

Når danske myndigheder taler forskelligt dansk: Begrebsafklaring med terminologiske ontologier

Bodil Nistrup Madsen, Copenhagen Business School
Hanne Erdman Thomsen, Copenhagen Business School

Abstract: Terminological Ontology combines traditional terminology work with typed feature structures, thus providing a systematic and controlled method for concept clarification and definition writing. Terminological ontologies include a formalization of the inheritance of characteristics and a number of constraints which help experts formulate consistent intensional definitions. In this paper we illustrate the usefulness of Terminological Ontology with several examples from a scenario where various Danish authorities need to exchange digitized information. The authorities involved include units under the Ministry of Justice (the Prison and Probation Service, the Police, the Prosecution Service in Denmark, and the Danish Court Administration), the National Board of Social Services, and the Danish Agency for Digitisation. The examples evolve around the central concept of ‘case’, and show how concept clarification and harmonization is necessary in order to enable a seamless exchange of digitized information between these authorities, and in essence, between organizations in general. On this basis, it is concluded that there is a need for a shared, cross-ministerial concept and knowledge database with appointed terminologists to collaborate with and coordinate contributions from the interested parties in ministries, public agencies and other authorities.

Keywords: Terminologi, terminologiharmonisering, begrebsafklaring, begrebssystem, ontologi, termbase, vidensbase, national termbank.

1. Introduktion

Hvornår er en sag en *sag*? Det er en række danske myndigheder ikke altid enige om. Selv inden for samme sprogområde kan folk, der kommer fra forskellige kulturer og faglige domæner, operere med vidt forskellige begrebsverdener. De taler med andre ord *ikke* ’samme sprog’. Artiklen giver en række autentiske eksempler på dette og viser, hvordan man gennem systematisk begrebsafklaring vha. såkaldte terminologiske ontologier kan bygge bro mellem forskellige begrebsverdener. De basale principper for terminologiske ontologier omtales i afsnit 2.

Eksemplerne stammer dels fra et samarbejde mellem fire myndigheder under Justitsministeriet: *Kriminalforsorgen*, *anklagemyndigheden*, *domstolene* og *Politiet*, som havde til formål at udvikle et fælles format til udveksling af digitale informationer (Hoffmann et al. 2012), dels fra det begrebsafklaringsarbejde, som udføres af hhv. *Digitaliseringsstyrelsen* (tidligere IT- og Telestyrelsen) og *Socialstyrelsen*. Arbejdet med begrebsafklaring i Justitsministeriets fire institutioner blev baseret på et arbejde, der oprindeligt var blevet påbegyndt i Kriminalforsorgen, som også internt havde oplevet problemer med at blive enige om betydningen af en række centrale begreber.

Samarbejdet mellem de fire myndigheder under Justitsministeriet afslørede, at der i mange tilfælde eksisterede vidt forskellige opfattelser af centrale begreber vedrørende *sag* og *sagsdokumenter*, hvilket ville have givet problemer i forbindelse med udarbejdelsen af et fælles udvekslingsformat. Digitaliseringsstyrelsen og Socialstyrelsen har også arbejdet med at fastlægge definitioner af *sag* og relaterede begreber. Eksemplerne i denne artikel viser med al tydelighed, at der er behov for fællesoffentlig begrebsafklaring.

De omtalte myndigheder er alle medlemmer af Forum for Videnmodellering i Offentligt Regi, FORVIR, jf. <http://blog.cbs.dk/forvir/>, som bl.a. arbejder på at fastlægge en række fællesoffentlige begreber. Inden for rammerne af FORVIR er der udarbejdet en topontologi¹ for generelle begreber,

1 Topontologien kan tilgås på <http://forvir.item.dk> (login og password: FORVIR).

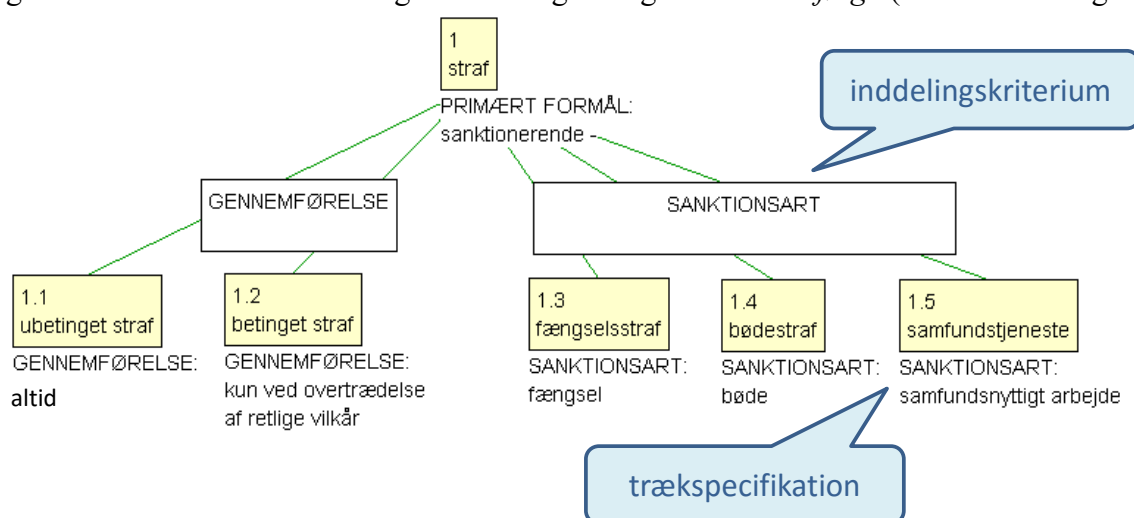
som for eksempel *aktivitet, hændelse, proces, redskab, aktør* og *information*. Topontologien beskrives kort i afsnit 2. Som det vil fremgå nedenfor, er der imidlertid et stort behov for at udvide dette arbejde til at omfatte mere specifikke, tværoffentlige begreber.

2. Begrebsafklaring vha. terminologiske ontologier

2.1. Basale principper for terminologiske ontologier

Til arbejdet i Kriminalforsorgen blev der nedsat en række arbejdsgrupper med deltagere fra forskellige afdelinger og medarbejdere fra DANTERMcentret, Copenhagen Business School, og indledningsvist blev der udarbejdet en terminologisk ontologi *Retsfølge* med informationer om forskellige straf typer, som fungerede som model for arbejdsgrupperne (Kriminalforsorgen n.d.). Et udsnit er vist i diagrammet i Figur 1. Vi vil anvende dette til at illustrere nogle af de basale principper for terminologiske ontologier.

Figur 1: Udsnit af den terminologiske ontologi i diagrammet *Retsfølge* (Kriminalforsorgen n.d.)

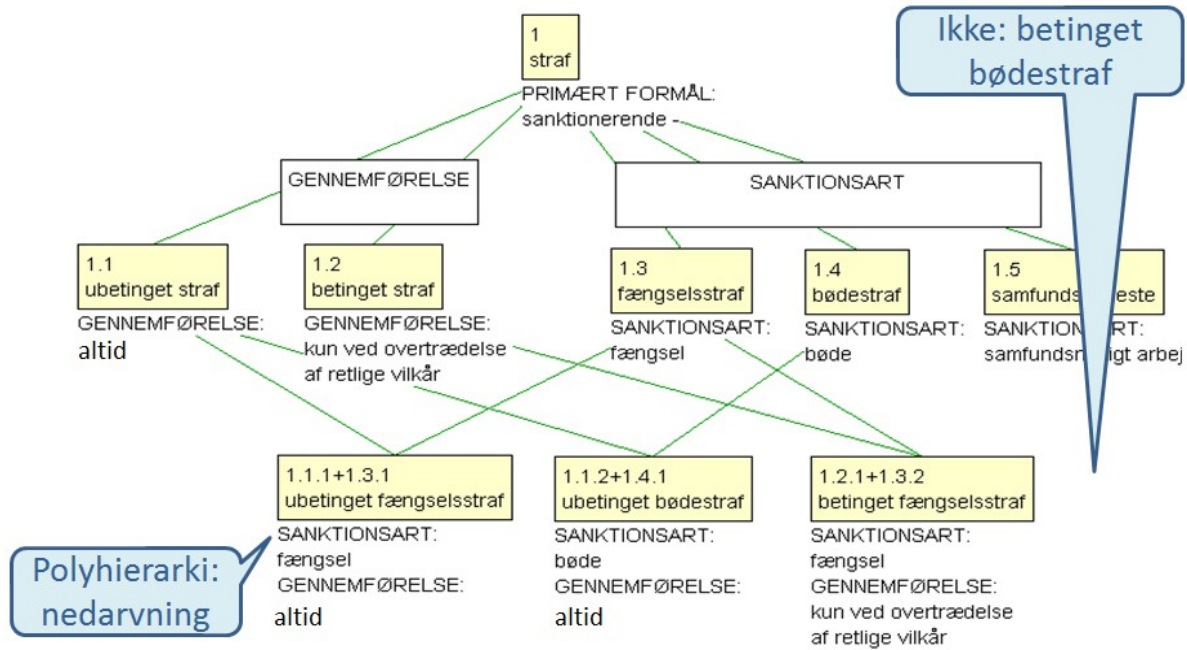


Principperne for terminologiske ontologier er baseret på principperne for terminologiske begrebssystemer, jf. ISO 704 (2009), og er udviklet af en gruppe forskere ved Copenhagen Business School (Madsen et al. 2004, Madsen & Thomsen, 2006, 2009, 2015a). I terminologiske ontologier er begrebernes karakteristiske træk anført i form af formelle trækspecifikationer (Madsen 1998: 339-348; Thomsen 1997: 21-36; Thomsen 1998: 349-359), svarende til attribut-værdi-par i “Typed Feature Theory” (Carpenter 1992). Dette er illustreret i Figur 1, hvor de farvede bokse repræsenterer begreber, og de karakteristiske træk er vist under begreberne. De grønne linjer mellem begreberne repræsenterer typerelationer. Andre relationsarter kan anvendes og vises ved hjælp af andre linjetyper. De hvide bokse med store bogstaver repræsenterer inddelingskriterier. I det følgende introduceres de vigtigste principper.

