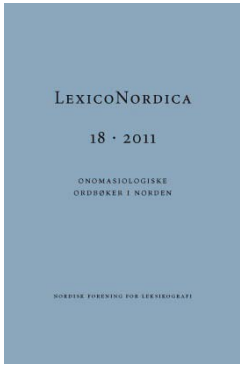


# LexicoNordica

Titel:	Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger	
Forfatter:	Sanni Nimb	
Kilde:	LexicoNordica 18, 2011, s. 135-156	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive">http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive</a>	

© LexicoNordica og forfatterne

## 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.

# Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger

*Sanni Nimb*

This article discusses the synergies between thesauri, wordnets and semasiological dictionaries, describing how the lexical data from a wordnet (DanNet), as well as the data from a monolingual dictionary (The Danish Dictionary), are being used to establish a Danish thesaurus. In opposition to other thesauri, this one's main organizing principle is the grouping of words by semantic criteria. It thus combines words from different parts of speech in different types of groups. Furthermore, a number of formalized data on the words' semantic relations leave us with a large number of possibilities when it comes to using thesaurus information in the online presentation of The Danish Dictionary. It also means that the thesaurus can be used to extend the number of semantic relations in DanNet.

## 1. Indledning

En ny dansk begrebsordbog er under udarbejdelse ved Det Danske Sprog- og Litteraturselskab (DSL, 2010-13). Projektet er finansieret af Carlsbergfondet og skal munde ud i en trykt ordbog, men ordbogens data vil også blive tilgængelige på nettet. I selve den ontologiske inddeling af betydningsområder i begrebsordbogen tages der udgangspunkt i 906 kapiteloverskrifter oversat fra *Der deutsche Wortschatz nach Sachgruppen* (Dornseiff 2004). Trap-Jensen (2011) diskuterer begrebsordbøger generelt samt de overvejelser DSL har gjort sig mht. det metodiske grundlag for en ny dansk begrebsordbog. Den diskuterer også behovet for begrebsordbogen i forhold til online-udgaven af Den Danske Ordbog

(DDO). Ordbogen benytter i dag DanNet-data til visninger af beslægtede ord, hvilket ikke i alle tilfælde fungerer optimalt.

Indeværende artikel beskriver mere detaljeret hvordan begrebsordbogen i første omgang udnytter data fra det danske wordnet DanNet og fra DDO og efterfølgende beriger disse data med yderligere formaliseret semantisk information der kan genanvendes i de selvsamme ordbøger. Begrebsordbogens mikrostruktur bliver beskrevet i detaljer, bl.a. hvad angår de semantiske metadata der tildeles de forskellige ordgrupperinger.

## 2. DDO, DanNet og begrebsordbogen – og synergien imellem dem

Begrebsordbogen bygger som nævnt direkte på data i DDO og DanNet, som vi derfor beskriver indledningsvis. DSL udarbejdede i perioden 1992 til 2005 DDO og arbejder fortsat på at udvide ordforråd og betydningsantal i ordbogen, der i dag indeholder ca. 91.500 lemmaer og 116.000 betydningsbeskrivelser. Ordbogens betydningsinddelinger er grundstammen i det danske wordnet DanNet, som DSL udvikler sammen med Københavns Universitet (Pedersen 2010; Lorentzen & Nimb 2010; Pedersen et al. 2009).

DanNet indeholder pr. 1. juli 2011 65.000 af DDO's betydninger, struktureret i over-underbegrebs-hierarkier via relationen 'has\_hyperonym' samt forbundet med ca. 75.000 andre indbyrdes semantiske relationer, bl.a. via nedarvning – og endnu flere hvis relationerne beregnes som transitive, jf. hjemmesiden *andreord.dk* der visualiserer wordnettet (Johannsen & Pedersen 2011) – her angives tallet for relationer til ca. 321.000. Synonyme betydninger er i DanNet knyttet sammen i det som er et wordnets grundenhed, et synset (fx {ben, bentøj, pusselanke, skank, stang, underekstremitet}); {} bruges i denne artikel til at markere begreber, modsat udtryk eller lemmaer). Hvert synset er tildelt en ontologisk type

(fx for {kvinde} [Human], for {kasse} [Container] og for {ben, bentøj, pusselanke...} [Bodypart]).

DanNet udmærker sig i forhold til de fleste andre wordnets ved at tage udgangspunkt i betydningsinddelinger for sproget selv – i dette tilfælde dansk – frem for at anvende oversatte betydninger og dermed også automatisk arvede relationer fra det engelske Princeton WordNet (Fellbaum 1998), en metode som fx det finske wordnet har anvendt (Lindén & Carlson 2010). Desuden er DanNet suppleret med nogle i vores øjne meget vigtige semantiske relationer for menneskeskabte ting der ikke anvendes i andre wordnets, inspireret af SIMPLE-projektet (Lenci et al. 2000) og Pustejovskys qualia-struktur (Pustejovsky 1995). Nær-synonymi er også beskrevet i DanNet, både mellem ord fra samme ordklasse ('near\_synonym') og på tværs af ordklasser. Derudover er der en lang række relationer der beskriver hvem eller hvad der er involveret i en handling (fx for {hamre}: involved\_agent {person}), ligesom der er en række relationer der bruges til at beskrive delhelhed. Også en mere bred relation 'concerns' er tilknyttet en lang række begreber.

