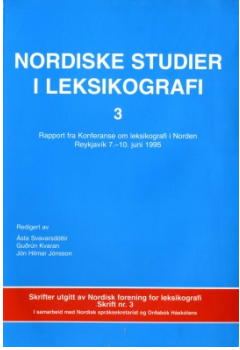


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Korpusbaserede relationsoplysninger og lemmatisering af flerordsforbindelser	
Forfatter:	Berit Heer Henriksen	
Kilde:	Nordiske Studier i Leksikografi 3, 1995, s. 195-203 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995	
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.

## Korpusbaserede relationsoplysninger og lemmatisering af flerordsforbindelser

Of the approximately 100 000 entries selected for *The Danish Dictionary*, not all will be one-word entries. Some of them will be subentries consisting of word groups. These subentries are described in detail as if they were ordinary entries, but we also have other and shorter ways of presenting information about phraseology and valency. Bearing these different possibilities in mind, we examine the concordances and word distribution reports generated from a text corpus of 40 million running words — and make the selection of subentries for the dictionary.

*Den Danske Ordbog* bliver en stor, ny ordbog over nutidsdansk, og den udkommer i seks bind i 1998–99.

Nogle af ordbogens ca. 100 000 opslagsord er hovedlemmaer, andre er sublemmaer. Et hovedlemma er fx et ord som *båd*, der kan slås op på ordets alfabetiske plads i ordbogen. Et sublemma er derimod en gruppe af ord, fx *gå i bådene*. Det enkelte sublemma er placeret under et relevant hovedlemma.

Sublemmaerne i *Den Danske Ordbog* er kombinationer af ord, der altid optræder sammen. De danner begrebslige helheder med en konkret eller overført betydning. Det kan dreje sig om talemåder som fx *brændt barn skyr ilden*, eller det kan være idiomer som fx *være i samme båd*, eller det kan være andre kollokationer som fx *bryde ud, sætte i gang, til fods, tak for mad, både . . . og* — inklusive termer som fx *tungt artilleri*.

Men ud over sublemmaerne har ordbogen oplysninger om valensmønstre og om typiske samforekomster. Oplysningerne om syntaktiske mønstre er især relevante for verbernes vedkommende, mens vi om et ord som fx *båd* vil give den oplysning, at man typisk træffer samforekomsterne *lille båd og om bord i/på en båd*.

For ordbogsredaktøren kan oplysninger om valenser og en angivelse af typiske samforekomster somme tider fungere som pladsbesparende alternativer til lemmatisering af en ordgruppe. Dét forhold vil jeg beskæftige mig med i det følgende. Jeg vil gennemgå mulighederne for at medtage samforekomster og valenser i *Den Danske Ordbog* — herunder oprettelsen af sublemmaer. I den forbindelse vil jeg vise nogle af de udnyttelsesmuligheder, der ligger i vores store elektroniske tekstkorpus. Til sidst vil jeg pege på konkrete problemer, der kan opstå under selektionen af sublemmaer.

### 1 Sublemmaer, samforekomster og valenser

Vi har forskellige muligheder, når vi vil have en samforekomst eller en valens med i *Den Danske Ordbog*. Vi skelner mellem

- sublemma (fx *båd* [. . .] *være i samme båd*)

- valens (fx *spise* [...] [*NGN spiser NGT*])
- typisk samforekomst (fx *chef* [...]~ *for*)
- typisk kerne (fx *garvet* [...] *garvede politifolk*)
- typisk adled (fx *rød* [...] *blussende rød*)
- citat (fx *eg* [...] *som gave til mor havde han lavet et solidt sybord i eg*)

De forskellige oplysningstyper kan kombineres. Behandlingen af et sublemma kan fx godt indbefatte oplysninger om valenser og typiske samforekomster.

### 1.1 Sublemmaer i Den Danske Ordbog

Sublemmaerne i *Den Danske Ordbog* behandles lige så grundigt som hovedlemmaerne, så lemmatisering af en ordgruppe er en pladskrævende løsning.

Alle sublemmaer får en betydningsbeskrivelse, og i ordbogen fremstår sublemmaet med halvfed skrift e.l. Desuden angives trykforhold, hvis det er afgørende for betydningsdifferentieringen — fx er der forskel på '*skrive* *sammen* 'brevveksle' m.m. og *skrive* *sammen* 'sammenskrive'. Idiomer og talemåder lemmaklassemarkeres, men ikke kollokationer — således står der *talemåde* efter sublemmaet *man skal ikke smide barnet ud med badevandet*.