Ifølge terminologisk teori (ISO 704 2009, Madsen 1999, m.fl.) nedarves et karakteristisk træk til dets underbegreber, og det samme er tilfældet med trækspecifikationer i Carpenters (1992) teori, så princippet om nedarvning af karakteristiske træk realiseres uden videre ved at anvende “Typed Feature Theory”. Nedarvning er ikke vist i Figur 1, men ses i Figur 2, hvor for eksempel underbegrebet *ubetinget fængselsstraf* arver trækket *SANKTIONSART: fængsel* fra overbegrebet *fængselsstraf* og trækket *GENNEMFØRELSE: altid* fra overbegrebet *ubetinget straf*. Et andet princip er, at et overbegreb kan have mere end et underbegreb. Dette omtales som polyhierarki og vises i den terminologiske ontologi i Figur 2, som også illustrerer, at ikke alle begreber under de to inddelingskriterier *GENNEMFØRELSE* og *SANKTIONSART* kombineres i det danske retssystem.

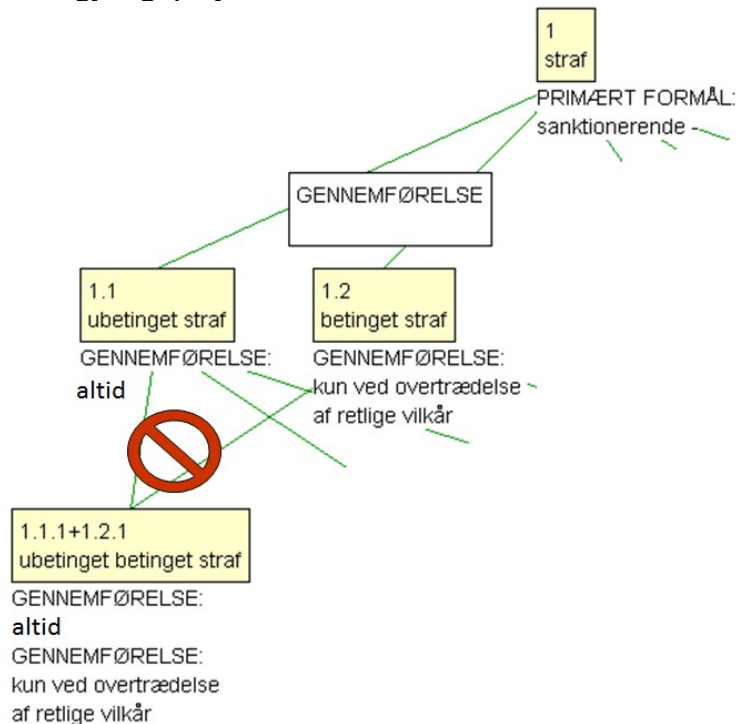
For eksempel eksisterer begrebet *betinget bødestraf* ikke.

Figur 2: Polyhierarki med mere viden om straf typer



For at undgå konflikt siger et tredje princip, at en attribut på et begreb kun kan have én værdi. I Figur 3 illustreres et ugyldigt polyhierarki, med et fiktivt begreb *ubetinget betinget straf*, hvor attributten *GENNEMFØRELSE* har to værdier.

Figur 3: Ugyldigt polyhierarki med konflikt mellem attributværdier

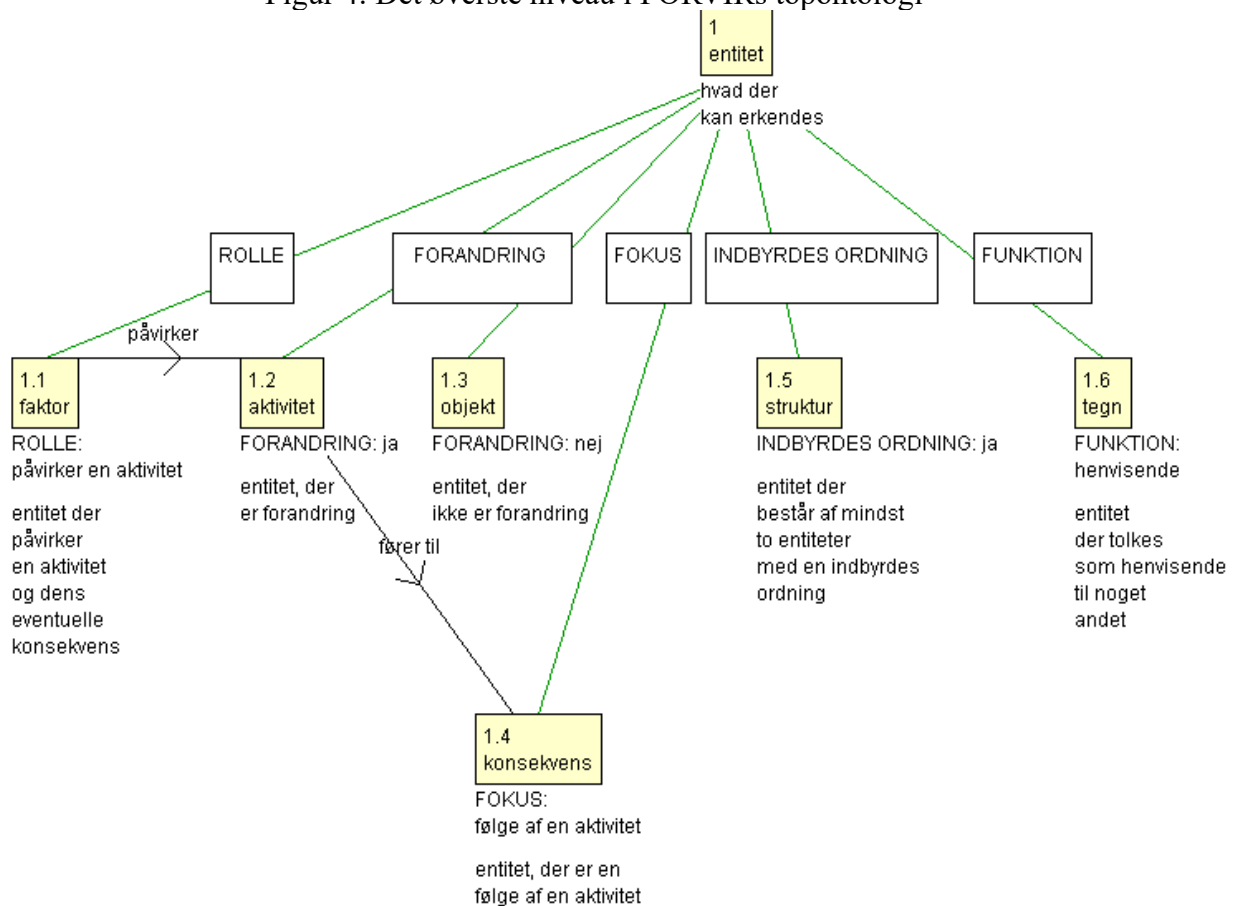


På basis af terminologiske ontologier er det nemt at udarbejde intensionale definitioner, dvs. definitioner, hvor der først anføres nærmeste overbegreb og dernæst det træk, som adskiller et begreb fra dets sideordnede begreber. Den intensionale definition tilskrives den græske filosof Aristoteles (384 – 322 f.v.t), den anbefales af både leksikografer og terminologer og er foreskrevet i internationale standarder, jf. ISO 704 (2009). I Thomsen (2016) gøres der nærmere rede for sammenhængen mellem terminologiske ontologier og intensionale definitioner, og det illustreres hvordan terminologiske ontologier er et godt grundlag for at udfærdige korte, klare og konsistente definitioner, som for eksempel definitionen af betinget straf: ”straf der kun skal gennemføres hvis retlige vilkår overtrædes” (Kriminalforsorgen n.d.).

2.2. FORVIRs topontologi

I afsnit 2.1 Basale principper vises nogle udsnit af terminologiske ontologier, hvori der udelukkende indgår typerelationer (grønne, skrå linjer), der danner hierarkier. I Figur 4 vises det øverste niveau i FORVIRs topontologi.

Figur 4: Det øverste niveau i FORVIRs topontologi



Her ses en anden relationsart, den associative, eksemplificeret ved hjælp af hhv. *påvirker* og *fører til*. En associativ relation repræsenteres af en sort linje med relationsnavn og en pil, der viser relationens retning.

Ved udarbejdelsen af topontologien blev der hentet inspiration i projektet Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE n.d.), især mht. inddelingen i hhv. *aktivitet* og *objekt*. I en kommentar i topontologibasen forklares denne inddeling nærmere: Udgangspunktet for definitionerne er, at *aktivitet* er ”noget der sker” og *objekt* er ”noget der er”.

