


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Fra alfabet til begrep: Bokmålsordboka og NorNet	
Forfatter:	Ruth Vatvedt Fjeld, Rune Lain Knudsen & Julie M. Torjusen	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 190-201 Rapport från Konferens om lexicografi i Norden, Lund 24.-27. maj 2011	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i lexicografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Fra alfabet til begrep: Bokmålsordboka og NorNet

Ruth Vatvedt Fjeld, Rune Lain Knudsen & Julie M. Torjusen

This article discusses the use of a monolingual dictionary (*Bokmålsordboka*) as basis for a wordnet for Norwegian (NorNet). The first part of the article describes the method for constructing the wordnet, and discusses the structure of NorNet with regard to synsets, unique beginners and top ontology. The second part of the article discusses some of the problems that arise due to inconsistencies in the dictionaries, and how these are solved in NorNet. The selection of entries for the wordnet is also discussed. Finally, the article describes plans and future developments. ction of entries to the wordnet is also discussed. Finally, the article accounts for plans and future developments.

NØKKEWORD: ordnett, merge-tilnærming, ontologi

1. Ordnett generert fra en ordbok – og ordbok rettet ved hjelp av et generert ordnett

Et ordnett er en datamaskinleselig beskrivelse av de semantiske relasjonene mellom et utvalg ord. Et tidlig moderne ordnett er Princeton WordNet (PWN) for amerikansk-engelsk, seinere er EuroWordNet (EWN) utviklet, der mange europeiske språk er koblet sammen via engelsk, og det er et mål at også det danske DanNet skal gå inn i EWN. I Sverige holder man på å utvikle ordnettet Swesaurus, i Finland har man oversatt PWN til finsk, og på Island er man også i gang med å lage et ordnett. Denne artikkelen skal handle om arbeidet med å utvikle et norsk ordnett kalt NorNet. Ordnettet utvikles ved Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo, avdeling for bokmålsleksikografi.

I Nygaards hovedoppgave *Frå ordbok til ordnett* (Nygaard 2006, Fjeld & Nygaard 2009) og ble semantiske relasjoner mellom substantiver automatisk generert ved hjelp av enkle regler som analyserte definisjonene i *Bokmålsordboka* (BOB) (Wangensteen 2005). Tabell 1 viser noen eksempler.

Regel	Om definisjonen består av ett enkeltstående substantiv eller flere enkeltstående substantiv i lemmaform adskilt med komma, er disse substantivene synonym til definiendum.
Eksempel	vidde [...] <u>område</u> , <u>areal</u>
Resultat	vidde SYNONYM_FOR område vidde SYNONYM_FOR areal
Regel	Dersom definisjonen har andre elementer enn enkeltstående substantiv, er det første substantivet i lemmaform overbegrep (hyperonym) til definiendum, med mindre dette substantivet står i en stoppliste (en liste med ord som ikke skal være underbegreper (hyponym)).
Eksempel	bistro m1 (fr 'vertshus(holder)') liten <u>restaurant</u>
Resultat	restaurant HYPERNYM_FOR bistro

Tabell 1: Eksempler fra det opprinnelige regelsettet og informasjonen som blir trukket ved hjelp av dette. Ord som fanges opp av regelen og får tildelt relasjoner er understreket.

NorNet ble automatisk generert i 2008, med utgangspunkt i et sett av slike regler som i Tabell 1 og definisjonene i BOB. Foreløpig inneholder ordnettet 37787 substantiv, 15516 overbegreps-/underbegrepsrelasjoner og 7172 synonymirelasjoner. Siden ordnettet er automatisk generert, inneholder det en del feil og mangler, og gjennomgår derfor en grundig redigeringsprosess. Når det arbeidet er ferdig, vil vi ha et godt grunnlag for videre utvidelse av ordnettet, og vi kan begynne å teste ut nye metoder for automatisk uttrekning av semantiske relasjoner. Vi vil også få et håndfast mål på hvor godt reglene som genererte ordnettet, hentet ut de ønskede relasjonene fra substantivsdefinisjoner.