Et sublemma er som hovedregel placeret under første betydningsbærende substantiv. Fx placeres talemåden *brændt barn skyr ilden* under *barn*. I de tilfælde, fx ved partikelverber, hvor ordgruppen ikke indeholder et betydningsbærende substantiv, må vi bruge andre anbringelsesprincipper, men det vil føre for vidt at gennemgå dem her.

Vi håber, at ordbogens brugere efterhånden vil lære sig, hvor de skal lede efter sublemmaerne. Desværre læser ordbogsbrugere så godt som aldrig ordbøgers indledning, hvor principperne kan være gennemgået. Vi ved på forhånd, at brugere af ordbøger ikke altid forventer at finde sublemmaerne under første betydningsbærende substantiv eller substantivering. Det ved vi bl.a. fra en lille spørgeskemaundersøgelse, der er foretaget af Henrik Lorentzen, en af redaktørerne på *Den Danske Ordbog*.

I nogle tilfælde vælger vi at komme ordbogsbrugeren i møde ved at opføre et sublemma i flere artikler og så henvise fra den ene artikel til dén artikel, hvor sublemmaet er behandlet. Henvisningen kan foregå eksplicit eller via henvisningsmærke. I *brænde*-artiklen finder man således den implicitte henvisning *brændt* ▷ *barn skyr ilden*. I artiklen om *badevand* finder man imidlertid henvisningen *man skal ikke smide barnet ud med badevandet, se barn*, fordi *barn* ikke står i sin grundform i talemåden.

Alt i alt må man sige, at det er en pladskrævende løsning, vi vælger, når vi vælger en lemmatisering. Ikke mindst, hvis vi lemmatiserer udtrykket flere steder i ordbogen som en service til brugeren.

Men der er ofte også andre præsentationsmuligheder end lemmatisering, og det kan være vanskeligt at afgøre, hvornår en alternativ mulighed er at foretrække. Meget afhænger af den undersøgelse, vi foretager i vores store elektroniske tekstkorpus.

	mut.inf.	coocc.	mean-pos	dev-pos
fods	247.02	[10]	-1.00	0.00
ombord	122.30	[9]	-2.56	0.25
cup	57.65	[8]	-1.13	0.11
bord	56.66	[27]	-2.33	0.22
bunden	51.99	[16]	-2.06	0.06
vandet	13.26	[10]	-2.60	0.44
bil	11.85	[9]	-1.78	0.17
pr	10.50	[11]	-1.00	0.00
meter	9.81	[8]	-1.75	0.44
samme	9.10	[45]	-1.07	0.11
manden	8.01	[8]	-2.13	0.11
lille	6.62	[33]	-1.06	0.06
danske	5.83	[27]	-1.15	0.27
m	5.18	[11]	-1.73	0.38
ude	5.05	[10]	-2.20	0.16
ned	4.72	[25]	-2.20	0.16
egen	4.55	[9]	-1.33	0.44
sin	4.15	[45]	-1.36	0.45
anden	4.05	[22]	-1.36	0.60
en	3.79	[399]	-1.47	0.44
min	3.68	[26]	-1.23	0.18
ligger	3.46	[9]	-2.11	0.32
stor	3.42	[13]	-1.46	0.56
mand	3.22	[11]	-2.64	0.23
alle	3.15	[34]	-2.26	0.90
ny	3.13	[11]	-1.27	0.38
din	3.07	[8]	-1.13	0.11
i	2.70	[489]	-1.77	0.63
fra	2.68	[78]	-1.77	0.74
tilbage	2.59	[11]	-2.09	0.08
mens	2.58	[10]	-1.60	0.84
omkring	2.46	[8]	-1.38	0.23
op	2.27	[30]	-2.27	0.20
denne	2.25	[16]	-1.25	0.31
den	2.23	[171]	-2.01	0.29
kom	2.20	[15]	-2.47	0.38

Figur 1. Den mutuelle statistik er baseret på Den Danske Ordbogs elektroniske tekstkorpus. Statistikken viser, hvilke ord der overhyppigt optræder på første, anden og tredje plads til venstre for ordet båd.

## 2 Korpus som kildemateriale

*Den Danske Ordbog* bygger på mange kilder. En af de vigtigste er et elektronisk tekstkorpus, der indeholder ca. 40 mio. løbende ord, hentet fra danske aviser, bøger, blade, radio, tv, brochurer, reklamer, tegneserier, breve, skolestile, lejlighedsviser osv. Fra dette korpus kan vi hente konkordanslister og statistiske oplysninger, som vi bruger under artikelredigeringen.