Aktivitet er defineret med inspiration fra begrebet *perdurant*:

Also known as occurrent, accident or happening. Perdurants are those entities for which only a part exists if we look at them at any given snapshot in time. When we freeze time we can only see a part of the perdurant. Perdurants are often what we know as processes, for example 'running'.

Definitionen af *objekt* er inspireret af forklaringen af begrebet *endurant*:

Also known as continuant, or in some cases 'substance'. Endurants are those entities that can be observed/perceived as a complete concept, at no matter which given snapshot of time. Were we to freeze time we would still be able to perceive/conceive the entire endurant. Examples are material objects, such as an apple or a human, and abstract 'fiat' objects, such as an organisation or the border of a country.

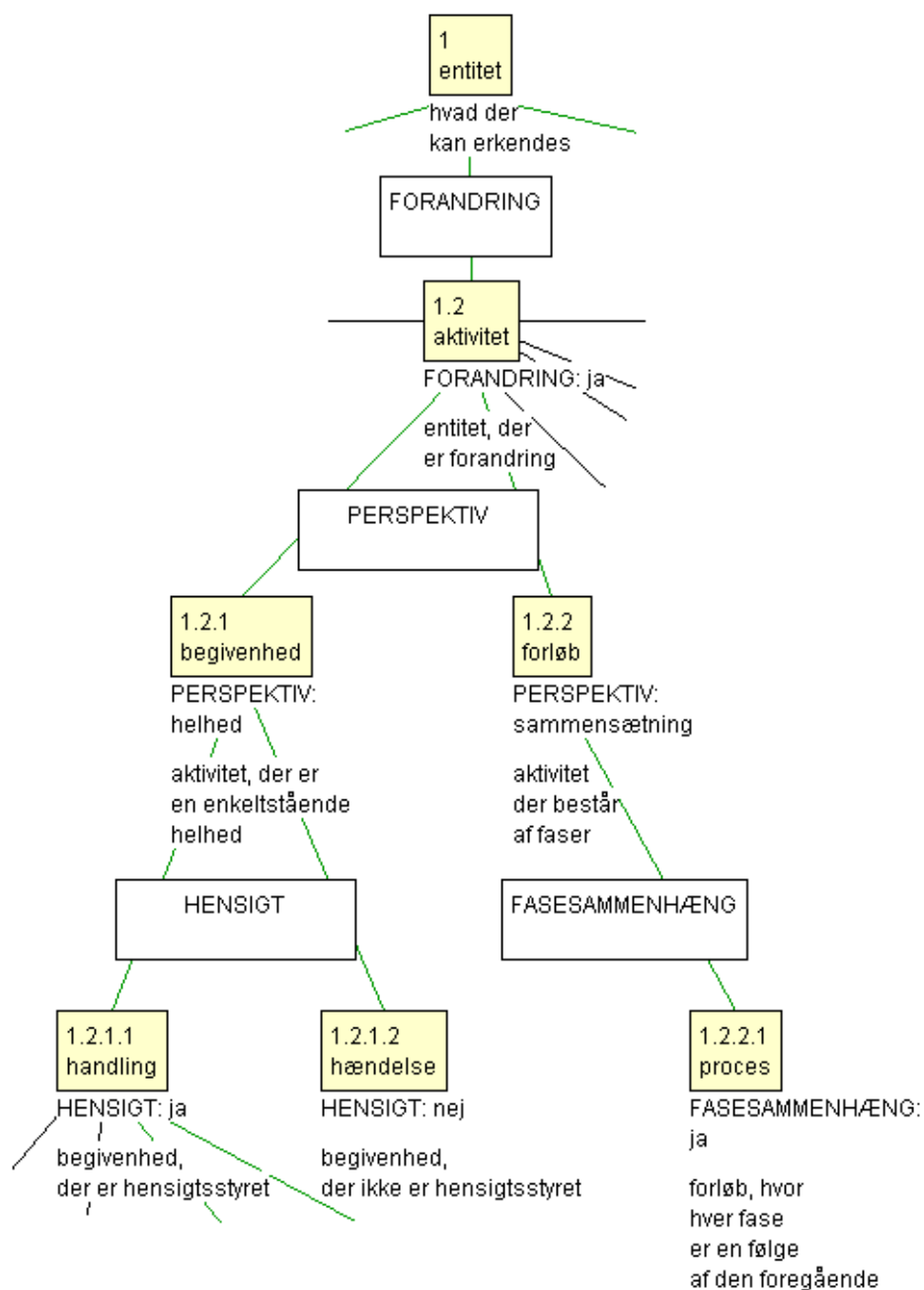
I Figur 5 vises et udsnit af topontologien med forskellige underbegreber til begrebet *aktivitet*. Det er svært at give en præcis faglig definition af disse begreber, da udtrykkene også bruges i almindeligt sprog. Men det bemærkes, at begreberne i denne sammenhæng betragtes som generelle, tværoffentlige begreber, hvis definitioner ikke altid svarer til almensprogets, som ofte er mere upræcise. Man kan i visse tilfælde ligefrem tale om homografer, dvs. udtryk som skrives ens, men har forskellige betydninger. I topontologien defineres *begivenhed* som: "aktivitet, der er en enkeltstående helhed", mens et *forløb* defineres som "aktivitet der består af faser". Den Danske Ordbog, som giver almensproglige definitioner, har følgende definition af *begivenhed*: "noget der sker eller finder sted, især noget vigtigt eller bemærkelsesværdigt". Den almensproglige betydning fokuserer ikke på perspektivet "helhed" (i modsætning til "sammensætning") og afviger altså i dette tilfælde fra den tværoffentlige.

De generelle, tværoffentlige begreber anvendes ofte, når forskellige arbejdsgrupper i de danske myndigheder udarbejder definitioner af mere specifikke begreber, og det er derfor vigtigt, at der er enighed om dem. For eksempel er det vigtigt at være opmærksom på forskellen mellem *handling* og *hændelse*, som begge er en type *begivenhed*, og som adskilles med hensyn til, om *begivenheden* er hensigtsstyret eller ej.

3. Begrebsafklaring i Justitsministeriets institutioner – begrebet *stævning*

Formålet med begrebsafklaringen i Justitsministeriet (Hoffmann et al. 2012) er "primært at forstå de centrale fælles begreber i de 4 myndigheder, synliggøre denne forståelse for alle (herunder it-leverandører) og dermed danne grundlaget for en fælles begrebskultur, entydighed ved dataudveksling og bedre datakvalitet" (Justitsministeriet n.d.). I forbindelse med udveksling af informationer mellem Justitsministeriets institutioner (Politiet, anklagemyndigheden, domstolene og Kriminalforsorgen) er der ca. 75 centrale begreber vedrørende *sager* og ca. 80 centrale begreber vedrørende *sagsdokumenter*. Et af dem er *stævning*, som vi vil bruge som eksempel i det følgende. I Figur 6 vises de forskellige opfattelser af *stævning* hos Politiet, anklagemyndigheden og domstolene.

Figur 5: Udsnit af FORVIRs topontologi – *aktivitet* og underbegreber



For det første ses det i Figur 6, at der er anvendt forskellige overbegreber: *henvisning*, *sagsdokument* og *indkaldelse*. *Stævning* har to betydninger, dels handlingen at stævne nogen, dels det dokument, som anvendes i forbindelse med denne handling. Ved fastlæggelse af de forskellige sagsdokumenter er det den sidste, der er den relevante, jf. anklagemyndighedens oprindelige definition. Af den fælles terminologiske ontologi, der er udarbejdet efter drøftelser mellem Justitsministeriets fire institutioner, fremgår det, at der er der ni forskellige hovedtyper af sagsdokumenter: *begærings-*, *erklærings-*, *hørings-*, *anmeldelses-*, *indkaldelses-*, *meddelelses-*, *efterforsknings-*, *sagsfremstillings-* og *retsafgørelsesdokument* (Justitsministeriet n.d.). En stævning er et *indkaldelsesdokument*, jf. Figur 7, som viser et udsnit af Justitsministeriets terminologiske

ontologi i diagrammet *JM-Sagsdokumenter der udveksles mellem Justitsministeriets myndigheder*². Begrebet *indkaldelsesdokument* er derfor det nærmeste overbegreb, som bør indlede definitionen af *stævning*, hvilket ikke er tilfældet i nogen af definitionerne i Figur 6. Et *indkaldelsesdokument* defineres i det fælles diagram i Figur 7 som et ”sagsdokument med besked om at give møde på et angivet sted og et angivet tidspunkt” (Justitsministeriet n.d.). Bemærk at den røde, knækkede linje repræsenterer en del-helhedsrelation mellem *sagsdokument* og *fysisk sag*.

Figur 6: Forskellige definitioner af begrebet *stævning*
(ikke publiceret arbejdsmateriale)

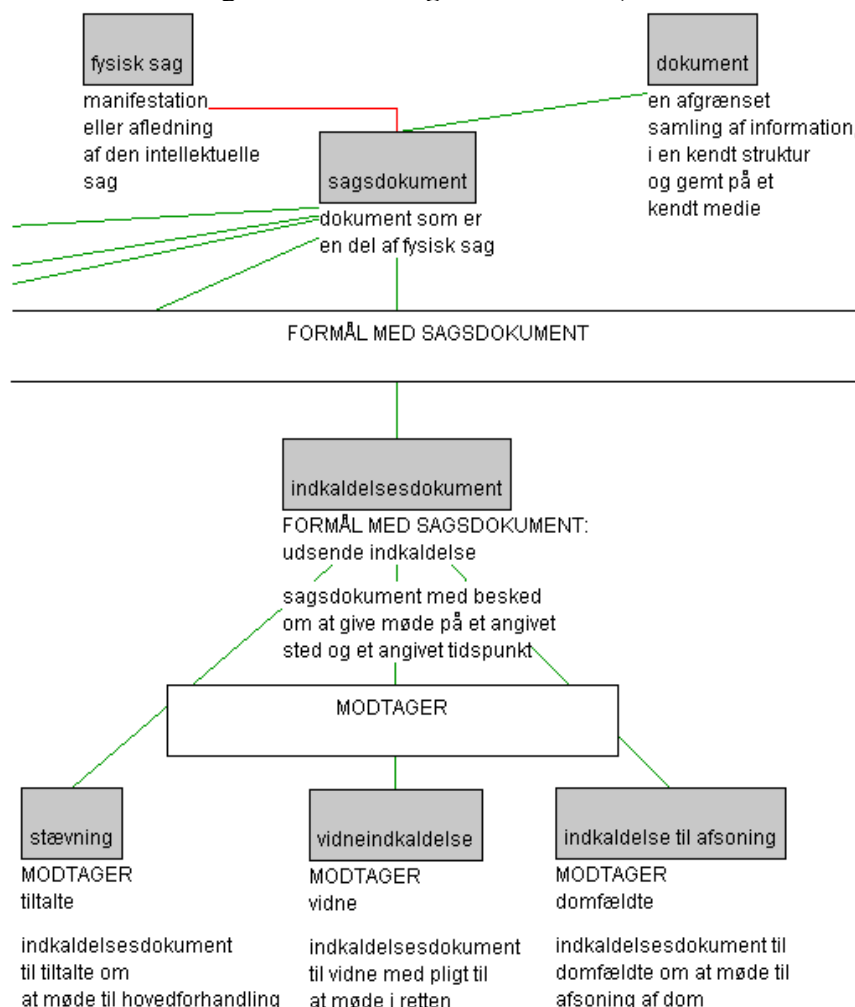
- Politiets oprindelige definition:
henvisning til anklageskrift
- Anklagemyndighedens oprindelige definition:
sagsdokument hvorved anklagemyndigheden
indkalder tiltalte til hovedforhandling
- Domstolenes oprindelige definition:
Anklagemyndighedens indkaldelse af tiltalte til
retsmøde

For det andet er der ikke enighed om det adskillende karakteristiske træk. Ved analyse af begreber kan man ofte identificere flere karakteristiske træk. I terminologiske ontologier udpeges ét adskillende træk, som skal indgå i definitionen, mens de øvrige træk er supplerende, og for eksempel kan medtages i en kommentar til begrebet.

Politiets definition adskiller sig fra de to øvrige ved at omtale begrebet *anklageskrift*. Et *anklageskrift* er ifølge Justitsministeriets diagram *JM-Sagsdokumenter* et ”sagsfremstillingsdokument med beskrivelse af tiltaleforhold” (Justitsministeriet n.d.). Et *sagsfremstillingsdokument* er et ”sagsdokument der indeholder kortfattet redegørelse for sagens faktiske omstændigheder” (Justitsministeriet n.d.). Det er rigtigt, at en *stævning* henviser til et *anklageskrift*, men det primære formål med en *stævning* er at indkalde tiltalte. At der henvises til *anklageskriftet* er således et supplerende træk, som ikke skal indgå i definitionen. Dette er klart i tilfældet her, men det er ikke altid indlysende, hvilket træk der er det adskillende.

Både anklagemyndighedens og domstolenes definitioner indeholder trækket: *indkalde tiltalte*, men til hhv. *hovedforhandling* og *retsmøde*. Ifølge Den Store Danske (n.d.) er *hovedforhandling* ”den afsluttende del af en retssags behandling ved domstolene, hvorefter der afsiges dom; indtil 1.1.2007 benævnt domsforhandling”, mens *retsmøde* er ”ethvert møde for en dommer i forbindelse med sagers efterforskning, forberedelse, domsforhandling, domsafsigelse mv.” Ved hjælp af en *stævning* indkaldes der ikke nødvendigvis til den afsluttende behandling, så *retsmøde* er mere korrekt, jf. domstolenes oprindelige definition. Dette underbygges af Retsplejeloven (Justitsministeriet 2014), §348: ”Sag anlægges ved indlevering af *stævning* til retten” samt af definitionen i Den Danske Ordbog betydning 2: ”indkalde eller få indkaldt nogen til at møde for retten som sagsøgt”. Retsplejeloven indeholder ikke en egentlig definition af *stævning* ud over den, som kan udledes af §348.

² I det følgende omtalt som *JM-Sagsdokumenter*.

Figur 7: Udsnit af diagrammet *JM-Sagsdokumenter* (Justitsministeriet n.d.)

I diagrammet *JM-Sagsdokumenter*, jf. Figur 7, findes den definition af *stævning*, som de fire institutioner blev enige om: ”indkaldelsesdokument til tiltalte om at møde til hovedforhandling” (Justitsministeriet n.d.). Som anført ovenfor er denne definition dog for snæver, da en stævning ikke blot anvendes til at indkalde til hovedforhandling, og Den Store Danske (n.d.) har da også følgende definition af *stævning*: ”formel indkaldelse til at møde i retten”.

Det kan altså konstateres, at Politiet, anklagemyndigheden og domstolene havde forskellige opfattelser af begrebet *stævning*, hvilket kunne give problemer i forbindelse med deres udarbejdelse og anvendelse af et fælles format for udveksling af data. Herudover er det adskillende træk i den definition, der blev opnået enighed om, for snævert, hvilket kan give anledning til misforståelser. På basis af overvejelserne ovenfor foreslås følgende definition af *stævning*: *indkaldelsesdokument der indeholder en formel indkaldelse af tiltalte til retsmøde*.

4. Behov for fællesoffentlig begrebsafklaring

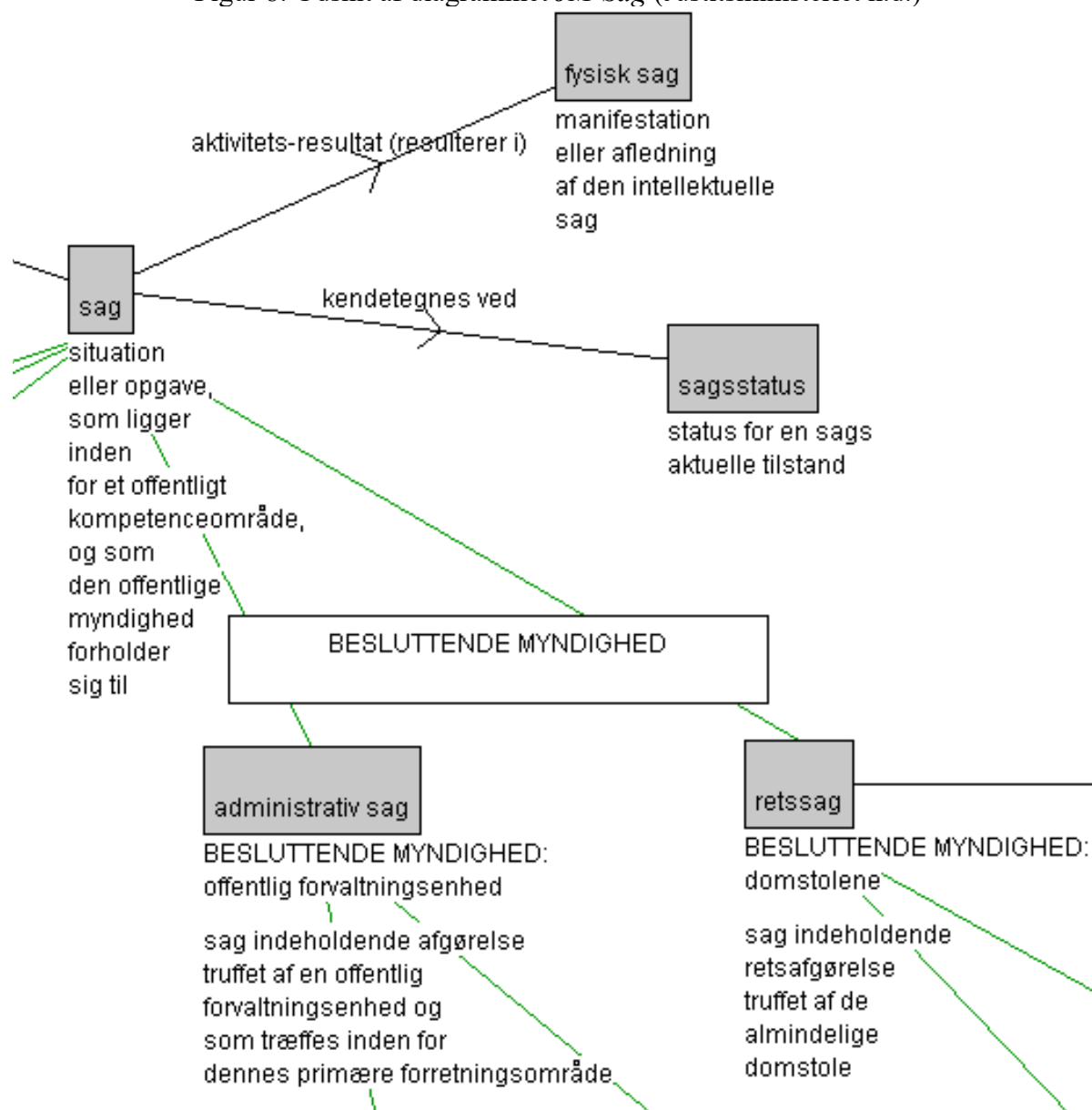
4.1. Begreberne *sag*, *administrativ sag* og *myndighedssag*

Mange myndigheder har arbejdet med at fastlægge en definition af begrebet *sag*. Det gælder for eksempel Justitsministeriets ovennævnte institutioner samt Socialstyrelsen og Digitaliseringsstyrelsen (tidligere IT- og Telestyrelsen). I Figur 8 vises et udsnit af Justitsministeriets diagram *JM-Sag* (Justitsministeriet n.d.). Det ses her, at der skelnes mellem begreberne *sag* (som i artiklen i begrebsbasen har synonymet *intellektuel sag* jf. Figur 9) og *fysisk*

sag.