Alle lemmaene i ordnettet er knyttet til ordboka gjennom et felles ID-nummer. Redigeringsarbeidet gir også som resultat at vi kan forbedre ordboka. Mange av feilene som forekommer i det uredigerte ordnettet, skyldes inkonskvenser i definisjonene i BOB. Ordboka er redigert med semasiologisk tilnærming og alfabetet som sorteringsprinsipp. Resultatene fra redigeringa av NorNet vil dermed kunne brukes til å gjøre ordboka mer konsekvent og systematisk. Dette kommer vi tilbake til i punkt 5. Arbeidet med redigeringa har også medført en rekke observasjoner av hvordan algoritmene og implementasjonen av dem har fungert.

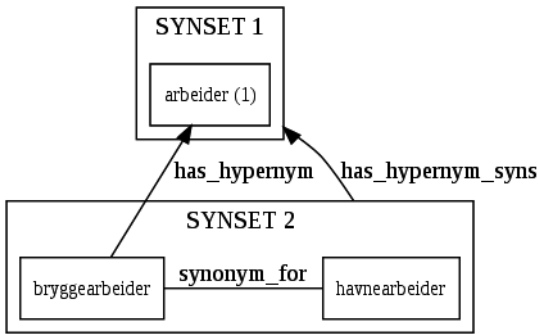
Foreløpig er ca. 25 % av under-/overbegreps- og synonymirelasjonene mellom delbetydningene rettet. Rettinga har i hovedsak blitt utført av Julie M. Torjusen. I første omgang har hun systematisk kontrollert den automatiske tilde-
linga av relasjoner mellom delbetydninger som ble konstruert ved hjelp av det opprinnelige regelsettet og slettet ukorrekte relasjoner som programmet lagde. En vanlig feil er at reglene ikke har skilt mellom homonymer, slik at for eksem-

pel adjektivet *kort* har blitt tolka som substantivet *kort*. Det har ført til at ord som *andakt* med definisjonen ”kort tale med utgangspunkt i en bibeltekst” har fått *kort* i betydninga ” lite rektangulært blad av papir, papp el. plast som brukes til forskjellige bestemte formål” som overbegrep. Til nå har det blitt utført i alt 3846 rettinger. 1788 av disse er korrigeringer av genererte relasjoner, de resterende 2058 rettingene er innsetting av manglende relasjoner. Dette arbeidet har vist at samsvaret mellom de automatisk genererte relasjonene og den manuelle redigeringa er på under 70 %, altså at godt over en fjerdedel av de automatiske uttrekkingene må rettes på et eller annet nivå. I tillegg indikerer antall innsettinger av manglende relasjoner en noe lav dekningsgrad. Anslaget for samsvar kan virke noe lavt, men siden det gjenstår ca. 75 % som vi ikke har undersøkt, er dette foreløpig ikke godt nok grunnlag for å vurdere algoritmens kvalitet. Det er grunn til å tro at resten av de automatisk genererte relasjonene vil ha omtrent like mange feil, men vi kan ikke si noe sikkert om dette før vi har gransket en større del av materialet.

2. Synsetta i NorNet

Synonyme delbetydninger har en symmetrisk relasjon SYNONYM_FOR mellom seg (se figur 1). Den samlede mengden av synonymer for et begrep utgjør ett synsett. En delbetydning kan ha overbegrepsrelasjon til en annen delbetydning, slik at den ene delbetydninga er overbegrep for den andre. Dette danner grunnlaget for to lag med overbegrepsrelasjoner: mellom delbetydninger isolert og mellom synsettene de er en del av. Overbegrepsrelasjonene mellom synsetta vil utledes fra overbegrepsrelasjonene funnet for de enkelte delbetydningene i et synsett. Overbegrepsrelasjonen for et synsett vil velges ut ifra hvilken delbetydning som oftest er angitt som overbegrep for de synonyme delbetydningene. En delbetydning som ikke har fått tildelt noen egne overbegrepsrelasjoner, vil dermed kunne få dette ved hjelp av synsettet den er en del av. Denne formen for synsettklassifisering vil vise nyanser mellom synonymer som får ulike overbegreper, samt kartlegge inkonsistens og konflikter i overbegreps-hierarkiene. Et eksempel på overbegrepsrelasjoner mellom delbetydninger i forhold til overbegrepsrelasjoner mellom synsett er gitt i figur 1.

