mugg-ramnsvert
H	3 flusker-gigla	104	Q	0	
I	4 gigle-harlemmerolje 5 harm - jåttut	315	Ra-	27	

## Definisjonsstatistikk

Eit viktig kriterium for å skilja underoppslag frå redaksjonelle døme, er om ordsambandet syner polysemi. Det er derfor interessant å sjå på kor mange tydingar underoppslaga har.

I alt har dei cirka 8 000 underoppslaga om lag 12 000 definisjonar, og om lag 80 prosent har éin definisjon, noko som svarer til 60 prosent av definisjonsmengda totalt. Underoppslag med ein eller to definisjonar utgjer 92 prosent av underoppslaga og har meir enn 75 prosent av definisjonane.

I den andre enden av skalaen har vi nokre tital underoppslag med ti eller fleire definisjonar. Dei utgjer litt over 3 promille av totalmengda artiklar, men dei har dryge 3 prosent av definisjonane totalt, og i omfang høyrer desse underoppslaga til dei større artiklane i *Norsk Ordbok*, endå om dei er plasserte inn i ein annan artikkel.

Dei aller fleste av underoppslaga med mange tydingar er av typen verb + partikkel, og dei er høgfrekvente i norsk, både i skrift og tale: *koma til*, *leggja opp*, *draga på*, *gjera av*. Some av dei har sterkt grammatikalisererte funksjonar: *koma til* er såleis den vanlegaste markøren for framtid i verbsystemet. Men

nokre ordsamband som for det meste fungerer adverbialt kjem også høgt i tal definisjonar: *døme kan vera i lag og opp og ned.*

Det er tydeleg omvendt proporsjonalitet mellom tal ord i underoppslaget og tal tydingar. Alle underoppslag med fem eller fleire tydingar er etter måten stutte, med tre eller færre ord. Orda i desse ordsambanda er også stutte – i underoppslag med sju tydingar eller meir er det ingen ord på meir enn seks bokstavar. Av dei (til no) opp mot 1 000 underoppslag med fire eller fleire ord er berre 88 førde med meir enn ei tyding.

Eit spørsmål til stadig drøfting er om ein skal føra opp den bokstavlege tydinga av eit ordsamband i tillegg til den leksikaliserte. Plassomsyn i høve til den prenta utgåva av *Norsk Ordbok* dreg i ei lei, medan det å syna samanheng og gje eit fullstendig bilete av tydingsspekteret dreg i motsett lei. Ofte blir den bokstavlege tydinga sløyfa av di ho alt er uttrykt i sjølve underoppslaget: *henta jordmor* er dermed definert som 'ha hast' medan tydinga 'skaffa hjelp til ei fødande kvinne' er underforstått.

## Kva slag struktur har underoppslaga?

Hovudregelen i *Norsk Ordbok* er at ordsamband som blir førde som underoppslag skal ha struktur som setningsledd, medan fulle setningar med fast form normalt blir rekna som ordtak. Ordtak blir førde som sitat, oftast med kjelde. Det er ikkje sett av rom – eller tid – til å fastsetja funksjonell ordklasse for underoppslaga. For det første veit ein av røynsle at same ordsambandet kan ha fleire ordklasserfunksjonar, alt etter kontekst. For det andre har dette ikkje vori gjort før i stor skala for norsk, og det er lett å sjå at fastsetjing av funksjonell ordklasse i breitt omfang vil krevja mykje utgreiingsarbeid. Dermed ville oppgåva blitt for stor, gjevi at *Norsk Ordbok* har ei snau tidramme for å bli ferdig.

Det ein lettare kan fastslå, er kva slags hypotagme underoppslaga representerer. Som venta dominerer verbfrasane, med 64 prosent av totalmengda. Deretter kjem preposisjonsuttrykka (preposisjon pluss nominal) og nominalfrasane (for det meste i form attributt pluss nominal) med 13–14 prosent kvar.