I definitionen af *sag* er der anført to overbegreber: *situation eller opgave*, hvilket er uheldigt, da det er upræcist. Det er dog vanskeligt at finde et passende overbegreb til *sag*, hvilket også ses, hvis man slår op på artiklen *sag* i Den Danske Ordbog (n.d.), hvor man finder definitionen, som er vist i Figur 10, og som indledes med ”helhed af faktorer og omstændigheder der knytter sig til en hændelse, en situation eller et forløb”. Det, der beskriver en *sag* i resten af definitionen, favner ligeledes meget bredt. I Den Danske Ordbog findes endvidere en underbetydning 1d. med definitionen: ”(de samlede) dokumenter, breve m.m. vedr. et bestemt, ofte juridisk eller administrativt forhold.” Denne svarer til *fysisk sag* i Justitsministeriets diagram *JM-Sag*.

Figur 8: Udsnit af diagrammet *JM-Sag* (Justitsministeriet n.d.)



Figur 9: Artikel med synonymet *intellektuel sag* (Justitsministeriet n.d.)



Danish: **sag**

Danish: **intellektuel sag**

General definition: situation eller opgave, som ligger inden for et offentligt kompetenceområde, og som den offentlige myndighed forholder sig til

Figur 10: Definition af *sag* i Den Danske Ordbog (n.d.)Du er her: [Forside](#) / [Den Danske Ordbog](#) / [Ordbog](#)Vis forkortet  **sag** substantiv, fælleskønVis overblik 

BØJNING -en, -er, -erne

UDTALE ['sæʔ']  OPRINDELSE oldnordisk *sök*, oldengelsk *sacu* 'strid, proces'

Betydninger



1. helhed af faktorer og omstændigheder der knytter sig til en hændelse, en situation eller et forløb, og som nogen forholder sig til og tænker over, typisk for at nå frem til et bestemt resultat eller løsningen på et problem

SE OGSÅ [forhold](#) | [anliggende](#) | [affære](#)

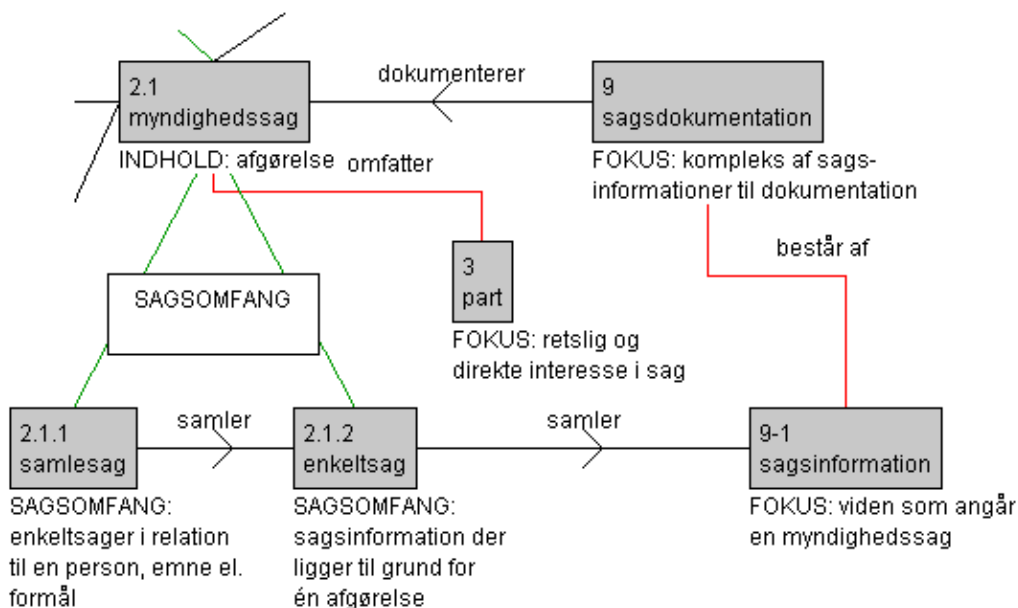
Justitsministeriets definition af *sag* er, som det fremgår af Figur 8: ”situation eller opgave, som ligger inden for et offentligt kompetenceområde, og som den offentlige myndighed forholder sig til.” Her er begrebet *sag* indsnævret vha. trækkene ”ligger inden for et offentligt kompetenceområde” og ”som den offentlige myndighed forholder sig til” (Justitsministeriet n.d.).

I Socialstyrelsen (n.d.) findes følgende definition af *sag*: ”situation eller opgave, som ligger inden for en offentlig forvaltningsenheds kompetenceområde, og som den offentlige forvaltningsenhed forholder sig til”. I Socialstyrelsens definition er begrebet *sag* altså indsnævret til en *forvaltningssag*, hvilket også er anført som synonym til *sag* i begrebsdatabasen. Det vil altså sige, at Justitsministeriets og Socialstyrelsens definitioner af *sag* adskiller sig mht. aktøren: *offentlig myndighed* eller *offentlig forvaltningsenhed*. Læseren kan således ikke afgøre, om definitionerne er ens, uden at undersøge om *offentlig myndighed* og *offentlig forvaltningsenhed* er det samme. Ser man i Socialstyrelsen (n.d.) er *offentlig myndighed* og *offentlig forvaltningsenhed* to forskellige begreber, og i Justitsministeriet (n.d.) findes ingen af termene (vi begrænser os her til myndighedernes definitioner).

I Socialstyrelsen (n.d.) findes *myndighedssag* som underbegreb til *sag* og med definitionen ”sag, hvori der træffes en afgørelse”, se også Figur 11. I Justitsministeriets diagram findes underbegrebet *administrativ sag*, jf. Figur 8, med definitionen: ”sag indeholdende afgørelse truffet af en offentlig forvaltningsenhed og som træffes inden for dennes primære forretningsområde” (Justitsministeriet n.d.). Det er således uklart, hvilket af Socialstyrelsens to begreber *sag/forvaltningssag* eller *myndighedssag* Justitsministeriets begreb *administrativ sag* svarer til, da

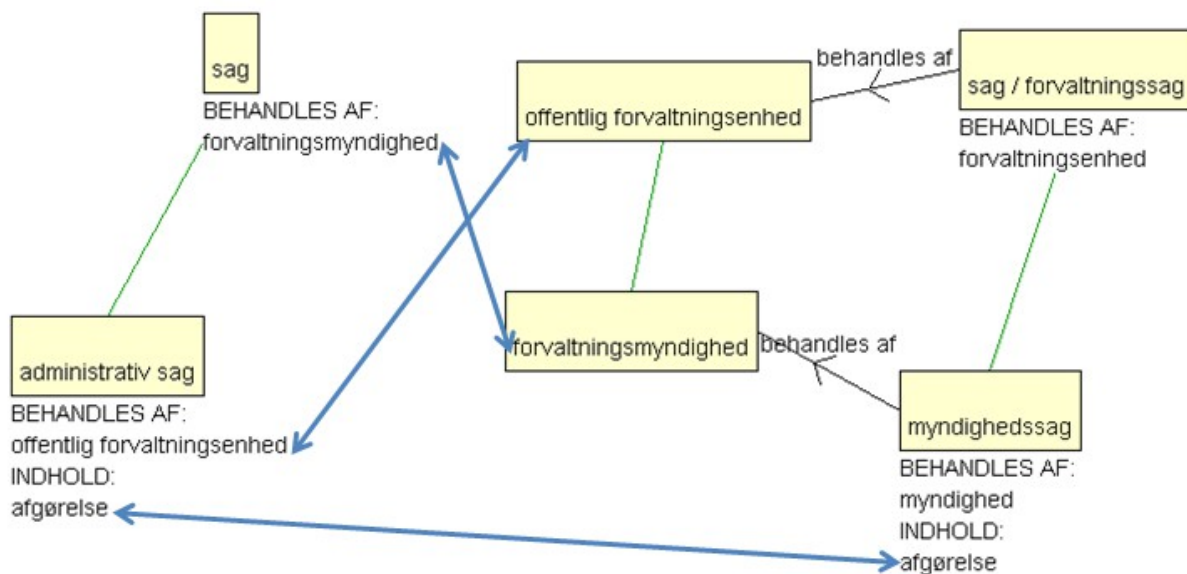
det på den ene side indeholder en afgørelse, men på den anden side behandles af en offentlig forvaltningsenhed.

