

Peter Widell og Ulf Dalvad Berthelsen (udg.):
11. Møde om Udforskningen af Dansk Sprog
Århus 2006

DANNET – ET LEKSIKALSK-SEMANTISK WORDNET FOR DANSK

Af Sanni Nimb (Center for Sprogteknologi, Københavns Universitet) og Nicolai Hartvig Sørensen (Det Danske Sprog- og Litteraturselskab)

1. Indledning

*”[En grundlæggende hypotese inden for psykolingvistikken] er at menneskers viden om enkelte ords betydning, deres semantiske hukommelse, har form af et **netværk** der omfatter samtlige de semantiske relationer der kan påvises mellem et givent ord og andre ord i sproget. Man inddrager her [mange relationsarter], hvilket resulterer i ét stort flerdimensionalt netværk, som kæder alle ord i sproget sammen ”på kryds og tværs” ” (Herslund og Smith 2003).*

DanNet-projektet har som sit ambitiøse mål at opbygge et sådant netværk for de danske almensproglige ord. Projektet der foregår i et samarbejde mellem Center for Sprogteknologi på KU og Det Danske Sprog- og Litteraturselskab, er støttet af Forskningsrådet for Kultur og Kommunikation og løber som udgangspunkt i perioden 2005-2007.

Målet for projektet er konkret at udvikle et såkaldt wordnet for dansk på mindst 40.000 ‘begreber’,¹ lagret i en bestemt databasestruktur som sikrer at nettet er kompatibelt med både Princeton WordNet (Fellbaum 1998) og med EuroWordNet (Vossen 1998). Af de 40.000 begreber vil cirka 5.000 blive kædet til det såkaldte InterLingual Index (se afsnit 3).

I det følgende vil vi først beskrive hvad et wordnet er, dernæst vil vi beskrive hvordan det danske wordnet, DanNet, opbygges og tre særlige udfordringer vi er blevet stillet over for i forbindelse med at benytte en eksisterende ordbog som udgangspunkt for DanNet. Til sidst vil vi give eksempler på hvordan en ressource som DanNet kan anvendes.

2. Wordnets, synsets og nedarvningshierarkier

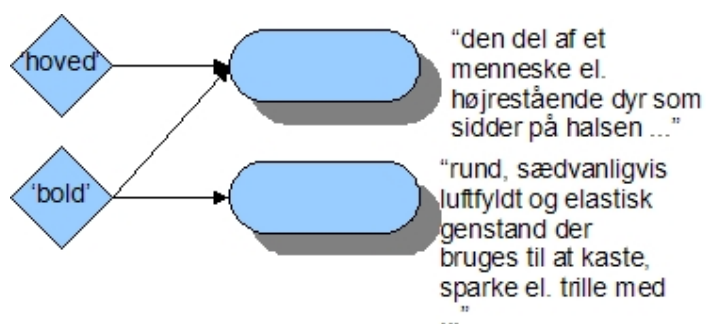
Det første wordnet, Princeton WordNet, blev udviklet ved Princeton universitet af blandt andre Christiane Fellbaum og George A. Miller fra 1985 og frem (Fellbaum 1998). Der bliver

¹ Hvad der i denne sammenhæng forstås ved et ‘wordnet’ og et ‘begreb’ beskrives i næste afsnit.

stadig udviklet på projektet, og en version 2.1 udkom i 2005.

Grundprincippet for et wordnet er at det skal afspejle det mentale leksikon som antages at være organiseret i begreber, og at det er begreber, og ikke leksikalske udtryk (fx lemmaer), der er forbundet gennem semantiske relationer. Et begreb kaldes i wordnets for et "synset", der er udledt af "a set of synonyms", dvs. "en mængde af synonymymer", fordi det samme begreb (synset) kan udtrykkes ved mange forskellige leksikalske udtryk. For eksempel kan begrebet MOTORCYKEL, dvs. "større tohjulet motorkøretøj med kraftig motor og med en egenvægt på højst 400 kg", udtrykkes med lemmaet 'motorcykel', men også med lemmaerne 'mc' og 'kværn' (i en af dette lemmas betydninger). Brugen af 'kværn' om en motorcykel tilhører naturligvis et andet stilleje end de to andre, men dette ser man bort fra i synset-sammenhæng med den begrundelse at udtrykkene henviser til samme begreb i verden. På samme måde kan begrebet HOVED udtrykkes med lemmaerne: 'hoved', 'knop', 'kasse', 'bold', 'byld', 'bøtte' m.fl. (oplysningen om stilleje findes dog i DanNet således at lemmaer kan frasorteres ud fra dette om nødvendigt).

Når et lemma har flere betydninger, kan det tilknyttes flere begreber, som i Figur 1. Figuren er forenklet idet både 'hoved' og 'bold' i DanNet har flere betydninger og dermed er tilknyttet flere synsets end vist, og tilsvarende er der flere lemmaer der peger på de to stiliserede synsets i figuren.

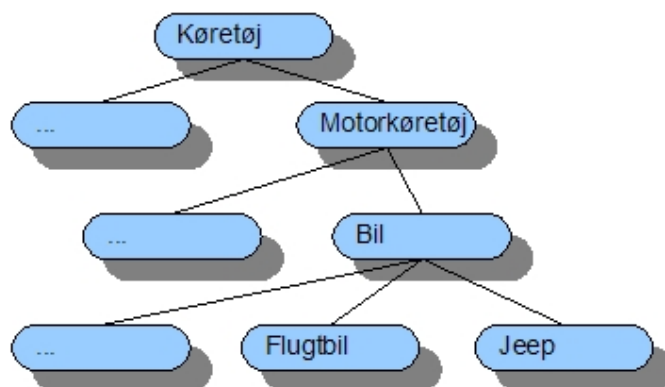


Figur 1 Flere begreber til et lemma

I DanNet har vi fået tilladelse til at tage udgangspunkt i Den Danske Ordbog (Hjort og Kristensen 2003-5), og vi følger nøje betydningsopdelingen i denne ordbog. DDO er korpusbaseret, og vi opnår derved, indirekte, at begrebsopdelingen i DanNet tilsvarende har belæg i korpusforekomster.

Et synset beskrives med relationer, fx relationerne *Has_meronym* som angiver om et begreb er en del af et andet begreb (fx er 'næse' en del af et 'hoved'), og *Has_holonym* som angiver at et begreb har et andet begreb som en del af sig (fx har 'motorkøretøj' nødvendigvis

en 'motor' som en del af sig). Som værdi for disse relationer angives altid et andet begreb (et synset), og wordnet'et bliver på den måde kædet sammen på kryds og tværs. Den mest betydningsfulde relation er dog relationen *Has_hyperonym* som angiver overbegrebet til et synset, idet det er grundliggende for wordnet-tankegangen at begreber er organiseret i et nedarvningshierarki i kraft af over-/underbegrebsstrukturen (Miller 1990:10f). Alle begreber – undtagen et abstrakt synset 'øverst' i hierarkiet – siges at have et overbegreb,² som angives som et andet synset. På den måde opstår et ordnet nedarvningshierarki hvor synsets 'længere nede' arver deres 'forældres' egenskaber (dvs. semantiske relationer): Hvis begrebet MOTORKØRETØJ beskrives med relationen *Has_meronym* 'motor', vides det umiddelbart at det må gælde at alle dets underbegreber (fx 'bil' og i næste led fx 'flugtbil' og 'jeep') også må have en motor.³ Det er ikke centralt for begrebet FLUGTBIL at det indeholder en 'motor', men den viden *nedarves* gennem over-/underbegrebshierarkiet (det er derimod helt centralt for FLUGTBIL at den bruges til at 'flygte' i, og denne information nedarves ikke fra noget overbegreb – den relation (*For_purpose_of*) må tilføjes på selve begrebet FLUGTBIL).



Figur 2 Uddrag af over-/underbegrebshierarki

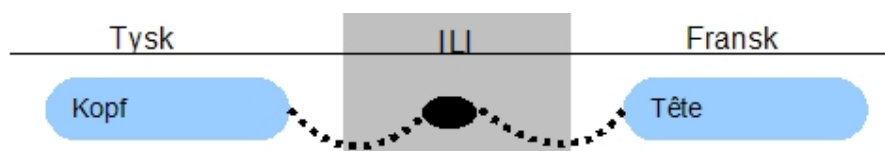
3. EuroWordNet

I slutningen af 1990'erne blev der etableret wordnets i stil med Princeton WordNet for seks europæiske sprog: Hollandsk, italiensk, spansk, tysk, fransk og estisk. Dette skete i projektet EuroWordNet. Ud over de monolingvale net blev de seks sprog kædet sammen gennem det såkaldte Inter-Lingual Index (ILI). Grundidéen i ILI'en er meget simpel: Begreber (synsets)

²I sjældne tilfælde kan det synes relevant at angive to overbegreber til et synset, fx er begrebet HAVESTOL både en slags STOL og en slags HAVEMØBEL.