### 2.1 Eksemplet *båd*

Da jeg skulle skrive ordbogsartiklen om ordet *båd*, lavede jeg først en mutuel statistik. En mutuel statistik er en undersøgelse af, hvilke ord der overhyppigt optræder sammen med et bestemt ord. Edb-programmet sammenligner sandsynligheden for at møde ordene hver for sig med sandsynligheden for at møde dem sammen. Mutuelle statistikker giver både et indtryk af et ords betydning/betydninger og af de samforekomster og valensmønstre, ordet indgår i. Statistikkerne laver vi kun på ord, der er relativt frekvente i korpus; der skal helst være mindst et par hundrede forekomster.

Corpus-Bench - BAADGAES.CBC		
File	New	Edit
View	Options	Help
CONC: båd (inf)		
<p>Ler fort under alle forhold. Og hvor er det dejligt at være om bord i en båd, hvor man stadig kan  Ler n &lt;/fremhæv&gt; forstærker følelsen af at være om bord i en sikker båd. Vindskærmen er ikke  IFHZ nle. ☐ Få minutter senere hev Svend Kloster sin ven ombord i sin båd. Da var Hans Mortens  IFHZ k Inga og Svend Kloster også egenhændigt bjærgt ombord i deres båd, men hans liv stod he  Jediga dem og var ved at skylle deres garn. Og han vil om bord i en af bådene, den, der tilhørte  Jediga han øje på to både. En af ligger ved søen. Jesus går om bord i en af bådene. Den tilhører fiske  KmO den i den hensigt at tage dit eget liv ved at springe over bord fra båden i løbet af natten.☐E  KmO ss havde virket fuldstændig naturlig, da han hjalp hende om bord i båden, han havde småsnak  KmO er var der ingen radiosender eller -modtager om bord i vor beskedne båd. Der var proviant til f  KmO iam Bligh og hans "loyalister" havde de tvunget ombord i en lille båd og ladet dem tilbage i  LanX Nykøbing Falster, når han besøger os, og springer adræt om bord i båden, siger Tom. - Han  Mide r trukket dybt ned i Middelhavet.☐Der er altid en læge om bord på båden, når Angela og andr  Ozw bevidsthed.☐En søndag var jeg med min far og morbror ombord i båden, på kryds ud af Sor  Ppn jazz og sejlbåde. En af hans mest prominente gæster ombord i sin båd var den brasilianske r  Plim lov til at gå til kajs i havnen ved Adriater-havet.☐Ombord i en af bådene var den jugoslavisk  EDT ernerlagt, at han intet kunne huske. Kun, at han var gået ombord på båden, og at han havde va  RØ chef for Nationalteatret i Oslo. Og ideen, som han fik om bord på båden, var den helt oplagt  FSD 20. august ved 6-tiden sejlede vi ud af Plymouth havn om bord på båden "Mistress" med sk  FSD s 6 ounces og var helt tilbage fra 1976 og også sat ombord på en båd fra Plymouth ude i Ke  GmØ lkvædet vise, og så taler de så myndigt. Specielt om bord på en båd, hvor der skal handles  HUr ns motorbåd, og undrede sig over hvad Fingal ville ombord på den båd, hvor han jo intet hav  Hig e ved havnen, hvor postbåden netop havde lagt fra kaj. Om bord på båden befandt sig flere al  Hig ☐Da han vågnede, befandt Jeroen og Kurt sig om bord på Tobias' båd på vej ud af indseilin</p>		
Total words: 0	Node words: 0	Matches: 38
Report updated.		

Figur 2. Konkordanslisten afslører bl.a. en uenighed blandt sprogbrugerne mht., hvorvidt om bord skal skrives i et eller to ord.

Den mutuelle statistik i figur 1 er en undersøgelse af konteksten til venstre for ordet *båd*. Som man ser, står ordet *fods* øverst på listen i statistikken. Mellem *båd* og *fods* er der

altså relativt stor gensidig tiltrækningskraft; der er et højt tal i statistikens første kolonne (*mut. inf.*). Så længe dette tal er over 1, er der tale om en overhyppighed.

