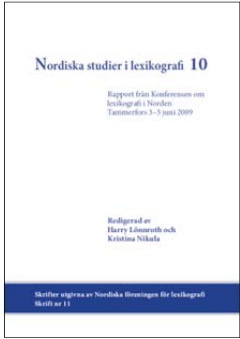


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Fra ordbog til wordnet - Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse?	
Forfatter:	Henrik Lorentzen & Sanni Nimb	
Kilde:	Nordiska Studier i Lexikografi 10, 2010, s. 329-344 Rapport från Konferens om lexicografi i Norden, Tammerfors 3.-5. juni 2009	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

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 lexicografi (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.

HENRIK LORENTZEN & SANNI NIMB

Fra ordbog til wordnet Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse?

In this article we discuss the problems involved in transforming traditional dictionary definitions into formalised descriptions meant for computational use. The starting point is the making of the Danish wordnet, DanNet, which is based on a large monolingual dictionary. In the process of extracting semantic information from the entries it turns out that much information is already present in the dictionary text, but a lot of useful information is only implicitly there or presupposed. The article falls into three parts: firstly, the principles of lexicographic definitions are discussed, in particular the ones used in The Danish Dictionary (*Den Danske Ordbog*). Secondly, it is demonstrated how definitions are “translated” into semantic relations in the wordnet, and examples of how missing information is added are given. Finally, the idea of making general dictionaries and wordnets in one operation is put forward and discussed.

Nøgleord: wordnet, definitioner, semantiske relationer, præsupposition

1. Indledning

I 2009 blev den første version af et dansk wordnet, DanNet, offentliggjort. DanNet, der er blevet til i et samarbejde mellem Center for Sprogteknologi ved Københavns Universitet og Det Danske Sprog- og Litteraturselskab, er opbygget på grundlag af en større monolingval dansk ordbog, *Den Danske Ordbog* (DDO). Brugen af DDO til dette formål har tydeliggjort nogle interessante problemstillinger, naturligvis fordi der er forskelle i beskrivelsestilgangene – onomasiologisk i wordnettet, semasiologisk i ordbogen – men også fordi almindelige ordbøger som DDO er udformet til mennesker med en forhåndsviden om verden og om de begreber der beskrives i definitionerne. Ladislav Zgusta (1988: 252) skriver at ”the dictionary relies on the user’s knowledge of the world and on his presuppositions”, og også Bo Svensén (2004: 264, 298) slår fast at man i ordbogsdefinitioner bygger på læserens præsuppositioner og viden om verden. Med et wordnet der er tænkt til computerbrug, er det anderledes: For at computeren kan ”forstå” en betydning, skal al viden konkretiseres ved hjælp af links til andre betydninger i det semantiske net. Hidtidige forsøg på at udnytte ordbøger til automatisk at opbygge formelle semantiske leksika har af samme grund mødt

mange vanskeligheder (jf. fx Ide & Véronis 1995, Kokkinakis et al. 2000, Ide & Wilks 2007). I DanNet har man da også, efter nogle indledende forsøg med semiautomatisk udtræk af relationer (Pedersen et al. 2009: 287 ff.), valgt at arbejde manuelt med at beskrive begreberne ud fra DDO's definitioner. I denne artikel vil vi diskutere hvilke oplysningstyper vi har medtaget i DanNet på trods af at de ikke er ekspliciteret i DDO, og vi vil vise eksempler på hvor tit ordbogsdefinitioner rent faktisk læner sig op ad den menneskelige brugers forhåndsviden og evne til at se betydningsrelationer mellem ord selvom disse ikke ekspliciteres.

Efter en kort gennemgang af selve projektet med at opbygge wordnettet ud fra ordbogen vil vi først diskutere principperne for intensionelle definitioner i almenordbøger i bred forstand samt mere specifikt de principper der ligger bag DDO's definitioner. Dernæst vil vi vise hvordan definitionerne er oversat til semantiske relationer i DanNet og give eksempler på tilføjelse af information der mangler i DDO. Endelig vil vi komme ind på hvordan man måske kan – og bør – samtænke udarbejdelsen af almenordbøger og wordnets, idet de to tilgange i høj grad supplerer hinanden og sikrer en mere fuldstændig betydningsbeskrivelse.

2. Den Danske Ordbog som udgangspunkt for det danske wordnet

Sjette og sidste bind af DDO udkom i 2005. Ordbogen, der er korpusbaseret, indeholder i den trykte version cirka 100 000 betydningsdefinitioner fordelt på 63 000 lemmaer beskrevet i en xml-struktur hvor genus proximum er opmærket i et særligt element, genprox. Sammen med det faktum at DDO's redigeringsprincip så vidt muligt var at udforme betydninger ved hjælp af såkaldte intensionelle definitioner (genus proximum + differentia specifica), gjorde de opmærkede genus proximum-oplysninger det oplagt at forsøge at udnytte ordbogen til at opbygge et dansk wordnet i lighed med det engelske Princeton WordNet (Fellbaum 1998). I et wordnet er hvert enkelt begreb knyttet til sit overbegreb. Endvidere er begreberne knyttet til hinanden med forskellige andre relationer der for eksempel udtrykker del-helhed (kop/hank) og antonymi (sur/sød). På trods af at DDO er semasiologisk i sin tilgang – hvert enkelt lemmas betydninger er først og fremmest beskrevet i forhold til hinanden, og betydninger inden for samme semantiske område er forholdsvis sjældent udformet efter en ensartet model – forelå der via genus proximum-opmærkningen data der principielt kunne indgå direkte som overbegreber i et sådant hyponymi-hierarki. Disse data skulle blot entydiggøres for så vidt angik tilfælde af homografi og

polysemi. Resten af definitionen, den del der udgør *differentia specifica*, skulle så omsættes til de øvrige semantiske relationer i så høj grad det var muligt.