³Det kan naturligvis være nødvendigt at *blokere* en relation, således at hvis 'flugt' beskrives som værende i stand til at 'flyve', så kan man ved underbegrebet 'pingvin' forhindre denne relation i at gælde for dette begreb.

med samme betydning i de forskellige sprog knyttes til samme knudepunkt i ILI'en. På den måde kan begreber i det ene sprog – løseligt – oversættes fra et sprog til et andet, som skitseret i Figur 3.

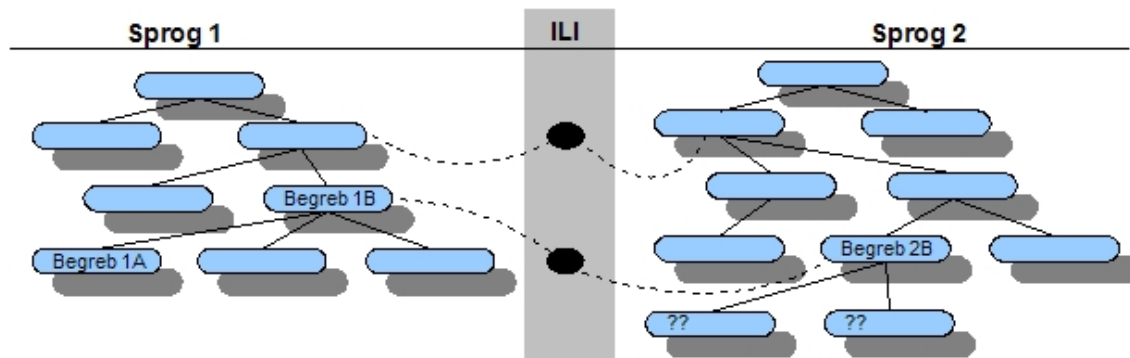


Figur 3 Oversættelse af tysk 'Kopf' til fransk 'tête' vha. InterLingual Index

Det velkendte 'uoverensstemmelsesproblem' ved end sådan Interlingua-agtig tilgang til oversættelse, dvs. at der for eksempel ikke er nogen direkte, entydig oversættelse af det engelske begreb WALL til dansk (det skal oversættes til enten 'mur' eller 'væg' afhængig af konteksten), løses ikke, men omgås ved at sammenkædningen mellem et begreb i et sprog og ILI'en ikke behøver at være direkte: Man kan angive at det synset man hæfter på ILI'en er et overbegreb eller et underbegreb til knuden i ILI'en. Idet ILI'en i realiteten består af synsets fra Princeton WordNet (fordi dette wordnet var det eneste der fandtes da man påbegyndte udviklingen af EuroWordNet), og der derfor findes en knude for engelske WALL, men ikke for MUR og VÆG, vil det være nødvendigt at kæde de to danske begreber til knuden for WALL, men som underbegreber.

Det er dog vigtigt at bemærke at det langtfra er alle synsets i et sprog der på denne måde knyttes til ILI'en. Oprindeligt blev der udvalgt en række begreber som fandtes i de monolingvale net i alle seks sprog. Det blev til cirka 1000 såkaldte Base Concepts i alt, så der er nødvendigvis tale om temmelig løst sammenkædede leksikalske net hvor det kun er muligt at foretage direkte oversættelse af begreber i relativt få tilfælde, som illustreret i Figur 4. Men ved at udnytte over- og underbegrebsnavigation er det muligt i det mindste at få en hjælp til en oversættelse. I Figur 4 er begrebet A ikke tilknyttet ILI'en, og der findes derfor ikke en direkte oversættelse i sprog 2. Men ved at gå et niveau *op* i hierarkiet finder vi at dets overbegreb som vi her kalder B, *er* oversat. Vi kan nu forvente oversættelsen af A skal findes blandt C's underbegreber i sprog 2. Vi har ikke fundet en oversættelse af begrebet A ved hjælp af ILI'en, men vi har fundet information om oversættelsen.