Underoppslaga som er verbfrasar syner eit rikt utval av underliggjande kollokat. Ikke uventa er det mest av typen verb pluss partikkel, men det er og mange leksikaliserte verb–objekt–samband. Dei høgfrequente verba som no er redigerte, innleier meir enn dobbelt så mange underoppslag som dei ein finn under sjølve ordartikkelen, 150–200 underoppslag i alt for dei best representerte (*ha, koma, liggja, gjera, gå*) medan verb som *stå* og *slå*, som enno ikkje er redigerte, innleier mellom 60 og 70 underoppslag kvar seg.

## Nokre ulike typar

Når ein ser på litteraturen om faste ordsamband, kan ein finna mykje om pene og ryddige hypotagme av typen *kasta vrak på, med eit naudskrik, magnetisk jordakse, mannen i månen* og liknande. Men i tilfanget som no samlar seg opp i databasen for *Norsk Ordbok* finn ein og typar som ikkje blir så ofte nemnde. Det er til dømes 1) parataktiske ordsamband, 2) personifiseringar, 3) syntagme – småsetningar som ikkje let seg skriva om til setningsleddsform, 4) samanlikningar og 5) utfyllingsskjema.

Om lag 400 underoppslag er parataktiske, bundne saman med *og* eller *eller*. Dei fleste bind saman substantiv eller verb (*under lås og slå, takk og pris*), men ikkje så få opptrer som frekvente funksjonsord: *ein eller annan, før eller seinare, att og fram, som seg hør og bør*. Det slående er mengda av dei, kombinert med den manglande omtalen. Vonleg kan lett vint tilgjenge – som *Norsk Ordbok*-databasen no byr på – endra på dette.

Det eg kallar personifiseringar, er ordsamband som inneheld namn. Mellom underoppslaga møter ein *Per inkjevetta, Pål skyttar, Rasmus kneppneve* og *Ola dunk*. På spinnesida finn ein *Glo-Maria, ho Kari Bjørka, Lussi langnatt* og *Mari vassause*. Slike personifiseringar av ting, eigenskapar, fenomen, merkedagar og mangt anna har eldgamle røter og er enno til stades i norsk språk. Når denne prosessen råkar namn som einskildord kan vegen til appellativ vera kort (*boykott, quisling*) og namnekarakteren går tapt. Namn som får eit karakteriserande tilleggsord held på personpreget og tek steget over i allegorien, og blir altså samla opp i *Norsk Ordbok*.

Småsetningane ligg nær ordtaka, og somme av dei er nok reduserte ordtak, til dømes *enkje vil ha make, fuglen er flogen*. Andre er munhell som anten ikkje let seg skriva om til setningsleddform eller som mest prototypisk har setningsform. Døme er *klokka er slegen* 'tidspunkt då det er for seint', *ikkje reint lite* 'ein god del', *her er det inga kjære mor* 'det hjelper ikkje å be for seg'. Andre igjen har nærast adverbfunksjon og står ofte som apposisjon: *dei kallar* 'som folk seier', *var det likt* 'det er ikkje anna å vente', *best råd er* 'så godt som mogleg'.

Haldninga til samanlikningar som underoppslag er restriktiv, for ei samanlikning er jo ikkje ein fullt utvikla metafor, og dermed kan ein stilla seg kritisk til leksikaliseringsgraden. På den andre sida finst det samanlikningar som er faste i form, tyding og bruk, og ein del av dei er registrerte som underoppslag. Det finst svært ulike døme, frå *som kjent* 'som dei fleste veit' til *som eit olja lyn* 'sers snøgt'. Dei fleste samanlikningane blir likevel førde som redaksjonelle døme.