Figur 11: Udsnit af diagrammet *Forvaltningsbegreber* (Socialstyrelsen n.d.)



I Figur 12 vises til venstre de to begreber fra Justitsministeriet (n.d.) og til højre fire begreber fra Socialstyrelsen (n.d.), idet nogle træk er hentet ud af definitionerne i de to baser. Figuren illustrerer begrebsforvirringen, og det er tydeligt, at der er behov for afstemning af definitionerne i de to myndigheders begrebsbaser. Dette arbejde kunne foregå inden for rammerne af Forum for Videnmodellering i Offentligt Regi, FORVIR, jf. <http://blog.cbs.dk/forvir/>, hvor de omtalte myndigheder er medlemmer.

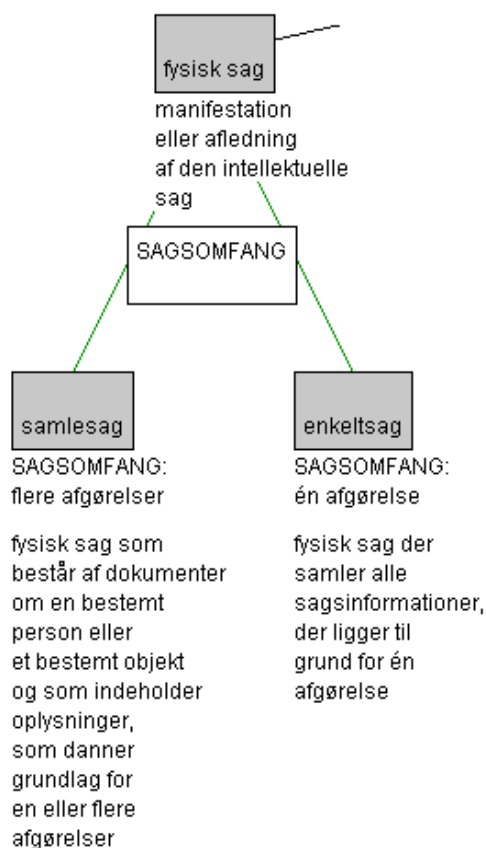
Figur 12: Sammenstilling af begreber fra Justitsministeriet (n.d.) og Socialstyrelsen (n.d.).



Figur 13: To artikler med definitioner og synonym fra diagrammet *JM-Fysisk sag* (Justitsministeriet n.d.)

Dansk:	samlesag
Dansk:	dossiersag
Generel definition: fysisk sag som består af dokumenter om en bestemt person eller et bestemt objekt og som indeholder oplysninger, som danner grundlag for en eller flere afgørelser	
Dansk:	enkeltsag
Generel definition: fysisk sag der samler alle sagsinformationer, der ligger til grund for én afgørelse	

Figur 14: Begreberne *samlesag* og *enkeltsag* i diagrammet *JM-Fysisk sag* (Justitsministeriet n.d.)



4.2. Begreberne *samlesag* og *enkeltsag*

IT- og Telestyrelsen (2008) har følgende forklaring af *sag*: ”En sag samler forskellige oplysninger og dokumenter, der vedrører et bestemt formål. Begrebet kan sammenlignes med et fysisk omslag.” Der er altså her tale om *fysisk sag*, hvis vi sammenligner med Justitsministeriets definition af dette

begreb. Der er ikke tale om en intensional definition med nærmeste overbegreb og adskillende træk, men en forklarende tekst.

I IT- og Telestyrelsen (2008) findes flere typer sager med følgende forklaringer: "... enkeltsag (de oplysninger, der ligger til grund for en afgørelse), dossiersag (en sag om en bestemt person eller objekt), samlesag (oplysninger om et bestemt emne) ...". Begreberne *samlesag* og *enkeltsag* findes også i Justitsministeriets diagram *JM-Fysisk sag* som underbegreber til *fysisk sag*, men *dossiersag* er synonym til *samlesag*, jf. Figur 13 og 14. På basis af definitionerne kan trækspecifikationerne ændres til:

samlesag

RELATERET TIL: en bestemt entitet

INDHOLD: oplysninger som danner grundlag for afgørelser

ANTAL AFGØRELSER: en eller flere

enkeltsag

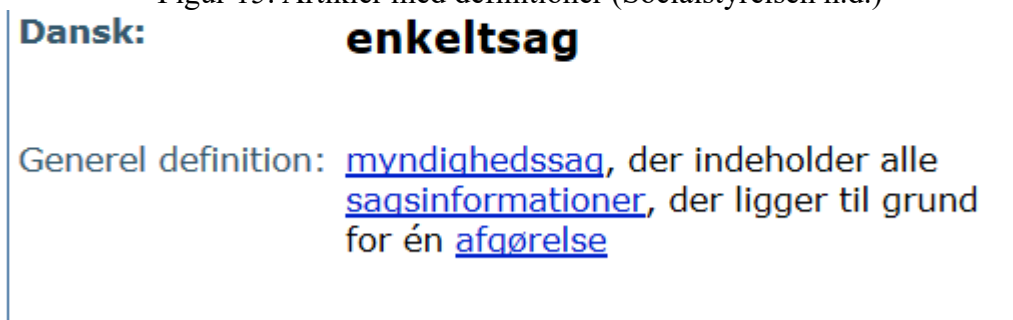
INDHOLD: oplysninger som danner grundlag for afgørelser

ANTAL AFGØRELSER: én

Både *samlesag* og *enkeltsag* har trækket *INDHOLD: oplysninger, som danner grundlag for afgørelser*, og dette træk bør derfor flyttes til overbegrebet *sag*.

I Socialstyrelsen (n.d.) findes ligeledes to begreber *samlesag* og *enkeltsag* som underbegreber til *myndighedssag*, jf. Figur 15 og 16.

Figur 15: Artikler med definitioner (Socialstyrelsen n.d.)

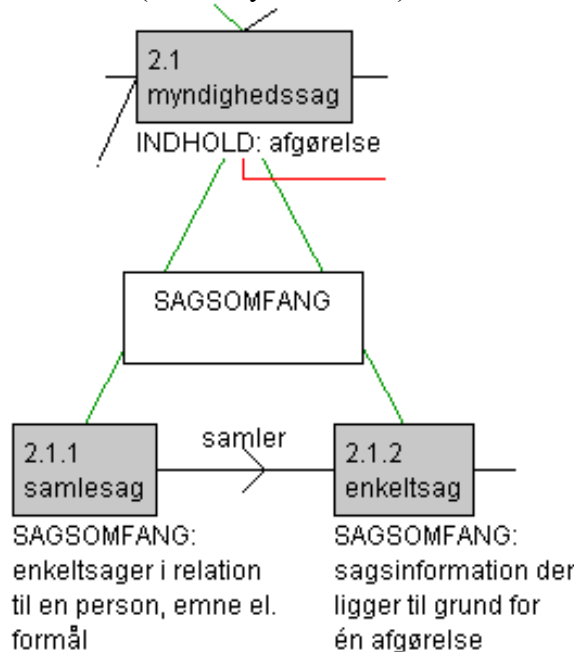


I Socialstyrelsens definitioner er der for *samlesag* fokuseret på, at sagen indeholder enkeltsager, der er relateret til en bestemt entitet (person, emne eller formål), medens der i definitionen af *enkeltsag* er fokuseret på, at den indeholder alle sagsinformationer, der ligger til grund for én afgørelse. Men da *samlesagen* indeholder enkeltsager, må den også indeholde oplysninger, som ligger til grund for flere afgørelser.

I Justitsministeriets definition af *samlesag* er der to træk: sagen indeholder dokumenter relateret til en bestemt entitet (person eller objekt), og den indeholder oplysninger, som danner grundlag for en eller flere afgørelser.

Der er altså igen her forskelle i formuleringerne af definitionerne. Formuleringerne er ikke klare. Baseret på Socialstyrelsens definition finder vi, at det der adskiller *enkeltsag* og *samlesag* er, om sagsinformationerne danner grundlag for en eller flere afgørelser, og det er det, der skal anføres som adskillende træk og danne grundlag for definitionen.

Figur 16: Begreberne *samlesag* og *enkeltsag*
(Socialstyrelsen n.d.)



I Digitaliseringsstyrelsen (2013) findes imidlertid ændrede definitioner af *enkeltsag* og *samlesag*:

Enkeltsag: En samling af sammenhørende dokumenter og øvrige sammenhørende oplysninger, der i sit hele anvendes til at dokumentere en arbejdsproces, typisk til administrative formål, herunder til at træffe afgørelser.

Samlesag: En samlesag består af dokumenter om en bestemt person eller et bestemt objekt som fx en person, et hus, en vej el.lign. En samlesag kan indeholde oplysninger, som danner grundlag for en eller flere afgørelser. Samlesag er synonym med dossiersag, men det er valgt i standarden at benytte begrebet³ 'samlesag'.