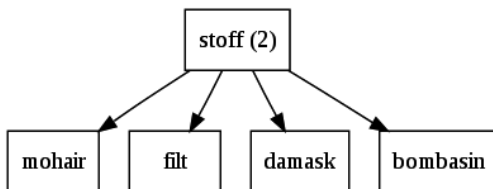
I synsett 2 er det bare *bryggearbeider* som har fått tildelt overbegrepsrelasjon. Basert på dette defineres synsett 1 som overbegrep for synsett 2. Underbegrepet *havnearbeider* i synsett 2 blir inkludert i overbegreps-hierarkiet fordi den er del av samme synsett som *bryggearbeider*.



Figur 1: Eksempel på relasjoner mellom delbetydninger i forhold til relasjoner mellom synsets.

3. Unike begynnere

Unike begynnere er betegnelsen for de lemmaene som står øverst i ordnettets overbegreps-hierarki, dvs. lemma som ikke selv har noen overbegreper (Miller 1998:28). Dette er illustrert i figur 2. NorNet har til nå i alt 22887 substantiv som ikke er tildelt noen overbegreper i ordnettets. Av disse begrepene har 20810 heller ikke noe underbegreper, altså delbetydninger som ikke har fått tildelt overbegreps-/underbegrepsrelasjoner og dermed er helt isolerte i forhold til ordnettets. Disse klassifiseres ikke som unike begynnere før de har fått tildelt én eller flere relasjoner. Av disse har imidlertid 5043 fått synonymirelasjoner og får derfor enten overbegrep ved hjelp av ett eller flere synsett-hierarkier, eller de blir en del av et synsett som er klassifisert som en unik begyner.



Figur 2: Eksempel på en unik begyner. Enkelte underbegreper er utelatt på grunn av plass-hensyn.

Når vi trekker fra antall lemmaer som verken har overbegreps- eller underbegrepsrelasjoner individuelt, står vi igjen med 2077 unike begynnere med ett eller flere underbegreper.

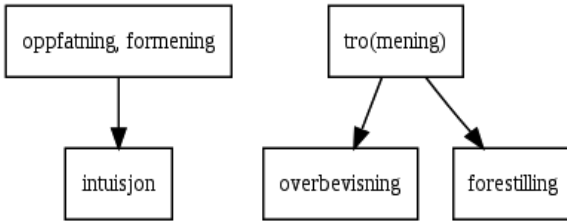
Lemmaer med høyest antall overbegreper er listet opp i Tabell 2.

Hyp.fr	Lemma.
1114	person
76	grunnstoff
72	stoff (1)
64	lyd
56	område
53	tid, ord
44	måte
42	stoff (2)
41	bevegelse, gjenstand
38	tilstand
36	myntenhet
34	enhet
33	musikkstykke
31	rekke
30	handling

Tabell 2: Liste over unike begynnere med 30 eller flere underbegreper. De med likt antall underbegreper står på samme rad adskilt med komma

Her ser vi at lemmaet *person* skiller seg ut med hele 1114 underbegrepsrelasjoner. Dette ser ut til å samsvare med en undersøkelse av samme lemma i DanNet (Pedersen og Braasch 2009), der 4246 lemmaer har *person* som nærmeste overbegrep. Siden underbegrepsfrekvensen til *person* er såpass høy, og siden begrepet har vært gjenstand for studier i DanNet, er personbetegnelser et semantisk felt vi ønsker å fokusere ekstra på, særlig for å kunne sammenligne med beskrivelsen i DanNet. Blant annet vil vi undersøke i hvor stor grad vi kan redusere antall umiddelbare underbegreper ved å innordne dem i et mer detaljert hierarki.

Isolerte begrepsøyer er klynger av begreper som er knyttet sammen ved hjelp av semantiske relasjoner seg imellom, men som mangler relasjoner som knytter de opp mot resten av ordnettet, jf. figur 3.