Corpus-Bench - BAADGAES.CBC	
File New Edit View Options Help	
CONC: båd (inf)	
IFCN	n det er vel et eller andet med, at du skal føle dig i samme båd som Stallone? — Nej, det er
IFCN	som Stallone? — Nej, det er ham, der skal føle sig i samme båd som mig (skraldgriner) ... b
19LE	hun blev udsat for i den forbindelse. Jeg har været i samme båd. Jeg var 17 år, da jeg blev
19LE	ne, holder vagt, o.s.v. Når det brænder på, er vi alle i samme båd, og det giver forståelse. —
12Kt	t af det intime samvær, der var mellem mennesker i samme båd på sygehuset, men alligevel
12Kt	nok." — Lone sukkede. Stakkels Jens. Han var altså i samme båd som viceværten. Jens kunne
KoRf	nsen — det navn der står på bl.a. visen "Vi er alle i samme båd". — Når far fik så mange opg
PaLH	ge Kjærulf Schmidt </fremhæv> synge "Vi er alle i samme båd". — Det, denne artikel vil pre
RHaF	at at være bange, for der var rigtig mange, der var i samme båd som os, fortsætter Merry. —
KWf1	ikke, vi har ikke råd, / midt om natten stod vi ud i samme båd / farligt vand at ligge og ko
K21V	ene op og binde sig sammen for at sige: vi er alle i samme båd. — Det var netop spontant. —
K21V	gumpe-tung hurtigleber! — Her får du to båd-typer i samme båd. — Som begge fungerer opti
100F	aja: Åhr hold kæft. — Elsebeth: Vi er jo alle sammen i samme båd, er vi ikke? — Maja: Hvorfor
100F	unden for det gamle ordsprog som siger: Vi er alle i samme båd. — </Tekst> </Korpusenhed>
01Kt	bet det gynger — og alle vi synger — for vi er jo alle i samme båd. — 2. Her har vi Richard — h
10SC	da det satte ind med storm. De fiskede også lidt fra samme båd, men det var jo ikke så far
01Vf	ed forskellige typer, der i "rejsetiden" er, om ikke i samme båd, så i samme tog. — Her var
10Vf	kan jeg i øvrigt tage Per Bregengaard fra VS med i samme båd, hvis de kunne forholde sig
01Vf	hold mellem passagererne ud fra devisen: Vi er alle i samme båd! — Er myldretids-toggangen t
EA0r	er samfundsforskere og politikere mere eller mindre i samme båd. — Men nu er det jo langt fra
EA0r	teresser, som er placeret i flere europæiske lande, i samme båd. — Der resterer stadig 500 p
05Jt	til nu uden held — og ved, at mange andre unge er i samme båd som jeg. Min generation lær
05Jt	er Folketinget ved det kommende valg. — For Ninn er i samme båd som Schlüter og H.P. Claus
Total words: 0   Node words: 0   Matches: 43	
Report updated.	

Figur 3. I samme båd kan man være — konkret, men også overført.

Af næste kolonne (*coocc.*) kan vi se, at der er 10 samforekomster med *fods* og *båd* i korpus. At ordet *fods* står højt på listen betyder altså ikke, at der er mange samforekomster — kun at der er relativt mange. Der er fx mange flere samforekomster med *den* — hele 171 — men denne samforekomst rangerer meget lavere på listen, for *den* er ikke meget mere tiltrukket af ordet *båd* end af så mange andre substantiver.

Tredje kolonne (*mean-pos*) i den mutuelle statistik fortæller, at ordet *fods* optræder (gennemsnitligt) på den første plads til venstre for *båd* — ikke på den anden eller tredje plads. Den sidste kolonne (*dev-pos*) angiver, at afvigelsen fra denne position er nul, dvs. *fods* står hver gang umiddelbart til venstre for *båd*.

Et kig i korpus afslører, at statistikens oplysninger dækker over eksempler i stil med *en 50fods båd*. *Fod* bliver altså overhyppigt brugt som længdemål om både, men dén oplysning vil jeg ikke bruge i min ordbogsartikel om *båd*, for jeg synes ikke, det er interessant nok.

Mere interessante er de ord der står længere nede i den mutuelle statistik, men som samtidig repræsenterer mange samforekomster, bl.a. *bord*, *samme*, *lille* og *i*. Disse samforekomster undersøger jeg nu nærmere gennem konkordanslister. Jeg skal finde ud af, om samforekomsterne skal med i ordbogen — og i hvilken form.

Figur 2 viser en konkordansliste, hvor jeg har isoleret de linier, der har ordet *bord* eller *ombord* i *båds* kontekst. Vi kan se, at der tegner sig en samforekomst i retning af *være om*

*bord i en båd* eller *være om bord på en båd*. Det kan være interessant for især brugere med dansk som andetsprog el. fremmedsprog at vide, at man kan bruge begge præpositionerne, *i* og *på*, i forbindelse med en båd. På den anden side er det nok ikke nødvendigt ligefrem at lemmatisere samforekomsterne med *om bord*, for betydningen giver sig selv, hvis man kender betydningerne af henholdsvis *om bord* og *båd*. I den færdige artikel har jeg derfor valgt en angivelse af typisk samforekomst i stedet for oprettelsen af et sublemma.

Endnu et kig i korpus gør det klart, at ordet *lille* i ordbogsartiklen kan behandles som et typisk adled til *båd*, men at der så heller ikke er grund til at gøre mere ud af det.

Af den næste figur, figur 3, kan vi se baggrunden for, at ordet *samme* stod højt på listen i den mutuelle statistik. Der lader til at være en etableret brug af udtrykket *være i samme båd* i overført betydning. Konkordanslisten har flere eksempler på, at *alle* er i samme båd, men jeg opretter et sublemma, der lyder: *være i samme båd* og betragter det som en ordgruppe med overført betydning som et idiom.

Corpus Bench - BAADGAES.CBC

File New Edit View Options Help

CONC: båd (inf)

104 om en stand på land. <fremhæv>Finansfirmaer går i bådene </fremhæv> <fremhæv>Sej  
 111 havde en slagside på 45 grader, og at besætningen gik i bådene. <fremhæv>De fem søfolk, der anses fo  
 112 ig foran skibet og skød på os, så såre vi var kommet i bådene. Ingen blev ramt, men vi roed  
 113 et på bagbord side. Den 32-mand store besætning gik i bådene, og "Kharq 5" drev siden ind  
 114 kkelser. <fremhæv>I BÅDENE</fremhæv> <fremhæv>Alle mand i bådene! blev der råbt, da der udbrød  
 115 "et skib på kollisions-kurs", og at han selv "er gået i bådene for at samle overlevende op."  
 116 </Spb> </Header> <Tekst ID=EBPf> <fremhæv>Alle mand i bådene </fremhæv>145 søfolk og r  
 117 hikske. Udfør Nordkap gik 16 besætningsmedlemmer i bådene, inden de tilbageværende sejl  
 118 <replik>.. Spagelser</replik> <replik>Alle mand i bådene!</replik> <replik>Hjælp!  
 --- NO MORE MATCHES ---

Total words: 0	Node words: 0	Matches: 9
----------------	---------------	------------

Report updated

Figur 4. Korpus viser variationer over sublemmatet *gå i bådene*. Her er fx konstruktioner med verberne *komme* og *sende*, og her er eksempler på kommandoen: *alle mand i bådene!*

Ordet *i* er der næsten 500 samforekomster med, viste den mutuelle statistik, og jeg er interesseret i at vide, om der er bestemte verber, man bevæger sig i både med. Min sprogfornemmelse siger mig, at der er noget, der hedder *gå i bådene*, hvor *bådene* betyder 'skibets redningsbåde'. Det bekræftes af konkordanslisten i figur 4, der viser samforekomster

med *båd* og *i*. Jeg opretter herefter et sublemma, der lyder: *gå i bådene*, og jeg betragter det som en kollokation, fordi betydningen er konkret.

Mit færdige udkast til artikel om *båd* kom til at se ud som i figur 5. Artiklen angiver et typisk adled, en typisk samforekomst og to sublemmaer. Begge sublemmaer har betydningsbeskrivelse og citater. Citatet under sublemmaet *være i samme båd* skal bl.a. illustrere, at man typisk taler om, at alle er i samme båd.

**båd** sb. -en, -e, -ene

**1** mindre fartøj, der drives frem vha. årer, sejl el. motor typisk lille ~; om bord i/på en båd □ han [havde] allerede fået sportsvogn og ridehest og båd gå i bådene gå om bord i et (synkende) skibs redningsbåde el. -flåder □ coasteren havde en slagside på 45 grader, og .. besætningen gik i bådene være i samme båd ofø være i samme (problemfyldte) situation □ Vi var alle i samme båd, på godt og ondt. Det styrkede fællesskabet.

**2** større skib, der sejler i fast rutefart typisk om bord i/på en båd □ fyren .. var blevet sat på englandsbåden på sin polterabend, og på båden møder han sit livs kærlighed.

**3** trekantet stykke af rund frugt el. grøntsag, hvor stykkets ene side er frugtens buede overflade □ Skræl æblerne og skær dem i både.