Samlesag har her samme forklaring som Justitsministeriets ovennævnte definition, svarende til trækspicifikationerne: *RELATERET TIL: en bestemt entitet* og *ANTAL AFGØRELSE: en eller flere*. Forklaringen af *enkeltsag* har ligesom Justitsministeriets definition kun trækket: *INDHOLD: oplysninger som danner grundlag for afgørelser*, men herudover et træk *FORMÅL* med værdien: *dokumentere en arbejdsproces*. Ovenfor så vi at trækket *INDHOLD* bør flyttes til overbegrebet (*fysisk sag*) og det samme er formentlig tilfældet med *FORMÅL*. *Enkeltsag* er således ikke adskilt fra sit overbegreb i Digitaliseringsstyrelsen (2013).

Eksemplet viser klart det store behov for fællesoffentlig begrebsafklaring.

5. En fællesoffentlig begrebsdatabase?

Begrebsafklaring er et ressourcekrævende arbejde. Flere myndigheder har ansat begrebsmedarbejdere, men det er svært at finde tid til en mere omfattende harmonisering og standardisering på tværs af myndighederne. Det ville derfor være en god idé, hvis der blev afsat

³ Her bør hellere anvendes 'betegnelsen'.

ressourcer til opbygning af en fællesoffentlig begrebsdatabase med et større antal begreber. En sådan begrebsdatabase ville ikke blot være til stor nytte for myndigheder, men også for borgerne i al almindelighed, som her ville kunne finde klare definitioner af de mange begreber, som bliver brugt på tværs af de forskellige områder for eksempel i forbindelse med opslag på borger.dk og diverse digitale selvbetjeningsløsninger eller ved direkte kommunikation mellem myndigheder og borgere.

Vi har tidligere (Madsen & Thomsen 2015b) beskrevet fordelene ved at have adgang til struktureret viden i form af terminologiske ontologier med præcise definitioner i stedet for at skulle læse de mange forskellige tekster med ustrukturerede informationer, som kan findes på nettet. Ofte må man læse længere tekster og sammenligne flere hjemmesider for at opnå en forståelse af det emne, der søges viden om, og ofte bliver man ledt på afveje af tvetydigheder.

Informationerne i en fællesoffentlig begrebsdatabase kan også indgå i en større terminologi- og vidensbank med begreber inden for mange, for eksempel tekniske og økonomiske, emneområder på både dansk og engelsk. Når man samler begreber fra mange fagområder, vil det samme udtryk optræde flere gange. Hvis et fagudtryk bruges med forskellig betydning inden for to forskellige fagområder, vil der være tale om to forskellige begreber med homografe termudtryk, og begge vil blive vist ved søgning.

For at drive og udvikle en sådan terminologi- og vidensbank er der brug for en organisation med fastansatte begrebsmedarbejdere, som kan være drivkraft i udarbejdelse af terminologiske ontologier med tilhørende definitioner, og som kan samarbejde med fageksperter inden for de forskellige emneområder. Man kunne også tænke sig, at der til en sådan organisation blev knyttet en slags helpdesk i stil med Dansk Sprognævns svartjeneste, men blot med fokus på faglige begreber.

For at kunne etablere en terminologi- og vidensbank af et tilstrækkeligt omfang og med indhold af en tilstrækkeligt høj kvalitet er der brug for metoder, der kan automatisere nogle af processerne i begrebsarbejdet, for eksempel ved at muliggøre automatisk ekstraktion af viden fra tekster og automatisk opbygning af udkast til terminologiske ontologier. I DanTermBank-projektet ved Institut for International Business Communication, Copenhagen Business School, har vi i nogle år arbejdet på dette, og der eksisterer i dag prototyper til automatisk opbygning af tekstkorpora samt til automatisk ekstraktion af viden om begreber fra tekster og til opbygning af råskitser til ontologier, jf. nedenfor. Der mangler imidlertid ressourcer til at drive en terminologi- og vidensbank, hvilket nok, jf. eksemplerne ovenfor, bør være en tværministeriel opgave.

6. Prototyper til automatisering af terminologiarbejde

I den første fase af DanTermBank-projektet (2011-2014) har vi udviklet avancerede metoder og prototyper til ekstraktion af viden om begreber fra tekster, samt til opbygning og validering af terminologiske ontologier. Projektet er nærmere beskrevet i fx Lassen et al. (2013: 176-182) og Lassen et al. (fc) samt på projektets websted DanTermBank (n.d.).

Prototyperne kan anvendes sekventielt eller enkeltvis til de forskellige faser af terminologiarbejde: indsamling af tekster, identifikation af termkandidater, opbygning af udkast til ontologier ved ekstraktion af relationer mellem begreber og validering af ontologierne. Nogle af prototyperne er som udgangspunkt tilpasset dansk sprog (ordklasseopmærkning og termekstraktion), men er forberedt til at kunne benyttes til andre sprog ved tilføjelse af sprogspecifikke modeller og regler, de øvrige kan umiddelbart bruges sproguafhængigt. I det følgende bringer vi en kort oversigt over prototyperne, en mere udførlig gennemgang findes i Thomsen et al. (fc).

6.1. Korpuskompilering med DanTermBank-crawleren, dtC

DanTermBanks web-crawler, dtC, er et værktøj som kan samle et tekstkorpus til terminologiarbejde. dtC kan søge efter tekster som 'ligner' en eller flere tekster man har i forvejen,

de såkaldte frøtekster⁴. Det sker ved at fremsøge nøgleord fra frøteksterne og bruge nøgleordene til videre søgning på internettet. Som alternativ til frøtekster, kan brugeren selv levere en liste med frøtermer.

6.2. Ordklasseopmærkning med DanTermBank-taggeren – dtTag

DanTermBank-taggeren udfører ordklasseopmærkning med TreeTagger (Schmid 1994, 1995, n.d.) trænet for dansk (Lassen 2012: 218-230). Taggeren opmærker danske tekster i txt-format med ordklassetags og oplysning om lemma. Vi bruger tags fra Penn Treebank (The Penn Tree Bank Project 1999), som angiver både ordklasse og bøjningsform, for eksempel:

- VB** - Verb, base form
- VBD** - Verb, past tense
- VBP** - Verb, non-3rd person singular present
- VBZ** - Verb, 3rd person singular present

I DanTermBank-projektet er taggeren desuden udvidet med analyse og opmærkning af komposita med sammensætningsdelenes ordklasser og fugeelementer (Lassen et al. *fc*). Analysen af komposita udnyttes både i forbindelse med termekstraktion og ekstraktion af begrebsrelationer.

6.3. Termekstraktion med dtX

Prototypen dtX finder termkandidater i det ordklasseopmærkede korpus ved hjælp af morfosyntaktiske mønstre og statistiske metoder. Et morfosyntaktisk mønster består af en række af ordklassetags som man ved en term kan bestå af. For eksempel er mange termer af formen adjektiv + substantiv, og mønstret JJ-NN kan derfor bruges til at finde en del termkandidater. Det samme gælder komposita bestående af to substantiver. De statistiske beregninger supplerer mønstrene. Hvis beregningerne viser at et ord eller en ordsekvens optræder hyppigere i den givne tekst end i et almensprogligt korpus, eller optræder hyppigere sammen end hver for sig, er sandsynligheden for at der er tale om en term større.

6.4. Terminterne relationer med dtR-intra

Første fase i fremsøgning af begrebsrelationer består i at finde term-interne relationer, deraf navnet 'intra'. Ud fra mønsterregler som fortæller hvilke relationer der er mellem delene i en sammensat term, finder dtR-intra relationer i flerordstermer og komposita på en liste med termer som den der kommer ud af dtX.

Reglerne kan for eksempel se således ud:

- JJ-NN →
1. SEQ:REL:0;
 2. SEQ:ISA:1;

Disse regler fortæller systemet hvordan en ontologi skal opbygges for termer, der matcher et givent term-mønster.

Reglen ovenfor specificerer, at for termer der matcher mønsteret *JJ-NN* (adjektiv og substantiv særskrevet, som for eksempel '*speciel fuldmagt*'), eksisterer der en ikke nærmere specificeret associativ relation mellem begrebet '*speciel fuldmagt*' og det begreb, der udgøres af JJ '*speciel*', samt en typerelation mellem '*speciel fuldmagt*' og det begreb, der udgøres af NN '*fuldmagt*'.

4 Betegnelserne *frøtekst* og *frøterm* henviser til at det er som frø man sår, for derefter at høste nye tekster.

6.5. Relationer mellem termer med dtR-inter

Prototypen dtR-inter består af to dele, en mønsterfinder og en relationsfinder. Mønsterfinderens tager som input tripler bestående af to termer og en relation, hvor det er kendt at relationen findes mellem de to begreber, og leder derefter i korpus efter tekststrengene (mønstre) som svarer til de givne relationer.

Den anden del af dtR-inter, relationsfinderens, finder nye relationer mellem par af termer i et korpus. Resultatet fra dtR-inter kombineres med resultatet fra dtR-intra til et udkast til en samlet terminologisk ontologi, som der kan arbejdes videre med.

6.3. Validering af ontologier med dtV

Den sidste prototype, dtV, validerer terminologiske ontologier, det vil sige undersøger om de overholder de principper som gælder for blandt andet nedrivning af karakteristiske træk jf. ovenfor. Hvis det er muligt, rettes der automatisk i ontologien, så principperne overholdes. Enkelte rettelser, som nedrivning af karakteristiske træk, er implementeret, mens andre afventer mere forskning på dette felt, herunder eksperimenter med forskellige kombinationer af fejltyper.

7. Konklusion

Som det fremgår af ovenstående gennemgang, er der behov for fællesoffentlig begrebsafklaring i forbindelse med begrebet *sag* og relaterede begreber. Disse begreber anvendes af en lang række myndigheder udover de her nævnte, og må betragtes som meget centrale.

