

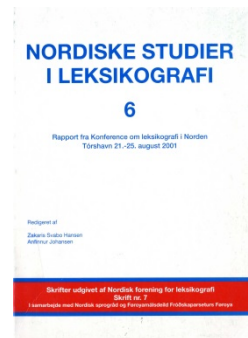
# NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Leksikalsk materiale fra sekundære kilder og ophavsret

Forfatter: Rita Lenstrup

Kilde: Nordiske Studier i Leksikografi 6, 2003, s. 203-208  
Rapport fra Konference om leksikografi i Norden, Tórshavn 21.-25. august 2001

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

## Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

## Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

RITA LENSTRUP

## Leksikalsk materiale fra sekundære kilder og ophavsret

This subject is important in the light of the growing number of machine-readable dictionaries because they are secondary sources that are easy to copy from for inclusion in new dictionaries (Svensén 1993:62). This paper will take its starting point in the question of how much lexical material can be drawn from secondary sources such as other dictionaries for the inclusion in a new dictionary without risking to infringe upon the copyrights of others; but in particular it will focus on how we can represent what and how much we want to draw, or how much we think might have been drawn from another secondary source for the use in a new dictionary? Two Danish English LSP dictionaries have been used for the illustration of practical attempts at such comparison.

### Baggrund

Emnet har fået ny interesse i lyset af det voksende antal maskinlæsbare ordbøger (Svensén 1993:62), og det bør derfor også indgå i undervisning i leksikografi. Emnet rejser mange spørgsmål, herunder de to spørgsmål, som mit indlæg især vil fokusere på, nemlig 1) Hvor meget kan der hentes fra sekundære kilder til brug for en ny ordbog uden at det krænker andres ophavsret? og 2) Hvordan kan der gøres rede for hvad der kan hentes eller kan være hentet fra en sekundær kilde til brug for en ny ordbog? Disse spørgsmål belyses med et praktisk eksempel fra to fagsproglige bilinguale ordbøger.

### Fagsprogslemmaer som et særligt udvalg

Bo Svensén siger (ibid:60) at det er ordbogsforfatterens overlevelsesinstinkt der gør at han konsulterer flest mulige eksisterende ordbøger, og Svensén berører spørgsmålet, men uden at gå i detaljer, om hvad og hvor meget leksikalsk materiale der kan inddrages fra eksisterende ordbøger i et ordbogsprodukt uden at krænke andre forfatters ophavsret.

Svensén siger (ibid:61) at ord er facts og facts er ikke beskyttet. Vil det så sige at ord, f.eks. opslagsord, uden videre kan hentes fra sekundære kilder? Nej, Svensén fortsætter (ibid:62) at hvor en eksisterende ordbog udgør "a certain selection of words and expressions [et særligt udvalg af opslagsord; min oversættelse]", et udvalg som er "defined in a certain way [defineret på en bestemt måde; min oversættelse]" dér må en sådan ordbog betragtes som et stykke originalt arbejde.

Og den tanke er jeg gået videre med idet jeg antager at en fagsproglig ordbog må være et udvalg af opslagsord, et udvalg som netop er "defineret på en bestemt måde". Altså en klart kvalitativ synsvinkel. Den kvantitative synsvinkel er også klar hvis det drejer sig om at genbruge en komplet ordliste fra en eksisterende ordbog. Så må genbrugeren ifølge Svensén (ibid:62) "come to an agreement with the publisher as to the right to use the dictionary in question as a basis [træffe en aftale med udgiveren om retten til at benytte den pågældende ordbog som forlæg; min oversættelse]." Emnet berøres af Bergenholtz der imidlertid ikke

diskuterer det men blot refererer til begrebet "ordbogskriminalitet", idet han siger at dette begreb i den metaleksikografiske litteratur bruges hvor en allerede forældet ordbog genoptrykkes uden henvisning til den tidligere udgave (Bergenholtz&Tarp1994:99)

Men hvis det kun drejer sig om genbrug af uddragne dele af en lemmaliste i en sekundær kilde, er der så ikke længere tale om krænkelse af ophavsret? Eller mere præcist – hvor lille en procentdel af en sekundær kilde skal man begrænse sig til for at sikre sig mod at begå krænkelse? Og har det nogen betydning om genbruget kun udgør en lille del i det produkt det er genbrugt i?