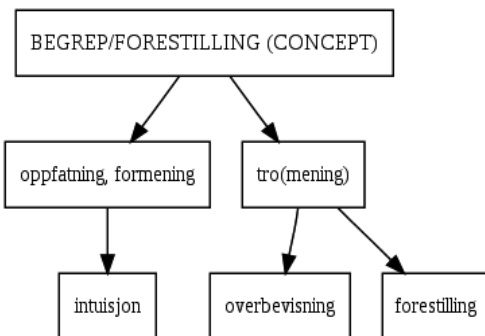


Figur 3: Eksempel på isolerte begrepsøyer/unike begynnere

Dette gjør disse klyngene usynlige i mange søk, siden stien av relasjoner fra ett begrep til neste blir avbrutt. Arbeidet med å redusere antall unike begynnere ved å få dem ordnet i et bedre hierarki er derfor av høy prioritet. Et lavt antall unike begynnere er en god indikasjon på et lavt antall isolerte begrepsøyer, og dermed et “tettere” nett av begreper som gir mer komplett informasjon ved søk i ordnettet.

4. Toppontologi

En toppontologi for ordnett forsøker å beskrive overbegreper som samler så mange underbegreper som mulig under seg i en hierarkisk struktur. Å bygge opp en dekkende og veldefinert toppontologi er viktig for å redusere antall isolerte begrepsøyer i ordnettet. EWN har utviklet en toppontologi som er basert på sammenligninger av de ulike toppontologiene i alle EWN-ordnetta. Vi mener at denne toppontologien er et godt utgangspunkt for å strukturere Nor-Net. Med en slik felles toppontologi blir det enklere å slå sammen, sammenligne og kommunisere mellom ordnett for forskjellige språk. Vi har derfor tatt utgangspunkt i EWNs toppontologi og bygger en kjerneontologi for NorNet som er mest mulig i samsvar med den, jf. figur 4.

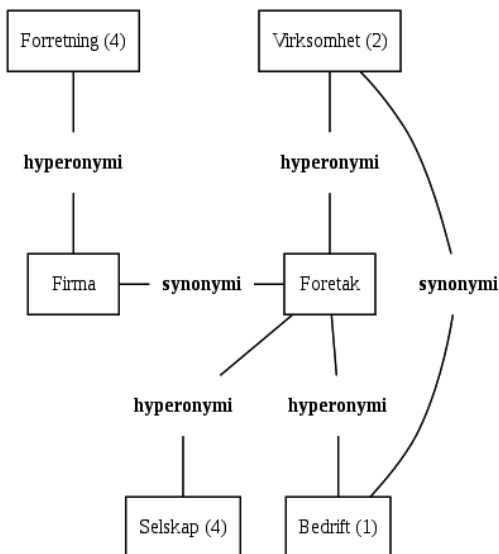


Figur 4: Eksempelen fra figur 3 knyttet sammen ved hjelp av et begrep fra en toppontologi.

5. Redigeringsarbeidet

Som nevnt i punkt 1 er NorNet basert på BOB. Det er en enspråklig allmennordbok som først og fremst er ment for menneskelige brukere, og det har ført til enkelte problemer når definisjonene i BOB har blitt lest av et dataprogram. I dette punktet skal vi gå gjennom noen av disse problemene og hvordan vi har prøvd å løse dem. Vi skal også sammenlikne problemene vi har hatt i arbeidet med NorNet med problemer som har dukka opp i arbeidet med DanNet (Pedersen m.fl 2009). Det gjelder inkonsekvens i valg av overbegrep, ettordsdefinisjoner som ikke er synonymmer og ad hoc-sammensetninger. Vi skal også diskutere lemmaseleksjon i ordnettet (jf. punkt 6).

Et av problemene når ordnett konstrueres ut fra definisjonene i en ordbok, er at ord som tilhører samme semantiske felt, ofte kan ha ulike overbegreper i definisjonene. Dette ser vi f.eks. ved synsettet {firma; selskap (4); bedrift (1); virksomhet (2); forretning (4); foretak}¹. Ifølge definisjonene i BOB er *firma* underbegrep til *forretning* og synonymt med *foretak*, mens *selskap* er underbegrep til *foretak*. Det samme gjelder *bedrift*, men *bedrift* er i tillegg synonymt med *virksomhet*. *Virksomhet* er i tillegg overbegrep til *foretak* (se figur 5).