Ved at anvende principperne for terminologiske ontologier, bliver det lettere at afklare begrebs karakteristiske træk og at udforme korte, klare og konsistente definitioner. Når man skal anføre karakteristiske træk i form af korte trækspecifikationer og udvælge det adskillende træk, 'tvinges' man til at fokusere på det, der adskiller relaterede begreber, som illustreret med definitionerne af begrebet *stævning*.

Ved udformning af en definition er det vigtigste princip netop, at den opbygges med det nærmeste overbegreb og det karakteristiske træk, som adskiller det pågældende begreb fra sideordnede begreber. Dette princip overholdes i forbindelse med en lang række begreber både i Justitsministeriet (n.d.) og i Socialstyrelsen (n.d.), men i forbindelse med definitionen af *sag* er det ikke gennemført, jf. afsnit 4.1. Her anføres både to overbegreber (*situation* og *opgave*) og to karakteristiske træk: "som ligger inden for et offentligt kompetenceområde" og "som den offentlige myndighed forholder sig til" (Justitsministeriet n.d.). Det er altid vanskeligt at fastlægge det nærmeste overbegreb i forbindelse med begreber, som anvendes i flere sammenhænge. Men da der i denne sammenhæng er tale om et mere specifikt begreb end det, der for eksempel defineres i Den Danske Ordbog (n.d.), bør der også findes et mere specifikt overbegreb. Endvidere bør det ene træk udpeges som adskillende, hvis det andet kan siges at være supplerende.

Som det formodentlig er fremgået af vores artikel, er begrebsarbejde tidskrævende, og der er derfor behov for, at der afsættes ressourcer til opbygning og drift af en fællesoffentlig begrebsdatabase med et større antal begreber, som ville være nyttige for myndigheder, men som også kunne være tilgængelige for offentligheden.

Referencer

borger.dk (n.d.). URL: <https://www.borger.dk>. Sidst tilgået: 2. september 2015.

Carpenter, Bob (1992). *The Logic of Typed Feature Structures*. Cambridge: Cambridge University Press.

DanTermBank (n.d.). URL: www.dantermbank.dk. Sidst tilgået: 2. september 2015.

Den Danske Ordbog (n.d.). Det Danske Sprog- og Litteraturselskab. URL: <http://ordnet.dk/ddo>. Sidst tilgået: 2. september 2015.

- Digitaliseringsstyrelsen (2013). *Specifikation af Serviceinterface for sag - OIO-Godkendt* [vs. 1.2].pdf – 1.oktober 2013. URL: <https://digitaliser.dk/resource/1567587>. Sidst tilgået: 31. august 2015.
- Den store Danske (n.d.). *Gyldendals åbne encyklopædi*. URL: <http://www.denstoredanske.dk/>. Sidst tilgået: 31. august 2015.
- DOLCE (n.d.) *Descriptive Ontology for Linguistic and Cognitive Engineering*. URL: <http://www.loa.istc.cnr.it/old/DOLCE.html>. Sidst tilgået: 8. januar 2016.
- Hoffmann, Pia, Charlotte Pedersen & Bodil Nistrup Madsen (2012). 'Begræbsmodellering i Justitsministeriet. Modeller for samarbejde om begrebsafklaring'. I *NORDTERM 17 Samarbetet ger resultat: från begreppskaos til överenskomna termer*. Vaasa, Finland, 7.-10. juni 2011. Helsinki. 35-38.
- ISO 704 (2009). *Terminology work - Principles and methods*. International Standards Organisation, Geneva.
- IT- og Telestyrelsen (2008). *Referencearkitektur for sags- og dokumentområdet (ESDH)* version 3.0, 3. oktober 2008, [ref: 31] URL: <http://www.digst.dk/Arkitektur-og-standarder/Standardisering/Datastandardisering/Sag-og-dokumentomraadet/Referencearkitektur-for-sags-og-dokumentomraadet>. Sidst tilgået: 31. august 2015.
- Justitsministeriet (n.d.). *Begrebsafklaring under Justitsministeriet*. URL: <http://jm.item.dk/>. Sidst tilgået: 31. august 2015.
- Kriminalforsorgen (n.d.). *Kriminalforsorgens begrebsafklaring*. URL: <http://kriminalforsorgen.item.dk/search/>. Sidst tilgået: 31. august 2015.
- Lassen, Tine (2012). 'A corpus compilation and processing prototype for terminology work'. I Guadalupe Aguado de Cea, Mari Carmen Suárez-figueroa, Raúl García-Castro & Elena Montiel-Ponsoda (red.), *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)*, Madrid, Spanien, 19.-22. juni 2012. 218-230.
- Lassen, Tine, Bodil Nistrup Madsen & Hanne Erdman Thomsen (2013). 'Automatisk opbygning og validering af terminologiske ontologier'. Indlæg på konferencen: *Från förarbete till förvaltning – terminologiarbete steg för steg. NORDTERM 2013*, Stockholm.
- Lassen, Tine, Bodil Nistrup Madsen & Hanne Erdman Thomsen (fc). 'Automatisk opbygning og validering af terminologiske ontologier'. *Proceedings fra NORDTERM 2013*, Stockholm.
- Lassen, Tine, Bodil Nistrup Madsen, Louise Pram Nielsen, & Hanne Erdman Thomsen (2013). 'The DanTermBank Project: Automatic knowledge extraction and knowledge structuring for a national term bank'. *Poster session Terminologi - ansvar og bevissthed*, Språkrådet, Oslo, Norge, 11.-12. Oktober 2012.
- Madsen, Bodil Nistrup (1998). 'Typed feature structures for terminology work - Part I'. I Lita Lundquist, Heribert Picht & Jacques Qvistgaard (red.), *LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes*. Frederiksberg: Copenhagen Business School. 339-348.
- Madsen, Bodil Nistrup (1999). *Terminologi 1: Principper og metoder*. København: Gads Forlag.
- Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2006). 'Terminological ontologies in normative terminology work'. *TSTT'06: Proceedings of the International Conference on Terminology, Standardization and Technology Transfer*. Beijing: Encyclopedia of China Publishing House. 122-133.
- Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2009). 'CAOS – A tool for the construction of terminological ontologies'. I Kristiina Jokinen & Eckhard Bick (red.), *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. NEALT Proceedings Series, Vol. 4 2009. 279-282. Publiceret af Northern European Association for Language Technology (NEALT) URL: <http://omilia.uio.no/nealt>. Sidst tilgået: 9. februar 2016. Elektronisk publiceret ved Tartu University Library (Estland) (Vol. 4:[52], ISSN 1736-6305).

- URL: <http://hdl.handle.net/10062/9807>. Sidst tilgået: 9. februar 2016.
- Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2015a). 'Concept modeling vs. data modeling in practice'. I Hendrick J. Kockaert & Frieda Steurs (red.), *Handbook of Terminology*. Vol. 1. John Benjamins Publishing Company, Amsterdam. 250-275. DOI:10.1075/hot.1.con1
- Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2015b). 'Fra tekst til struktureret viden'. I Dorthe Duncker (red.), *Rette ord. Festskrift til Sabine Kirchmeier-Andersen i anledning af 60-årsdagen*. Sprognævnets skriftsserie 46. Frederiksberg: Dansk Sprognævn. 227-246.
- Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner (2004). 'Principles of a system for terminological concept modelling'. I Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (red.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation. ELRA*. 15-19.
- Justitsministeriet (2014). Retsplejeloven . URL: <https://www.retsinformation.dk/forms/r0710.aspx?id=164280>. Sidst tilgået: 31. august 2015.
- Schmid, Helmut (1994). 'Probabilistic part-of-speech tagging using decision trees'. *Proceedings of International Conference on New Methods in Language Processing* Manchester, Storbritannien.
- Schmid, Helmut (1995). 'Improvements in part-of-speech tagging with an application to German'. I *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Schmid, Helmut (n.d.). 'TreeTagger - a language independent part-of-speech Tagger'. URL: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/treetagger.html>. Sidst tilgået: 15. oktober 2015.
- Socialstyrelsen (n.d.). Socialebegreber. URL: http://www.socialebegreber.dk/begrebsarbejde/begrebsdatabasen/find_begreb. Sidst tilgået: 31. august 2015.
- The Penn TreeBank Project. (1999). URL: <https://www.cis.upenn.edu/~treebank/>. Sidst tilgået: 14. november 2015.
- Thomsen, Hanne Erdman (1997). 'Feature specifications applied to the field of life insurance'. *Terminology Science and Research - Journal of the International Institute for Terminology Research*, 8(1/2): 21-36.
- Thomsen, Hanne Erdman (1998). 'Typed Feature Structures for Terminology Work - Part II'. I Lita Lundquist, Heribert Picht & Jacques Qvistgaard (red.), *LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes*. Frederiksberg: Copenhagen Business School. 349-359.
- Thomsen, Hanne Erdman (2016). 'The intensional definition'. I Rute Costa & Christophe Roche (red.), *The definition in Terminology*. Berlin: Frank & Timme, Verlag für wissenschaftliche Literatur.
- Thomsen, Hanne Erdman, Bodil Nistrup Madsen & Tine Lassen (fc). 'Automatisk vidensindsamling til DanTermBank'. *NORDTERM 2015*, Reykjavik, 9.-12. Juni 2015.