Version 1 af det danske wordnet DanNet indeholder cirka 125 000 relationer mellem 41 000 begreber, organiseret i såkaldte ”synsets” som består af sæt af synonymmer (eller blot et enkelt lemma). Heraf er cirka 34 000 synsets fra substantiver (fx [*gulvtæppe; tæppe*]; [*antenne*]; [*gave; present*] og [*dame, damemenneske, kvinde, fruentsimmer, kvindemenneske*]). Derudover er der 6 000 synsets fra verber og cirka 1 000 fra adjektiver, og antallet udbygges løbende frem til 2011 inden for projektet DK-CLARIN. Opbygningen af et wordnet ud fra en monolingval sprogbeskrivelse, som det blev valgt i det danske tilfælde, er ikke den mest udbredte metode. I wordnet-projekter for andre sprog tager man ofte direkte udgangspunkt i Princeton WordNet, idet de engelske begreber oversættes til begreber i målsproget, hvorefter hele netstrukturen, det vil sige alle de semantiske relationer der er specificeret for de engelske begreber, direkte kan overtages. Fordelen ved i stedet at anvende en korpusbaseret monolingval ordbog som DDO var dels at man var sikker på at de betydninger der kom med i DanNet, rent faktisk var udbredte i dansk, dels at hele begrebsstrukturen var intuitivt korrekt for en dansk sprogbruger og ikke overtaget fra et fremmedsprog.

Lad os slå fast med det samme at anvendelsen af DDO til opbygningen af DanNet var meget udbytterig, idet man med en kombination af semiautomatiske metoder (udtræk af begreber med samme genus proximum) og manuelt arbejde på forholdsvis kort tid etablerede et hyponymi-hierarki med som nævnt 41 000 begreber og derudover mere end 80 000 andre dels manuelt indsatte, dels nedarvede relationer. Men naturligvis var der store udfordringer. På nogle punkter var vanskelighederne forudsigelige. For eksempel er nogle ordklasser, for eksempel adverbier og interjektioner, vanskelige at indpasse i wordnettet, og denne del af ordforrådet er da heller ikke medtaget i DanNet i første omgang. Men også betydninger beskrevet med et i princippet velegnet genus proximum i DDO voldte problemer. Definitionsmetoden medfører nemlig uundgåeligt cirkulære definitioner, et fænomen der anerkendes som et generelt problem når det gælder leksikografiske definitioner (Zgusta 1988: 248). I den logisk opbyggede træstruktur som et wordnets hyponymi-relationer udgør, med de mindst specifikke begreber øverst og flere og flere specifikke træk tilføjet for hvert niveau af underbegreber, er cirkularitet ikke mulig. At ’egetræ’ er en slags ’træ’ og ’træ’ en slags ’plante’, udelukker for eksempel at ’træ’ samtidig kan være en slags ’egetræ’ eller ’plante’ en slags ’træ’, og en mekanisme i DanNet-værktøjet advare derfor om rekursion i hierarkiet. Ved cirkularitet blandt de meget centrale, overordnede begreber er man derfor tvunget til at udvælge ét der udgør toppen af hyponymi-hierarkiet (eller hvis man skønner at de er synonymmer, at lade dem

indgå i samme synset i toppen). For eksempel har vi i DanNet ladet 'område' være det øverste begreb for alle arealer, såvel store som ganske små, og dermed benyttet det som overbegreb for 'sted', selvom *område* i DDO netop defineres som 'sted som er (geografisk afgrænset) i forhold til noget andet', og sted modsat defineres som 'område med bestemt beliggenhed'. I tilfælde af cirkulære definitioner eller synonymdefinitioner på lavere niveauer i begrebshierarkiet, typisk anvendt i DDO ved mindre frekvente lemmaer, er situationen lettere at håndtere i wordnettet, idet de to ord udpeger samme begreb og derfor kan indgå i samme synset med fælles overbegreb. I DDO henvises man for eksempel fra *lastvogn* til *lastbil*, hvor der gives en udfoldet genus proximum-definition, samtidig med at lastvogn gives som synonym. Lemmaerne bliver i wordnet-systemet beskrevet som to udtryk for samme begreb, det vil sige de bliver medlemmer af samme synonymysæt med overbegrebet 'bil'.

Et måske mere overraskende problem var at der skulle ændres på en hel del af de overbegreber der kunne udtrækkes automatisk via DDO's genus proximum-udtryk. I mange tilfælde er der brugt et mere generelt og almindeligt udtryk i ordbogen end det systematisk set nærmeste overbegreb. Omvendt er cirka 10 procent af de anvendte genus proximum-udtryk ved substantiverne i DDO slet ikke lemmaer i selve ordbogen og derfor umulige at udnytte direkte i DanNet. Det drejer sig om forholdsvis sjældne ad hoc-komposita som *guitarlyd*, *bakteriebelægning*, *rabarbersort* og *kulthandling* (se også Pedersen et al. 2009: 292). For den menneskelige læser af en ordbog er det naturligvis ikke noget problem at tolke sådanne sammensætninger, men da de ikke selv er defineret i DDO, må de enten oprettes som ekstra begreber i DanNet eller erstattes af et eksisterende begreb i ordbogen.

I DDO har man prioriteret brugervenlige definitioner i form af forholdsvis ukomplicerede sætninger med ikke for svære ord. Zgusta (1988: 252) påpeger en risiko for at en høj grad af brugervenlighed kun kan opnås på bekostning af præcision og eksplicitte definitioner. COBUILD's meget brugervenlige definitioner kan for eksempel ifølge ham kun fungere fordi leksikografen endnu mere end ved andre ordbøger bygger på ordbogsbrugerens forhåndsviden og på præsuppositioner. Ordbøger til computeranvendelse derimod, skriver han, skal nødvendigvis være både eksplicitte og præcise og vil have svært ved at håndtere den pragmatiske information der bygges på i ordbøger til mennesker. De to typer af leksikalsk beskrivelsesmetode kan derfor nemt bevæge sig i hver sin retning (Zgusta 1988).

At mange af DDO's definitioner også i høj grad bygger på den menneskelige læsers evne til tolkning, blev mere og mere tydeligt under arbejdet med at omsætte definitionernes *differentia specifica* til semantiske relationer i DanNet. Meget information om de enkelte betydninger som vi mente skulle medtages i

et wordnet, viste sig at være udeladt i DDO. I afsnit 3 vil vi fremlægge de overvejelser der ligger bag DDO's betydningsdefinitioner, og diskutere hvorvidt ønsket om læservenlighed har ført til mindre eksplicitering end man for eksempel finder i mindre brugervenlige ordbøger.

3. Definitioner i almenordbøger

I det følgende vil vi koncentrere os om genus proximum-definitionerne, som er den mest anvendte type i DDO og andre almenordbøger (Svensén 2004: 273), og som umiddelbart er oplagte at udnytte i opbygningen af et wordnet. De kaldes også *intensionelle* definitioner, det vil sige definitioner der retter sig mod begrebets indholdsside og angiver dets egenskaber, i modsætning til *ekstensionelle* definitioner der opregner et begrebs underbegreber (jf. NLO). De intensionelle definitioner tager udgangspunkt i en semantisk analyse af karakteristiske træk ved det begreb der skal defineres, og sammenligner det med beslægtede begreber. Resultatet bliver en formulering med genus proximum (nærmeste overbegreb) og et eller flere distinktive træk (*differentiae specificaе*), for eksempel *bil*: 'motordrevet køretøj med fire hjul', hvor "køretøj" er genus proximum, og "motordrevet" og "med fire hjul" er distinktive træk. Trækket "motordrevet" kan således adskille *bil* fra for eksempel *cykel*, og "med fire hjul" kan adskille *bil* fra *motorcykel*. Når en sådan definition overføres fra ordbogen til wordnettet, bruges de enkelte træk til relationer som HAS_HYPERONYM ("har som overbegreb") og HAS_MERO_PART ("har som del").

På trods af deres tilsyneladende enkle opbygning og velegnethed som udgangspunkt for et wordnet, frembyder genus proximum-definitionerne rent faktisk en række problemer. Først vil vi se på hvilke krav der kan stilles til udformningen af sådanne definitioner i ordbøger for mennesker, idet i hvert fald tre dimensioner kommer på tale:

1. Detaljeringsniveau: tilstrækkelig over for udtømmende
2. Typikalitet: typisk over for marginal
3. Komplexitet: enkel over for kompleks

3.1 Detaljeringsniveau

I tidligere tid var idealet for definitioner i større modersmålsordbøger at de skulle være udtømmende og tage højde for alle tænkelige forhold. Målet var

fuldstændighed, og konsekvensen af dette blev ofte nogle meget komplekse betydningforklaringer. Det er artiklen sofa i ODS et glimrende eksempel på:

Sofa .. et af den østerlandske divan (3) udviklet bænk lignende, i alm. polstret, med fjedrende sæde forsynet møbel (til at ligge ell. sidde paa), i sin europæiske form tidligst med hoved- og endestykke, men ingen ryg (jf. Ottoman samt Løjbenk), nu (som allr. i 18. aarh.) i alm. med rygstød (jf. Kanapé), anv. (tidligere spec. i fornemme ell. borgerlige hjem) som vægmøbel (sjældnere fritstaaende) med siddeplads til to (ell. flere) personer (ogs. om tilsvarende magelige siddepladser i jærnbanevogn, biografteater olgn.)

Her får vi en lang række semantiske oplysninger om genstanden kombineret med mere og mindre synonyme ord, med forbehold og præciseringer og suppleret med historiske informationer; der er virkelig tale om det Svensén (2004: 275) kalder ”encyklopedisk slagsida”. Det er værd at bemærke at traditionen lever videre i SAOB, som har en lige så omstændelig beskrivelse af *soffa* i bind 28 fra 1981, men her er det selvfølgelig afgørende at SAOB føler sig forpligtet til at følge de én gang vedtagne redaktionsprincipper for at sikre værkets homogenitet. Visse af ODS’ redaktører er herostratisk berømte for at skrive grundige og derfor vanskeligt tilgængelige definitioner (jf. Andersson 2006: 36 f.), og da DDO skulle redigeres, stod det klart at idealet nu var et andet. Her defineres *sofa* således:

sofa .. polstret, flerpersoners siddemøbel som har ryg- og armlæn, og som typisk står i en stue, fx sammen med et sofabord og en el. flere lænestole

Der angives færre træk end i ODS, men til gengæld er de tilstrækkelige til at udpege genstanden sofa ude i verden uden at alle forhold tages i betragtning. For eksempel nævnes det ikke at man også kan ligge på en sofa, eller at sofaen ofte står op ad en væg.

3.2 Typikalitet

I forlængelse af spørgsmålet om fuldstændighed ligger spørgsmålet om det typiske over for det marginale. Det er en klar tendens hos de seneste årtiers britiske leksikografer at det gælder om at fremhæve det typiske frem for det marginale: ”what is normally the case rather than what is necessarily the case” (Hanks

1987: 118), eller som det udtrykkes i følgende gode råd fra B. T. Sue Atkins og Michael Rundell (2008: 430): "Don't attempt to account in your definition for every conceivable instantiation of the LU [: Lexical Unit]." Disse betragtninger har været afgørende for udformningen af definitionerne i DDO og har medført en praksis der i mange tilfælde udmøntes i en relativt kort definition suppleret med en tilføjelse der lægger vægt på det prototypiske. Disse tilføjelser indledes ofte med ordene *typisk* eller *især* og svarer dermed til Patrick Hanks' *normally*. Her ses et par eksempler på denne fremgangsmåde:

- alkoholvaner** .. et menneskes el. en social gruppes vaner mht. indtagelse af alkohol, især forbrugets omfang
- fernisering** .. officiel åbning af en kunstudstilling, typisk med indbudte gæster og servering af mad og drikke

På denne måde lykkes det at give brugeren en forestilling om det prototypiske begreb uden at udelukke andre muligheder.

3.3 Komplexitet

En tredje dimension er den konkrete formulering af betydningsbeskrivelserne som i takt med den stigende bevidsthed om brugervenlighed er gået fra det komplekse til det enkle. Atkins og Rundell (2008: 451) udtrykker det således: "A definition should contain no more words than necessary, consistent with the demands of intelligibility and information-transfer", og Svensén (2004: 300) er på linje hermed når han skriver: "I en allmänordbok gäller det därför att anpassa definitionerna efter användarnas kompetens och behov och beskriva begreppet på ett sätt som är begripligt och användbart för lekmannen". Disse principper er også søgt gennemført i DDO og eksplicit udtrykt i projektets interne redaktionsregler, hvor der står: "Vi skal forsøge at forfatte den denotative betydningsbeskrivelse så enkelt og ligefremt som muligt." Dette betyder at der er lagt en klar distance til ODS' (og tilsvarende værkers) ofte komplicerede syntaks idet ønsket om enkelhed medfører et fravalg af det marginale og et fokus på det centrale. Denne prioritering har som præmis at det er unødvendigt at insistere på det indlysende, men det har store konsekvenser for brugen af ordbogens data til en sprogresurse som DanNet. Mennesker ved noget som maskiner ikke ved; det er en banal konstatering, men den indebærer en manglende eksplicitering af mange oplysninger.

4. Betydningsbeskrivelse i DanNet

I DanNet er kravet ikke enkelhed men derimod en mere udtømmende beskrivelse hvor så meget prototypisk viden om et begreb som muligt omsættes til relationer til andre begreber, idet intet naturligtvis kan forudsættes bekendt af computersystemet i den tiltænkte anvendelse af wordnettet til informationssøgning, automatisk oversættelse, tekstresumering og så videre.

Betydningsbeskrivelsen for hvert enkelt begreb tager udgangspunkt i et antal relationstyper der som udgangspunkt defineres for hver af de cirka 200 ontologiske typer i DanNet-ontologien (tilpasset fra EuroWordNet; jf. Vossen 1998). Nogle eksempler er [Artifact+Object] der skal dække menneskeskabte genstande, modsat naturskabte ting der tildeles typen [Natural+Object]. Spiselige dele af dyr, for eksempel kød, tildeles på lignende vis den ontologiske type [Animal+Comestible+Part], og alt tøj tildeles for eksempel typen [Garment+Artifact+Object]. Relationerne kan vælges blandt 18 forskellige der beskriver aspekter som forskellige former for del-helhed (meronymi, fx stof/bikini; bikini/bh; folketingsmedlem/folketing og oase/ørken), begrebets funktion (skrue/fæstne), begrebets tilblivelse (bikini/sy), antonymi (sur/sød), nærsynonymi (bog/hæfte; handling/handle), personer der er involveret i begrebets anvendelse (musikinstrument/musiker), genstande der er involveret i en anvendelse (spille/musikinstrument), og en persons aktive rolle (stewardesse/betjene) eller passive rolle (interviewoffer/interviewe).

Så snart et begreb er tildelt en ontologisk type, bliver det via typens tilhørende relationer forsynet med en slags specifikation for hvordan det bør beskrives, og det er netop denne fremgangsmåde der fører til opdagelsen af ”huller” i DDO’s definitioner, det vil sige påviser på hvilke områder de enkelte definitioner ikke har været semantisk udtømmende.

