

ALTING PÅ SIN PLADS OG PLADS TIL ALTING. OM AT ORDNE OG UDNYTTE VIDEN OM VERDEN

Bodil Nistrup Madsen
Handelshøjskolen i København

Abstract

After a brief introduction to various types of knowledge structuring and the differences and similarities between them, the paper focuses on terminological concept systems, ontologies and data models. In connection with ontologies, the use of domain specific and general ontologies will be discussed as well as the possibility of using one common ontology to solve the problems of communication between experts from different domains.

1. Indledning

Det er et velkendt faktum, at mennesket til alle tider har ønsket at ordne viden om verden. Filosofer har udarbejdet ontologier, ofte også kaldet metafysikker (Aristoteles, Immanuel Kant, Charles Sanders Pierce m.fl.), se f.eks. Sowa (1997). Dokumentalister og fageksperter har udarbejdet emneklassifikationer (Melvil Dewey, Paul Otlet, Henry la Fontaine m.fl.), se f.eks. BS 1000M: Part 1 (1985) og tesaurusser, se f.eks. Soergel (1974), Spang-Hanssen (1985) og Vickery (1960). Filologer har udarbejdet begrebsordbøger (Peter Mark Roget og Harry Andersen), se f.eks. Vickery (1960) og Ruus (1988). Terminologer har udarbejdet begrebssystemer, se f.eks. Arntz & Picht (1989), Nuopponen (1994), og Madsen (1996 og 1999).

Ordning af viden har gennem tiden haft forskellige formål. Generelle ontologier har tidligere primært tjent til en systematisering af viden om verden på en filosofisk baggrund. Emneklassifikationer er blevet udarbejdet til brug for ordning af objekter, typisk ordning af publikationer i et bibliotek eller museumsgenstande i et museum. Snævert forbundet med emneklassifikationer er tesaurusser, som anvendes inden for informations- og dokumentationssøgning til indeksering og genfindning af publikationer mv. Tesaurusser har til formål at hjælpe brugeren til at komme fra et ord til et indhold, eller snarere til den deskriptor, som

er valgt til at beskrive et givent indhold, typisk indholdet i en publikation. Begrebsordbøger¹ derimod har til formål at hjælpe forfattere med at komme fra ide, dvs. indhold, til ord. Begrebsordbøger og tesaurusser har mange lighedspunkter, men adskiller sig mht. formål. Begrebsordbøger har endvidere mange lighedspunkter med de almene ontologier, som i de seneste år er blevet udarbejdet, f.eks. Cyc, som omtales nedenfor, se f.eks. Sowa (1997), Vickery (1997). Både tesaurusser, begrebsordbøger og ontologier er ligesom begrebssystemer baseret på en begrebsorienteret systematik, men begrebssystemer udarbejdes i forbindelse med terminologiarbejde for at registrere relationer mellem faglige begreber og at fastlægge definitioner af sådanne begreber inden for et afgrænset emneområde m.h.p. begrebsafklaring og fastlæggelse ækvivalens mellem begreber på forskellige sprog.

I forbindelse med anvendelsen af datamatiske systemer er der ligeledes behov for en strukturering af den information, som skal benyttes i form af data i sådanne systemer. Dette omtales f.eks. i forbindelse med databaser som datamodellering eller semantisk modellering, se f.eks. Date (1990: 579f), og består typisk i udarbejdelse af entitets/relation-diagrammer. I dag anvendes i stigende grad objektorienteret datamodellering, OMT, *Object-oriented Modeling Technique* (Coad & Jourdon 1991), og UML, *Unified Modeling Language* (Eriksson & Penker 1998).

Med den øgede anvendelse af datamatiske systemer, f.eks. til opbygning af vidensbaser, udarbejdelse af natursprogsgrænseflader og avancerede søgemetoder, er strukturering af viden, udarbejdelse af præcise begrebsdefinitioner og datamodellering blevet endnu mere aktuell, og grænserne mellem datamodellering og begrebsstrukturering samtidig mere udflydende.

I nærværende indlæg vil en række centrale problemstillinger vedrørende ontologier, begrebssystemer og datamodeller blive behandlet. Pladsen tillader ikke at komme ind på en nærmere diskussion af tesaurusser og emneklassifikationer, hvorfor der henvises til Spang-Hanssen (1985) og Madsen (1999).

¹ På engelsk bruges 'thesaurus' både om informationssøgningstesaurusser og begrebsordbøger.

2. Ontologier

En ontologi er en systematisk specification af viden om klasser af entiteter i verden og relationerne mellem disse. Oprindeligt omfattede ontologier kun generiske relationer, men i dag indgår ofte andre relationsarter, især del-helheds-relationer. Hver knude i en ontologi repræsenterer en klasse af entiteter. Der findes forskellige typer ontologier, som har forskellige formål.

Sowa (1997: 36) siger:

"Philosophers usually build their ontologies from the top down. They start with grand conceptions about everything in heaven and earth. Programmers, however, tend to work from the bottom up. For their database and AI systems, they start with limited ontologies or microworlds, which have a small number of concepts that are tailored for a single application."

Guarino (1998:4) skelner overordnet mellem ontologier, som er udarbejdet af filosoffer og er baseret på en bestemt verdensanskuelse, og ontologier, som anvendes i forbindelse med kunstig intelligens til at beskrive en del af virkeligheden. En 'filosofisk' ontologi er sprog-uafhængig, den gengiver en bestemt 'vision of the world' (Guarino 1998:4), som er den samme uanset hvilket sprog den beskrives i, hvorimod ontologier, som bruges som grundlag for datamatiske systemer, er sprogafhængige. Herudover skelner Guarino (1998:9–10) mellem typerne 'top-level ontology' (generelle begreber uafhængige af emne eller opgave), 'domain ontology' (emnespecifik), 'task ontology' (opgavespecifik) og 'application ontology' (både emne- og opgavespecifik). Endelig skelnes mellem formelle og ikke formelle ontologier, jf. f.eks. Smith (1998:19f). På basis af ovennævnte opstiller jeg følgende inddeling af ontologier² (jf. også oversigten i **Appendiks 1**):

² Der mangler gængse udtryk for nogle af begreberne, og ikke alle begreber omtales eksplicit i de her nævnte kilder (Sowa 1997, Vickery 1997, Guarino 1998). Valg af benævnelser og udarbejdelse af definitioner vanskeliggøres af upræcise definitioner af udtryk som *generel*, *almen*, *specifik* og *speciel*.

mht. anskuelsesmåde:

filosofisk ontologi: baseret på en bestemt verdensanskuelse
 pragmatisk ontologi: ikke baseret på en bestemt verdensanskuelse

mht. niveau:

top-niveau-ontologi: omfatter kun generelle begreber på et overordnet niveau
 speciel ontologi: omfatter specielle begreber på lavere niveauer
 universel ontologi: omfatter både generelle og specielle begreber

dækningsgrad mht. emne:

domænespecifik ontologi: omfatter et specifikt emneområde
 almen ontologi: omfatter almindelige (dagligdags) begreber

dækningsgrad mht. formål:

opgavespecifik ontologi: udarbejdes til et specielt formål³
 opgaveafhængig ontologi: udarbejdes ikke til et specielt formål
 ontologi:

dækningsgrad mht. sprog:

sproguafhængig ontologi: ikke relateret til et bestemt sprog
 sprogafhængig ontologi: relateret til et bestemt sprog

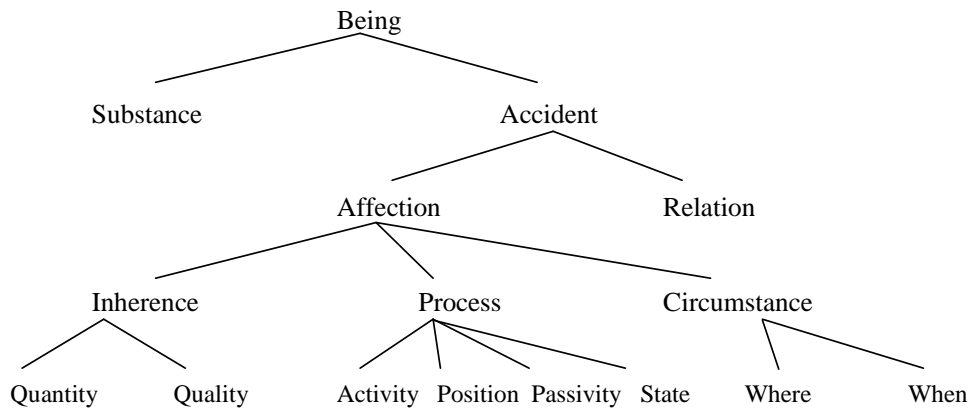
mht. formalisering:

formel ontologi: omfatter formalisering af relationer (og evt. aksiomer)
 ikke formel ontologi: omfatter ikke formalisering

Et eksempel på en *filosofisk ontologi* er Franz Brentanos træstruktur omfattende Aristoteles' ti kategorier plus 6 andre kategorier taget fra Aristoteles forskellige værker, jf. Brentano (1862). Denne ontologi er vist i **figur 1**.

Denne ontologi er samtidig en *sproguafhængig, opgaveafhængig, ikke formel, top-niveau-ontologi*. Det er et spørgsmål, om man skal anse denne, og andre filosofiske ontologier, for at være *almen* eller *domænespecifik* (fagområde: filosofi?). Begreberne repræsenteres ganske vist af almindelige udtryk, men de kan formodentlig ikke siges at referere til almene, dagligdags begreber, da de tillægges en særlig betydning i ontologien, bl.a. i kraft af de relationer, de indgår i.

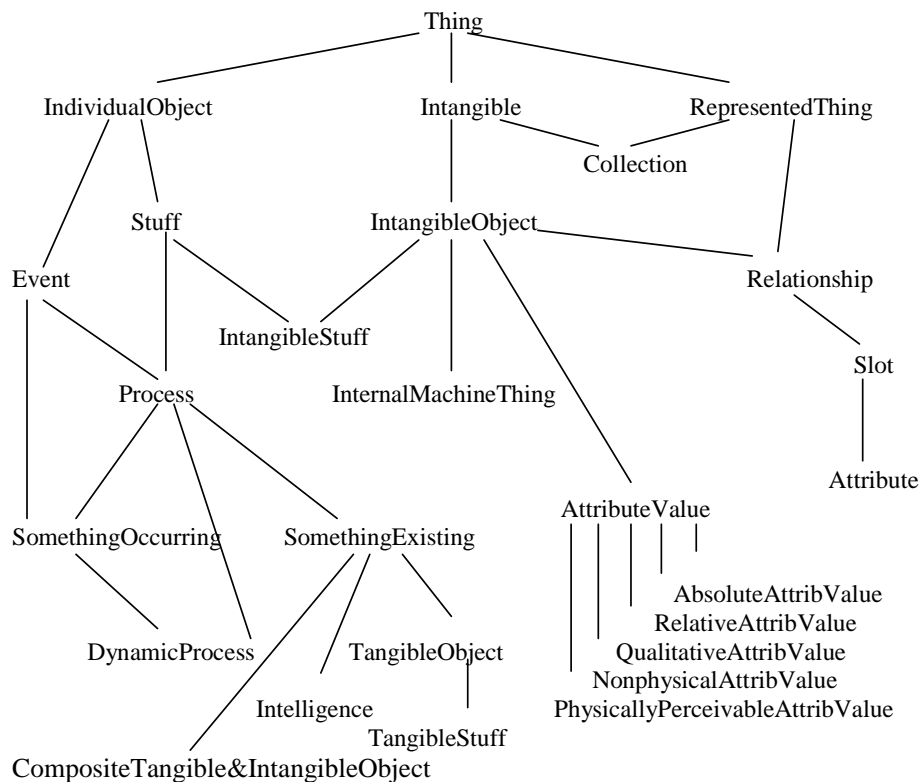
³ En ontologi, som er både domæne- og opgavespecifik kalder Guarino (1998:10), som ovenfor nævnt, 'application specific'.



Figur 1. Brentanos træstruktur over Aristoteles' kategorier

Et nutidigt eksempel på en *pragmatisk ontologi* er Cyc, hvis navn er afledt af 'encyclopedia', jf. f.eks. Lenat (1995), Sowa (1997). Cyc er udviklet af Cycorp, Austin Texas og findes på <<http://www.cyc.com>>. Cyc har gennemgået mange ændringer i tidens løb. I **figur 2** gengives en tidligere version af Cyc's 'top-level categories', som omfatter ca. 3.000 begreber, jf. Sowa (1997). Under denne findes ca. 10.000 begreber.

Cyc er samtidig et eksempel på en *opgaveafhængig, universel ontologi*, da den omfatter både generelle og specielle begreber. Top-level-kategorierne er *generelle* og *sprogafhængige*, medens begreberne på de lavere niveauer må betragtes som *specielle* og *sprogafhængige*, men dette er der formodentlig delte meninger om. Ligesom ved Aristoteles' begreber kan det diskuteres, om der er tale om almene (dagligdags) begreber. F.eks. er begreberne under f.eks. 'Attribute Value' ikke særlig dagligdags. Cyc er *formel*, idet den omfatter en formel beskrivelse af begrebsrelationer.

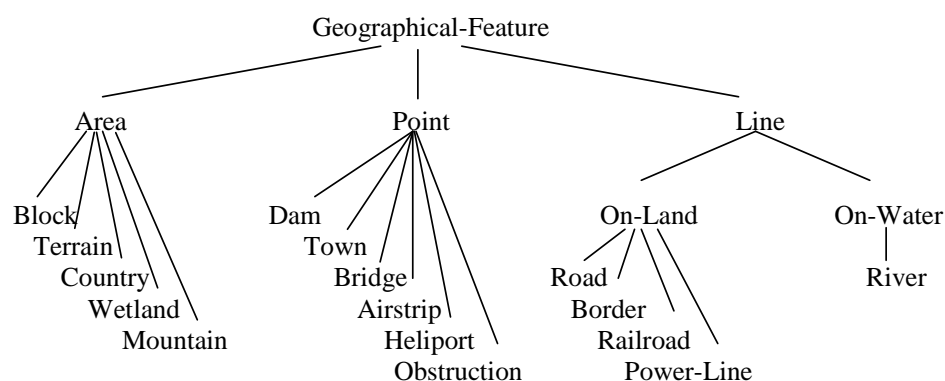


Figur 2. Cyc's øverste kategorier

Et eksempel på en *domænespecifik ontologi*, som ligeledes omtales i Sowa (1997) er 'Chat-80 ontologien', som omfatter geografiske kategorier, jf. også Warren & Pereira (1982: 110–122). Den er *pragmatisk* og *speciel*, men det kan diskuteres, om den er *sprogafhængig*. Chat-ontologien er udviklet til 'Chat-80 question-answering system', og er derfor endvidere et eksempel på en *opgavespecifik ontologi*. Chat-80-kategorierne anvendes flere steder i Chat-80-systemet:

"for reasoning, they support inheritance of properties from supertypes to subtypes; for queries, they map to the fields and domains in the database; and for language analysis, they determine the constraints on permissible combinations of nouns, verbs, and adjectives" (Sowa 1997: 36).

Chat-ontologien vises i **figur 3**.



Figur 3. Chat-ontologien

Sowa (1997) fremhæver, at Chat-ontologien er udviklet specielt til Chat-80-applikationen, hvilket bl.a. fremgår af at f.eks. 'River' og 'Road' er underbegreber til 'Line', og 'Town', 'Bridge' og 'Airstrip' behandles som en slags 'Point'. Dette stemmer meget godt overens med den opfattelse, som kommer til udtryk i følgende citat. Mahesh (i: Principles for 'ontological engineering', citeret fra Vickery 1997: 283) siger: "there is no unique ontology of the world (or even of a narrow domain). Ontologies are not natural entities to be discovered, they are artifacts to be constructed (and their structure will depend on their intended use)". I en anden domænespecifik ontologi inden for samme emneområde ville 'Line' og 'Point' ikke nødvendigvis være overbegreber.

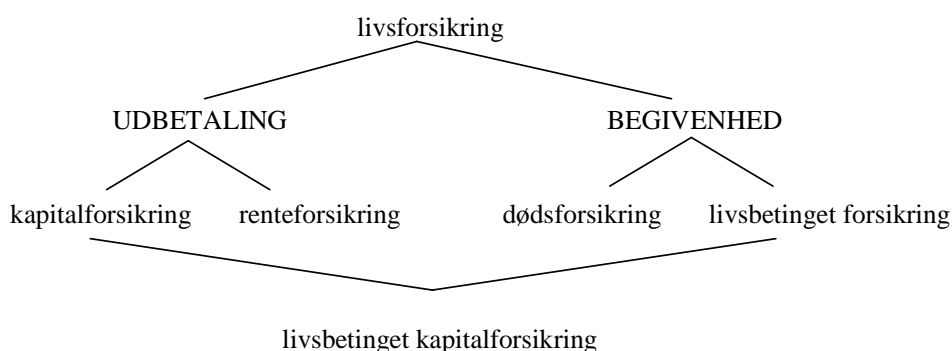
3. Begrebssystemer

Et begrebssystem er en systematisk specifikation af begreber inden for et specifikt emneområde, som fastlægger relationerne mellem begreberne. Formålet med et begrebssystem er, som nævnt i indledningen, afklaring og definition af begreber, og i oversættelsessammenhæng endvidere fastlæggelse af ækivalensrelationer mellem begreber på forskellige sprog. Begrebssystemer omfatter mange forskellige begrebsrelationer, jf. f.eks.

Arntz & Picht (1989: 79ff.) og Nuopponen (1994, 1997: 142–160). De indeholder ofte inddelingskriterier, og de kan være polyhierarkiske (dvs. at et underbegreb kan have flere overbegreber).

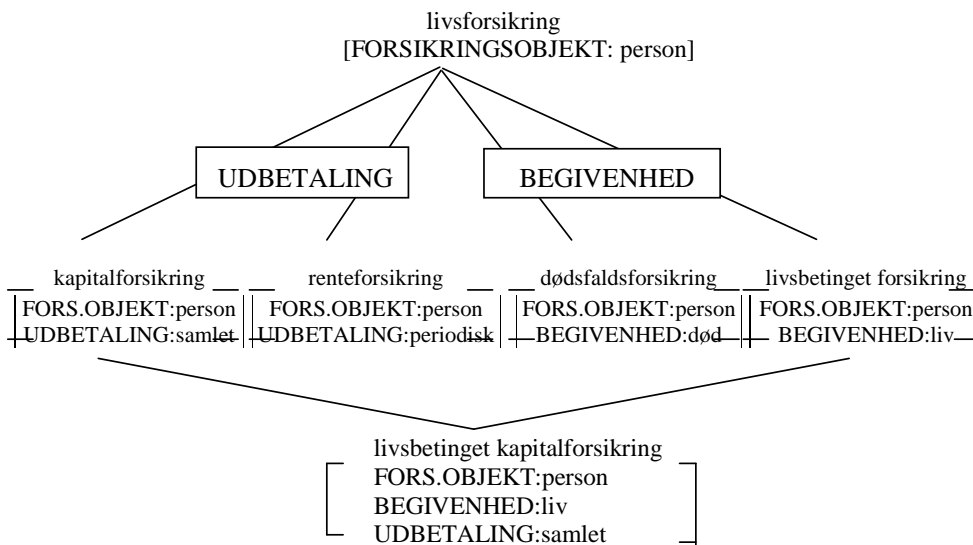
Hver knude i et begrebssystem repræsenterer et begreb, bestående af en mængde af karakteristiske træk. Denne definition af begreb findes hos Madsen (1991, 1998). En tilsvarende definition gives af Dahlberg (1985: 141) og af Picht (1997: 112) som, anfører at karakteristiske træk er det enkelte begrebs byggesten.

I **figur 4** vises et udsnit af et begrebssystem inden for livsforsikring, jf. Thomsen (1998a, 1998b). I dette system findes to inddelingskriterier, markeret med versaler: UDBETALING og BEGIVENHED. Inddelingskriterierne svarer til de adskillende træk. Inddelingskriterierne er naturligvis egentlig også begreber, men de betragtes ikke som sådanne i dette begrebssystem, da de kun har til funktion at gøre begrebssystemet mere overskueligt. Endvidere er der tale om polyhierarki, idet *livsbetinget kapitalforsikring* er underbegreb til to overbegreber: *livsbetinget forsikring* og *kapitalforsikring*.



Figur 4. Udsnit af et begrebssystem inden for livsforsikring

I overensstemmelse med definitionen af et begreb, kan man til hvert begreb i et begrebssystem anføre en trækstruktur, jf. Thomsen (1998 a, 1998 b), og Madsen (1998). En trækstruktur er en mængde af trækspesifikationer, bestående af træk-værdi par, jf. f.eks. Carpenter (1992). Dette eksemplificeres i **figur 5**.



Figur 5. Udsnit af begrebssystem med trækstrukturer

I **figur 5** ses, at underbegreber arver overbegrebers træk, og at inddelingskriterierne (de adskillende træk) fremgår af trækspicifikationerne. Madsen, Thomsen & Vikner (1999) foreslår, at den i **figur 5** anvendte metode til anførelse af inddelingskriterier anvendes, da den er mere korrekt end den, der traditionelt anvendes i terminologiske begrebssystemer. Når inddelingskriterierne er indsat som særlige hjælpeknuder, som i **figur 4**, kan læseren forledes til at opfatte dem som begreber på linie med de øvrige begreber i det pågældende system.

4. Forskelle og ligheder mellem begrebssystemer og domænespecifikke ontologier

Begrebssystemer og ontologier har mange lighedspunkter. De anvendes begge til strukturering af viden om verden. De adskiller sig imidlertid mht. det mere specifikke *formål* og *beskrivelsen* af arten af de elementer, de omfatter. Ontologier kan, som nævnt i 2, være enten *domænespecifikke* eller *almene*. Domænespecifikke ontologier og begrebssystemer har meget tilfælles, idet de anvendes til strukturering af viden inden for et specifikt

emneområde. I det følgende vil vi ved sammenligningen med begrebssystemer kun referere til *domænespecifikke ontologier*.

Den mest markante forskel er den måde, hvorpå knuderne beskrives: som hhv. *klasser af entiteter* (i ontologier) og *mængder af træk* (i begrebssystemer), hvilket imidlertid må siges at være to sider af samme sag (Madsen 1994 a: 99). Der er dog også en anden forskel, nemlig den, at en ontologi evt. indeholder nogle knuder, som ikke umiddelbart ville forekomme i et begrebssystem. Dette gælder f.eks. Chat-ontologien i **figur 3**, jf. f.eks. 'Line' og 'Point'. Man kunne dog evt. forestille sig, at disse ville danne grundlag for nogle inddelingskriterier i et begrebssystem.

I Madsen (1994 b: 86) redegøres der for, hvorledes inddelingskriterier i et begrebssystem evt. vil blive indført som selvstændige knuder i et *semantisk net*⁴, udarbejdet m.h.p. oprettelsen af en vidensbase. Det drejer sig om et net for aktietyper, udarbejdet af Jensen & Worsøe (1993), hvori der er indført nogle knuder, f.eks. *aktie mht. overdragelse*, *aktie mht. udbytte*, *aktie mht. stemmeret* og *aktie mht. aktiekapitalforhøjelse*, som er baseret på nogle inddelingskriterier i et tilsvarende begrebssystem for aktier: OVERDRAGELSE, UDBYTTE, STEMMERET og AKTIEKAPITALFORHØJELSE (Madsen (1994b: 80). Knuderne i nettet (*aktie mht. overdragelse* osv.) svarer altså ikke til egentlige begreber i det pågældende begrebssystem, men hvis man baserer en vidensbase på dette net, er det muligt at stille forespørgsler, f.eks. om en *fondsaktie* udstedes ved *aktiekapitalforhøjelse*, eller om en *A-aktie* er en aktie, som giver særlige rettigheder til *udbytte* eller *stemmeret*.

Arbejdet med flere ontologier vil vise, om denne type knuder, som ikke repræsenterer egentlige begreber i et begrebssystem, hyppigt vil forekomme i domænespecifikke ontologier.

⁴ Semantiske net og ontologier har mange lighedspunkter. Semantiske net indeholder mange forskellige relationer mellem knuderne, hvor ontologier tidligere typisk indeholdt generiske relationer, men denne forskel eksisterer ikke længere, idet ontologier kan indeholde andre relationer. Semantiske net og ontologier behandles derfor i denne sammenhæng under ét.

5. Forskellen på generiske og ontologiske relationer

Inden for den klassiske terminologilære inddeles begrebsrelationer i logiske og ontologiske begrebsrelationer, jf. Nuopponen (1994: 67–71, 1997: 142–160). Ifølge Wüster (1985: 9–13) er den *logiske begrebsrelation* baseret på lighed mellem begreber, medens den *ontologiske begrebsrelation* er baseret på en relation mellem de entiteter, som begreberne refererer til. Ontologiske begrebsrelationer er f.eks. *del-helheds-relationer*, *temporale relationer* og *kausale relationer*.

Logiske relationer er baseret på den logiske inklusionsrelation (Dahlberg 1985), og kaldes også generiske relationer. Sidstnævnte benævnelse foretrækkes her, da der strengt taget findes andre logiske relationer (identitet, eksklusion, intersektion og negation). Det der karakteriserer et underbegreb, som står i en generisk relation til et overbegreb, er, at underbegrebet arver alle overbegrebets karakteristiske træk, og derudover besidder mindst et yderligere træk, som adskiller det fra dets sideordnede begreb(er). Ud fra et intensionelt synspunkt udgør overbegrebets karakteristiske træk altså en delmængde af underbegrebets karakteristiske træk. Ud fra et ekstensionelt synspunkt udgør referenterne i underbegrebets ekstension en delmængde af referenterne i overbegrebets ekstension, jf. f.eks. også Andersen (1992: 38), Madsen (1994 a: 99). Hvis man i et begrebssystem for hvert begreb anfører en trækstruktur, som i **figur 5**, ses dette forhold mellem over- og underordnede begreber tydeligt.

Generiske relationer svarer til hyponymi-relationer i leksikalsk semantik (Ruus 1988: 118). I leksikalsk semantik opereres også med *taksonomier* (Cruse 1986) og *typehierarkier* (Copestake, Sanfilippo, Briscoe & de Paiva 1993). Disse omfatter kun generiske relationer, og i taksonomier må et begreb kun forekomme ét sted i hierarkiet og må kun have ét overbegreb.

Mellem begreber, som indgår i *ontologiske relationer*, består ikke et nedarvningsforhold. Et delbegreb har ikke træk fælles med helhedsbegrebet og to begreber i et temporalt forhold har heller ikke nødvendigvis fælles træk.

To sideordnede delbegreber har det tilfælles, at de entiteter, som de refererer til, begge indgår i den entitet, som helhedsbegrebet referer til. I del-helheds-relationer er der tale om en samling af delbegrebernes karakteristiske træk i helhedsbegrebet. Del-helheds-relationen svarer til meronymi-relationer i leksikalsk semantik (Ruus 1988: 118). Ruus (1995: 131) siger, at der foregår en egenskabstransport fra meronymet (delbegrebet) til holonymet (helhedsbegrebet): en *hoveddør* har fra meronymet *brevsprække* den egenskab 'at man kan putte breve igennem den', og fra meronymet *lås* har den egenskaben 'at den kan låses'. Ofte bruges *egenskab* og *karakteristisk træk* som synonyme, men f.eks. Madsen (1991: 71) og Nuopponen (1994: 61) skelner mellem disse: en *egenskab* er på 'entitetsniveau', medens et *karakteristisk træk* er på 'begrebsniveau', og et karakteristisk træk er i sig selv et begreb.

