

# **Fra begrebsordbog til sprogteknologisk ressource: verber, semantiske roller og rammer – et pilotstudie**

*Sanni Nimb & Bolette Sandford Pedersen*

This paper describes a method of compiling a lexicon of Danish semantic frames within the model of the Berkeley FrameNet (BFN). Large groups of near-synonymous verbs and verbal nouns, including multiword units, within the domains of communication and cognition are identified and extracted from the source manuscript of a newly published Danish thesaurus. Each word or expression is then assigned an appropriate frame from BFN. The fact that words within the same domain all belong to a manageable subset of frames in BFN makes it possible to map a high number of words to their corresponding frames simultaneously. In a forthcoming annotation project where words within the same two domains are already identified in the corpus, the idea is to pre-annotate with the frames in our lexicon, leaving afterwards human annotators to confirm the frame and test whether it is possible to identify the BFN semantic roles described for English in the Danish text. Our method reveals some interesting divergences between the semantic divisions established in the thesaurus in contrast to the ones found in BFN, showing that the two resources contribute with different types of linguistic information and thereby constitute a useful supplement to one another.

## **1. Baggrunden for pilotstudiet**

I forskningsprojektet *Semantic Processing across Domains*, et 3-årigt samarbejdsprojekt mellem Københavns Universitet og Det Danske Sprog- og Litteraturselskab, støttet af Det Frie Forskningsråd, eksperimenteres der med forskellige former for semantisk opmærkning af danske, genre-mæssigt meget forskellige tekster, som grundlag for en efterfølgende udvikling af automatiske processeringsmetoder til brug for forskellige former for sproglig betydningsanalyse. Der skal bl.a. eksperimenteres med opmærkning af semantiske rammer inden for den i princippet sproguafhængige model der er etableret i det amerikanske FrameNet-projekt *Berkeley FrameNet* (BFN), og

som allerede er bevist velegnet til overførsel til andre sprog, fx spansk, japansk, svensk og fransk, jf. Boas (2009). Fordelen ved at benytte en standard som BFN til opmærkning er at man dermed får adgang til at overføre de sprogteknologiske resultater indenfor automatisk betydningsanalyse der er opnået med anvendelsen af BFN på andre sprog. Denne artikel beskriver vores arbejde i forskningsprojektet med at etablere et BFN-rammeleksikon for dansk til brug for det planlagte opmærkningsarbejde. Vi beskriver hvordan rammeleksikonnet opbygges på basis af den semantiske gruppering af det danske ordforråd som man finder i Den Danske Begrebsordbogs grundmanuskript (DDB 2014). Efterfølgende skal leksikonnet bruges til præannotering af vores korpustekster således at annotøren kan nøjes med at godkende den automatisk tildelte ramme, evt. blandt flere mulige forslag, og derefter fokusere på at opmærke rammens tilknyttede semantiske roller sådan som de udmøntes i den danske tekst. Det særlige ved vores metode er at de mulige, semantiske roller som annotøren skal identificere i teksten, overtages direkte fra BFN's rammebeskrivelse. Vi har nemlig en hypotese om at det rolleinventar som er fundet i engelske tekster og derefter omhyggeligt beskrevet i BFN, sandsynligvis også vil dække de roller der er relevante at opmærke i danske tekster, og at man derfor kan spare den tidskrævende arbejdsgang det ville være at beskrive rolleinventaret for hvert verbum i selve det danske leksikon inden annoteringen påbegyndes.

### 1.1. Forskelle og ligheder mellem et FrameNet og en begrebsordbog

Både en begrebsordbog som DDB og et formelt semantisk leksikon som BFN beskriver handlinger og situationer. Et FrameNet udgør en formaliseret delmængde af en begrebsordbog, blot uden en overordnet emneinddeling. En såkaldt ramme (frame) i BFN, inklusive dens tilknyttede leksikalske enheder, svarer til den række af ord i et afsnit i en begrebsordbog der beskriver handlingerne eller hændelserne inden for det emne afsnittet omhandler. Omvendt mangler en begrebsordbog en præcis betegnelse for netop denne gruppering af ord – afsnittets navn er typisk bredere formuleret og dækker også andre former for begreber i afsnittet, ikke kun handlinger og hændelser. Et FrameNet indeholder desuden en formaliseret beskrivelse af disse handlinger og hændelser. For det første navngives hver enkelt handlings- og hændelsesgruppe med en præcis og entydig rammebetegnelse,