SMS: til bet. 1 bådehavn, bådmotor; \*motor-, \*ro-, \*sejlbåd

Figur 5. Et udkast til ordbogsartiklen om ordet *båd*.

### 3 Problemer

Men er det nu så indlysende, at artiklen om ordet *båd* skal se netop sådan ud? Man kunne vel tænke sig, at korpus' oplysninger om samforekomster havde været behandlet anderledes?

Fx kunne sublemmaet *gå i bådene* måske have været udeladt. I stedet kunne jeg have oprettet delbetydningen 'redningsbåd'. Jeg valgte en lemmatisering, fordi min sprogfornemmelse siger mig, at der er tale om et fast udtryk, bl.a. fordi der sker tryktab. Der er ikke tryk på verbet *gå*, men derimod på *bådene*, og dét er med til at cementere udtrykket. Desuden er *båd* i denne betydning altid i bestemt form pluralis, uanset hvor mange redningsbåde der er tale om. Var der kun én redningsbåd, ville man stadig råbe: *Alle mand i bådene! Kvinder og børn først!*

Man kan diskutere, om det var nødvendigt at lemmatisere *gå i bådene*, men man kunne modsat også hævde, at der burde have været flere sublemmaer i artiklen. Det gælder fx en ordgruppe, jeg har kategoriseret som en typisk samforekomst, *om bord i/på en båd*. Det gælder også samforekomsten *åben båd*, som man kunne vælge at opfatte som en term, der fortjener lemmatisering, men som jeg slet ikke har taget med.

Man kan også diskutere, om det var rimeligt at angive *lille båd* som en typisk samforekomst, når det nu gælder for vældig mange genstande, at de kan betegnes som små. Måske var det mere relevant at angive samforekomsterne med *fods*, selvom der var færre af dem.



Endelig kan man diskutere, om sublemmaerne i artiklen om *båd* i tilstrækkelig grad tager højde for de variationer, som korpus viste for udtrykkene *gå i bådene* og *være i samme båd*. Variationer behandler vi forskelligt alt efter, af hvilken art de er. Figur 6 viser nogle eksempler på, hvordan variationsproblemer kan løses.

Som vi har set, indgår der mange overvejelser i beslutningen om lemmatisering versus ikke-lemmatisering af en flerordsforbindelse.

Mutuelle statistikker skal holdes op mod konkordanslister, hvor de konkrete eksempler bag statistikkens tal findes — her afslører bl.a. homograferne sig. Konkordanslisterne skal så holdes op imod andre kilder, bl.a. andre ordbøger. Og ordbogsredaktøren må holde alle kilderne op mod sin egen sprogfornemmelse — og i sidste ende forsøge at forholde sig kritisk til samme sprogfornemmelse!

<b>spurv</b>	(..) skyde spurve/gråspurve med kanoner
<b>jod</b>	(..) rød(t) jod
<b>hals</b>	(..) få galt i halsen el. få i den gale hals
<b>gelé</b>	(..) have gelé i knæene (..) varierer i form <i>Hans sensuelle øjne og bløde basstemme gi'r pigerne gele i knæene</i>
<b>alter</b>	(..) ofre NGT på NGTs alter (..) <i>velfærdstanken .. er nu blevet ofret på overbudets og frådseriets alter</i>
<b>skrive</b>	(..) skrive 'af (..) [NGN skriver NGT af; NGN skriver af efter NGT/NGN; NGN skriver af fra NGT]
<b>luft</b>	(..) der er NGT i luften (..) <i>typisk der er forår i ~en, der er noget i ~en, der er romantik/øretæver i ~en □ Den 29-årige Esbjerg-spiller vendte ikke just den anden kind til, hvis der var lussinger i luften</i>

Figur 6. Et sublemma kan have variationer, og variationerne kan medtages i ordbogen på mere eller mindre pladsbesparende måder.

Vi har hele tiden ordbogens begrænsede omfang at tænke på, og lemmatisering er en pladskrævende løsning. I hvert enkelt tilfælde må vi tage stilling til, om der er tale om en begrebslig helhed, som brugeren skal kunne slå op — eller om der er tale om et valensmønster eller mere eller mindre tilfældige eller typiske samforekomster, der ikke behøver lemmatisering. Når vi tager stilling, må vi se på artiklen som helhed og vurdere de oplysninger, som artiklen i øvrigt giver om hovedlemmaet.

Vi skal finde den løsning, der giver størst mulig brugervenlighed på mindst mulig plads!