Antallet af Base Concepts i ILI'en er siden blevet udvidet i flere omgange af andre projekter, blandt andet af BalkaNet (Sofia et al. 2002) der har arbejdet med at etablere wordnets for sprogene bulgarsk, tjekkisk, græsk, rumænsk, tyrkisk og serbisk. I DanNet regner vi med at bruge disse udvidede Base Concepts, dvs. at vi vil knytte knap 5.000 synsets til ILI'en.



Figur 4 To monolingvale net knyttet til ILI'en

4. DanNet

Det danske wordnet-projekt DanNet bygger som tidligere nævnt på Den Danske Ordbog (DDO) som blev udarbejdet af Det Danske Sprog- og Litteraturselskab 1992-2004 (Hjort og Kristensen 2003-5) og som i alt indeholder ca. 100.000 betydninger. Inden for DanNets treårige projektperiode skal der opbygges et ordnet bestående af 40.000 synsets hvis tilhørende lemmer til dels udvælges efter frekvens i DDO's korpus.

De fleste af de 100.000 betydninger blev i forbindelse med udarbejdelsen af DDO forsynet med et overbegreb (genus proximum) som ikke er synligt i den trykte ordbog. Overbegrebet indgår som hovedregel også som del af selve ordbogsdefinitionen. Fx har lemmaet 'fluestang' i DDO definitionen "lang fiskestang der bruges ved fiskeri med flue" og overbegrebet 'fiskestang'. Overbegreberne er ikke entydiggjort i tilfælde af homografi og polysemi, men altså blot en tekststreng. Fx er ordene 'milt' og 'magistrat' i DDO begge blev angivet med genus proximumet (tekststrengen) 'organ', men der sigtes til to vidt forskellige betydninger. Alligevel er det i høj grad informationen om disse overbegreber der udnyttes i opbygningen af wordnet'et for dansk. Vi har i DanNet-projektet udviklet et it-værktøj der benytter oplysningen fra DDO til at foreslå et eller flere begreber som mulige overbegreber til et synset som derefter kan udvælges og godkendes.

Hyponymirelationerne mellem begreberne udgør som tidligere nævnt grundstrukturen i ordnettet og er de semantiske relationer der i første arbejdsfase fastlægges for det udvalgte ordforråd. Det kunne umiddelbart fremstå som et forholdsvis enkelt stykke arbejde når man på denne måde ved hjælp af et værktøj får adgang til DDO's overbegreber, men der ligger stadig et meget stort arbejde i at udvælge det helt rigtige overbegreb. Dette skyldes først og fremmest at der er rigtig mange polyseme overbegreber. Der skal i disse mange tilfælde for det første træffes en beslutning om præcis hvilken betydning af lemmaet det drejer sig om

blandt de (ofte mange) mulige i DDO. For det andet får man i den automatiske søgning på alle underbegreber til et lemma, i tilfælde af polysemi, en lang række forkerte betydninger med som efterfølgende skal frasorteres. Sidst kan også nævnes at flere forskellige synonyme udtryk kan være brugt som overbegreb. Disse skal i DanNet slås sammen og indgå i det samme synset.

Grundlæggende skyldes mange af vanskelighederne at DDO og DanNet bygger på to forskellige og i princippet modsatrettede leksikalske anskuelser. Ordbogen er naturligt nok udarbejdet efter den semasiologiske metode (Dirven og Verspoor 1998; Herslund og Smith 2003), dvs. med fokus på det enkelte lemmas interne betydningsstruktur og udvisning af polysemi (omend der selvfølgelig i definitionerne relateres til andre begreber i sproget, fx et overbegreb). Det er derimod den onomasiologiske beskrivelsesmetode (Dirven og Verspoor 1998; Herslund og Smith 2003) der anvendes i DanNet-projektet. Her har man udelukkende fokus på hver enkelt betydning eksterne relation til sproget øvrige begreber. DDO's definitioner har derudover et præcist formidlende formål; et formål der, naturligt nok, af og til er tilgodeset på bekostning af angivelsen af det strukturelt set nærmeste overbegreb i sproget. Lemmaet 'budcykel' angives fx i DDO til at have overbegrebet 'cykel' og ikke 'ladcykel' selv om 'ladcykel' i en formaliseret struktur er det nærmeste overbegreb når man nærlæser betydningsdefinitionerne for de to ord. Dette skyldes udelukkende formidlingsmæssige aspekter samt det faktum, at leksikografen skulle udvælge overbegrebet blandt ordene i definitionen. På grund af disse faktorer skal det automatisk foreslåede overbegreb derfor forholdsvis ofte forkastes og et nyt tildeles: enten skal der peges på en af de andre betydninger som det foreslåede lemma har, eller på et helt nyt lemma som er mere præcist når det gælder opbygningen af et begrebshierarki for dansk.

Ud over i første omgang at opbygge et ordnet ud fra ordbetydningens relation til det nærmeste sproglige overbegreb skal det enkelte begreb (synsetet) også i samme ombæring tildeles en såkaldt 'ontologisk type'. På den måde indplaceres de danske ord i en ontologi, dvs, en standardiseret struktur der inddeler alle begreber i verden i bestemte klasser med særlige egenskaber. Efterfølgende kan man udtrække alle ord fra samme klasse, fx alle 'instrumenter'. Disse ord deler nemlig ontologisk type 'Instrument+Artifact+Object' selv om de har forskellige overbegreber, såsom fx VÆRKTØJ, INSTRUMENT eller MASKINE. Den ontologi der anvendes, er fastlagt af EuroWordNet-projektet, og de enkelte klassers (typers) navne er komponeret på en struktureret måde ud fra de betydningselementer som de tilknyttede begreber indeholder (oprindelse (fx 'Natural' eller 'Artifact'), fysisk udformning

(Fx 'Substance' eller 'Object'), eventuel komposition (dvs. om de iboende er en del af noget (som fx BORDBEN – angives med 'Part'), eller om de modsat udgøres af dele eller medlemmer (som fx FLOK – angives med 'Group') og endelig funktion (fx 'Instrument', 'Vehicle' eller 'Container'). Begrebet CYKEL har fx den ontologiske type 'Vehicle+Object+Artifact'.

I anden runde af den leksikalske beskrivelse, dvs. når alle hyponymirelationerne er etableret i wordnet'et, skal flere relationer tildeles de enkelte betydninger (fx meronymi), og betydningerne skal sammenkædes med ILI'en (InterLingual Index), jf. ovenfor (afsnit 3).

5. Ikke-taksonomiske hyponymer, systematisk polysemi og abstrakte betydninger

Vi er i arbejdet med at opbygge hyponymirelationer i DanNet-projektet stødt på en række udfordringer som vi her kort vil beskrive.

For det første har vi konstateret at der bag hyponymirelationen gemmer sig to ret forskellige relationer mellem under- og overbegreb som vi i det danske wordnet ønsker at kunne skelne imellem, bl.a. for at komme referenceproblemer til livs. Den ene hyponymirelation er den prototypiske overbegrebsrelation som vi vil betegne den taksonomiske hyponymirelation. Den angiver med et sprogligt udtryk at 'X er en bestemt type Y'. Fx er LADCYKEL en bestemt type CYKEL, og RACERCYKEL er en anden bestemt type CYKEL, og de to ko-hyponymer udelukker gensidigt hinanden: en LADCYKEL kan ikke samtidig være en RACERCYKEL. Den anden hyponymirelation vil vi sprogligt definere som relationen 'X er en hvilken som helst type Y om hvilken det gælder at Z'. Et eksempel er begrebet HAVELÅGE ("gammel cykel i dårlig stand"). Om denne type underbegreb gælder det at et hvilket som helst af ko-hyponymerne i princippet samtidig kan være en sådan X, modsat hvad der er tilfældet for de taksonomiske underbegreber. Både en LADCYKEL og RACERCYKEL kan godt samtidig være HAVELÅGER. I princippet skal egenskaber derfor kunne nedarves eller overføres fra et ikke-taksonomiske hyponym til dets taksonomiske ko-hyponymer. Underbegreber der har den særlige egenskab at de ikke udelukker deres ko-hyponymer, får tildelt et særligt træk i databasen. Det vil dog udelukkende være førsteordensentiteter i DanNet – de konkrete entiteter – der vil blive beskrevet på denne måde (jf. Pedersen og Sørensen 2006).

Det andet problem er systematisk polysemi, altså samme slags flertydighed der går igen for mange ord inden for samme betydningsområde. Systematisk polysemi er i DDO af pladmæssige hensyn behandlet meget forskelligt, alt efter om der er tale om et højfrekvent eller et lavfrekvent lemma. I DanNet skal de derimod behandles ens, uanset frekvens. Fx er

lemmaet 'bog' i Den Danske Ordbog udfoldet helt i ordbogsindgangen, således at der er angivet to særskilte betydninger: en for 'bog' som konkret genstand ("trykte el. skrevne blade af papir" etc., som i eksemplet *der står en bog på reolen*) og en for 'bog' som tekstrepræsentation ("tekst der står på disse blade", som i eksemplet *bogen var meget underholdende*). Langt de fleste af 'bog's mange underbegreber har derimod kun én betydning i DDO der så ved at referere til 'bog' som overbegreb muliggør begge fortolkninger hos læseren. Dette gælder fx i de to ordbogsartikler 'digtsamling': "bog der indeholder digte, som regel skrevet af samme forfatter" og 'debatbog': "bog der sætter et aktuelt emne til debat". Da man i DanNet til enhver tid skal kunne fremfinde en bestemt betydnings under- og overbegreber, er det nødvendigt at lave to synsets i de mange tilfælde hvor der i DDO indirekte peges på to mulige overbegreber. DEBATBOG og andre underbegreber til 'bog' vil således få to synsets, et med overbegrebet BOG_1 (genstandsbetydningen) og et med overbegrebet BOG_2 (tekstbetydningen). Andre eksempler på udbredt systematisk polysemi i dansk er fx ord for bygninger der ofte også har betydningen 'gruppe personer' og 'institution', dette ses fx ved begreberne HUS og SKOLE og alle deres underbegreber. Vi følger SIMPLE-projektets (Lenci et al. 2000) retningslinjer for behandlingen af systematisk polysemi i datamatiske semantiske ordbøger og beskriver de grupper der findes i det danske SIMPLE-leksikon.

Størrelse, sb.

1. Fysisk udstrækning som noget el. nogen har i én, to el. tre dimensioner SYN: omfang
 - 1.1. Hvert af flere mål som noget, især tøj, fremstilles i, og som udtrykker hvor meget el. hvor stor en enhed det kan rumme
2. Omfang af et forhold, et fænomen el. en faktor
 - 2.1. Grad hvori noget forekommer el. gør sig gældende, sådan som graden kan måles og udtrykkes i tal bl.a. om økonomiske forhold
 - 2.2. Grad af vigtighed, betydning, storslåethed e.l.

omfang, sb.

1. En masses el. et legemes udstrækning i bredde, længde, tykkelse, omkreds el. rumfang SYN: størrelse
2. det som noget fylder el. rummer; det som noget omfatter, indbefatter, indebærer e.l. SYN: udstrækning
 - 2.1. en vis mængde, udstrækning el. grad

udstrækning, sb.

1. En vis mængde el. grad af noget SYN: omfang
 - 1.1. det som noget omfatter, indbefatter, indebærer e.l. SYN: omfang
2. Den størrelse, længde, masse e.l. som noget har JF: omfang

Figur 5 Betydningerne i Den danske Ordbog for 'størrelse', 'omfang' og 'udstrækning'

Et tredje, ikke overraskende, problem gælder opbygningen af hyponymirelationerne for de abstrakte begreber. Mange abstrakte begreber er beskrevet med adskillige hoved- og underbetydninger i DDO, og det er svært at holde de mange betydningsnuancer ude fra hinanden når ordnettet skal opbygges, ligesom det er svært at gennemskue hvornår begreberne er synonyme udtryk. Man kan få en ide om vanskelighederne ved at kigge på betydningerne beskrevet i DDO for substantiverne 'størrelse', 'omfang' og 'udstrækning' (Figur 5 forrige side).

En mulig simplificering som overvejes, er at se helt bort fra underbetydninger og nøjes med at lave synsets og semantiske relationer for hovedbetydningerne samt at lave en meget flad hierarkisk struktur hvor mange abstrakte betydninger deler samme overbegreb.

6. Mulige anvendelser af DanNet

DanNet-ressourcen vil i sig selv udgøre en elektronisk begrebsordbog for dansk. Men først og fremmest vil den kunne understøtte en lang række sprogteknologiske værktøjer inden for fx informationssøgning, leksikografi og NLP (dvs. "natural language processing"). Ved informationssøgning tænker vi fx på søgemaskiner som Google og Jubii der finder relevante hjemmesider baseret på et eller flere søgeord. Hver søgemaskine har sin metode til at udvælge hvilke hjemmesider der anses som mest relevante, men fælles for dem er at mindst ét af de ord man angiver, skal optræde som tekst på hjemmesiden. Søger man på 'hund', findes hjemmesider som indeholder netop det ord.⁴ Men i praksis behøver hjemmesiden ikke indeholde ordet 'hund' for at være interessant ved en sådan generel søgning. Typisk vil man også være interesseret i hjemmesider der handler om schæfere, gravhunde eller pitbullterriere. Det er næppe hensigtsmæssigt at opremse alle de hunderacer der findes i verden, men med en leksikalsk ressource som DanNet er det muligt automatisk (og eventuelt skjult for brugeren) at udvide søgningen til også at omfatte alle underbegreber – og evt. synonymer – til et søgeord. Man kan også forestille sig søgemaskiner der udnytter andre relationer i DanNet og gør det muligt for eksempel at søge efter alle redskaber (dvs. underbegreber til begrebet REDSKAB) der har et skaft (dvs. har relationen *Has_meronym* SKAFT) eller bruges til at skære med (dvs. har relationen *For_purpose_of* SKÆRE).

Sammenkædningen med ILI'en (InterLingual Index) gør derudover at det bliver muligt

⁴ Nogle søgemaskiner har dog indbyggede ordbøger som sørger for at søgningen også leder efter bøjningsformer af søgeordet.

at søge på flere sprog samtidig. En søgning på 'hund' vil uden videre kunne finde engelske dokumenter hvor 'dog' indgår (og underbegreber), franske dokumenter hvor 'chien' indgår (og underbegreber) osv.

Vi forestiller os desuden at et dansk tekstkorpus på sigt kan annoteres automatisk med semantiske oplysninger der bygger på DanNet-ressourcen på samme måde som korpora i dag annoteres automatisk med bl.a. ordklasse. Hvis korpusset for eksempel annoteres med ontologisk type (jf. afsnit 4 ovenfor), vil man kunne foretage specifikke søgninger på fx verbet 'køre' efterfulgt af et objekt hvor kernen er angivet som Vehicle, eller – for at lede efter overførte betydninger – hvor kernen *ikke* er angivet som Vehicle. På længere sigt vil sådanne udbygninger af korpusser, mener vi, kunne være første spæde skridt til en egentlig automatisk betydningsdisambiguering af korpusmateriale og en automatisk tildeling af seletionsrestriktioner for verber.

Litteratur

- Dirven, Rene & Marjolijn Verspoor (1998) What's in a word: Lexicology. *Cognitive Exploration of Language and Linguistics*. Amsterdam/Philadelphia: John Benjamin Publishing Company.
- Fellbaum, Christiane, red. (1998): *WordNet: An Electronic Lexical Database*. MIT Press.
- Herslund, Michael og Viktor Smith (2003). Semantik. M. Herslund & B. Lihn Jensen, red. *Sprog og sprogbeskrivelse*. København: Samfundslitteratur, 83-126.
- Hjort, Ebba & Kjeld Kristensen (2003-5): *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab/Gyldendal.
- Lenci, A, N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonowsky, I. Peters, W. Peters, N. Ruimy, M. Villegas, A. Zampolli (2000) SIMPLE: A General Framework for the Development of Multilingual Lexicons. *International Journal of Lexicography*, XIII (4):249-263.
- Miller, George A. (1990) Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography* 3 (4):245-264.
- Pedersen, Bolette S. og Nicolai Sørensen (2006) Towards sounder taxonomies in Wordnets. Alessandro Oltramari, Chu-Ren Huang, Alessandro Lenci, Paul Buitelaar, Christiane Fellbaum (eds.) *Ontolex 2006*, pp. 9-16. Genova, Italien.
- Sofia, S, O. Kemal, P. Karel, C. Dimitris, C. Dan, T. Dan, K. Svetla, T. George, D. Dominique og G. Maria (2002) Balkanet: A multilingual semantic network for the balkan languages. *Proceedings of the 1st Global WordNet Association conference*.
- Vossen, Piek, red. (1998) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht, Holland: Kluwer.