Figur 5: Synsettet {firma; selskap (4); bedrift (1); virksomhet (2); forretning (4); foretak}

¹ Vi bruker notasjonen {} for synsett (Miller 1998).

Selv om en slik betydningsbeskrivelse kan gjenspeile vanlige språkbrukeres forståelse og bruk av disse ordene og dermed fungere godt i en tradisjonell ordbok, er dette er en kaotisk løsning for et ordnett. Ved liknende tilfeller i DanNet har man valgt å slå sammen alle begrepene i ett synsett (Pedersen m.fl 2009:276). En slik løsning har vi også valgt for NorNet. Dermed vil f.eks. et søk på *firma* også gi treff på de andre lemmaene i synsettet.

En annen type inkonsekvens i definisjonene i BOB finner vi ved ettordsdefinisjoner, som ifølge vårt regelsett blir tolket som synonymdefinisjoner. Men i en del tilfeller viser det seg at det ikke stemmer, som i betydning 3 av *bue*, jf. figur 6:

<u>boge</u>	I bue el. I boge <u>m1</u> (da. <i>bue</i> , norr <i>bogi</i> , besl med <i>III bøye</i>)
<u>bue</u>	
	1 krumning, kurve <i>sirkelb-</i> , <i>ellipseb-</i>
	2 hvelving i byggverk <i>rundb-</i> , <i>spissb-</i> I <i>romanske og gotiske b-r</i>
	3 skytevåpen <i>pil og b-</i> / overf. <i>Spenne b-n for høyt sikte</i> for høyt, overvurdere seg selv
	4 del av strykeinstrument <i>fiolinb-</i>

Figur 6: Definisjon og betydningsinndeling av bue/boge i Bokmålsordboka

Logisk sett er *bue* underbegrep til skytevåpen, ikke synonym. Dette må rettes manuelt i NorNet. I denne retteprosessen oppdager vi altså en del oppslag i ordboka som bør endres og gjøres mer konsistente. I dette eksempelet ser vi at betydning 4 viser en meronymi-relasjon ("del av strykeinstrument"). Det burde betydning 3 også gjøre, for en bue er ikke et skytevåpen uten en pil. Viss denne definisjonen blir endra til "del av skytevåpen", vil den ikke lenger være en ettordsdefinisjon og ifølge vårt regelsett ikke et synonym.

BOB inneholder mange ad hoc-sammensetninger, dvs. sammensetninger som brukes som overbegreper i definisjonene, men som ikke er egne oppslagsord i ordboka. I og med at disse sammensetningene ikke fins i ordboka som egne lemmaer, har ikke programmet funnet noe overbegrep for dem. Dette ser vi i definisjonene til lemmaene *biedermeier*, *buttskjøt* og *bøffelgress*, jf. figur 7.

biedermeier	biedermeier <i>-en</i> (utt bidermeier el. <i>-maier</i> ; ty. 'spissborger', omdannet av <i>Biedermann</i> og <i>Bummelmaier</i> , to ty. besteborgertyper) møbelstil fra første halvdel av 188-tallet
buttskjøt	buttskjøt sveiseskjøt mellom jernplater som ligger butt i butt (I, 1)
bøffelgras	<i>-gras</i> el. <i>-gress</i> småvokst præriegress som lager tette tuer el. <i>Matter</i> , <i>Buchlöe</i>
bøffleggess	<i>dactyloides</i>

Figur 7: Eksempler på lemmaer uten overbegrep

Verken *møbelstil*, *sveiseskjøt* eller *præriegress* er egne oppslagsord i BOB. Dette skaper et hierarkisk problem, fordi vi får unike begynnere som ikke hører til høyt oppe i hierarkiet. Dette kan løses ved å gi lemmaene overbegrep etter det siste leddet i sammensetninga, som for de to siste eksemplene er identiske med sisteleddet i lemmaet. Ved *biedermeier* blir *stil* overbegrep, og dermed får man ikke noe mellomnivå mellom *biedermeier* og *stil*, men man slipper å få *biedermeier* som unik begyner.

Lorentzen og Nimb drøfter ad hoc-sammensetninger i ordbøker som danner basis for ordnett i sin artikkel om forholdet mellom tradisjonelle ordbøker og ordnett. De foreslår to mulige løsninger: å opprette sammensetningene som egne begreper i ordnettet, eller å erstatte dem med ord som allerede fins i ordnettet (Lorentzen og Nimb 2010:332). Når det er snakk om ad hoc-sammensetninger, er det trulig mindre hensiktsmessig å opprette dem som egne begreper i et ordnett. De står ikke på egne bein i samme grad som ord som har gjennomgått en utvelgingsprosess for lemmastatus i en ordbok. Dette kan eventuelt undersøkes ved hjelp av korpus.

6. Lemmaseleksjon

Lemmaseleksjonen i et ordnett kan gi litt andre konsekvenser enn lemmaseleksjonen i en vanlig ordbok fordi det er en viss forskjell på lemmaenes synlighet i dem. I en ordbok (særlig elektronisk) er det mange ord man sjelden vil se i bruk, med mindre man søker spesifikt etter dem. I et ordnett vil man derimot få vist mange ord man ikke har søkt etter, fordi de står som over- eller underbegreper til et ord man har søkt etter. I NorNet er lemmaseleksjonen basert på definisjonene av substantivene i BOB, og ofte er det ikke nærmeste overbegrep som står først i definisjonen, men et overbegrep lenger oppe i det ontologiske hierarkiet. Det fører i en del tilfeller til at enkelte lemmaer får veldig mange underbegreper (jf. tabell 2). Leksikografisk Bokmålskorpus (LBK) er et balansert, lingvistisk annotert korpus for moderne bokmål, som er en viktig ressurs for lemmaselek-

sjonen til NorNet. Et mye brukt overbegrep i BOB er for eksempel *evne* (1), som bare under bokstaven a har underbegrepene *absorberingsevne*, *absorpsjonsevne*, *abstraksjonsevne*, *akselerasjonsevne*, *anskuelsesevne*, *arbeidsevne* og *avdråttsevne*. Mange av disse lemmaene er marginale i faktisk språkbruk, ved søk i LBK finner man kun ett treff på *absorberingsevne*, ett treff på *absorpsjonsevne*, fire treff på *abstraksjonsevne*, to treff på *akselerasjonsevne* og tolv treff på *arbeidsevne*. De øvrige lemmaene får man ingen treff på. Hvorvidt så sjeldne ord skal tas opp som lemmaer i et ordnett, kommer an på hva ordnettet skal brukes til. Ved oppbygging av et nytt ordnett, mener vi det er viktigere å ta med mer sentrale deler av ordtilfanget først, så ord med svært lav eller ingen forekomst i LBK blir merket for undertrykkelse ved visning, men lemmatiseres i NorNet som kandidater for et fulldekkende ordnett i framtida. På denne måten bruker vi LBK som datagrunnlag for en statistisk analyse av delbetydningene for å anslå NorNets dekningsgrad og relevans for moderne bokmål.

7. Metodevalg

Eksempelene i punkt 5 viser at det er umulig å lage et ordnett med utgangspunkt i en ordbok uten å bruke tid på manuell retting. Et alternativ kan være å oversette et ordnett laget for et annet språk, noe som er ansett som en raskere og billigere metode, og som samtidig gjør at ordnett for forskjellige språk blir kompatible (Pedersen m.fl. 2009:271). Men ved å oversette fremmedspråklige ordnett risikerer man at strukturer som fins i målspåket går tapt, fordi de ikke fins i kildepråket. Viss vi skulle oversatt DanNet til norsk, ville det bl.a. bli problemer med de mange sammensetningene i norsk og dansk som har forskjellig bruk og betydning. I DanNet er *brusebad* et *bad*, mens en *dusj* er en *sprut* i NorNet, og *killling* er et *pattedyr* i DanNet, mens en *kattunge* er en *unge* (1). I NorNet kan *flom* bety en stor mengde, som i *tåreflom* og *ordflom*, men denne betydninga fins ikke i DanNets søkefunksjon andredord.dk. