


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	At have begreb skabt om noget – om Den Danske Begrebsordbog	
Forfatter:	Henrik Lorentzen, Sanni Nimb & Thomas Troelsgård	
Kilde:	Nordiske Studier i Leksikografi 12, 2013, s. 303-326 Rapport fra Konferanse om leksikografi i Norden, Oslo 13.-16. august 2013	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi 2014

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.

At have begreb skabt om noget – om Den Danske Begrebsordbog

Henrik Lorentzen, Sanni Nimb & Thomas Troelsgård

This paper describes the process of compiling a comprehensive Danish thesaurus at the Society for Danish Language and Literature (DSL) to be published in 2014. The thesaurus is mainly based on the sense descriptions and the lemma selection in our monolingual dictionary, Den Danske Ordbog (DDO). We will focus on the editing process and show how we use digital methods to establish annotated semantic groups which are afterwards automatically transferred to word class groups in the printed version of the thesaurus. The index constitutes an important part of the dictionary and will also be discussed, as will the future plans of making use of the thesaurus data, in particular in other dictionaries published by DSL.

En begrebsordbog over moderne dansk har været savnet i mange år. I denne artikel vil vi fortælle om baggrunden for og arbejdet med en ny dansk begrebsordbog der udkommer i begyndelsen af 2014. Artiklen er disponeret som følger. Først omtaler vi genren begrebsordbøger generelt, hvorefter vi går i detaljer med Den Danske Begrebsordbog (DDB). Fokus vil ligge på det praktiske arbejde med udtræk af ordstof fra andre baser og på sortering og opmærkning af materialet. Vi vil også komme ind på overvejelser vedrørende brugssituationer og de konsekvenser de har for præsentationen i den trykte bog, for dens layout og typografi og for indekset der udgør en væsentlig

del af bogen og dens tilgangsstruktur. Endelig afslutter vi med fremtidsperspektiverne, herunder især hvordan vi forventer at udnytte begrebsordbogens data på ordbogssitet <ordnet.dk> under Det Danske Sprog- og Litteraturselskab (DSL).¹

1. Begrebsordbøger

Slår man op i *Den Danske Ordbog* (DDO) under *begrebsordbog*, får man følgende definition: ‘særlig ordbog eller ordliste hvor ordene er ordnet efter betydning (og altså ikke alfabetisk)’, idet man oplyses om synonymet *thesaurus*. Man får altså det indtryk at den prototypiske ordbog er alfabetisk, og at begrebsordbogen således udgør en undtagelse. Det stemmer givetvis også med den moderne opfattelse (der dog er under forandring med de digitale ordbøger, jf. Trap-Jensen 2014), men i historisk perspektiv er tematisk opstillede ordlister ikke noget særsyn. Blandt andet af indlæringsmæssige hensyn er denne ordning mere velegnet og kendes tilbage fra oldtid, middelalder og renaissance (se hertil bl.a. Boeck 2009).

I 1800-tallet dukker en ny type tematisk ordbog op hvor ord med både konkrete og abstrakte betydninger ikke blot listes under et emne, men yderligere grupperes ud fra betydning. Den første og internationalt stadig mest kendte er for engelsk og udkom i 1852: *Roget's Thesaurus* med 15.000 ord og 1.000

1. Vi vil gerne takke vores kolleger Liisa Theilgaard og Lars Trap-Jensen for nyttige kommentarer til denne artikel.

emnegrupper i førsteudgaven. Først langt senere fik tysk sin begrebsordbog da Dornseiffs *Der deutsche Wortschatz nach Sachgruppen* udkom i 1934. Den bruger en lidt anden emneinddeling og lander på 906 grupper. Cirka 10 år senere kom så også en begrebsordbog for dansk: Harry Andersens *Dansk Begrebsordbog* (1945). Andersen inddeler ordstoffet i 892 grupper, idet han lægger sig tæt op ad Dornseiff, »hvis Inddeling under nummererede Grupper næsten helt er fulgt« (Andersen 1945:IX).

Roget's Thesaurus, der altså var pioner og forbillede på området, kan roligt regnes for en af ordbogshistoriens største salgssucceser. Siden førsteudgaven er den udkommet i 32 mio. eksemplarer og er mere end tyvedoblet i omfang: Den nyeste udgave (2002) indeholder ifølge bagsideteksten mere end 300.000 ord og udtryk. Denne sproglige rigdom, som også betegnelsen *thesaurus* (oldgræsk *thesauros* 'forråd', 'skatkammer') antyder, stilles til rådighed for den bruger der søger sproglig inspiration, ønsker at finde det helt rigtige ord. Skulle man have glemt det, kan man med fordel slå op i Roget under fx *forgotten* og der finde en lang række forslag til ord og udtryk der betyder noget af det samme:

Adj. *forgotten*, clean f., beyond recall; well forgotten, not missed; unremembered, left in limbo 458 *neglected*; disremembered, misremembered etc. vb. almost remembered, on the tip of one's tongue; in the recesses of one's mind, gone out of one's head, passed out of recollection; burried, suppressed ...

Ikke mindst udtrykket *on the tip of one's tongue* inkarnerer en væsentlig funktion for begrebsordbøger, nemlig at lokke det frem der venter på at blive brugt, at hjælpe brugeren med at komme i tanke om synonymmer og beslægtede ord og begreber.

2. Den Danske Begrebsordbog (DDB)

Nu er der så en ny dansk begrebsordbog på vej. Projektet og udgivelsen er muliggjort takket være en bevilling fra Carlsbergfondet og gives eksplicit til en trykt ordbog. Projektet løber over 3 år, fra 2010 til 2013, manuskriptudarbejdelsen nærmer sig i skrivende stund sin afslutning, og bogen forventes at udkomme i efteråret 2014. Redaktionen består af Sanni Nimb, Thomas Troelsgård, Liisa Theilgaard, Lars Trap-Jensen og Henrik Lorentzen. Titlen på den nye ordbog bliver *Den Danske Begrebsordbog*, DDB, og lægger sig derved i slipstrømmen af DDO.

2.1. Ønsker og omfang

Udover at der skulle komme en trykt ordbog på gaden, havde vi en række andre ønsker til projektet. Først og fremmest vil vi gerne fremtidssikre vores data i den forstand at den trykte ordbog blot er ét udtræk blandt flere mulige, og at det bliver muligt at anvende begrebsordbogens data i vores andre ordbøger. Det kan dels være i form af integration med de eksisterende ordbøger, dels i form af en selvstændig onlineversion på <ordnet.dk>.

Med hensyn til omfanget har vi haft som ideal at medtage alle ord og udtryk i DDO suppleret med nyt materiale. På nuværende tidspunkt estimeres det at vi når 80 % af DDO's materiale, dvs.

- 180.000 ord og udtryk (inkl. gengangere) \approx
- 100.000 forskellige ord og udtryk
- 5.000 nye ord og udtryk

Det store antal gengangere dækker over at mange ord og udtryk helt naturligt optræder i flere emnegrupper, fx findes ordet *ordbog* i øjeblikket i tre afsnit: *11.30 Kundskab, viden, erfaring;* *12.16 Betegnelse, ord;* *12.58 Bog, hæfte.*

2.2. Forudsætninger

Der er nogle gode forudsætninger for at dette projekt kan lade sig gennemføre på så kort tid. Begrebsordbogen ligger nemlig i klar forlængelse af DSL's seneste ordbogsarbejde for moderne dansk: DDO og det danske wordnet DanNet (se Pedersen mfl. 2009), og det ligger lige for at udnytte data fra disse to resurser.

DDO indeholder ca. 117.000 betydninger der er fordelt på 110.000 ord og faste udtryk og i mange tilfælde er forsynet med emnetilhørsforhold (som fx *militær* eller *håndarbejde*). Dertil kommer ca. 18.000 afledninger og sammensætninger uden betydningsbeskrivelse. I DanNet finder man cirka halvdelen af dette materiale organiseret i begrebshierarkier, hvor fx spil som 'håndbold' og 'volleyball' er placeret som underbegreber til 'boldspil', der igen er underbegreb til 'sportsgren'. Onlineudgaven af DDO giver allerede i dag brugeren mulighed for at se en del oplysninger om synonymer og andre beslægtede ord. Den trykte bog var forsynet med en del synonymer, antonymer, ord fra samme ordfelt osv. De fungerer som klikbare links i netversionen og kan sende brugeren på opdagelse i ordforrådet.

Begrebshierarkiet fra DanNet er også udnyttet i funktionen »Beslægtede ord« (se Sørensen & Trap-Jensen 2010), hvor man får oplysninger om over- og underbegreber og kan få dem visualiseret i klikbare figurer (stillet til rådighed på <www.andreord.dk>). I figur 1 ser vi således eksempler på synonymer og betydningsbeslægtede ord i artiklen **vigtig** i DDO:

vigtig adjektiv
BØJNING -t, -e || -ere, -st/(-est)

Betydninger

1. som har stor betydning eller relevans

SYNONYMER betydningsfuld | væsentlig SE OGSÅ magtpåliggende

For os er det meget vigtigt at understrege, at pasningsordningen ikke er en opbevaringsordning men skal indeholde stimulation for børnene DetGrøOmr86

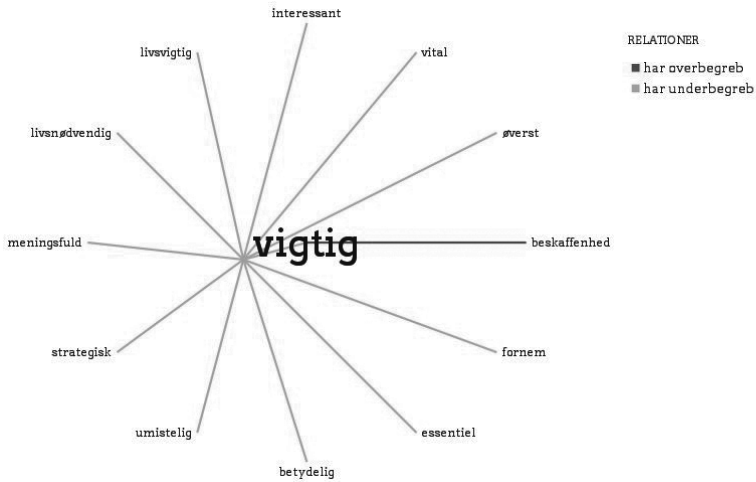
Figur 1: Uddrag af artiklen **vigtig** i DDO.

Tilsvarende ser vi i figur 2 en grafisk fremstilling af *vigtig*'s beslægtede begreber. Bemærk at man får serveret flere ord end i DDO-visningen ovenfor, fx *vital*, *interessant*, *betydelig*, *umistelig*, *essentiell* osv.

vigtig

ADJ

som har stor betydning el. relevans



Figur 2: Ordet *vigtig* med over- og underbegreber på <www.andreord.dk>.

Ingen af disse visninger er dog så detaljeret og omfattende som den kommende begrebsordbog, idet vi ikke bare medtager flere synonyme, men også semantisk relaterede ord og udtryk fra andre ordklasser. Vi kan her sammenligne med afsnittet *Væsentlighed* i DDB, hvor *vigtig* optræder sammen med ca. 250 andre ord og udtryk, her kun et mindre udsnit (fra redigeringsværktøjet, se figur 3).

·væsentlig, ·relevant, ·essentiell, substantiel, af interesse, ·bemærkelsesværdig, ·interessant, ·ikke uinteressant, aktuel, mere aktuel end nogensinde, ·tung, ·vægtig, ·tungtvejende, ·betydningsfuld, ·af stor betydning, ·vigtig, ·værdifuld, ·betydelig, ·toneangivende, ·fremherskende, ·betydende, ·vedkommende, ·meningsfuld, ·bærende, ·fundamental, ·grundlæggende, ·central, ·nøgle-, ·egentlig, ·strategisk, ·uundværlig, ·afgørende, ·skelsættende, ·udslagsgivende, ·vital;

Figur 3: Afsnittet *Væsentlighed* i DDB (uddrag).

I den del af afsnittet vi ikke ser her, findes også andre ordklasser end adjektiver. Her nævnes fx substantiver og substantiviske udtryk som *essens*, *sagens kerne* og verbaludtryk som *spille en rolle*, *gøre sig gældende*. Med lidt andre nuancer følger senere bl.a. *hovedperson*, *nøglefigur*; *vægte*, *prioritere* osv., forhåbentlig relevante forslag for den bruger søger hjælp til at udtrykke sig, jf. udvalget fra Roget i afsnit 1.

2.3. DDB overfor andre begrebsordbøger

Inden vi går over til at beskrive detaljerne i redigeringsarbejdet, vil vi kort sammenligne DDB med eksisterende begrebsordbøger. Hvad angår omfang, kommer DDB til at placere sig i den tungere ende. Med sine ca. 100.000 unikke ord bliver den ca. tre gange større end forgængeren *Dansk Begrebsordbog* (Andersen 1945, ca. 33.500 ord), den bliver væsentligt større end den nyeste større norske begrebsordbog *Ord og begreper* (Rosbach 2001, ca. 7.000 ord). DDB bliver også noget større end den

tyske Dornseiff (2004, ca. 90.000 ord inkl. gengangere), men til gengæld bliver den mindre end *Roget's Thesaurus* (2002, 300.000 ord inkl. gengangere).

Pladsen tillader ikke en grundig komparativ analyse, men det er oplagt at sammenligne lidt nærmere med den danske for-gænger Andersen (1945). Vores ordbog er for det første korpus-baseret, noget der ikke var muligt for næsten 70 år siden, hvor introspektion var det bedste man kunne gøre. Således indeholder den flere specifikke begreber, bl.a. i form af sammen-satte ord fra korpus. Disse er registreret i DDO og lette at få fat i, som vi ser i figur 4 med ordet *problem*:

▷·**problem**, ·**problematik**, ·**problemstilling**, ·**dilemma**,
·**vanskelighed**, ·**kattepine**, ·**knibe**, ·**klemme**, ·**suppedas**^{ufarm.}
...
samspilsproblem, problembarn; ▷**samfundsproblem**,
verdensproblem, beskæftigelsesproblem,
·**arbejdsløshedsproblem**, flaskehalsproblem, ·**miljøproblem**
, ·**affaldsproblem**, resurseproblem◄; ▷·**narkoproblem**,
·**narkotikaproblem**◄; ▷·**støjproblem**, ·**trafikproblem**◄;
strukturproblem; ▷**helbredsproblem**, vægtproblem,
spiseproblem, søvnproblem, ·**vejtrækningsproblem**,
hjerteproblem, ·**rygproblem**, potensproblem,
·**orgasme**problem◄; ▷·**misbrugsproblem**,
spiritusproblem, alkoholproblem◄; ·**ensomhedsproblem**;
▷·**pengeproblem**, likviditetsproblem◄; ·**transportproblem**

Figur 4: Uddrag af afsnittet *Problem* i DDB.

Foruden synonymer som *dilemma*, *vanskelighed*, *kattepine* osv. er der også nævnt mange specifikke problemer, fx samfundsproblemer som *beskæftigelsesproblem*, *miljøproblem*, *narkoproblem*. Andersen udelod med vilje sammensætninger ligesom han bevidst ikke gjorde meget ud af områder som dyr og planter

(Andersen 1945:IX); her tager vi meget mere med, igen i tråd med DDO's lemnaudvalg.

Der er heller ingen områder der på forhånd er erklæret tabu, så vi medtager også mange begreber indenfor områder som sex, racisme osv., felter der af forskellige grunde kunne være vanskelige at berøre i 1940'erne. Strukturelt gør vi også det at vi fortsætter den semantiske gruppering så langt ned vi kan komme, til laveste niveau (se Nimb 2013 for en nærmere beskrivelse), hvor Andersen på dette niveau lister begreberne alfabetisk, jf. figur 5 og 6:

A. aaben, fremkommelig, fri, gennemtrængelig, gisten, hul, hullet, læk, porøs, svampet, tilgængelig, utillukket.

Figur 5: Adjektivafsnittet under *Aabne* hos Andersen (1945).

åben, åben, ·åben, helt åben, vidtåben, pivåben^{uform.},
·åbentstående, utillukket, oplukket;
▷gennemtrængelig, ·fremkommelig, ·fri, tilgængelig◁;
·opknappet^{uform.}; ▷halvt åben, halvåben, på klem◁;
stå åben, stå på klem, stå på vid gab, ·gabe, ·gabe,
·lække→→→ Tivoli er åben^{spøg.}, oplukkelig
hullet, ·gennemhullet, ·gennembrudt, gennemstukket
; ▷læk, gisten^{gl.}◁; ·porøs, ·svampet

Figur 6: Grupperne med *åben* og *hullet* i DDB-afsnittet *Åben, åbning*.

På denne måde mener vi at sikre den størst mulige semantiske sammenhæng mellem et begreb og dets naboer i begrebsordbogen.

3. Det praktiske arbejde

3.1. Afsnit

DDB's overordnede struktur er baseret på Dornseiff (2004), dvs. med en opdeling i 22 kapitler, med i alt ca. 900 afsnit (ca. 20 i det mindste kapitel, over 80 i det største). Inddelingen begynder med verden omkring os og går derefter via handlinger, følelser, samfundsmæssige fænomener m.m. til kapitel 22: *Religion*. Figur 7 viser et udsnit af DDB's makrostruktur.

1	Natur og omverden	
2	Liv	
3	Rumlig dimension, bel...	1.1
4	Størrelse, mæ...	1.2
5	Væsen, forhold,	1.3
6	Tid	1.4
7	Synlighed, lys, farve,	1.5
8	Sted og stedforandring	1.6
9	Vilje og handling	1.7
10	Følelse, affekt, karakter	1.8
11	Tænkning	1.9
12	Tegn, meddelelse, spro	1.10
13	Videnskab	1.11
14	Kunst og kultur	...
15	Socialt liv	1.25
16	Mad og drikke	
17	Sport og fritid	
...		
22	Religion, oversanselig	

Figur 7: Udsnit af makrostrukturen i DDB med kapitler og afsnit.

Den overordnede struktur i DDB er overtaget fra Dornseiff, idet det dog – udover selve oversættelsen af afsnittenes overskrifter – har været nødvendigt også at foretage en del justeringer i strukturen på grund af sproglige og kulturelle forskelle. Således har det bl.a. vist sig at den »nedarvede« makrostruktur på nogle områder afspejler et ældre samfunds- og menneskesyn der ikke matcher ordforrådet i moderne dansk, ligesom den tekniske udvikling har medført at ord fra nogle afsnit i dagens sprogbrug er blevet perifere.

Som følge af sådanne forskelle har vi fx opdelt afsnittet *Spiele* i Dornseiff i to afsnit i DDB: *Spil* og *Leg*, og vi har nedlagt afsnit som *Vandvej* og *Bilindustri*, og et afsnit som *Arbejdsløshed* er blevet oprettet, hvor det hos Dornseiff blot står under *Untätigkeit*, bl.a. sammen med ord for ‘pause’ og ‘dovenskab’. Også afsnit som fx *Agtelse*, *Lydig*, *Kryberi*, *Servilitet*, *Skabagtighed*, *Soning*, *Ildsted*, *Ondsindet*, *Vogter*, *Barbari*, *Befaling* og *Tvang* er blevet nedlagt, idet de ord der ville stå i disse afsnit, har kunnet placeres i andre afsnit; fx kandiderer ord fra *Soning* til placering under *Dårlig samvittighed* og *Afsoning*, og ord fra *Ondsindet* kan placeres under *Had* eller *Fjendskab*.

3.2. Ordstoffet

Indhentning af egnede ord og udtryk sker fra flere kilder:

- DDO – via søgninger på ordformer, faste udtryk (sublemmaer), kollokationer, definitioner, genus proximum, fagmarkeringer m.v.
- DanNet – via søgning på overbegreb

- manuelle opslag, primært i Andersen (1945), men også forskellige synonymordbøger og fremmedordbøger
- introspektion

Tager vi et eksempel som *hat*, vil man i DDO – foruden selve ordet *hat* – kunne finde sammensætninger som *stråhat* og *tyrolerhat*, sublemmaet *høj hat* samt kollokationseksempler som *blød hat* og *bredskygget hat*.

Det er typisk under arbejdet med indhentning af egnet ordmateriale at det har vist sig om der var grundlag for at ændre et afsnits struktur eller overskrift (jf. 3.1. ovenfor).

3.3. Opmærkning og sortering

Hvert afsnit består af en række grupper defineret ved en formel relation der beskriver forholdet mellem ordene i gruppen og et eller flere nøglebegreber (se Nimb 2011, Nimb 2013), fx:

- genstande der har et fælles overbegreb
- personer der udfører en bestemt handling
- egenskaber der vedrører et bestemt fænomen
- handlinger der vedrører et bestemt fænomen

Ordene i hver gruppe ordnes med det mest overordnede først, efterfulgt af det mere underordnede og specifikke, ofte undergrupperet, så det i den trykte bog (ved hjælp af separatorer) kan markeres at en mindre gruppe ord er nært knyttet til hinanden semantisk. Endelig har vi givet mulighed for til sidst i en gruppe at placere nogle ord i en løsere tilknyttet »hale«, dvs. en

samling af ord der ikke dækkes af den formelle relation, men alligevel er relevante for den pågældende gruppe (og som ikke passer ind i andre af afsnittets grupper).

Grupperne ordnes så de mest centrale og overordnede udtryk placeres først; på grund af materialets heterogenitet har vi ikke fulgt et fast mønster, men taget udgangspunkt i de grupperinger der naturligt viste sig under redigeringen af det konkrete afsnit, idet semantisk nærhed har været det overordnede princip.

Figur 8 viser afsnittet *Utroskab* i DDB. Efter afsnittets nummer og navn følger tre henvisninger til relaterede afsnit, og herefter selve ordstoffet, ordnet i grupper. Først i hver gruppe er den formelle relation anført; således rummer den første gruppe ord og udtryk der betegner handlingen 'at være utro'. Næste gruppe rummer adjektiver og adverbier med betydningen 'som

15.008.Utroskab

↗10.052 *Kærlighed, forelskelse* ↗15.006 *Ægteskab, samliv* ↗15.036 *Prostitution, frigjorthed*

- ⊙ {08_Vb_SbAflædning/has_hyperonym: 'handle involved_agent: 'person}
 - ▷·være nogen utro, ·være utro mod nogen, ·bolle udenom^{uform.}, ·bolle ved siden af^{uform.}, ·snyde◄; ▷·begå ægteskabsbrud, ·bedrage, ·besvige◄; ▷·beligge^{gl.}, ·ligge med^{gl.}, ·ligge hos^{gl.}, ·ligge i med^{uform.}◄; ▷·bole^{gl.}, ·hore, ·bedrive hor^{gl.}◄; ▷·gøre til hanrej^{gl.}, sætte/give nogen horn i panden◄;
- ⊙ {01_Overbegreb/has_hyperonym: 'uægteskabelig}
 - ▷·uægteskabelig, ·førægteskabelig, ·udenoms◄; til venstre hånd
- ⊙ {01_Overbegreb/has_hyperonym: 'utroskab}
 - ▷·utroskab, ·ægteskabsbrud, ·sidespring, ·troløshed◄; ▷·bigami, ·flerkoneri◄; ▷·hor^{gl.}, ·utugt, ·usædelighed◄; ▷·fri kærlighed, ·partnerbytte, ·åbent ægteskab◄; ▷·trekantsdrama, ·trekantsforhold◄; ægteskabsdrama
- ⊙ {03_Kerne_Adled/has_hyperonym: 'person}
 - ægteskabsbryder, ·bigamist; ▷·elsker, ·elskerinde◄; ▷·maitresse, frille, ·konkubine◄; ▷·horebuk^{eds.}, ·horkarl◄; ·hore^{gl.}, ·hanrej^{gl.}, ·swinger↔↔↔↔
 - ▷·utro, ·troløs, ·svigefuld^{gl.}, ·polygam, ·utugtig◄;

Figur 8: Afsnittet *Utroskab* i DDB's redigeringsværktøj.

har med utroskab at gøre'. Derefter følger en gruppe af betegnelser for 'utroskab' (i forskellige former), og endelig er der en gruppe af personbetegnelser der både rummer udøvere af og ofre for utroskab, og som også (i en hale) har en undergruppe af adjektiver med betydningen 'som udøver utroskab'.

4. Den trykte bog

4.1. Ordklassesortering

Efter redigeringsfasen, hvor semantisk nærhed er det primære princip for gruppernes rækkefølge, følger til brug for den trykte bog en omgruppering hvor udtrykkene i hvert afsnit omstruktureres, idet:

- ordene grupperes efter ordklasse
- de oprindelige grænser mellem grupperne gengives ved separatorer
- nøgleord bevares og fremhæves; ord der kommer til at stå først i en gruppe, ændres til nøgleord
- dubletudtryk der kommer til at stå i samme gruppe, fjernes

Når vi har fundet det nødvendigt med denne omgruppering, skyldes det at vi mener at brugeren er bedst tjent med en ordklassevis præsentation af begreberne, især den bruger der

ønsker et paradigmatiske alternativ i form af det samme eller et beslægtet begreb manifesteret i samme ordklasse. Desuden bliver bogpræsentationen homogen da hvert opslag følger samme struktur (jf. diskussionen i Trap-Jensen 2011:194ff.). I figur 9 ses et eksempel på hvordan den bogtrykte version forventes at blive, her med de fire store ordklasser substantiver, verber, adjektiver og adverbier.

4.21 Fuld

4.10 Tyk; 4.20 Mange, meget; 4.41 Helhed

SB • *komplet samling*, den totale mængde • *menneskefyldt* • fuldt hus
• *overfyldte busser* • overflod, overmål • virvar.

VB • **fylde op**, fylde, udfylde, fylde ud, fylde helt op, stoppe (med noget), proppe (med noget), stoppe til, overfylde, belæse • **være fyldt med vand** • være fyldt til randen, være proppet til randen • **fuldende** • løbe fuld • stå som sild i en tønde.

ADJ • **fuld**, helt fuld, fyldt • stopfuld, propfyldt, propfuld, stopfyldt, overfyldt, stuvende fuld/fyldt • tungt lastet • pakket, tæt pakket, fuldt lastet ANT tom • **komplet**, fuldendt, hel, fuldstændig, total, fuldtallig ANT ukomplet • **fyldt**, fyldt med/af mennesker, fuld af mennesker • helt fuld, propfyldt, propfuld, stopfuld, stopfyldt, overfyldt, stuvende fuld, stuvende fyldt • pakket, tæt pakket, fuldt lastet • (alt) besat, fuldt optaget, udsolgt ANT mennesketom, tom • fuldt op af/med noget.

ADV • *som sild i en tønde*, så tæt som sardiner i en dåse

Figur 9: Foreløbig udgave af afsnittet *Utroskab* i bogversionen af DDB.

Efter omgrupperingen følger en slutredigering, hvor de nye afsnit gennemgås med henblik på omplacering eller sletning af udtryk der er kommet til at stå uhensigtsmæssigt efter ændringen.

4.2. Indeks

Omkring halvdelen af den trykte bog kommer til at bestå af et alfabetisk indeks der henviser til forekomsterne i selve indholdsdel af ordbogen. I indekset medtages de flerordsudtryk der er sublemmaer i DDO (under deres hovedlemma), samt alle enkeltord (og DDO-flerordslemmaer). For hvert henvisningsmål anføres i indekset:

- afsnitsnummer
- ordklassegruppe
- nærmeste foregående nøgleord

Figur 10 viser et udsnit fra bogstavet B i en foreløbig udgave af det alfabetiske indeks. Hvis vi ser på ordformen *bakke*, fremgår det at den bl.a. findes i afsnittene *1.15 Geologi* ('forhøjning i landskabet'), *3.12 Skrå* ('hældende underlag'), *3.15 Understøtte* ('plade til at bære noget på', sammen med bl.a. *fad* og *bageplade*), *8.8 Trafik* ('køre baglæns'), *8.9 Vej, gade* ('hældende vejstrækning'), *8.29 Tilbage, baglæns* ('køre baglæns'), *19.4 Støtte* ('del af en skruestik'), *19.5 Beholder til flydende ting* ('bæger/holder til madvarer'), *19.6 Beholder til faste ting* ('plade til at bære noget på'). Derefter følger en række henvisninger vedr. flerordsudtryk med *bakke*.

bajonetsav 19.11 SB sav
bak 8.14 SB *dæk*, 8.29 ADV *tilbage*
bakelit 1.22 SB *plaststof*
bakgear 8.29 SB *bakgear*, 19.21 SB *gear*
bakkanal 17.27 SB *raveparty*
bakke 1.15 SB *bakke*, 3.12 SB *skrånende grund*, 3.15 SB *bakke*, 8.8 VB *dreje*, 8.9 SB *vejstrækning*, 8.29 VB *gå baglæns*, 19.4 SB *klemme*, 19.5 SB *bakke*, 19.6 SB *bakke*
bakke op 9.68 VB *støtte*
bakke på en pipe 16.11 VB *ryge*
bakke ud 21.5 VB *fortryde*
det går ned ad bakke 5.261 VB *gå tilbage*
en bakke.. 4.17 SB *en skål suppe*
op ad bakke 9.38 ADJ *anstrengende*, 9.53
ADJ *strabadserende*
bakkebord 19.3 SB *skænk*
bakkedrag 1.14 SB *bakkelandskab*, 4.12 SB *højdedrag*
bakkehæld 1.15 SB *bakke*, 3.12 SB *skrånende grund*
bakkekam 1.15 SB *bakke*
bakkeland 1.14 SB *bakkelandskab*, 1.15 SB *bakke*, 4.12 SB *højdedrag*

Figur 10: Udsnit af bogstavet B i DDB's alfabetiske indeks.

5. Fremtiden

Data fra DDB vil – udover at kunne publiceres som selvstændig onlineordbog i en eller anden form – kunne udnyttes i flere sammenhænge:

- DDB-data kan vises integreret i DDO eller udnyttes til forbedring af DDO's oplysninger om synonyme og især »Beslægtede ord« (jf. Trap-Jensen 2011).
- DDB-data kan berige DanNet, idet nogle af de registre-

rede relationer vil kunne oprettes i DanNet og linkene overføres dertil. Alene det faktum at alle ord i et afsnit per definition vedrører det område der beskrives af afsnittets overskrift, kan medvirke til en finere underopdeling af nogle af DanNets største »søskendeflokke« (jf. Nimb & Pedersen 2012, Nimb et al. 2013).

- DDB-data vil kunne anvendes ved semantisk opmærkning af korpuser, idet man ud fra ordbestanden i et bestemt tekststykke vil kunne identificere tekstens emneområde i de tilfælde hvor mange af ordene optræder i et bestemt DDB-afsnit. Ved automatisk semantisk analyse af korpustekster vil en sådan opmærkning bl.a. kunne understøtte disambiguering af homografe former samt udnyttes ved betydningsbestemmelse af polysemer.
- Nyt ordstof fundet ved introspektion eller fra andre kilder end DDO er oplagte kandidater til nye ord i DDO (se Lorentzen & Nimb 2011).

Som nævnt ovenfor er det oplagt at udnytte begrebsordbogens data til at forbedre oplysningerne under »Beslægtede ord« i DDO. For tiden er disse oplysninger baseret på DanNet, der har sit hovedfokus på hierarkier af over- og underbegreber. Ved inddragelse af data fra DDB vil man kunne vise flere af de semantiske relationer et begreb indgår i. Figur 11 viser hvordan fx *røg*, udover sin placering i ontologien som 'partikelholdig luftform' (øverste linje i tabellen), med ændret fokus tillige kan tilknyttes en række andre emnegrupper, hver gang med nye ord som søskende.

<i>røg</i>	Luftform, fordampning	<i>em, os, dunst, uddunstning, ..</i>
	Ild	<i>os, røgsøjle, røgmasse, røgslør, aske, sod, ..</i>
	Uigennemsigtig	<i>tåge, damp, em, sky, ..</i>
	Let	<i>støv, luft, æter, ..</i>
	Tobak	<i>røgtåger, røgsky, tobaksrøg, piberøg, røgring, nikotin, tjære, kondensat, ..</i>

Figur 11: Fem mulige placeringer af *røg* med eksempler på søskendeord i den pågældende sammenhæng.

Sådanne oplysninger om et ords tilknytning til forskellige semantiske »klumper« bruges allerede i nogle ordbøger, fx hos Macmillan, hvor man ved de enkelte betydninger kan vælge tesaurus-visning af den pågældende betydning, dvs. at overskriften for den konkrete relation vises, efterfulgt af de første ord i emnegruppen. Macmillan giver endvidere brugeren mulighed for at klikke sig et skridt videre til en ren tesaurus-artikel med overskrift og alle gruppens ord.

Vi håber at kunne udnytte dataene fra DDB til en lignende berigelse af DDO-artiklerne på <ordnet.dk>, men om det bliver på samme måde, må tiden vise.

Litteratur

Alle webkilder er besøgt i oktober 2013.

Ordbøger

Andersen, Harry (1945): *Dansk Begrebsordbog*. København: Munksgaard.

andreord.dk = <www.andreord.dk> ved Anders Johannsen.

Viser data fra DanNet.

DanNet = Bolette Sandford Pedersen mfl.: *DanNet*, <wordnet.dk>.

DDO = Ebba Hjorth, Kjeld Kristensen mfl.: *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab & Gyldendal, online: <ordnet.dk/ddo>.

Dornseiff, Franz (2004): *Der deutsche Wortschatz nach Sachgruppen* (8. Auflage) (von Uwe Quasthoff). Berlin/New York: Walter de Gruyter.

Macmillan = *Macmillan Online Dictionary*, <www.macmillan-dictionary.com>.

Roget, Peter Mark (2002): *Roget's Thesaurus* (150th anniversary edition edited by George Davidson). London: Penguin.

Rosbach, Johan Hammond (2001): *Ord og begreper. Norsk thesaurus*. Oslo: Pax.

Anden litteratur

Boeck, Simon Skovgaard (2009): Etablering af dansk som kil-
desprog. I: *LexicoNordica* 16, 255–270.

- Lorentzen, Henrik & Sanni Nimb (2011): *Fra krydderkage til running sushi – hvordan nye ord kommer ind i Den Danske Ordbog*. I: Margrethe Heidemann Andersen & Jørgen Nørby Jensen (red.): *Nye ord – København 12.-13. april 2011, Sprognævnets konferenceserie 1*. København: Dansk Sprognævn, 69–85.
- Nimb, Sanni (2011): Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger. I: *LexicoNordica 18*, 135–154.
- Nimb, Sanni & Bolette Sandford Pedersen (2012): Towards a richer wordnet representation of properties – exploiting semantic and thematic information from thesauri. I: Nicoletta Calzolari (red.): *LREC 2012 Proceedings*. Istanbul: European Language Resources Association, 3452–3456.
- Nimb, Sanni (2013): Leksikalsk-semantisk information i en ny dansk begrebsordbog. I: Dorthe Duncker, Anne Mette Hansen & Karen Skovgaard-Petersen (red.): *Betydning & Forståelse. Festskrift til Hanne Ruus*. København: Selskab for Nordisk Filologi, 251–266.
- Nimb, Sanni, Bolette Sandford Pedersen, Anna Braasch, Nicolai Hartvig Sørensen & Thomas Troelsgård (2013): Enriching a wordnet from a thesaurus. I: Lars Borin, Ruth Vatvedt Fjeld, Markus Forsberg, Sanni Nimb, Pierre Nugues & Bolette Sandford Pedersen (red.): *Proceedings of the workshop on lexical semantic resources for NLP. NEALT Proceedings Series 19*. Linköping Electronic Conference Proceedings; Volume 85 (ISSN 1650-3740).
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen & Henrik Lorentzen (2009): DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation, Volume 43, Number 3*

/ *September*, 269–299.

- Sørensen, Nicolai Hartvig & Lars Trap-Jensen (2010): Den Danske Ordbog som begrebsordbog. I: Harry Lönnroth & Kristina Nikula (red.): *Nordiska studier i lexikografi 10. Rapport från Konferensen om lexikografi i Norden, Tammerfors 3–5 juni 2009*. Tammerfors: Nordiska föreningen för lexikografi, 164–179.
- Trap-Jensen, Lars (2011): Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica 18*, 183–197.
- Trap-Jensen, Lars (2014): Leksikografisk tradition og fornyelse: tre revolutioner på 100 år? I: Ruth E. Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.–16. august 2013* (dette bind).

Henrik Lorentzen
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
hl@dsl.dk

Sanni Nimb
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
sn@dsl.dk

LORENTZEN, NIMB & TROELSGÅRD

Thomas Troelsgård
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
tt@dsl.dk