For alle begreber der tildeles en ontologisk type indeholdende betegnelsen Artifact (’menneskeskabt ting’, fx [Artifact+Object] og [Garment+Artifact+Object]), er DanNet-leksikografen for eksempel tvunget til at tage stilling til hvordan genstanden er lavet, hvad formålet med genstanden er, og hvilke personer der er involveret i anvendelsen af genstanden, idet disse relationer automatisk følger med valget af Artifact-typerne. Samtidig indebærer nedrivning af de udfyldte relationer fra overbegrebet også at der skal tages stilling til om relationen skal udfyldes mere specifikt for netop dette underbegreb. Det kræver selvfølgelig at de øverste begreber i hierarkiet er omhyggeligt forsynet med relationer. For eksempel har man på øverste niveau beskrevet ’beklædningsgenstand’, der er af typen [Garment+Artifact+Object], med relationerne USED_FOR ’klæde’ (’bruges til at klæde’), USED_FOR ’dække’ og USED_FOR_OBJECT ’krop’ (’bruges til at dække kroppen’), MADE_BY ’sy’ (’lavet ved at sy’), HAS_

MERO_MADEOF 'stof' ("lavet af stof") og HAS_HOLO_MEMBER 'beklædning' ("er en del af beklædning") samt endelig INVOLVED_AGENT 'menneske' (den mest generelle udfyldning af denne rolle). Alle underbegreber til 'beklædningsgenstand' arver samtlige relationsværdier, og der kan så eventuelt foretages ændringer eller blokeringer af nedarvningen ved hvert enkelt underbegreb. På den måde er leksikografen for eksempel automatisk blevet bedt om at tage stilling til om et 'menneske' bruger en bikini, og har indsnævret begrebet til 'kvinde', også selv om det ikke er anført noget sted i DDO's artikel.

At kvinder ikke er nævnt i artiklen *bikini*, skyldes højst sandsynligt en ubevidst antagelse om at ordbogsbrugeren ved dette i forvejen. Men inden vi tager fat på denne problemstilling, vil vi lige nævne at den manglende uddybning af leksikalsk viden i DDO efter vores mening ofte skyldes de syntaktiske begrænsninger på definitionsudformningen. Der er ganske enkelt ikke plads inden for en enkelt sætning til at beskrive alle del-helhed-relationer for komplekse genstande. For eksempel er *bog* i DDO defineret som: 'trykte el. beskrevne blade af papir indbundet el. på anden måde sammenhæftet i rækkefølge så de danner en helhed, ofte en sammenhængende tekst, beregnet på at blive læst'. Der nævnes intet om at bøger har en titel, en ryg, for- og bagside, eller at de ofte handler om et bestemt emne. I DanNet har vi valgt så vidt muligt at tilføje relationer mellem helheder og dele når begge begreber optræder i wordnettet, så vi for eksempel skaber en forbindelse i form af en semantisk relation mellem 'bog' og 'side' og 'bog' og 'emne'.

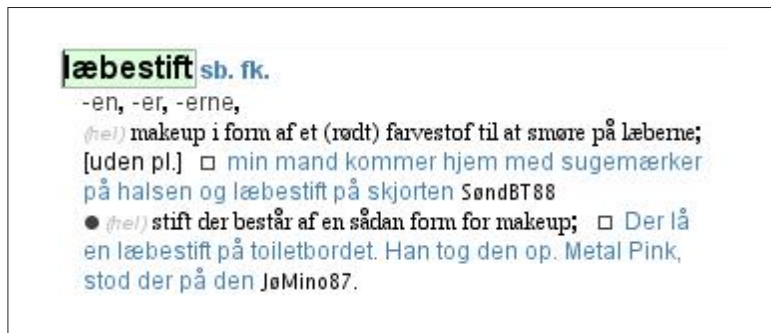
Der mangler derimod kun sjældent information om hvad en genstand bruges til i DDO. Man finder dog enkelte eksempler på det; det er for eksempel påfaldende at der i definitionen af *snekanon* ('maskine der kan fremstille sne og kaste det på det ønskede sted') intet står om at den typisk har relation til skiløb. Men gennemgående er det næsten altid beskrevet hvad ting bruges til i ordbogen, ligesom man ofte finder oplysninger om hvordan de er lavet. Indførelsen af relationerne MADE_BY og USED_FOR i DanNet, som reelt er en udvidelse i forhold til hvad andre wordnets (fx Princeton WordNet) indeholder af relationer, har vist sig at være en god beslutning. Den skyldes erfaringer fra det danske SIMPLE-projekt (Pedersen & Paggio 2004), hvor de udgjorde de grundlæggende relationer i to af de fire betydningsdimensioner som beskrivelsesmodellen var baseret på (inspireret af Pustejovsky 1995). At denne type information typisk er med i DDO, tyder på at den er et vigtigt element i ordforklaringer der ønsker at fokusere på det formidlingsmæssige aspekt.

Derimod er involverede personer i forbindelse med forskellige genstandes brug (eller frembringelse) som regel helt udeladt i DDO som vist ved eksemplet *bikini*. Vi mener at dette fortrinsvis skyldes at DDO, som andre almenordbøger, bygger på præsuppositioner og forventet forhåndsviden om hvordan verden er

indrettet. Nogle gange angives det i DDO at en genstand anvendes af mennesker generelt, for eksempel i form af et *man* i definitionen (*redskab*: 'genstand [som er] nyttig el. uundværlig når man skal udføre en bestemt handling el. et bestemt stykke arbejde'). Men i langt de fleste tilfælde er det naturligt nok underforstået at en genstand anvendes af mennesker eller er fremstillet af mennesker. For eksempel står der ikke i definitionen af *bog* noget om at den er skrevet af en forfatter og bliver læst af en læser; ej heller om *værktøj* at det typisk anvendes af en håndværker. Ved *budcykel* står der intet om at den person der bruger sådan en cykel, typisk er et bud – i citatet nævnes ordet *bud* heller ikke, men ordbogsbrugerens udleder nemt at jeg er buddet: *Jeg afventede min fars ordre til at samle varerne sammen og springe på budcyklen*. Ved *kundevogn* og *indkøbsvogn* står der intet om *kunde*; det gør der heller ikke ved *forretning* og *butik*. Faktisk er det kun ved 11 lemmaer (af mindst 150) med overbegræbet *butik* eller *forretning* at ordet *kunde* overhovedet nævnes i artiklen, enten i selve ordbogsdefinitionen, som kollokation eller i et citat. Ved *raket* står der intet om at personer om bord på en raket typisk er astronauter ('styrbart fartøj med raketmotor og med plads til mennesker og udstyr til rejser i rummet'); *rumfartøj*'s definition nævner heller ikke astronauter, og der er ingen forbindelse mellem lemmaet *astronaut* og de forskellige ord for rumfartøjer i DDO andet end i et enkelt citat til *rumskib*. Og i definitionen for *indlæggelsesseddel* ('dokument der er udstedt af en læge, og som foreskriver indlæggelse og behandling') nævnes der intet om at en patient er involveret, ej heller i citatet hvor den menneskelige læser dog nok vil forstå at Marie er patienten: *Doktoren gav Marie en indlæggelsesseddel til et sanatorium*. I alle disse tilfælde ekspliciteres informationen i DanNet.

Vi finder også tilfælde hvor den involverede bruger er nævnt på anden vis i ordbogsindgangen på grund af morfologisk samhørighed, og hvor en menneskelig læser nemt kan etablere forbindelsen mellem de to ord. Et eksempel er *apoteker*, der nævnes som orddannelse til *apotek* i DDO; i DanNet specificeres den semantiske relation mellem de to begreber, idet 'apotek' får relationen INVOLVED_AGENT 'apoteker'. Når lemmaet er et kompositum, indgår ordet for den involverede person i mange tilfælde som førsteled. *Budcykel* og *kundevogn* er allerede nævnt, men listen er lang over komposita dannet på denne måde: for eksempel *malergrej*, *bagerbutik*, *beduintelt*, *dykkerur*, *dametøj*, *matrostøj* og *hyrdestav*. Også i disse tilfælde mener vi at det er oplagt at præcisere den semantiske relation mellem begreberne i DanNet-resursen selvom DDO ikke omtaler det i betydningsdefinitionen. Der er en tæt, og ofte leksikalsk, sammenhæng mellem en genstand og den person der (lidt bredt forstået) er involveret i brugen eller frembringelsen af den, hvilket netop også kan ses ved de mange tilfælde af morfologisk slægtskab.

Hvad angår en markering af om det er mænd eller kvinder der anvender bestemte genstande, har man tydeligvis ikke ønsket at formidle alt for fordomsfulde oplysninger i DDO. For eksempel står der ikke at en barberkost, barbersprit og seler typisk anvendes af mænd, og heller ikke at makeup, læbestift, en underkjole og som tidligere nævnt en bikini typisk anvendes af kvinder. Og ved *læbestift* (se figur 1) kan kun ordbogsbrugerens viden om verden forhindre at han/hun ikke udleder af citaterne at det er mænd der er typiske brugere af læbestift.



Figur 1. Artiklen læbestift i DDO

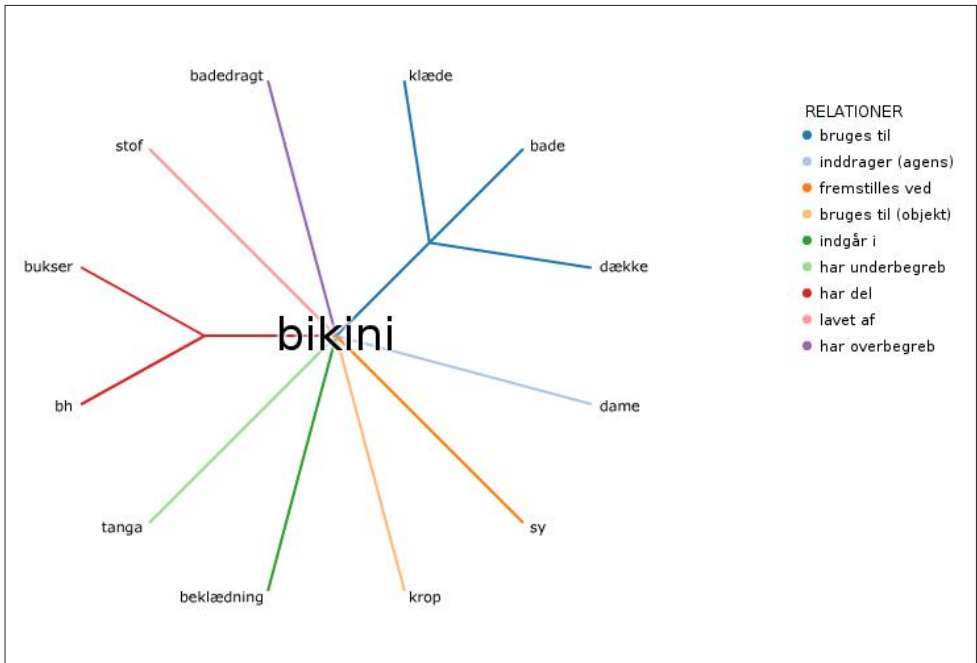
Dog er information om kønsspecifik anvendelse af og til med i DDO – og cirka dobbelt så mange gange (i 226 definitioner) anvendes ordet *kvinder* i forhold til de kun 116 gange *mænd* anvendes i beskrivelsen, hvilket man måske kan opfatte som tegn på at ”mandeanvendelse” ubevidst regnes for mest selvfølgelig.

Det skal dog siges at vi i DDO af og til finder eksempler på at den involverede person nævnes i definitionen, for eksempel når det er den eneste differentierende egenskab (*dametøj* og *herretøj*) eller i tilfælde af genstande med en meget begrænset type involverede personer: *jungletromme* (bruges af naturfolk), *catwalk* (bruges af mannequiner), *beduintelt* (bruges af arabiske nomadefolk), *dykkerur* (bruges af dykkere), *gliedermann* (bruges af kunstnere), *spejderhytte* (bruges af spejdere), *sommerfuglenet* (bruges af insektsamlere) og *træningsdragt* (bruges af sportsfolk). I alle disse tilfælde beskrives det også i DanNet. I tabel 1 kan man se yderligere eksempler på information om den involverede person der er tilføjet i DanNet.

Tabel 1. Tilføjede informationer om den involverede person

Lemma	DDO-definition	Værdi for tilføjet relation INVOLVED_AGENT i DanNet
flyvecertifikat	officielt bevis som berettiger indehaveren til at flyve et fly, en helikopter e.l.	pilot
letlæsningsbog	bog med en typografi og sproglig form der tager hensyn til modtagergruppens manglende læsefærdigheder	skoleelev
bronkoskop	(bøjeligt) rør med lys og optik der føres ned gennem lufrøret under lungeundersøgelser	læge
vielsesattest	dokumenter der bevidner at to personer er gift med hinanden	ægtepar
indkøbsvogn kunde vogn	firehjulet vogn af metal med en stor trådkurv og håndtag i hele kurvens bredde, beregnet til at lægge varer i under indkøb i fx et supermarked	kunde
ski	hvert af de to lange, flade og smalle stykker træ el. kunststof med opadbøjede spidser som spændes fast under fødderne så man kan glide på sne	skiløber

Hvis vi kort skal opsummere, finder man altså ofte i DDO oplysninger om en genstands formål og frembringelse, hvorimod man sjældent finder oplysninger om de involverede personer i brugen (eller evt. frembringelsen) af den, ligesom man sjældent finder alle dele af genstandene beskrevet i definitionerne. I DanNet tilføjes denne information, og i figur 2 vises et eksempel på visualisering af en udtømmende beskrivelse bestående af 1) nedarvning (klæde, bade, dække krop, beklædning, stof, sy), 2) information der er direkte oversat fra DDO-definitionen 'todelt badedragt bestående af bh og trusser' (bh, trusser), 3) information der er tilføjet (dame) samt endelig 4) begreber der har bikini som overbegreb (tanga) 'bikini' opnår på denne måde i alt 12 relationer til andre begreber i wordnettet.



Figur 2. Visualisering af begrebet 'bikini'. Figuren kan ses i farver på <http://andreord.dk>, hvor man også kan se illustrationer af alle DanNets andre begreber. Kun første lemma i et synset vises (fx dame i synsettet [dame, damemenneske, kvinde, fruentimmer, kvindemenneske])

5. Ordbog og wordnet i én arbejdsgang?

Set i bakspejlet må vi konstatere at de eksplicite oplysninger om genus proximum i DDO har været yderst værdifulde da de har dannet grundlag for etableringen af begrebshierarkiet i DanNet. Men samtidig vil vi til dels beklage at brugen af genprox-elementet ikke har været mere konsistent. Det skyldes at elementet var tænkt som et eksperiment uden en klar forestilling om hvad det kunne eller skulle bruges til. Opgaven for redaktørerne var at forfatte en mundret og for mennesker tilgængelig definition, hvorefter genprox-elementet forholdsvist mekanisk skulle udfyldes med ét og kun ét ord fra definitionen. Denne praksis har medført læservenlige definitioner, men bedre for wordnettet ville det være hvis man slet og ret havde en formel definition ved siden af den "menneskelige". En sådan formaliseret definition kunne i et standardiseret format eksplicite en lang række af de træk som vi har diskuteret i artiklen. Denne dobbelte tilgang ser vi som en mulighed for dem der begynder forfra på et tilsvarende ordbogs-

projekt, og den er faktisk også taget i anvendelse ved redigeringen af *Algemeen Nederlands Woordenboek* (jf. Moerdijk 2008).

Tanken er her at betydningsbeskrivelser netop skal være dobbelte og bestå dels af en definition i traditionel forstand, dels af et såkaldt semagram. Semagrammet er kort fortalt en formel semantisk beskrivelse med mange betydningstræk ekspliciteret, for drikkevarer for eksempel farve, lugt, smag, sammensætning, bestanddele, tilberedelse (Moerdijk 2008: 561). Eksemplet i tabel 2, ordet *cow* (nederlandsk *koe*), illustrerer tydeligt det høje detaljeringsniveau:

Tabel 2. Uddrag af semagram for *cow* (*koe*) (efter Moerdijk 2008: 566)

A cow	
Upper category:	is an animal # animal; mammal; ruminant
Category:	is a bovine (animal) # bovine; ruminant
Colour:	is often black and white spotted, but also brown and white spotted, black, brown or white # black and white; brown and white; red and white; spotted; black; blackspotted; white; brown; rusty brown
Size:	is big # big
Sound:	moows/lows, makes a sound that we imitate with a low, long-drawn "boo" # moo; moows; low; lows; boo
Parts:	has an udder, horns and four stomachs: paunch, reticulum, third stomach, proper stomach #
Build:	is big-boned, bony, large-limbed in build #
Function:	produces milk and (being slaughtered) meat #
Place:	is kept on a farm; is in the field and in the winter in the byre #
Age:	is adult, has calved # adult; calved
Gender:	is female # female
Property:	is useful and tame; is considered as a friendly, lazy, slow, dumb, curious, social animal #
Behaviour:	grazes and ruminates #
Treatment:	is milked every day; is slaughtered # milk; slaughter
Product:	produces milk and meat # milk; meat
Value:	is useful # useful

Teksten efter tegnet # vil ikke være synlig for brugeren (og er for overskuelighedens skyld udeladt flere steder i tabellen), men skal være søgbar som nøgleord for onomasiologiske søgninger hvor brugeren kan indtaste et eller flere karakteristiske træk ved det begreb der efterspørges (jf. Niestadt & Tiberius & Moerdijk i trykken). Denne samtænkning af traditionel menneskeordbog og formel maskinordbog virker som en meget tiltalende (men også arbejdskrævende og omkostningstung) vej at gå så det bliver muligt at foretage avancerede søgninger både fra udtryk til indhold og fra indhold til udtryk. Onlineversionen af DDO har taget et første skridt på vejen i retning af begrebssøgning (se Sørensen & Trap-Jensen i dette bind), idet funktionen ”Beslægtede ord” viser DanNet-data inden for rammerne af den traditionelle ordbogsartikel, og der gives mulighed for dels at navigere videre blandt de beslægtede ord, dels for at se begreberne i en visuel præsentation (jf. figur 2). I næste fase er tanken at give brugeren adgang til en specifik begrebssøgeside hvor data fra DanNet i endnu højere grad kan udnyttes som sorterings- og søgeparametre.

LITTERATUR

- Andersson, Henrik, 2006: ODS – træk af en historisk ordbogs historie. I: *LexicoNordica* 13. S. 25–39.
- Andre ord. Ved Anders Johanssen. <http://andreord.dk/>
- Atkins, B. T. Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- COBUILD = Collins COBUILD English Language Dictionary. London 1987.
- DanNet. <http://wordnet.dk/>
- DDO = Den Danske Ordbog 1–6. København 2003–2005. <http://ordnet.dk/ddo/>
- DK-CLARIN. <http://dkclarin.ku.dk/>
- Fellbaum, Christiane (ed), 1998: *WordNet – An Electronic Lexical Database*. Cambridge/London.
- Hanks, Patrick, 1987: *Definitions and Explanations*. I: *Looking Up. An Account of the COBUILD Project in Lexical Computing*, red. af John Sinclair. London. S. 116–136.
- Ide, Nancy & Véronis, Jean, 1995: *Knowledge Extraction from Machine-Readable Dictionaries: An Evaluation*. I: *Machine Translation and the Lexicon*. Third International EAMT Workshop, Heidelberg, April 26–28, 1993, Proceedings, red. af Petra Steffens. London. S. 19–34.
- Ide, Nancy & Wilks, Yorick, 2007: *Making sense about senses*. I: *Word sense disambiguation – Algorithms and applications*, red. af E. Agirre & P. Edmonds. Springer.
- Kokkinakis, Dimitrios, Maria Toporowska Gronostaj & Karin Warmenius, 2000: *Annotating, disambiguating & automatically extending the coverage of the Swedish SIMPLE lexicon*. I: *Proceedings from the Second International Conference on Language Resources and Evaluation*. Athens. S. 1397–1405.

- Moerdijk, Fons, 2008: Frames and Semagrams. Meaning Description in the General Dutch Dictionary. I: Proceedings of the XIII EURALEX International Congress, red. af Elisenda Bernal & Janet DeCesaris. Barcelona. S. 561–570.
- Niestadt, Jan, Carole Tiberius & Fons Moerdijk, i trykken: Searching the ANW Dictionary. I: Proceedings of eLexicography in the 21st century: New challenges, new applications, red. af Sylviane Granger & Magali Paquot. Louvain-la-Neuve.
- NLO = Nordisk leksikografisk ordbok. Oslo 1997.
- ODS = Ordbog over det danske Sprog 1–28. København 1919–1956.
- 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 43: 3. S. 269–299.
- Pedersen, Bolette Sandford & Patrizia Paggio, 2004: The Danish SIMPLE Lexicon and its Application in Content-based Querying. I: Nordic Journal of Linguistics 27: 1. S. 97–127.
- Princeton WordNet. <http://wordnet.princeton.edu/>
- Pustejovsky, James, 1995: The Generative Lexicon. Cambridge, Massachusetts.
- SAOB = Svenska Akademiens ordbok. Lund 1898–.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
- Sørensen, Nicolai Hartvig & Lars Trap-Jensen: Den Danske Ordbog som begrebsordbog. I dette bind.
- Zgusta, Ladislav, 1988: Pragmatics, lexicography and dictionaries of English. I: Word Englishes 7: 3. S. 243–253.
- Vossen, Piek (ed.), 1998: EuroWordNet, A Multilingual Database with Lexical Semantic Networks. Dordrecht.