I temporale relationer har de enkelte begreber det til fælles, at de indgår som delbegreber i det samlede forløb, som i denne sammenhæng kan betragtes som en helhed (Madsen 1996: 5).

Det kan umiddelbart forekomme forvirrende, at en *ontologi* typisk er bygget op omkring *generiske* relationer, og ikke *ontologiske* relationer. Betegnelsen ontologi hænger dog sammen med, at en ontologi beskriver det værende, dvs. (dele af) verden, og den har ikke noget at gøre med de relationsarter, som anvendes i denne beskrivelse.

6. Kan man drage fordel af generelle ontologier i forbindelse med opbygning af begrebssystemer og datamatiske systemer?

Når man i forbindelse med et terminologiprojekt udarbejder definitioner, anvender man den klassiske definitionsmetode, analysemetoden, til at udforme en indholdsdefinition, ifølge hvilken der anføres *nærmeste overbegreb* og *adskillende træk*. Denne metode kan dog ikke uden videre anvendes, når der er tale om et begrebssystem, som er baseret på ontologiske relationer, f.eks. del-helheds-relationer. I dette tilfælde kan man enten vælge at indlede definitionen med at fastslå, at delbegrebet repræsenterer en del af den entitet, som helhedsbegrebet referer til, eller man kan vælge at anvende et mere generelt overordnet begreb (Madsen 1999). Et eksempel er begrebet *bremse*, som kan defineres på to måder:

'del af cykel, som har til formål at nedsætte hastigheden'

'apparat, som har til formål at nedsætte hastigheden'.

Et andet tilfælde, hvor man ikke kan referere til et direkte overordnet begreb i en generisk relation, er ved det mest overordnede begreb i et begrebssystem, 'topbegrebet'. Det er jo netop karakteristisk for begrebssystemer, at de oftest starter med et begreb, som hører til i det specifikke emneområde, og derfor er der ikke umiddelbart et overbegreb, som kan indsættes i definitionen. Også i dette tilfælde kan man anvende et mere generelt begreb. Frandsen (1982: 118–120) nævner, at man inden for tekniske områder af og til er henvist til at bruge et generelt begreb, som f.eks. *anlæg*, og at dette ikke behøver at blive defineret, da det er alment kendt. Frandsen (1982: 28) anfører at en variant af analysemetoden består i, at man bruger et formelt kategoribegreb svarende til en af Aristoteles' kategorier, jf. Dahlberg (1976: 102). Se også Dahlberg (1978: 144f, 1981: 19f). I sådanne tilfælde kunne man forestille sig, at man kunne drage nytte af en generel ontologi, hvor man kunne bruge nogle af de overordnede kategorier som nærmeste overbegreb i definitioner.

I forbindelse med arbejdet inden for området 'medicinsk informatik' inden for den europæiske standardiseringsorganisation, CEN, har en arbejdsgruppe udarbejdet en standard, CEN prENV 12611 (1996), *Medical informatics – Categorial structure of systems of concepts – Medical Devices*, som foreskriver udarbejdelsen af en 'kategoristruktur' inden for et givent emneområde, som kan bruges ved udarbejdelse af definitioner inden for det pågældende område. Kategoristrukturen i *Medical devices* svarer til en slags ontologi, som omfatter en systematisering af både *generelle, overordnede begreber* ('base concepts'), f.eks. 'equipment', og *domænespecifikke begreber* ('target concepts'), f.eks. 'airway temperature monitor', samt *trækverdier* ('associated concepts'), f.eks. 'monitoring', 'controlling' og 'transporting'. Trækverdierne, som anvendes til at beskrive 'target concepts', indgår i kategoristrukturen som begreber. Madsen (1998) og Thomsen (1998 b) viser, hvorledes standardens begreber og trækverdier kan indkodes i et typehierarki i systemet Lexical Knowledge Base (LKB), jf. Copestake (1993). I LKB indkodes trækverdierne ligesom de generelle og de specifikke begreber som typer, dvs. som selvstændige begreber.

De *overordnede begreber* i CEN prENV 12611 (1996) omfatter kun begreber, som er relevante inden for medicinsk udstyr. Topbegrebet er 'device', dvs. at der er ikke tale om, at hierarkiet starter med 'being', 'entity' eller 'thing', som i top-niveau-ontologierne. Om det, f.eks. i forbindelse med natursprogsanalyse vil være hensigtsmæssigt at anvende en almen, top-niveau-ontologi i kombination med en domænespecifik ontologi, vil være genstand for nærmere undersøgelser i et kommende forskningsprojekt.

7. Hvad er forskellen på hhv. datamodellering og begrebsstrukturering?

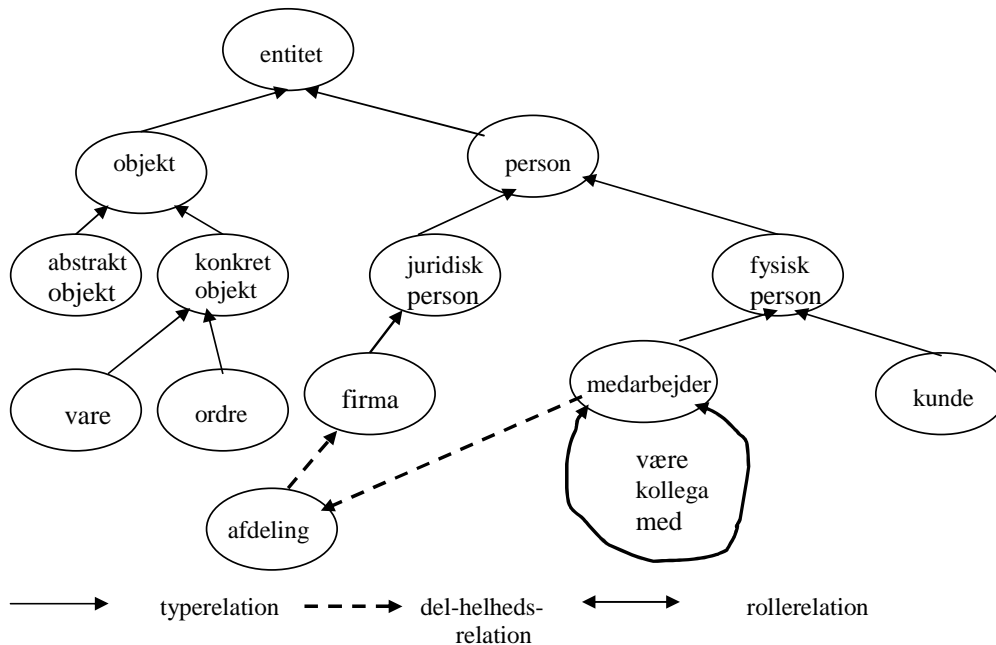
I mange sammenhænge anføres det, at en ontologi kan sammenlignes med en datamodel, som bruges til udarbejdelse af datastrukturen i en database, jf. f.eks. Guarino (1998: 14). Igennem de seneste år er der da også blevet udarbejdet domæne- og opgavespecifikke ontologier og semantiske net som grundlag for udvikling af vidensbaser og informations-systemer, som f.eks. vidensrepræsentationssystemet KL-ONE, se f.eks. Ahrenberg (1992) og Brachmann (1985).

Som nævnt i indledningen er grænserne mellem datamodellering og begrebsstrukturering i dag ikke så klare. Principper, som bruges i datamodeller, kan også anvendes i begrebssystemer. F.eks. kan oplysninger om relationsgrad og fakultativitet, som findes i entitets/rerelationsdiagrammer, med fordel anvendes i begrebssystemer, da det kan være relevant at vide, om f.eks. en maskine har én eller flere komponenter af en bestemt slags, og om en eller flere komponenter evt. kan udelades, jf. Madsen (1994 a:105).

Her skal argumenteres for den opfattelse, at selv om der ganske vist er mange lighedspunkter mellem begrebs- og datamodellering, er der dog også forskelle, som primært er relateret til *formålet* med anvendelsen af de forskellige modellerings- eller struktur-beskrivelsesværktøjer og den måde de beskrives på. I Madsen (1994 a: 84) er der redegjort for et entitets/rerelationsdiagram til en database med oplysninger om en virksomhed. I dette diagram indgår nogle entitetstyper som f.eks. AFDELING, MEDARBEJDER, VARE, KUNDE og ORDRE. Disse entitetstyper genfindes direkte som tabeller i databasen. Ved arbejdet med en natursprogsgrænseflade til databasen (Madsen, Stahél og Vikner 1997) viste det sig, at disse entitetstyper (tabeller) er tilstrækkelige til, at man vha. grænsefladen

kan formulere spørgsmål om f.eks. hvilken afdeling en medarbejder er ansat i, eller hvem der leder salgsafdelingen, da betydningen af substantiver og verber repræsenteres i grænsefladens leksikon af semantiske prædikater, som refererer til tabeller og kolonner, jf. Madsen og Vikner (1994).

Imidlertid kunne man ikke vha. grænsefladen formulere følgende spørgsmål: 'Hvem er kollega til Jørgen Hansen' eller 'Hvilke varer forhandler firmaet?' I det første spørgsmål skal grænsefladen vide, hvad *kollega* betyder, og i det andet spørgsmål skal det være kendt, at *firmaet* udgøres af afdelinger. Da der ikke eksisterede entitetstyper (tabeller) svarende til *kollega* og *firma*, var der heller ikke defineret prædikater for disse begreber. Der blev derfor udarbejdet et udkast til en domænemodel i form af et semantisk net, som skulle danne baggrund for udarbejdelse af flere prædikater. I **figur 6** vises et udsnit af dette net, som dog ikke er implementeret i grænsefladen.



Figur 6. Udsnit af semantisk net

Det semantiske net omfatter altså de nævnte begreber, som ikke findes i entitets/re-lationsdiagrammet, og derudover er der indført nogle få generelle begreber, som danner

top-niveauet i nettet. Der er dog kun indført sådanne begreber, som ville være nødvendige af hensyn til grænsefladen.

I forbindelse med udarbejdelse af modeller til elektroniske patientjournaler er spørgsmålet om forskelle og ligheder mellem begrebssystemer og datamodeller blevet meget aktuelt. I et *begrebssystem*, udarbejdet som grundlag for udvikling af en datamodel i forbindelse med elektronisk registrering af oplysninger om patienter, vil der typisk findes et *begreb*, som f.eks. *sygdom* med en række underbegreber, svarende til forskellige typer af sygdomme. I en *datamodel til en relationel database* vil der også findes en *entitetstype sygdom*, men de enkelte sygdomstyper vil ikke findes i modellen. I en database vil disse tilsammen udgøre det *domæne*, som er knyttet til et attribut (et felt) SYGDOM. Dette svarer til den opsplitning i **begrebssystem**, **model for journalsystem** og **patientdata i journalsystemet**, som findes i en svensk rapport (Hedin et al 1998: 92). Hvis der imidlertid ikke er tale om en datamodel til en relationel database, men f.eks. en datamodel til en vidensbase, vil der være større sammenfald mellem begrebs- og datamodel, og dette sammenfald er måske netop det, der kommer til udtryk i at ontologier (og semantiske net), anvendes ved både begrebs- og datamodellering. Dette spørgsmål vil være genstand for nærmere udforskning i forbindelse med udarbejdelse af datamodeller i forbindelse med udvikling af elektroniske patientjournaliseringssystemer.

8. Er det muligt at udarbejde en fælles ontologi for beslægtede områder?

I takt med det øgede behov for strukturering af viden og udarbejdelse af definitioner til brug for datamatiske systemer stiger behovet for afklaring og definerings af en lang række grundlæggende begreber, da kommunikationen mellem forskellige grupper af fagfolk ellers i stigende grad hæmmes. I foreningen Lykeion arbejder bl.a. en række danske forskningsinstitutioner og store virksomheder, specielt rådgivende ingeniørvirksomheder, på metoder og værktøjer til forbedring af kommunikation mellem forskellige faggrupper. Ved gennemførelse af store projekter (f.eks. bygning af et hospital eller en bro) viser det sig ofte, at de forskellige parter, som er impliceret i et projekt, har svært ved at forstå hinanden. Dette skyldes, at et og samme udtryk ofte tillægges forskellige betydninger, og at der samtidig anvendes forskellige udtryk om samme begreber. Som eksempler kan blot nævnes nogle

gængse udtryk som *model*, *system*, *attribut*. Dvs. at der er brug for klare definitioner af en lang række begreber, som anvendes i forbindelse med *analyse*, *design*, *konstruktion* og *anvendelse af projektstyringsystemer*. I Lykeion diskuteres det, om det vil være muligt og hensigtsmæssigt at udarbejde et sæt af fælles grundbegreber med en generel definition for hvert begreb, dvs. en fælles ontologi. Det er dog et spørgsmål, om der ikke snarere er behov for en eller flere *domænespecifikke ontologier*. Hvis det rent faktisk kan konstateres, at der er tale om flere forskellige domæner, vil det formodentlig være en mere farbar vej, at udarbejde specifikke definitioner for begreberne inden for de enkelte (del)domæner, med henvisninger til forskelle og ligheder inden for disse områder.

9. Sammenfatning

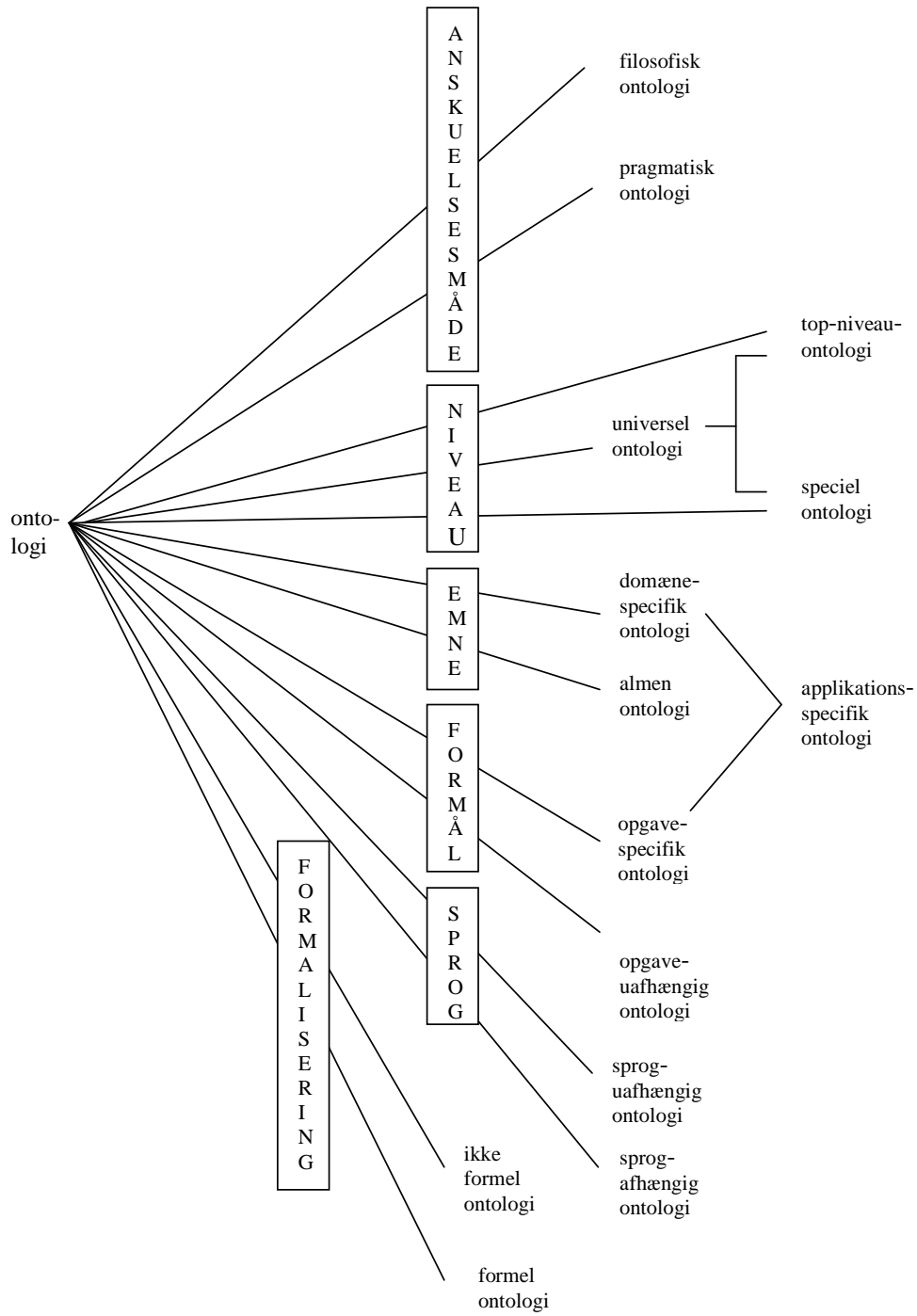
I dette indlæg er langt fra alting sat på plads. Der er opstillet en foreløbig inddeling af ontologier, som der skal arbejdes videre med, og der er redegjort for ligheder mellem begrebssystemer og ontologier (og lignende strukturbeskrivelser, f.eks. semantiske net), men der er også gjort opmærksom på forskelle, som skal udforskes nærmere. Generiske og ontologiske begrebsrelationer er diskuteret specielt i forhold til ontologier, og ligeledes nytten af at anvende generelle ontologier i forbindelse med begrebssystematisering og data-modellering. Det konkluderes, at der er forskel på begrebs- og datamodellering, men i denne sammenhæng udgør ontologier (og semantiske net) måske en slags hybrid form, idet de har træk af både begrebs- og datamodellering. Endelig sættes der spørgsmålstegn ved, om det vil være muligt at udarbejde en fælles ontologi for en række grundbegreber, som anvendes i forbindelse med *analyse*, *design*, *konstruktion* og *anvendelse af projektstyrings-systemer*.

Referencer

- Ahrenberg, Lars (1992). Kunskapsrepresentation – ett datalingvistisk perspektiv. I: *Terminologi, edb & vidensteknik, 2. nordiske symposium*, 78–103. Ed. Bertha Toft. Varde: Handelshøjskole Syd.
- Andersen, Harry (1945). *Dansk Begrebsordbog*. Paa Grundlag af Viggo Petersens efterladte Manuskript. København: Munksgaard.
- Andersen Øivin (1992). Begrepshierarkiers logiske og empiriske egenskaper. In: *Terminologi, edb & vidensteknik, 2. nordiske symposium*, 34–54. Ed. Bertha Toft. Varde: Handelshøjskole Syd.
- Arntz, Reiner & Heribert Picht (1989). *Einführung in die Terminologearbeit*. Studien zu Sprache und Technik, Band 2, Hildesheim: Georg Olms Verlag.
- Brachmann, Ronald J. & James G. Schmolze (1985). An overview of the KL-ONE Knowledge Representations System. *Cognitive Science*, 9 (2), 171–216.
- Brentano, Franz (1862). *Von der mannigfachen Bedeutung des Seienden nach Aristoteles*, translated as *On the Several Senses of Being in Aristotle*, Berkeley: University of California Press, 1975. Citeret fra Sowa 1997.
- BS 1000M: Part 1 (1985). *Universal Decimal Classification*. International Medium Edition, English text, FID publication no. 571.
- Carpenter, Bob (1992). *The Logic of Typed Feature Structures*. Cambridge, Mass.: Cambridge University Press.
- CEN prENV 12611 (1996). *Medical Informatics – Categorial Structure of Systems of Concepts – Medical Devices*. European prestandard. Brussels: European Committee for Standardization, Oct. 1996.
- Coad, P. & E. Yourdon (1991). *Object-Oriented Analysis*. Prentice Hall.
- Copestake, Ann (1993). *The Compleat LKB*. Technical Report No. 316, University of Cambridge Computer Laboratory.
- Copestake, Ann, Antonio Sanfilippo, Ted Briscoe & Valeria de Paiva (1993). The ACQUILEX LKB: An Introduction. In: *Inheritance, Defaults and the Lexicon*, 143–163. Eds T Briscoe, V de Paiva & A Copestake. Cambridge: Cambridge University Press.
- Cruse, D.A. (1986). *Lexical Semantics*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Dahlberg, Ingetraut (1976). Über Gegenstände, Begriffe, Definitionen und Benennungen. Zur möglichen Neufassung von DIN 2330. *Muttersprache* 2, 81–117.
- Dahlberg, Ingetraut (1978). A referent-oriented, analytical concept theory for INTERCONCEPT. In: *International Classification* 5 No. 3, 142–151.
- Dahlberg, Ingetraut (1981). Conceptual Definitions for INTERCONCEPT. In: *International Classification* 8: 1, 16–22.

- Dahlberg, Ingetraut (1985). Begriffsbeziehungen und Definitionstheorie. In: *Terminologi und benachbarte Gebiete: 1965–1985*, 137–148. Wien: Infoterm.
- Date, C J (1990). *An Introduction to Database Systems*, Volume I, Fifth Edition. Reading: Addison-Wesley Publishing Company.
- Erikson, Hans-Erik & Magnus Penker (1998). *UML Toolkit*. New York: Wiley.
- Frandsen, Lene (1982). *Definition. Objekter, metoder og regler*. ARK 15, Sprog-institutternes Arbejdsrapport. København: Handelshøjskolen i København.
- Guarino, Nicola (1998). Formal Ontology and Information Systems. In: *Formal Ontology in Information Systems, Proceedings of the First International Conference (FOIS'98)*, 3–15. Ed. Nicola Guarino. Amsterdam: IOS Press.
- Hedin, Anita, Lena Jernberg, H. C. Lennér, Torsten Lundmark & Sven-Bertil Wallin (1998). *Hälso- och sjukvårdens begreppssystem. Stömfåra, kontrapunkt eller blå dunster?* Skåne Term, Region Skåne.
- Jensen Pia Dorrit & Joan Worsøe (1993). *Semantisk netværk og rammer i et vidensbaseret system*, Speciale i datalingvistik, Institut for Datalingvistik, Handelshøjskolen i København.
- Lenat, Douglas B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38:11, 33–38. Citeret fra (Sowa 1997).
- Madsen, Bodil Nistrup (1991). In terms of concepts. *Copenhagen Studies of Language* 14, 67–91.
- Madsen, Bodil Nistrup (1994 a). Begrebssystemer og vidensmodellering. I: *Fagsproglig kommunikation. Jubilæumsskrift Bind 3*, 82–106. Ed. Annelise Grinstead. Kolding: Handelshøjskole Syd, Samfundslitteratur.
- Madsen, Bodil Nistrup (1994 b). Begrebssystemer og vidensmodellering. I: *ARK 76, LEKSIKON OG VERDEN, Sprog-institutternes Arbejdsrapport*, 71–91. Ed. Finn Sørensen. København: Handelshøjskolen i København.
- Madsen, Bodil Nistrup (1996). Ikke-generiske begrebsrelationer. *LAMBDA* Nr. 21, 1–13. Handelshøjskolen i København.
- Madsen, Bodil Nistrup (1998). Typed feature structures for terminology work – Part I. In: *LSP – Identity and Interface – Research, Knowledge and Society*. Proceedings of the 11th European Symposium on Language for Special Purposes, 339–348. Ed. L. Lundquist, H. Picht & J. Qvistgaard. Copenhagen Business School.
- Madsen, Bodil Nistrup (1999 b). *Terminologi, Principper og metoder. Terminologi 2 Øvelser og eksempler*. København: Gads Forlag.
- Madsen, Bodil Nistrup, Annie Stahél & Carl Vikner (1997). *A Database Query Generator. Conversion of Natural Language Semantic Representations into Database Queries*. Department of Computational Linguistics, Copenhagen Business School.
- Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner (1999). The project 'Computer-Aided Ontology Structuring' (CAOS). Udkommer i særnummer af *Copenhagen Studies in Language*, rapport om projekt OMNIS.

- Madsen, Bodil Nistrup & Carl Vikner (1994). Leksikalsk flertydighed ved forespørgsler til en database. I: *ARK 76, LEKSIKON OG VERDEN, Sproginstitutionernes Arbejdsrapport*, 93–108. Ed. F. Sørensen. København: Handelshøjskolen i København.
- Nuopponen, Anita (1994). Begreppssystem för terminologisk analys. *Acta Wasaensia*, No 38, Språkvetenskap 5. Vasa universitet.
- Nuopponen, Anita (1997). Begreppsrelationer och begreppssystem. I: *Terminologi som vetenskapsgren*, 142–160. Eds Laurén, Christer, Johan Myking & Heribert Picht. Lund: Studentlitteratur.
- Picht, Heribert (1997). Genstand og begreb. I: *Terminologi som vetenskapsgren*, 96–122. Eds Laurén, Christer, Johan Myking & Heribert Picht. Lund: Studentlitteratur.
- Ruus, Hanne (1988). General purpose semantics: Hyponymy and generic relationships. *Nordic Journal of Linguistics* 11, 111–128.
- Ruus, Hanne (1995). *Danske Kerneord. Centrale dele af den danske leksikalske norm*. København: Museum Tusulanums Forlag.
- Smith, Barry (1998). Basic concept of formal ontology. In: *Formal Ontology in Information Systems, Proceedings of the First International Conference (FOIS'98)*, 19–28. Ed. Nicola Guarino. Amsterdam: IOS Press.
- Soergel, Dagobert (1974). *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville Publishing Company.
- Sowa, John F. (1997). *Knowledge Representation. Logical, Philosophical and Computational Foundations*. Boston, Massachusetts: PWS Publishing Company. Draft of a book in preparation, udkommer maj 1999.
- Spang-Hanssen, Henning (1985). Terminologi og I&D. I: *Nordisk terminologikursus II*, del 2, 581–594. Handelshøjskolen i København.
- Thomsen, Hanne Erdman (1998 a). Feature specifications applied to the field of life insurance. In: *Terminology Science & Research – Journal of the International Institute for Terminology Research*, vol. 8 (1997), 1/2, 21–36. Wien: TermNet.
- Thomsen, Hanne Erdman (1998 b): Typed feature structures for terminology work – Part II. I: *LSP – Identity and Interface – Research, Knowledge and Society*. Proceedings of the 11th European Symposium on Language for Special Purposes, 349–359. Copenhagen Business School.
- Vickery, B C (1960). Thesaurus – A new word in documentation. *Journal of Documentation* 16: 4, 181–189.
- Vickery, B C (1997). Ontologies. In: *Journal of Information Science* 23: 4, 277–286.
- Warren, David H. D. & Fernando C. N. Pereira (1982). An efficient easily adaptable system for interpreting natural language queries. *American Journal of Computational Linguistics* 8: 3–4, 110–122.
- Wüster, Eugen (1985). *Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie*. The LSP Centre, The Copenhagen School of Economics.



APPENDIKS 1. Oversigt over typer af ontologier