## Litt semantisk grovinndeling

Gjennom underoppslaget har *Norsk Ordbok* dokumentasjon på og redaksjonell handsaming av ordsamband som opptrer som fleirordige

1. funksjonsord (preposisjonar, konjunksjonar, pronomen): *all den tid, i og med, er så at, kven som helst*
2. Adverbial: *av og til, kor som er, for den del, til liks med, i eigen person, her på berget*
3. Verbsamband: *leggja i hop, koma ut for, halda seg for god til, berga livet, rasla med sabelen*
4. Kollokasjonar til adjektiv og substantiv: *klin kokos, fallande stjerne, menneskeleg svikt, natta som var*
5. Interjeksjonar og helsingar: *god dag, Gud betre, velkomen i klubben*
6. Pragmatiske formalar: *ja då, takk likeins, det skulle berre mangla*
7. Reduserte ordtak og munnhell: *gjort er gjort, Gud eig alla dagar*
8. Termar og artsnamn: *n fakultet, identiske likningar, leska kalk, jomfrua i det grøne*
9. Idiom: *brød og vatn, det glatte lag, gå den vegen høna sparkar, telja på fingrane*
10. Ord med venger: *fylka til strid, møllestein om halsen, enno er ikkje alle jomsvikingar døde*

Inventaret er størst på punkt 2, 3 og 4, og truleg vanskelegast å fanga opp på punkt 1, 2 og 3 av di ein der har så mange kombinasjonar av svært frekvente ord. Korpusbruk er til svært stor hjelp når det gjeld å finna og dokumentera desse ordsambanda.

*Norsk Ordbok* skal dekkja både skrift og tale. Dialekttilfanget er det ein lykkeleg skyldnad å ta seg av, for det er *Norsk Ordbok* aleine om. Samstundes står det ikkje til å nekta at den nynorske skriftheimen har mindre spesialisert faglitteratur enn bokmål. I tillegg har nynorsk skriftkultur ein tradisjon for å dyrka ein talemålsnær stil også i meir formelle sjangrar. Det er ikkje usannsynleg at desse faktorane verkar inn på inventaret av ordsamband, og gjer sitt særleg til at verbsambanda blir ei rikholdig gruppe.

## Konklusjonar

Attende til innleiinga: Stemmer det at inventaret av ordsamband i eit språk har språkspesifikke drag? Det er vel for vågalt å seia ja generelt. Men det er ikkje tvil om at den sterkt kombinatoriske verbsemantikken i nynorsk skrift og tale har stor innverknad på inventaret av faste ordsamband. Det er vidare heva over tvil at dei vidgreina språksamlingane, kombinert med systematisk bruk av nynorsk-korpuset, har vori og er heilt nødvendige for å fanga opp dei ordsambanda av funksjonsord og høgfrequente innhaldsord som pregar norsk kvardagspråk i skrift og tale. Korpusutviding og innføring av korpusstatistikk (med opplæring) står derfor på arbeidsplanen.

Idiom og litterære ord med venger fangar blikket og gjev språket farge, og *Norsk Ordbok* skal ha med alt som det er dekning for å ta med. Det same gjeld den terminologien som er blitt del av kvardag og ikkje minst skule.

Har det vori fornuftig å vera varsam med å setja redaksjonelle rammer for kva slags ordsamband som skal få plass i underoppslaget? Ja, eg meiner det. Nynorsk er enno ikkje eit fullt standardisert språk, men redigeringa av *Norsk Ordbok*, og ikkje minst handsaminga av nynorske ordsamband, inneber eit langt steg på vegen mot å bli det. *Norsk Ordbok* har knytt til seg eit stort korps med høgt skulerte lingvistar som redaktørar for at dei skal sikta underlagstilfanget etter beste skjønn, og drøfta fram gullet. Da må dei få rom til å utøva skjønn bygt på fagkunnskap. At ikkje alle tenkjer likt heile tida, bør reknast som ein stimulerande føremon.

## LITTERATUR

- Farø, Ken & Lorentzen, Henrik, 2009: De oversete og mishandede ordforbindelser. Hvilke, hvor og hvorfor? I: *LexicoNordica* 16. S. 75–102.
- Fjeld, Ruth Vatvedt, 2009: Leksikografisk dokumentasjon av flerordsenheter i norsk. I: *LexicoNordica* 16. S. 103–118.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.