DanNet indeholder endnu ikke tilbundsående beskrivelser af samtlige begreber, og ligesom vi ønsker at udnytte begrebsordbogens data i DDO (jf. Trap-Jensen 2011), ser vi også en stor mulighed i på sigt at kunne udnytte dem til en udvidelse af DanNet. Genbrug af data på tværs af DSL's ordbøger er mulig fordi id-numre fra hver af DDO's betydninger (fx *ben*, bet. 1, ID 21006106) går igen i DanNet, idet de er knyttet til hvert af de udtryk der indgår i et synset (ID 21006106 er tilknyttet udtrykket *ben* og indgår i synset 2109 sammen med 6 andre lemmaers id-numre (lemmaerne {ben, bentøj, skank, stang, underekstremitet}). *Bens* andre betydninger (knogle; stoleben m.fl.) indgår tilsvarende i andre synsets. I opbygningen af begrebsordbogen anvendes de samme id-numre. *Ben* indgår her i kapitel 2.24 *Legemsdele*. De fælles id-numre sikrer at den berigelse med semantisk information der nedfældes i be-

grebsordbogen, vil kunne udnyttes til at supplere både DanNet og DDO med oplysninger. Men inden vi vender tilbage til dette, vil vi se på hvordan DanNet og DDO i første omgang udnyttes i selve arbejdet med at opbygge begrebsordbogen.

### 3. Data fra DanNet og DDO til begrebsordbogen

Den grundlæggende over-underbegrebs-struktur i et wordnet som DanNet har et klart fællestræk med begrebsordbøgers traditionelle måde at gruppere begreber på. Fx indeholder både Harry Andersen (1945) og Dornseiff (2004) mange ordgrupper hvis fællestræk udgøres af overbegrebet (se figur 1, der viser borde og skuffemøbler; andre eksempler er tøj, skydevåben, værktøj og beholdere).

#### **Bord – skuffemøbel**

**Substantiver:** anretterbord, boghylde, bogreol, bogskab, bord, buffet, chatol, chaffioniere, dagligstuebord, disk, dragkiste, etagere, hjørneskab, hylde, kateder, kiste, klapbord, klædeskab, kommode, konsol, pengeskab, pult, reol, rygebord, sekretær, servante, skab, skrivebord, skrivepult, skuffemøbel, skænk, spisestuebord, toilet møbel.

Figur 1: Borde og skuffemøbler i Harry Andersen (1945). Overbegreber er fællestræk for ordene i gruppen.

Det er derfor meget oplagt at udtrække grupper af underbegreber fra DanNet til brug for en dansk begrebsordbog. Et udtrækningsprogram udviklet til formålet<sup>1</sup> gør redaktøren i stand til at vælge et udtryks underbegreber i DanNet og indsætte dem direkte i begrebsordbogen. Figur 2 illustrerer hvordan man ud fra DanNet, og i øvrigt i dette tilfælde uden særlig meget efterredigering, kan lave et lignende, moderne afsnit *bord – skuffemøbel*. Det skal nævnes at figurerne i denne artikel er taget fra redigeringsværktøjet og derfor

<sup>1</sup> Udarbejdet af Nicolai H. Sørensen, DSL.

ikke illustrerer hvordan den færdige begrebsordbog kommer til at se ud. Fx kan metadata ses på nogle af figureerne.

☉ **bord** ▶ træbord, fyrretræsbord, mahognibord, teaktræsbord, plankebord-; ▶ kakkelbord, glasbord, marmorbord-; ▶ klappbord, udtræksbord-; sofabord, spisebord, langbord, spisestuebord, køkkenbord; ▶ aftensbord, morgenbord, middagsbord, frokostbord, kaffebord, tebord, fødselsdagsbord, julebord, taffel-; ▶ campingbord, klappbord-; ▶ gavebord, fødselsdagsbord-; ▶ bakkbord, indskudsbord, anretterbord-; sidebord, hjørnebord, rygebord; ▶ rullebord, trådvogn-; telefonbord, sybord  
 ☉ **arbejdsbord**, bænk; ▶ filebænk, hølebænk-; slagtebænk, vaskebænk; ▶ omstillingsbord, telefonbord-; operationsbord, stinkskab, linnedskab, medicinskab; ▶ det grønne bord, eksamensbord-; kateder; ▶ *syn*: skranke, retsskranke-; vidneskranke, talerstol  
 ☉ **opbevaringsmøbel** ▶ **kommode**, skuffemøbel, skuffedarium-; chiffonniere, rokokokommode; ▶ **skænk**, købmændsdiske, dækketøjsskab, sælvetøjsskab, glasskab, vitrine, buffet-; ▶ **chatol**, sekretær-; ▶ **kiste**, skibskiste, dragkiste, kistebænk, brudekiste-; ▶ kommodeskuffe, skuffe-;  
 ☉ **skab**, jalousiskab, glasskab, vægskab; ▶ *syn*: garderobeskab, klædeskab-; garderobe, karlekammerskab, hjørneskab, hjørneskab; ▶ svaleskab, isskab, varmeskab-; ▶ kosteskab, værktøjsskab-; skabelement, skabsvæg  
 ☉ **reol**, bogreol; ▶ byggereol, stigereol, ståltreol-; ▶ rack, chassis, kabinet, tv-tårn-; hjemmebar; ▶ barskab, bar, hjemmebar-; vinreol; ▶ hylde, boghylde-; bogkasse; ▶ amagerhylde, kaminhylde, konsol, etagere-; ▶ glashylde, hattehylde-;

Figur 2: Borde og skuffemøbler (opbevaringsmøbler) i den nye danske begrebsordbog – opstået som udtræk fra DanNets møbel-hierarki og endnu ikke forsynet med metadata (se også *møbel*, [www.andreord.dk](http://www.andreord.dk)).

Typisk forestår der dog et efterfølgende finredigeringsarbejde. DanNet dækker kun 65.000 af DDO's 116.000 betydninger, og derudover kan den samme betydning være relevant at nævne i flere kapitler i begrebsordbogen; derfor bliver underbegrebshierarkierne suppleret med yderligere data fra DDO. Til dette arbejde har vi oprettet et særligt xml-dokument<sup>2</sup> der indeholder data fra DDO vedrørende ordenes semantik: definitioner, konnotative betydninger og encyklopædiske oplysninger, genus proximum, fag- og domæneoplysninger samt oplysninger om ordforbindelser, synony-

<sup>2</sup> Udarbejdet af Thomas Troelsgård, DSL.

mer, antonymer og nærsynonymer. Xml-dokumentet er desuden for hver betydning suppleret med digitaliserede oplysninger om synonymer taget fra en dansk synonymordbog (Schultz 1997) til yderligere inspiration når der skal findes udtryk til et kapitel i begrebsordbogen. Via disse supplerende søgninger i DDO opdager redaktøren ofte endnu flere udtryk eller kommer selv i tanke om ord som endnu ikke er medtaget i DDO. Også en tidligere dansk begrebsordbog, Harry Andersen (1945), bliver konsulteret i arbejdet med at finde udtryk. Udtryk der ikke står i DDO i forvejen, medtages hvis de er udbredte på internettet og markeres samtidig som lemmakandidater til DDO. I arbejdet med foreløbig fem kapitler i begrebsordbogen er der tilføjet 437 lemmakandidater samt i øvrigt også knap 400 ordforbindelser der ikke er registreret i DDO. Begrebsordbogen kan derfor 'føde tilbage' til DDO med typiske ordforbindelser og ord af nyere dato i forhold til DDO's korpus (fx *simremad*, *øhop* og *running sushi* – første belæg i Infomedia henholdsvis 1997, 1998 og 2009) samt med huller i DDO's ordforråd. Dette gælder fx ord som *krydderkage*, *kanelbrød* og *dåseskjul*, der kan dateres tilbage til DDO's periode eller før.

### 3.1. Ikke-anvendelige hierarkier fra DanNet

Begrebsordbogens kapitloverskrifter lægger op til en beskrivelse af domænespecifikke situationer og deres tilhørende handlinger, egenskaber, aktører og rekvisitter – og dette går i mange tilfælde på tværs af de sprogligt set korrekte overbegreber man finder i DanNet. Derfor skal ko-hyponymer ofte fordeles på forskellige kapitler i begrebsordbogen. Fx har {maske} i DanNet følgende underbegreber: {ansigtsmaske}, {muddermaske}, {iltmaske}, der i begrebsordbogen placeres under 2.23 *Sundhed; hygiejne*; {fægtemaske}, der placeres under 17.19 *Kampsport* i begrebsordbogen; {dykermaske}, der placeres under 17.20 *Vandsport*, og {fastelavnsmaske}, der placeres under 17.27 *Selskabelighed*. Et andet eksempel

er {pind}, der i DanNet har underbegreberne {kniplepind}, {kødpind}, {pøsepind}, {røgelsespind} og {trommestik}. {Kødpind} og {pøsepind} hører i begrebsordbogen under *16.10 Madlavning*, {trommestik} under *14.15 Musikinstrumenter* og {kniplepind} under *19.08 Stoffer, tekstiler*. Især når der er tale om meget flade, ordrige hierarkier i DanNet, er dette et problem. Nogle eksempler er konkrete overbegreber som {person} og abstrakte begreber som {følelse} og {tidsperiode}, der alle har et meget stort antal direkte underbegreber. Ex er {fritid}, {fyringssæson} og {halvleg} alle direkte underbegreber til {tidsperiode}. Udtrækkene kan i stedet fungere som oversigtslister, idet især udtryk fra de abstrakte betydningsområder heller ikke er lette at udtrække vha. søgninger i DDO. Lister over bestemte semantiske områder kan i øvrigt også i visse tilfælde fremfindes i DanNet ud fra den ontologiske type; fx har alle tidsbetydninger typen [TIME].

Alt i alt kan man konkludere at DanNet er særdeles velegnet til hurtigt at indsætte forholdsvis store mængder data i begrebsordbogen og dermed skabe en grundstruktur i et kapitelafsnit. Særligt de tilfælde hvor der er relativt mange, men ikke uoverskueligt mange underbegreber, og hvor overbegrebet selv hører til under det pågældende afsnit i begrebsordbogen, er velegnet til genbrug. I det videre redigeringsarbejde suppleres disse grundlæggende data med udtryk fra DDO.

Når et overbegreb er meget generelt i forhold til den domæneorienterede begrebsordbog, forestår et større arbejde med at sortere underbegreberne på forskellige kapitler. Denne sortering i domæner kan senere anvendes til at forbedre DDO's visning af beslægtede ord. Oplysning om domæne-tilhørsforhold kan ligeledes indsættes direkte i DanNet som en form for ekstra 'concerns'-relation og supplere de nuværende relationer. Og selve suppleringen med flere udtryk fra DDO vil mere eller mindre automatisk kunne anvendes til at udvide dækningsgraden i DanNet, ligesom den vil sikre at beslægtede ord-funktionen i DDO tilbydes ved



endnu flere opslagsord. Der kræves dog en række forholdsvis nemt indsatte metadata i begrebsordbogen for at genbruget kan foregå automatisk; dette beskriver vi i næste afsnit.

#### 4. Semantisk struktur i begrebsordbogen

I modsætning til andre begrebsordbøger der præsenterer ordene efter ordklasse (jf. Trap-Jensen 2011), arbejder vi i den nye danske begrebsordbog med semantisk definerede grupperinger på første opdelingsniveau inden for et kapitel. Også andre semantiske grupperinger end den ovenfor beskrevne (underbegreber til samme overbegreb) er mulige. Semantisk gruppering i skrivefasen letter arbejdet med at skrive ordbogen da redaktøren i arbejdsprocessen 'tænker' i betydninger og leder efter udtryk ud fra semantiske kriterier – ikke ud fra ordklasser. Samtidig giver den semantiske gruppering os mulighed for at specificere en lang række underliggende metadata i form af semantiske relationer der gælder mellem ordene i gruppen. Disse metadata er som nævnt nødvendige hvis ordbogens oplysninger skal genbruges i DDO og DanNet. Andre begrebsordbøger (fx Dornseiff 2004, Harry Andersen 1945, Rogge's Thesaurus (2002) og Norsk Tesaurus 2001) indeholder ofte en implicit semantisk gruppering på 2. niveau, fx ved at nævne et overbegreb fulgt af dets underbegreber inden et nyt overbegreb nævnes, eller ved at præsentere en række adjektiver der beskriver samme type kerne, inden en ny række adjektiver nævnes. I den nye danske begrebsordbog ekspliciteres denne viden i form af metadata der tilknyttes de enkelte grupperinger.

I modsætning til de omtalte begrebsordbøger indeholder den desuden også grupper med ord fra flere ordklasser. Det har den klare fordel at ord der ellers ville mangle kontekst fordi det semantiske område er ordfattigt, kan hægtes på en gruppe ord fra en anden ordklasse med beslægtet betydning. De fire omtalte be-

grebsordbøger har i vores øjne også det problem at de har for vage markeringer af de centrale begreber. I den nye danske begrebsordbog ønsker vi en tydelig markering af nøgleord (fx med fed skrifttype) på et forholdsvis detaljeret niveau. Fx markerer vi ofte et overbegreb inden underbegreberne nævnes. Ligeledes fremhæves det første synonym ud af en række synonymer typisk.

Vi etablerer så store semantiske grupperinger at man opnår den ideelle gensidige informationseffekt fra naboordene. Men da vi samtidig gerne vil undgå at grupperingerne skal blive for store og uoverskuelige, har vi udarbejdet forskellige ret fleksible regler for tilladte semantiske grupper. Man kan som redaktør vælge hvilken gruppering man vil benytte, afhængigt af antallet af ord i det givne domæne. Ved meget ordrige semantiske områder benytter redaktøren typisk en gruppering udelukkende med ord fra samme ordklasse (ofte en over-underbegrebs-gruppe), mens man ved mindre ordrige områder grupperer på tværs af ordklasser for at opnå en tilpas stor og informativ gruppe af ord. På denne måde tilgodeser vi både ordrige og mere ordfattige semantiske områder af sproget.

De semantiske grupperinger der er tilladt, er opstillet på baggrund af de intuitive grupperinger der blev foretaget i to gennemarbejdede kapitler (ca. 10 % af begrebsordbogen). Alle grupperinger er gjort fleksible idet de indeholder mulighed for en 'hale', hvor nært beslægtede udtryk der ikke overholder gruppens overordnede systematik, kan nævnes. Og alle grupperinger indeholder oplysninger om hvilken semantisk relation der gælder mellem ordene i gruppen, fx præcise oplysninger om hvad der er det fælles overbegreb eller hvem/hvad der har den egenskab der beskrives i gruppen. Vi forsyner desuden de semantiske grupperinger med forklarende overskrifter til brugeren.

Metadata-inventaret der anvendes, består af nogle af de semantiske relationer der allerede bruges i DanNet, men der er også indført nye relationer, inspireret af EuroWordNets og SIMPLE's forslag til semantiske relationer (Vossen 1998, Vossen 1999; Len-

ci et al. 2000). Ud over at metadataene, som tidligere nævnt, vil kunne bruges til at indsætte supplerende relationer i DanNet, vil de også muliggøre særlige præsentationer af ord både i ordnet.dk/ddo og på tværs af kapitlerne i begrebsordbogen.

På trods af den semantiske tilgang ønsker vi at have mulighed for at præsentere ordene på samme måde som andre begrebsordbøger, dvs. ordklassesorteret. Den ordklassesorterede visning har traditionen bag sig (jf. Trap-Jensen 2011), men den semantiske visning har også fordele, idet det jo netop er de omgivende ords semantik i en begrebsordbog der gør at man overhovedet får information om det enkelte ord. Jo færre semantisk beslægtede ord i en gruppe, jo mindre information om de enkelte ord, kunne man sige. Ordklassemarkering på hvert udtryk sikrer mulighed for ren ordklassesortering hvis det foretrækkes til den trykte ordbog eller til en supplerende digital visning i den elektroniske udgave af ordbogen. Vi vender tilbage til de forskellige typer af grupperinger i næste afsnit idet vi først vil se på den semantiske mikrostruktur.

#### 4.1. Semantisk mikrostruktur

En semantisk gruppering består af flere 'bundter' af særligt tæt beslægtede ord. Selve grupperingen kan indledes af et nøgleord som i figur 2 (fx *bord* og *arbejdsbord*), som samtidig udgør overbegrebet for alle ord i gruppen (dette præciseres i metadata). Selve bundterne består typisk af nær-synonymer (hvis andet ikke er markeret for gruppen) eller af synonymer (markeret med *syn* i redigeringsværktøjets visning af data, se fx figur 3, de sidste fire udtryk i halen).

◊Om person {04\_Egenskaber/has\_hyperonym: Ilgevægtig property\_of: person}
   
 ▶ **rolig og afbalanceret**, afslappet ◄; ▶ i balance med sig selv, i balance, harmonisk,
 afbalanceret, velafbalanceret, hvilende i sig selv ◄; let til mode; ▶ besindig, nede på
 jorden, stenrolig, tilbagelænet, adstadig, uimponeret ◄; ▶ **ligevægtig**, robust, stoisk,
 ◊ sindsligevægtig ◄; ▶ **svær at bringe ud af fatning**, svær at ryste, ubevægelig,
 uforstyrrelig ◄; ▶ **usårlig**, selvberende, støt ◄; ▶ **behersket**, besindig, rolig, rolig og
 fattet, diplomatisk, stoisk, koldsindig, stabil, tilregnelig ◄; ▶ **medgørlig**, skikkelig,
 godmodig, tålmodig, overbærende, fredsommelig, sindig, langmodig ◄; skædesløs;
 ▶ **uanfægtet**, upåvirket, uberørt, uberørt af situationen, upåvirket af situationen,
 bleg, men fattet ◄; ▶ **på højde med situationen**, cool, all right, fit for fight, fuld af
 gåpåmod, vel til mode, uforfærdet, ufortrøden, uforsagt, kæk, frejdig, ubekymret, tryk
 ◄; ▶ **overfrisk**, friskfyragtig, letsindig, lalleglad ◄; → → → ▶ syn: **med sindsro**, uden at
 blinke, uden at fortrække en mine, gladelig ◄;

Figur 3: Et udsnit af kapitel 10.08 *Ufølsomhed*; *sindsro*'s 230 udtryk med både enkeltord og ordforbindelser.

Men også to antonymer kan indgå i en undergruppe, og endelig kan den bestå af et overbegreb med dets underbegreber. Det centrale og indledende udtryk i hvert enkelt bundt af ord fremhæves som nøgleord når der er tale om et større betydningsskift i forhold til de ord der står umiddelbart før (se fx *misdannelse* og *plet* i figur 4).

◊Grim genstand {01\_Overbegreb/has\_hyperonym:
   
 rædsel}
   
 ▶ **rædsel**, monstrum, stavær, skrummel ◄; skændsel;
   
 ▶ karikatur af/på nogen/noget, vrængbillede af/på
   
 noget/nogen ◄; ▶ øjebæ, betonkasse ◄; misfarving,
   
 hadegave, klatmaleri
   
 ◊Grim legemsdel {01\_Overbegreb/has\_hyperonym:
   
 legemsdel}
   
 ▶ **misdannelse**, abnormitet, deformitet, monstrøsitet
   
 ◄; ▶ pukkelryg, pukkel, udvækst, gevækst, vansiret
   
 ansigt, skæmmende ar, hæsligt/grimt fjæs, tud ◄;
   
 ▶ **plet**, udslæt, rødt udslæt, vorte, byld, bums, filipens
   
 , akne ◄;
   
 ◊Grim person; grim skabning {01\_Overbegreb
   
 /has\_hyperonym: grimrlan}
   
 ▶ **grimrlan**, abemenneske, skabilken, skabilkenhoved
   
 ◄; ▶ kvabso, smatso, sæk, tudse ◄; langt spektakel:
   
 ▶ pak, rakkerpak ◄; ▶ **vanskabning**, vantrivning,
   
 misfoster, krøbling, pukkelryg ◄; ▶ **bæst**, troid, heks,
   
 uhyre, uhyre, udyr, monster ◄;

Figur 4: Grupper bestående af underbegreber hvor nøgleord er markeret.

Ordene efter nøgleordet optræder i logisk rækkefølge. Nøgleordet er, som i Roget's Thesaurus, typisk det mest frekvente eller almin-

delige udtryk (vurderet alene ud fra introspektion) (fx <*sejlsport, søsport, sejlads*> og <*ro, padle, sejle i kajak, ro i kajak*>. Hvis undergruppen består af et overbegreb og dets underbegreber, vil det være overbegrebet der er markeret, fx <*gangart, pasgang, trav, galop, galopade*>. Nogleord kan række hen over flere undergrupper hvis der ikke er stort betydningsskift. Dette gælder fx undergruppen <*pukkelryg, pukkel, udvækst, gevækst, vansiret ansigt, skæmmende ar, hæsligt/grimt fjæs, tud*> i figur 4. Det indledende og markerede ord i synonym-gruppen umiddelbart før, *misdannelse*, fungerer som en overskrift for alle disse udtryk.

Der bliver ikke foretaget nærmere sproglige undersøgelser for at udfylde markeringer af undergrupperne; de beror udelukkende på subjektive skøn eller på oplysninger om synonymi og ordfelter der allerede er anført i DDO.

#### 4.2. Semantiske metadata på grupperne

En overordnet semantisk gruppes overbegreb angives som metadata i starten af gruppen (se figur 4) i form af et typenummer (*01\_Overbegreb*), et udtryk (*rædsel*) med tilhørende id-nummer og et attribut der angiver relationstypen ('has\_hyperonym'). Id-nummeret sikrer som tidligere nævnt at oplysningen kan genbruges i DanNet og DDO. Det kan også være holonymet (den helhed som noget er del af) der indsættes som metadata, eller det kan være oplysning om hvem egenskaber i gruppen gælder (fx *property\_of {person}* i figur 3). Ved grupper uden systematisk relation mellem ordene anvendes den lidt løsere relation 'concerns'. Fx har gruppen <*fitness, gå til fitness, fitnessinstruktør, fitnesscenter*> kun oplysningen 'concerns {fitness}' som metadata. Også andre relationer kan indsættes, fx hvem der udfører handlingerne i gruppen ('involved\_agent').

De fleste grupper indeholder et afsluttende felt, en hale, hvor de relationer der angives som metadata, ikke længere gælder, og

hvor semantisk beslægtede udtryk alligevel kan nævnes. Se fx figur 3, hvor adverbialerne <med sindsro, uden at blinke, uden at fortrække en mine, gladelig> nævnes til sidst.

### 4.3. Grupperinger på tværs af ordklasser

Grupperinger der består af flere ordklasser (også når halen ikke medregnes), bruges i to tilfælde. For det første når der inden for et bestemt betydningsområde er tale om forholdsvis få ord fra mindst en af ordklasserne – så få at de, hvis de står alene, udgør en meget lille og derfor nemt lidt intetsigende gruppe. Ved at hægte de få ord på en større ordgruppe af over- og underbegreber får man både fortalt noget om den semantiske relation mellem ordene fra de to ordklasser (idet hver type gruppering bærer et typenummer), og man får etableret en mere informationsrig gruppering af en vis størrelse. Fx er en lille gruppe adjektiver ofte placeret efter den gruppe af substantiver som de typisk er adled til. Eks. fra kapitel 16.06 *Mad og madretter*:

*frugt, eksotiske frugter, sydfrugter, kernefrugt, stenfrugt, modne frugter, runde frugter, nedfaldsfrugt, nedfalden frugt, nedfaldsæble – moden, solmoden, plukkemoden, frisk, nedfalden*

Andre tilfælde hvor det er tilladt at hægte forholdsvis små ordgrupper på en over-underbegrebs-gruppering, er:

- Verber og de udråb der bruges i den givne situation (*væmmes / bvadr*).
- Substantiver og de verber der udføres af substantivet (*dommer / dømme; cyklist / cykle*).
- Substantiver og de verbaludtryk der har dem som objekt (*pibe / stoppe; håndbold / gå til håndbold*).
- Substantiver der angiver sted, og de handlinger der foregår det pågældende sted (*mål / komme i mål*).

#### TEMATISKE BIDRAG

For det andet laves der grupperinger på tværs af ordklasser når der er tale om manifestationer af det samme begreb. Nogle af disse grupper er rene afledningsgrupper, hvoraf især én er blevet hyppigt anvendt i det hidtidige arbejde. Den udgøres af verber og deres modsvarende verbalsubstantiver. Også her angives der metadata for gruppens ord, idet der gives information om hvad overbegrebet er til verberne henholdsvis verbalsubstantiverne. Desuden gives der information om hvem der udfører handlingen, udløser begivenheden etc. En sådan gruppe kan fx se ud på følgende måde:

*reagere, gå fra koncepterne, gå fra snøvsen, flippe ud, knække totalt sammen, bryde sammen, bryde grædende/hulkende sammen, gøre noget overilet, forhaste sig, forivre sig, forglemme sig (selv), fare op, bruse op etc. reaktion, respons, anfald, angstreaktion, nervesammenbrud, den umiddelbare reaktion, negative reaktioner etc.*

I dette tilfælde indeholder den bagvedliggende struktur metadata om at den involverede agent er {person}, og at overbegreberne er {reagere} (for verberne) og {reaktion} for substantiverne.

En anden gruppering på tværs af ordklasser hvor der er tale om nærtbeslægtede begreber, er en gruppering af verber (samt eventuelt også de tilhørende verbalsubstantiver) og de adjektiver der betegner en efterfølgende tilstand forårsaget af den behandling som verberne udtrykker. Ofte vil adjektivet være en participiumsform af verbet og være beskrevet under verballemmaet i DDO, men der kan også være tale om ikke-afledte adjektiver. Et eksempel fra kapitel 16.07 *Krydderi* er:

*krydre, smage til, tilsmage, runde af, tilsætte, iblande, blande i, komme i, pebre, salte, krydring, tilsmagning, tilsætning, iblanding, krydret, pebret, pikant, hot, stærk*

Til sidst vil vi nævne en gruppering på tværs af ordklasser der har vist sig meget anvendelig i det foreløbige arbejde med begrebsordbogen – en semantisk gruppering der samler egenskaber på tværs af ordklasser (fx *begeistret, begejstring, i højt humør, ikke være til at styre på en hel tønde land*). Egenskaber udgør en forholdsvis stor del af ordforrådet. De beskriver ofte enten en person (som i eksemplet ovenfor), en genstand eller en hændelse, og vi ønsker at inddele egenskaberne i tydelige grupper ud fra dette. Desuden ønsker vi at præcisere hvad egenskaberne lægger sig til, både i form af en eventuel overskrift til brugeren og i en formaliseret form som metadata (sammen med de sædvanlige overbegreber). Figur 3 ovenfor er et typisk eksempel. Ofte kræver inddelingen information om ordene som ikke er ekspliciteret i DDO, når man som redaktør skal tage stilling til hvem eller hvad egenskaben gælder. Fx skelnes der i et andet kapitel 10.16 *Beundring, skønhed* mellem udtryk der beskriver smukke personer, og udtryk der beskriver smukke genstande (se figur 5):

Om person (04\_Egenskaber/has\_hyperonym: smuk  
property\_of: person):  
 >smuk, skøn, flot<; >fin, gracios, køn, vakker,  
 henrivende, yndig, fager<; >syn: underskøn, billedskøn,  
 gudeskøn, ualmindelig(t) smuk, ufattelig(t) smuk,  
 usædvanlig(t) smuk<; >smuk og dejlig, unig og smuk,  
 smuk og tiltrækkende, dårende dejlig<; velplejet; >syn:  
 velbygget, velskabt, veldrejet<; kvielejet; >syn: lækker,  
 lækker<; >syn: smart, fiks, chik, elegant<; >hyp:  
 skønhed, smukt udseende, kvindelig skønhed, ydre  
 skønhed<; >se godt ud, virke tiltrækkende<; >lyse op,  
 stråle<; blomstre<--->skønhedsideal, man må lide for  
 skønheden



Om *genstand* {04\_Egenskaber/has\_hyperonym: flot  
property\_of: genstand}  
 ▶ **flot**, skøn, smuk◄; ▶vækker, yndig, fager◄; ▶syn:  
**ualmindelig(t) smuk**, underskøn, ufattelig(t) smuk,  
 usædvanlig(t) smuk◄; ▶syn: abeskøn, skæppeskøn,  
 dødsbøn, superskøn◄; smuk og harmonisk; ▶vidunderlig,  
 paradisisk, overjordisk, underbar, underfuld◄;  
 ▶**overdådig**, prægtig, prud◄; ▶syn: **smart**, fiks, chik,  
 elegant, raffineret◄; ▶syn: lækker, fjong, laber◄; ▶super,  
 superb◄; ▶**glansfuld**, glitrende, strålende◄;  
 ▶**farvestrålende**, farveprægtig◄; flatterende, æstetisk,  
 ren, rigt udsmykket, smukt udsmykket; ▶**skønhed**, pragt,  
 pyd, pænhed◄; ▶ynde, gratie◄; æstetiske kvaliteter,  
 ydre skønhed; ▶**glans**, stjernestøv, glamour,  
 overdådighed◄; naturskønhed; ▶**se pæn ud**, pryde,  
 pynte, passe perfekt/fint◄; ▶**stråle**, lyse op, funkle,  
 funkle, tindre◄; ▶funkle, opalisere◄; ▶**klæde nogen  
 godt**, klæde nogen◄; ▶**æstetik**, æstetik, æstetisk  
 sans, skønhedssans; ▶**æstet**, skønånd, æstetiker◄;  
 ingen roser uden torne

Figur 5: Kapitel 10.16 i begrebsordbogen. Ord der udtrykker skønhed ved personer henholdsvis genstande.

## 5. Begrebsordbogens omfang

Ordbogen vil komme til at indeholde et forholdsvis højt antal udtryk, væsentligt flere end Norsk Tesaurus (der kun indeholder ca. 7.000 udtryk), men også flere end både Harry Andersen (en skønsnæssig optælling af den alfabetiske ordliste giver ca. 33.500 forskellige udtryk) og Dornseiff (ca. 56.000 udtryk, også skønsnæssigt optalt). Den vil snarere nærme sig Rogets størrelse (over 300.000 udtryk ifølge oplysninger på omslaget).

Vi kan også konstatere at den nye danske begrebsordbog bliver ca. tre gange så omfangsrig som Harry Andersen, alene ved at se på antallet af udtryk i de færdiggjorte kapitler. Fx indeholder kapitel 10.08 *Ufølsomhed; sindsro* (figur 3) 230 udtryk i sammenligning med Harry Andersen, der har 75 udtryk i det tilsvarende kapitel. Dette er et meget gennemsnitligt eksempel på forskellen og svarer også meget godt til at DDO som leverer udtryk til begrebsordbogen, indeholder ca. 102.000 lemmaer og faste udtryk.

Begrebsordbogen kommer desuden til at indeholde et stort antal ordforbindelser, enten taget fra DDO's i alt ca. 32.500 oplysninger om typiske adled, kerne og andre naboord eller tilføjet som helt nye data via introspektion og frekvenstjek på internettet. Sådanne typiske ordforbindelser vil give god vejledning i sproglig formulering til brugere der ønsker at bruge begrebsordbogen til skrivehjælp, men de vil også kunne finde anvendelse inden for det sprogteknologiske fagområde, hvor de traditionelt betegnes *bi-grams*, *trigrams*, *n-grams*, idet de i begrebsordbogen er semantisk entydiggjort. I figur 4 ses et udsnit af kapitel 10.08 *Ufølsomhed*; *sindsro*'s 230 udtryk, hvor man ud over enkeltord og faste forbindelser med egne betydningsdefinitioner i DDO kan se adskillige typiske ordforbindelser (*i balance med sig selv*, *hvilende i sig selv*, *let til mode*, *svær at ryste* etc.). Begrebsordbogen angiver i øvrigt også typiske bøjningsformer, fx *afbalanceret*, der ikke er selvstændigt opslagsord i DDO, men beskrives som participium under lemmet *afbalancere*.

## 6. Konklusion og videre perspektiver

Redaktionsreglerne for de semantiske grupperinger er blevet til i et forsøg på at skabe klare linjer i de mange måder man kan gruppere ord på inden for et betydningsområde. Reglerne sikrer at man griber sorteringen af ord an på en nogenlunde ensartet måde fra kapitel til kapitel samt redaktørerne imellem. Det at man er tvunget til at tænke over hvilken type gruppe man er ved at etablere, samt hvilke metadata der knytter sig til den, sikrer også at grupperne bliver stringente. Samtidig gør halerne og 'nødløsningen' i form af en tilladt gruppe uden fast struktur det muligt også at beskrive udtryk der er svære at indplacere i en systematisk gruppe.

Grupperne og de forskellige muligheder for at angive metadata om relationerne mellem ordene er lavet ud fra et ønske om

på et senere tidspunkt at kunne anvende de meget strukturerede data i andre sammenhænge. Begrebsordbogens struktur bygger på de semantiske relationer der allerede bruges i DanNet, og det er derfor muligt at udnytte oplysningerne i den til at udvide antallet af eksisterende relationer i DanNet. Man vil fx kunne anvende de mange oplysninger om overbegreber til at supplere, evt. korrigere hierarkier i wordnettet. Særligt tilføjes der mange funktionelle over-underbegrebs-grupper i begrebsordbogen i forhold til hvad man finder i DanNet (fx grupper af fodboldudstyr, haveting osv.). Og fra mange af grupperne der indeholder ord fra mere end én ordklasse, vil man automatisk kunne supplere DanNet med bestemte semantiske relationer der gælder for de enkelte ord i gruppen ved at anvende de formaliserede oplysninger om typenummer og relationer i metadata.

Der er også indført relationer i begrebsordbogen som ikke anvendes i DanNet, og som det er meget oplagt at udvide DanNet med. Den oplysning der tegner til at kunne bidrage med mest ny information i DanNet, er relationen 'property\_of'. Egenskaber er ikke behandlet særligt detaljeret i DanNet, og det vil være en klar forbedring af wordnettet inden for dette betydningsområde hvis en relation der beskriver hvad egenskaben lægger sig til, kan indsættes automatisk ud fra begrebsordbogens metadata. Fx vil alle ord i figur 3 automatisk kunne tilskrives relationen 'property\_of {person}'. Der eksperimenteres i øjeblikket med at integrere disse data automatisk fra begrebsordbogen til DanNet.

Hvad angår brug af data fra begrebsordbogen i DDO, har et af formålene med begrebsordbogen været at opnå leksikalske data der kunne forbedre beslægtede ord-funktionen på ordnet.dk/ddo, jf. Trap-Jensen (2011). Dette vil absolut være muligt pga. begrebsordbogens detaljeringsgrad: Man vil på det snævrreste semantiske niveau kunne vise kun de øvrige udtryk i samme undergruppe. Man vil også kunne vælge at præsentere hele grupperingen med alle dens undergrupper og den efterfølgende hale, og man vil end-

da kunne vise hele kapitlet som ordbetydningen optræder i. For nogle ords vedkommende kan der være tale om flere tilhørsforhold og dermed flere visninger.

De forskellige metadata vil også muliggøre særlige visninger af DDO's data på tværs af kapitlerne i begrebsordbogen. Endvidere får vi en meget interessant mulighed for at forklare et ords betydning alene ved at præsentere dets 'omgivelser' i begrebsordbogen, en visning der vil svare til en synonymdefinition. Ca. 18.000 lemmaer i netudgaven af DDO har nemlig endnu ikke definitioner da de i den trykte første udgave af DDO blot var nævnt som gennemskuelige sammensætninger eller afledninger under andre lemmaer. I den digitale udgave er de ophøjet til selvstændige lemmaer med morfologiske oplysninger. Mange adjektiver mangler fx definitioner; i det øjeblik disse adjektiver er medtaget i begrebsordbogen, er vi i stand til at give en synonymdefinition af dem, en definitionstype der i forvejen er brugt for adjektiver i DDO. Fx står adjektivet *uforsagt* der ikke er defineret i DDO endnu, i følgende undergruppe i begrebsordbogen <*på højde med situationen, cool, all right, fit for fight, fuld af gåpåmod, vel til mode, uforfærdet, ufortrøden, uforsagt, kæk, frejdig*>. Visningen af disse omgivende ord vil give en udmærket forklaring på betydningen af ordet.

Som det kan ses, åbner arbejdet med begrebsordbogen op for mange muligheder inden for det digitale leksikografiske arbejde der udføres på DSL og præsenteres via ordnet.dk.

## Litteratur

### Ordbøger

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

- DDO = Ebba Hjorth, Kjeld Kristensen m.fl. 2003-2005: *Den Danske Ordbog*. København: Gyldendal og Det Danske Sprog- og Litteraturselskab. Onlineversion: <http://ordnet.dk/ddo>.
- Dornseiff, Franz 2004: *Der deutsche Wortschatz nach Sachgruppen*, 8. Auflage, Berlin/New York: Walter de Gruyter.
- Roget, Peter Mark: *Roget's Thesaurus*, 150<sup>th</sup> anniversary edition edited by George Davidson 2002. London: Penguin.
- Schultz: Ulla Albeck, Mikal Rode og Erik Timmermann 1997: *Dansk Synonymordbog*. København: J.H. Schultz Information A/S.
- Norsk Tesaurus: Johan Hammond Rosbach 2001: *Ord og begreper Norsk tesaurus*. Oslo: Pax Forlag A/S.

#### Anden litteratur

- Fellbaum, Christiane 1998 (ed.): *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Johannsen, Anders & Bolette S. Pedersen 2011: "Andre ord" – a wordnet browser for the Danish wordnet, DanNet. I: *Proceedings from 18th Nordic Conference of Computational Linguistics, NODALIDA 2011*, Riga, Latvia. Northern Association for Language Technology, University of Tartu, Vol. 11, 295-298.
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowsky, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, Antonio Zampolli: 2000. SIMPLE: A General Framework for the Development of Multilingual Lexicons. I: *International Journal of Lexicography*, 13(4), 249-263.
- Lindén, Krister & Lauri Carlson 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119-140.
- Lorentzen, Henrik & Sanni Nimb 2010: Fra ordbog til wordnet. Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse? I: *Nordiska studier i lexikografi* 10, 329-344.

- Pedersen, Bolette S. 2010: Semantiske sproressourcer – mellem sprogteknologi og leksikografi. I: *LexicoNordica 17*, 163-180.
- Pedersen, Bolette S., S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen 2009: The challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation, Computational Linguistics Series*, Volume 43, Number 3: 269-299.
- Pustejovsky, James 1995: *The Generative Lexicon*. Cambridge: MIT Press.
- Trap-Jensen, Lars 2011: Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica 18* (i dette bind).
- Vossen, P. 1998 (ed.): *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht.

## Internethenvisninger

- andreord.dk ved Anders Johannsen. Viser data fra DanNet.  
DanNets hjemmeside: wordnet.dk
- Vossen, P. 1999 (ed.): *EuroWordNet General Document, Version 3 Final* <http://www.illc.uva.nl/EuroWordNet/finalresults-ewn.html>

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