og denne suppleres yderligere med oplysninger om de elementer der kan indgå i den i form af en række semantiske roller fra et afgrænset inventar, fx en 'Speaker', en 'Message' og en 'Addressee'. Derudover listes der ved hver ramme de leksikalske enheder der kan have den pågældende rammebetydning (svarende til rækken af ord og udtryk i gruppen i DDB). Et eksempel er BFN-rammen 'Statement' der har den leksikalske enhed *talk* tilknyttet. De leksikalske enheder kan selvfølgelig være polyseme inden for BFN, dvs. have flere rammebetydninger, fx har den engelske verbum *talk* udover 'Statement' også fire andre rammer, fx 'Discussion' (svarende til at verbet *tale* står i mange forskellige afsnit i DDB). Men der er også tilfælde hvor en leksikalsk enhed kun har én rammemulighed, således har *think up* kun rammen 'Coming\_up\_with'.

I BFN-projektet (der fortsat udbygges med rammer) fokuseres der primært på den detaljerede opbygning af rammer og roller for det engeske sprog, og det har derfor en forholdsvis lav leksikalsk dækningsgrad (13.099 ord), men til gengæld 1073 rammer med tilhørende formaliserede rollebeskrivelser. I DDB fokuseres der lige modsat på selve det danske ordforråd. Vores metode går ud på at kombinere de to ressourcer i opbygningen af et dansk FrameNet, så vi opnår det bedste af to verdener. Vi udnytter det store inventar af allerede beskrevne rammer fra BFN, men lader ordforrådet i DDB udgøre de leksikalske enheder, idet vi i øvrigt overfører alle roller og relationer som er beskrevet i rammerne, direkte fra BFN. Vi tager udgangspunkt i DDB's tematiske inddelinger og rækker af nærsynonyme ord og udtryk idet vi udnytter at begrebsordbogens grundmanuskript indeholder formaliserede oplysninger om ordenes semantiske kategori (dvs. om de fx er handlinger eller hændelser). I det svenske FrameNet-projekt (se Friberg & Dannells 2015) benytter man også et semantisk leksikon som udgangspunkt, nemlig SALDO; herfra linkes der på samme måde til BFN. Men SALDO er organiseret efter indbyrdes semantiske relationer mellem ordene, ikke efter temaer; på den måde adskiller det sig en del fra en begrebsordbog, jf. Borin & Forsberg (2009). Bicks udvikling af en dansk framenet-lignende ressource (se Bick 2011) bygger heller ikke på et tematisk udgangspunkt, men på et constraint-grammar-baseret valens-leksikon suppleret med information fra det danske WordNet DanNet. I ressourcen defineres et sæt af i alt 38 semantiske roller som ikke er direkte kompatible med BFN's roller.

## 2. Udvalgelse af ordforråd til pilotstudiet

Afgrænsningen af ordforrådet til vores pilotleksikon bygger på de resultater vi har opnået i den første del af forskningsprojektet, nemlig den del der omhandler leksikalsk korpusopmærkning med betydningskategorier. Udover en eksperimentel opmærkning med et lille udvalg af Den Danske Ordbogs (DDO) og det danske WordNet DanNets finkornede betydningsinventar (se Pedersen et al. i dette bind), har vi også foretaget opmærkning af 1500 sætninger fra samme korpus (DK-Clarín (dkclarin.ku.dk)) med et sæt af 58 grove betydningskategorier, de såkaldte supersenses der udgør en international standard, fx 'person', 'food', 'time', 'communication' etc., se Martínez et al. (2015) for flere detaljer. Et af de resultater der kom ud af opmærkningsarbejdet, er en statistik over hvilke af de 58 supersenses der er mest udbredte. Blandt de 6 hyppigste kategorier var substantiver og verber med supersense 'communication', og verber med supersense 'cognition' var syvendehyppigst. Ved at etablere et rammeleksikon der dækker de centrale rollebærende ord inden for netop disse to frekvente betydningsområder, regner vi med at der vil være forekomster nok i vores 1500 supersense-opmærkede sætninger til at rammeleksikonnet, og at vores hypotese i første omgang kan afprøves på de samme sætninger. Senere skal leksikonnet udvides til at omfatte i princippet alle betydningsområder i et nyt forskningsprojekt der er bevilget af Carlsbergfondet i 2016-2017, baseret på de resultater og metodiske erfaringer vi har opnået i indeværende forskningsprojekt.