Sådanne spørgsmål er spørgsmål for jurister, men de har vakt min interesse fordi jeg gerne vil inddrage emnet *genbrug og ophavsret* i min undervisning efter at være kommet på sporet af det gennem daglig benyttelse af bilinguale ordbøger, for at forstå hvordan studerende i oversættelse og ordbogsbenyttelse kan være kommet frem til - jævnlige - skæve oversættelsesforslag, .

Jeg er nemlig stødt på en ny ordbog på et fagsprogligt område som bestandig giver samme svar som en anden tilsvarende ordbog giver, der som maskinlæsbar har været på markedet i nogle år. Den gamle, Dansk-Engelsk Økonomisk ordbog (Svensson 1996) kalder jeg herefter A og den nye, Dansk-engelsk Erhvervsordbog (Andersen et al.1999) kalder jeg B. Og jeg spørger selvfølgelig mig selv om det kan tænkes at B, der makrostrukturelt er helt forskellig fra A, alligevel i udstrakt grad kan have trukket på den maskinlæsbare A.

At ordbog B i et eller andet omfang har trukket på ordbog A fremgår af en oplysning i forordet til B om at "Ordbogens empiriske basis udgøres først og fremmest af eksisterende ordbøger samt forfatterens eget materiale fra mange års undervisning og translatørvirksomhed" (ibid:8), sammenholdt med at ordbog A optræder i bibliografien i ordbog B. Eller er den fagsproglige verdens natur sådan at to ordbøger produceret med få års mellemrum af forskellige lige kompetente forfattere på det samme felt uvægerligt vil komme til at fremtræde med stor konsistens i deres udvalg af opslagsord, ordforbindelser og ækvivalenter, siden parallelle opslag igen og igen giver ens svar.

### Praktisk sammenhold af to fagsproglige ordbøger

Men tanken om konsistens i udvalget af fagsproglige lemmaer er ren spekulation. Derfor har jeg valgt at se på den praktiske situation jeg havde foran mig for at finde frem til hvordan jeg overhovedet kunne fremlægge det der gav mig en fornemmelse af gentagelse i to ordbøger som umiddelbart for øjet er forskellige. Det var kun det leksikalske materiale der interesserede mig - ikke struktur, emnemarkører, symboler og lignende.

Jeg så først på hvad der teoretisk kunne have været hentet fra A idet jeg så på første side for hvert af bogstaverne **l**, **o** og **p** i ordbog B. For at illustrere hvordan jeg er kommet frem til den kvantitative præsentation i Fig. 2 gengiver jeg her i fig.1 et udsnit fra begyndelsen af bogstavet **o**, idet de dele i B der er identiske med dele i A er forsynet med grafisk markering i form af grå baggrund. Ordbogsudsnittet er genskrevet for at muliggøre elektronisk transport af den grå baggrundsfarve. Enkelte skrifttegn eller symboler kan derfor af tekniske grunde afvige i forhold til originalen. Jeg mener dog at det kun gælder symbolet  $\Rightarrow$ , der i originalen fremtræder med sort udfyldt spids.

o.a. (og andet) etc  
 o.lign. (og lignende) and the like, etc  
**objektiv** ADJ **objective**  
 σ *~vurdering* unbiased opinion; *~t ansvar* strict liability  
 ⇒  
**obligation** SUB  
 1. **bond**  
 2. (UK, selskabsobligation, også) **debenture**  
 σ *dollardenominerede* ~ dollar-denominated  
 bonds; *fastforrentede* ~ fixed-interest bonds;  
*forfaldne ~er* mature/redeemable bonds; *indfri en*  
*~* to redeem a bond; *indløse en ~* to redeem a bond;  
*konvertible ~er* convertible bonds; *kort ~* short  
 bond; *nul-kupon ~er* zero-coupon bonds; *~med en*  
*løbetid på 10 år* bond with a maturity of 10 years;  
*~er i stykker på kr. 100* bonds in denominations of  
 DKr100; *~er med variabel rente* floating rate  
 notes; *toneangivende* = (UK) mortgage bench-  
 mark bond; *toneangivende* ~ (US) bellwether  
 bond; *udstede ~er* to issue bonds; *udtrukket ~*  
 drawn/called bond

Fig. 1.

Det fremgår at f.eks. *fastforrentede* ~ "fixed-interest bonds" ikke er markeret med gråt, og altså ikke er talt med i diagrammerne i fig. 2, skønt *fastforrentet* som lemma og med ækvivalenten "fixed-interest" figurerer i A. Dette skyldes naturligvis at A ikke har medtaget *fastforrentet obligation* som "typisk ordforbindelse" (Andersen et al. 1999: 22) således som B har gjort det. Tilsvarende betragtning gælder ved ordforbindelsen med *variabel rente*. Det fremgår også at *konvertible ~er* med ækvivalenten convertible bonds er medregnet, her på trods af at forbindelsen ikke er i flertalsform i A. Det skyldes at jeg ikke har anset det for leksikalsk væsentligt om der benyttes ental i den ene ordbog's lemma og flertal i den anden's. Derfor er flertalsendelsen heller ikke regnet med i det grå felt, dog er endelsen -ble regnet med, idet den i antal typeenheder svarer til entalsendelsen -bel.

Den gråskraverede forbindelse *toneangivende* ~ og ækvivalentdelen "benchmark bond" er ikke regnet med, men jeg har overvejet at medregne dem. Det har jeg fordi lemmaet *toneangivende* ~ med ækvivalenten "benchmark bond" optræder i lemmalisten i såvel A som B, begge steder som sublemma. Men jeg har så alligevel fravalgt medtælling fordi

en total optælling af B ville medregne dette lemma og ækvivalenten dér hvor lemmaet er placeret i B.

Metoden med baggrundsmarkering rejser mange smålige spørgsmål, som jeg i optællingen har forholdt mig til ét for ét, og jeg håber at jeg hermed har antydet hvilke typer overvejelser jeg har gjort for at løse dem.

### Sammenfald i tal

Den grafiske gråmarkering giver et vist indblik og et visuelt indtryk af sammenfald, men er svær at få hold på, antagelig fordi den hverken er formaliseret eller kvantificeret. En optælling af typeenheder, uden medregning af mellemrum, for den første ordbogsside af hvert af bogstaverne l, o og p viser fordelingen mellem de gråmarkerede typeenheder, der viser identitet med A, og de ikke grå, der viser andet leksikalsk materiale, fig. 1.:

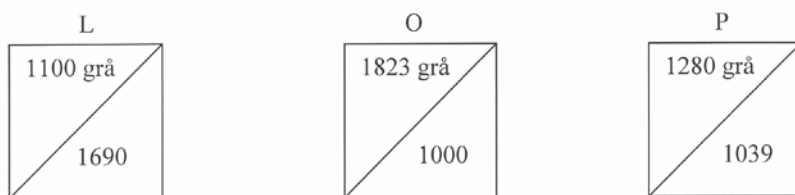


Fig. 2.

Dette billede er et udtryk for kvantitet, men om det er et udtryk for en uacceptabel grad af sammenfald, kan ikke umiddelbart siges. Billedet er nemlig måske et udtryk for at enhver ordbogsforfatter på området ville komme frem til en meget høj grad af ens leksikalsk materiale. At det samlede tal optalte typeenheder for de enkelte sider er ret forskelligt skyldes naturligvis at lange artikler indeholder mange typeenheder, medens korte artikler giver mere tomrum på papiret.

### Sammenfald i lemmaselektion og ordforbindelser

Hvis vi går tættere ind på hvad det grå består af, kan vi benytte en metode som Bergenholtz, Dam og Henriksen har benyttet i en artikel i *Hermes* (1990:127-136), hvor de sammenligner en ny fagsproglig ordbog med en allerede eksisterende almen bilingual ordbog på samme sprogpår.

Nu og her drejer det sig ganske vist om sammenligning af to fagsproglige ordbøger, men det interessante og for mig brugbare ved Bergenholtz et al.'s diskussion er at de peger på at en sammenligning kun er rimelig hvis der ses på såvel lemmaselektion som kollokationer, som i den pågældende fagsproglige ordbog anføres at være hovedsageligt sublemmaer (ibid:128). En sådan sammenligning indgår for så vidt også i ovenstående optælling, men ikke som *enheder* af lemmaselektion og kollokationer. Her ses derfor nu på lemmaer og kollokationer idet jeg ved afgørelsen af hvad der kan siges at være kollokationer holder mig til B's egen forklaring i ordbogens brugervejledning:

Ved mange opslagsord er der i et særligt felt givet eksempler på oversættelsen af typiske ordforbindelser, de såkaldte kollokationer. Dette felt er typografisk markeret

med en trekant(▲). De enkelte ordforbindelser er adskilt med semikolon (;), og der anvendes hverken store begyndelsesbogstaver eller tegnsætning. Den danske ordforbindelse er skrevet i kursiv. (Andersen et al.1999:22).

Tydligere kan det vel ikke siges, og den praktiske udmøntning må være at selv en ordforbindelse som *obligationer med en løbetid på 10 år* i B (se fig. 1) betragtes som en kollokation, og den tælles derfor her som en enhed. Desuden har jeg medtalt ”standardfraser” som ifølge vejledningen er ”typiske sætninger” (ibid:23). Jeg tager ligesom Bergenholtz et al. de 50 første af disse forekomster fra bogstaverne l og derudover også som ovenfor fra bogstaverne o og p og viser hvor mange der er ens i B og A ud af de første 50 i B. Kriterierne for enshed har ikke voldt problemer ud over dem i den kvantitative undersøgelse ovenfor. Resultatet fremgår af fig. 3:

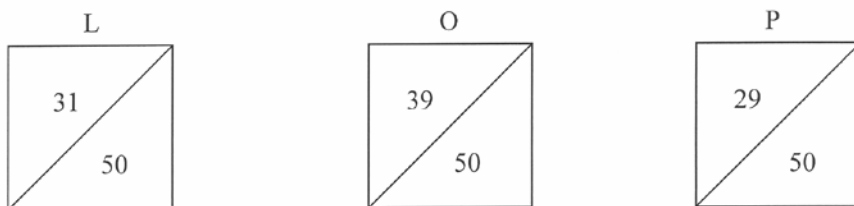


Fig. 3.

Hvis tallene havde været 50/50 eller måske blot 48-49/50 så ville der antagelig være tale om en ”komplet” ordliste (Svensén 1993:63) og derfor med Svenséns ord noget der er omfattet af ophavsret (forudsat ordlisten selv er ”the result of a creative contribution on the part of the copyright-owner [udgør resultatet af en kreativ indsats fra ophavsmandens side; min oversættelse]” (ibid:62)). Som det er her, er der måske snarere tale om at det fagsproglige område markerer sig så konsistent at der altid vil være så stor en grad af sammenfald mellem ordbøger af tilsvarende kvalitet som her. Dette spørgsmål vil man måske kunne få belyst ved at sammenholde ordbøgerne A og B med andre bilinguale ordbøger med samme sprogpar og retning, og på samme fagsproglige område hvis sådanne ordbøger findes, og ellers i mangel af bedre sammenligne med andre bilinguale ordbøger med samme udgangssprog og på samme fagsproglige område.

### Sammenfald på ækvivalentsiden

De elementer vi netop har set på her er forholdsvis korte tekststrengene, nemlig opslagsord eller forbindelser. Strengene bliver straks længere hvis vi medtager ækvivalenter. Men her opstår nye praktiske problemer, for mange af de optalte lemmaer har forskelligt antal ækvivalenter i de to ordbøger – der kan være flere hos A eller flere hos B. Desuden er rækkefølgen af ækvivalenterne jævnlige forskellig, hvilket skyldes forskellig opfattelse eller forskellig beregning af ækvivalenternes hyppighed. Et tilfælde med fem ækvivalenter i den ene ordbog og præcis de samme fem, men i anden rækkefølge i den anden ordbog ville være ekstremt, men der findes let tilfælde hvor en ækvivalent i B mangler i A, eller en anden er med i A,

hvor der er forskellig rækkefølge eller forskellig form for betydningsdifferentiering, altså punkter hvor det er svært at isolere det leksikalske fra struktur og principper.

Derfor stillede jeg et anderledes spørgsmål, nemlig: er der tilfælde hvor et opslagsord eller en forbindelse i B har en ækvivalentside, der på ingen måde gentager ord eller forbindelser fra A, og som derfor klart må have haft en anden kilde? Men det er der ikke blandt de optalte lemmaer og ordforbindelser fra bogstaverne l, o og p. For den mistænksomme bestyrker det mistanken om genbrug, men kunne det ikke lige så godt være et udtryk for konsistens på det fagsproglige område? Vi ved det ikke.

### Afsluttende betragtninger

Er det overhovedet noget problem at en stor del af et materiale i en ny ordbog er identisk med materiale i en ældre ordbog hvis den nye er struktureret efter andre principper? Nej, ikke hvis materialet er facts som ingen kan have ophavsret til.

Men hvornår er så de tekststrengene i en ordbog – lemmaer, lemmaer med ækvivalenter, lemmaer med forklaringer og ækvivalenter, og kollokationer med ækvivalenter - kollokationer som er sat sammen af de facts der også hedder ord - så kvalitativt markante, og hvornår er disse markante strenges selektion så kvantitativt sammenfaldende mellem en ny ordbog og en allerede eksisterende, at der er grund til at vogte sig for at krænke ophavsret, eller grund til med rimelighed at mene at der foreligger krænkelser af ophavsret? Også selv om den nye ordbog har en anderledes struktur?

Det er jurister der kan svare på disse spørgsmål, men det indblik i leksikografiens verden som juristerne skal have, det skal leksikografiske fagfolk levere. Og hvordan skal det se ud? Hvordan skal det gribes det an? Jeg har leveret forslag, som jeg vil kunne bruge i undervisningen til sammenligning af ordbøger, men måske vil den leksikografiske verden sige, at der skal noget helt andet til.

### Litteratur

- Andersen, Birger/Karin Balsgart/Morten Pilegaard/Sven Tarp 1999: *Dansk-engelsk erhvervsordbog*. Århus: Systime.
- Bergenholtz, Henning/Helle Dam/Torben Henriksen 1990: Udarbejdelse af en spansk-dansk juridisk ordbog: overvejelser og løsninger: *Hermes, Tidsskrift for sprogforskning* 5. Århus: Handelshøjskolen i Århus, 127 – 136.
- Bergenholtz, Henning/Sven Tarp 1994: Manual i fagleksikografi. Udarbejdelse af fagordbøger. Problemer og løsningsforslag. Herning: Systime.
- Svensén, Bo 1993: *Practical Lexicography. Principles and Methods of Dictionary Making*. Oxford:Oxford University Press.
- Svensson, Annemette Lyng 1996: *Dansk-Engelsk Økonomisk ordbog*. København: Samfundslitteratur.