Rammeleksikonnets leksikalske udgangspunkt, DDB, indeholder næsten 200.000 danske ord og udtryk (heraf 112.000 forskellige) sorteret i rækkefølge ud fra deres betydning og placeret under 888 afsnit og 22 kapitler med emneoverskrifter. I det grundmanuskript der ligger bag den trykte bog, er ordforrådet yderligere inddelt i 8.300 grupper opmærket med formaliserede semantiske oplysninger, og man kan derfor direkte udtrække de ca. 1700 grupper af verber og substantiver der beskriver en handling eller hændelse. Derved afgrænses netop det ordforråd som det er mest relevant at tildele BFN-rammer med henblik på at kunne udnytte det tilhørende engelske semantiske rolleinventar. Emneoverskrifterne i DDB gør det desuden muligt at identificere de ord der har kommunikations- eller kognitionsbetydning. Kapitel 11 "Tænkning" med 48 afsnit omfatter således kognition, mens kommunikationsordforrådet skal findes inden for de 49 afsnit i

kapitel 12 ”Sprog og kommunikation” samt i et udsnit af de i alt 58 afsnit i kapitel 15 ”Socialt liv” (fx i afsnit som ’Anerkendelse’, ’Smiger’ og ’Skælde ud’). En gruppe der er relevant for os, er fx gruppen af verber og substantiver der beskriver handlingen ’skælde ud’ i afsnit 15.24, se Figur 1.

```
{08_Vb_SbAflledning/has_hyperonym: 'vise sin vrede has_hyperonym: 'vredesudbrud involved_agent:
'person involved_patient: 'person}
> skælde ud, 'skrue bissen på, skænde, skælde, skælde (ud) for, 'tordne, 'tale dunder, tale med
store bogstaver, skælde og smælde, 'udskælde, 'gennemhegle, 'give (med) grovfilen, give en gang lak,
hegle igennem, 'sige et par borgerlige ord (til), give tørt på, skælde hæder og ære fra, skælde bælgen
fuld, skælde huden fuld, rive hovedet af nogen, 'tage nogen i skole, 'slå i bordet, 'bruge mund, 'herse,
overfuse, rise, dænge til<; > få luft for sin vrede, 'få afløb for sin vrede, 'bande nogen langt væk,
'rase ud<; > give luft for sin vrede, 'rase, 'fråde, 'se rødt, 'gå amok, 'springe/ryge i luften, 'koge over,
'eksplodere<; > snerre, 'bide ad, 'hvæse, 'spytte sætningen ud, 'sige vredt/bittert, 'råbe vredt<;
'komme efter nogen, 'småskænde på, 'vreden løb af med ham, gl'udøse sin vrede, 'skamme ud;
> 'komme med tilråb, 'fare i blækhuset, 'hvæsse/spidse pennen, 'hvæsse pennen<; 'hvæsse klørerne,
'forløbe sig; > vredesudbrud, raserianfald, udfald, 'vredesskrig, 'vris, 'snøft, 'gnaveri<;
> udskældning, 'skældud, 'udskæld, 'skænd, opsang, 'formaning, 'pegefinger, en sang fra de varme
lande uform<; > 'irettesættelse, 'røffel, 'gardinprædiken, 'moralprædiken eds<; > 'overhaling, 'skideballe
uform, 'sviner uform, '-et ordentligt pulver, 'møgfald, 'bredsider, 'det glatte lag, 'balle, 'overfald, 'hak i tuden<;
> 'hårde ord, 'knubbede ord, 'salut, 'svada eds, 'salve, 'dundertale, 'tordentale, 'afskedssalut<;
'forløbelse; > skænderi, 'større skænderi, 'ophidset diskussion, 'hidsig diskussion, 'skændsmål gl<; >
{syn} 'heftigt skænderi, 'kæmpeskænderi<; 'familieskænderi
```

Figur 1: Gruppen af ord og udtryk med betydningen ’skælde ud’ i Den Danske Begrebsordbogs grundmanuskript.

De 1700 grupper i DDB indeholder i alt mere end 40.000 verber og verbalsubstantiver. Det høje tal skyldes et stort antal flerordsudtryk, herunder mange faste udtryk, men også en række udfoldninger af samme ord der i kombination med forskellige andre ord får netop den pågældende betydning, i Figur 1 fx *sige vredt/bittert* og *råbe vredt*. Desuden skyldes det gengangere på tværs af DDB’s manuskript, svarende til tildelingen af flere rammer til samme udtryk i BFN-modellen, jf. engelsk *talk* ovenfor. Men der er rent faktisk også en stor mængde forskellige ord, for ligesom ord med kommunikationsbetydning og kognitionsbetydning er hyppige i vores opmærkede tekster, er også DDB’s afsnit om disse betydningsområder meget ordrige. I alt udgør antallet af verber og verbalsubstantiver i vores udvalgte afsnit fra de tre kapitler, i alt 63 afsnit, ca. 7000 ord og udtryk, svarende til ca. 16 % af det samlede antal verber og verbalsubstantiver i begrebsordbogen.

### 3. Fremgangsmåde

Da alle ord og udtryk i DDB via fælles betydnings-id-numre i grundmanuskriptet er knyttet til deres betydningsbeskrivelse, valens-angivelser mm. i

DDO, kan de udtrukne lister af ord og udtryk fra DDB suppleres med valensmønstre til brug for præcis identifikation af hvilken handling eller situation der egentlig er tale om når man oversætter de danske ord til engelsk og skal finde ækvivalentens ramme i BFN. I overvejelserne indgår også både rammens andre leksikalske enheder og dens semantiske rolleinventar. I Figur 2 ses et udsnit af verber og verbaludtryk, igen fra begrebsområdet 'skælde ud', med tilknyttede valensmønstre fra DDO. Vi tager typisk udgangspunkt i det mest generelle udtryk i dansk når den rette ramme skal findes, fx for 'skælde ud'-gruppen verbet *skælde ud*, og vi anvender en digital tosprogsordbog for dansk-engelsk fra Gyldendal til oversættelsesarbejdet. Når en ækvivalent er fundet (*skælde ud = to scold*), søger vi efter det engelske ord *scold* i Lexical Unit-indekset i BFN og ser hvilken ramme det hører ind under. I tilfældet *scold* er det 'Judgment\_direct\_adress', og da beskrivelsen af den situation rammen dækker over, inkl. de roller den omfatter, passer fint til det vi forstår ved det danske begreb 'skælde ud', og også til de valensmønstre ord inden for begrebet typisk er beskrevet med i DDO, sættes denne ramme på både *skælde ud* og alle de nærsynonyme ord og udtryk fra DDB der passer på rammen.

Afsnit i DDB	Udtryk i DDB	Lemma i DDO	Valensmønster i DDO
15.24 Skælde ud	<i>skælde ud</i>	<i>skælde</i> , fast udtryk	NOGEN skælder ud på NOGEN NOGEN skælder NOGEN ud (for at../NOGET)
15.24 Skælde ud	<i>skrue bissen på</i>	<i>bisse</i> , fast udtryk	NOGEN skruer bissen på (over for NOGEN)
15.24 Skælde ud	<i>skænde</i>	<i>skænde</i> ., bet. 1	NOGEN skænder (på NOGEN) NOGEN skænder REPLIK
15.24 Skælde ud	<i>skælde</i>	<i>skælde</i> , fast udtryk	NOGEN skælder (på NOGEN)
15.24 Skælde ud	<i>tordne</i>	<i>tordne</i> , bet. 2	NOGEN tordner REPLIK NOGEN tordner mod NOGET/NOGEN
15.24 Skælde ud	<i>tale dunder</i>	<i>dunder</i> , fast udtryk	NOGEN taler dunder (til NOGEN)
15.24 Skælde ud	<i>skælde og smælde</i>	<i>skælde</i> , fast udtryk	NOGEN skælder og smælder (over NOGET/at../..)
15.24 Skælde ud	<i>gennemhegle</i>	<i>gennem-hegle</i> , bet. 1	NOGEN gennemhegler NOGEN

Figur 2: Ord og udtryk fra en handlingsverbalgruppe i afsnit 15.24 "Skælde ud" i DDB, suppleret med valensmønstre fra DDO.

#### 4. Resultater og eksempler

I pilotstudiet fik de mere end 7000 verber og verbalsubstantiver på få måneders arbejde tildelt en rammebetegnelse fra BFN vha. den skitserede metode. I alt blev der anvendt 77 forskellige rammer af BFN's i alt 1073. Da de danske verber allerede er listet ud fra betydning og derfor typisk i én omgang kan oversættes til samme ramme, og da den rette ramme ofte findes inden for et stærkt reduceret antal af de over 1000 mulige i BFN fordi man bevæger sig inden for et overordnet betydningsområde (kommunikation eller kognition), var arbejdsmetoden meget effektiv. BFN har ikke defineret brede betydningsområder svarende til kapitlerne i begrebsordbogen, og man ville derfor ikke kunne indkredse hvilke rammer af de 1073 beskrevne der præcis dækker et bestemt betydningsområde som fx kommunikation. De rammer der er blevet anvendt ved oversættelse af de danske verber som vi som udgangspunkt definerede som hørende under emnet kommunikation i vores pilotstudie, blev dog typisk tildelt rammer fra BFN der umiddelbart også er genkendelige som hørende til dette tematiske område (fx 'Discussion', 'Become\_silent', 'Bragging', 'Chatting' og 'Mention'). Inden for rolleinventaret i de anvendte rammer fandt man også mange fællestræk, fx omfatter de typisk en 'Speaker' eller en 'Communicator', men nogle af de anvendte rammer har et andet inventar, fx 'Interlocutor' (ved rammen 'Discussion') eller 'Authority' (ved rammen 'Deny\_permission'). Det var dog også nødvendigt at benytte rammer der dækker mere end blot kommunikation, især til verber fra kapitel 15 'Socialt liv' i DDB, da afsnittene i dette kapitel ikke kun omfatter deciderede kommunikations-verber. Fx udgør *anerkende, værdsætte, skatte, agte og berømme* en ordgruppe i afsnittet 'Anerkendelse'. De fire første tildeles rammen 'Judgment'; kun det sidste, *berømme*, rammen 'Judgment\_communication'. Som et eksempel på BFN's skelnen mellem rammer anføres det i rammebeskrivelsen for 'Judgment' at: "*This frame is distinct from the Judgment\_communication frame in that this frame does not involve the Cognizer communicating his or her judgment to an Addressee*". Enkelte gange var der fuld overensstemmelse mellem et afsnits verber og en ramme i BFN. Fx svarer handlinger i afsnittet 'Fortale sig, sladre' i DDB ret præcist til rammen 'Reveal\_secret' i BFN. I meget få tilfælde fandtes der ingen passende ramme i BFN; begrebet 'hemmeligholde' er fx (endnu) ikke dækket af nogen ramme. Vi har også noteret os enkelte rammer

i BFN der burde være anvendt i vores pilotstudie hvis man vil opnå fuld dækning af området kommunikation. Dette gælder fx 'Communication\_means', og det skyldes at de danske verber står i et afsnit under et helt andet kapitel i DDB, nemlig det der beskriver apparater og teknik (i afsnit 19.17 'Telefon, e-mail'). Pilotstudiet giver således også feedback til arbejdet med DDB idet det er oplagt også at bringe ordene under kapitel 12 'Tegn, meddelelse, sprog'.

I langt de fleste tilfælde var der mange krydsninger mellem handlinger og hændelser i DDB's afsnit og rammer i BFN. Fx svarer det der kun udgøres af én ramme i BFN, 'Respond\_to\_proposal', til ord fra to afsnit i DDB, nemlig både 'Bekræfte' og 'Benægte'. Eller det omvendte var tilfældet: der er flere rammer i spil samtidig for en gruppe danske ord fra et afsnit i DDB, idet BFN afgrænser rammer fra hinanden ud fra betydningskriterier der ikke har været afgørende i DDB. Ved gruppen af danske 'skælde ud'-ord i DDB skelnes der fx ikke mellem at skælde ud direkte til en person eller skælde ud over vedkommende til en anden person. Men når BFN-rammerne skal tildeles de enkelte ord i afsnittet, skal man tage stilling til dette, dvs. om der er tale om rammen 'Judgment\_direct\_address' (der passer på de danske udtryk *skælde ud på nogen, skælde nogen ud, tale med store bogstaver, skælde nogen hæder og ære fra, give nogen tørt på, overfuse nogen, gennemhegle nogen, komme efter nogen, småskænde på nogen*), eller om der i stedet er tale om Judgment\_communication hvor der ikke tales til den person man er vred på (fx *bande nogen langt væk*). Når det danske verbum har begge betydninger (evt. i forskellige konstruktioner: *rase mod nogen, rase over noget*), vælges den bredeste ramme 'Judgement\_communication'. Til gengæld dækker både 'Judgment\_communication' og 'Judgment\_direct\_adress' bredere mht. et andet betydningsaspekt end de modsvarende danske afsnit i DDB gør, idet der i BFN ikke skelnes mellem at bedømme positivt (fx *rose*) og bedømme negativt (fx *skælde ud*). BFN forudsætter i stedet at oplysninger ved selve ordene skal belyse om der er tale om negativ eller positiv bedømmelse. Figur 3 viser hvordan inddelingskriterier krydser hinanden i BFN og DDB.



DDB, afsnit	15.19 Anerkendelse (positiv, både direkte og ikke direkte til person)	15.24 Skælde ud (negativ, både direkte og ikke direkte til person)
BFN, ramme		
Judgement_communication (både positiv og negativ, ikke direkte til person)	<i>berømme</i>	<i>bande langt væk</i>
Judgement_direct_adress (både positiv og negativ, direkte til person)	<i>komplimentere</i>	<i>overfuse, gennemhegle</i>

Figur 3: BFN og DDB benytter forskellige kriterier i deres inddeling af ordforrådet.

#### 4.1. Planer for brug af rammeleksikonet

Det rammeleksikon vi opbygger, vil som nævnt blive brugt til at forhåndsopmærke ord med rammer i vores 1500 supersenseopmærkede sætninger, vel at mærke i de tilfælde hvor ordet i korpusset er opmærket med supersense 'communication' eller 'cognition'. Ved den efterfølgende håndopmærkning af teksten vil annotørerne via oplysninger om den engelske rammes tilknyttede semantiske roller vide at de ved udtrykket *skælde ud* med rammen 'Judgment\_direct\_adress' skal lede efter og annotere en mulig 'Communicator', en 'Addressee', en 'Reason' eller et 'Topic' i den danske sætning. Vi har udvalgt et par eksempler på opmærkede sætninger (Figur 4).

Communicator	Addressee	Reason
Den ungarske landstræner	havde talt med store bogstaver til sine spillere	i pausen
Jeg skælder hende ud	for at være groft uansvarlig	
I debatten tordnes der løs	mod Det kgl. Teaters repertoire	

Figur 4: Semantiske roller for rammen 'Judgment\_direct\_adress' og eksempler på danske, opmærkede sætninger (i form af citater fra DDO).

Leksikonet indeholder langt flere ord og udtryk til hver ramme end man typisk finder i BFN, særligt flere flerordsudtryk. Nogle er kontinuerte, andre ikke, nogle kan varieres, andre ikke. I det kommende annoteringsarbejde vil det vise sig om flerordsudtrykkene rent faktisk kan genkendes i den automa-

tiske morfologiske analyse og dermed også udnyttes i den automatiske præannotering. Flerordsudtrykkene vil også kræve særlig opmærksomhed fra annotørernes side idet det semantiske rolleinventar der beskrives for rammen i BFN, mange gange vil være at finde i udtrykket selv i det danske leksikon fordi det af formidlingshensyn i DDB er nævnt i form af et pronomen eller en typisk leksikalsk udfyldning (fx i *skælde ud på nogen* og *skælde nogen ud*). I andre tilfælde er det i stedet tale om et fast udtryk, dette gælder fx *tale med store bogstaver*; her skal 'med store bogstaver' ikke tildeles nogen rolle i opmærkningen.

## 5. Konklusion

Pilotstudiet har bekræftet vores hypotese om at man med fordel kan kombinere DDB's inddeling af verber og verbalsubstantiver i emnegrupper og FrameNets rammer. De to semantiske ressourcer har begge en tematisk inddeling af handlinger og hændelser som noget meget centralt i beskrivelsen af ordforrådet, og det betyder dels at man kan koble store grupper af verber til en passende ramme i én arbejdsgang, dels at man kan fokusere på ordforrådet inden for et bestemt betydningsområde hvilket letter arbejdet betydeligt. Men man kan ikke overraskende konstatere at de lægger snittet mellem emnerne lidt forskelligt. Nogle gange omfatter en ramme fra BFN derfor verber og verbalsubstantiver fra flere afsnit i DDB, nogle gange kun en delmængde af verberne og verbalsubstantiverne i ét afsnit. Der er for så vidt ikke en af inddelingerne der er mere korrekt en den anden, men de rummer supplerende typer af information der med fordel kan bringes sammen. Den konkrete anvendelse af leksikonnet i vores kommende opmærkningsarbejde vil vise om vores hypotese om at de engelske rollebeskrivelser også passer for dansk, holder, eller om vi i stedet er nødsaget til at udvide det danske rammeleksikon vi opbygger på basis af DDB's ordforråd, med roller specifikt for dansk.

## Litteratur

### Ordbøger

BFN = *Berkeley FrameNet*: <framenet.icsi.berkeley.edu>.

DDB (2014) = S. Nimb, H. Lorentzen, L. Theilgaard & T. Troelsgård: *Den Danske Begrebsordbog*. København: Det Danske Sprog- og Litteraturselskab.

DDO = ordnet.dk/ddo – E. Hjort, K. Kristensen et al. (2003-2005): *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab.

#### Anden litteratur

Bick, Eckhard (2011): A FrameNet for Danish. I: B. S. Pedersen, G. Nešpore & I. Skadina (red.): *Proceedings of NODALIDA 2011*. Tartu, 34-41.

Boas, Hans C. (red.) (2009): *Multilingual FrameNets in computational lexicography: Methods and applications*. Trends in linguistics, Studies and Monographs 200. New York/Berlin: Mouton de Gruyter.

Borin, Lars & Markus Forsberg (2009): All in the Family: A Comparison of SALDO and WordNet. I: *Proceedings of the NODALIDA 2009 Workshop on WordNets and other Lexical Semantic Resources - between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*. NEALT Proceedings Series. Göteborg: Språkbanken, Göteborgs Universitet, 7-12.

Friberg, Karin Heppin & Dana Dannells (2015): Polysemy and questions of lumping or splitting in the construction of Swedish FrameNet. I: *Proceedings of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities*. Linköping: Linköping University Electronic Press, 12-20.

Martinez, Hector, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Sjøgaard & Bolette Sandford Pedersen (2015): Supersense tagging for Danish. I: *Proceedings of the 20th Nordic Conference on Computational Linguistics*. NODALIDA 2015. Linköping: Linköping University Electronic Press, 21-29.

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

Bolette S. Pedersen  
professor, institutleder  
Nordisk Forskningsinstitut  
Københavns Universitet  
Njalsgade 136, bygn. 27. 2  
DK-2300 København S  
bspedersen@hum.ku.dk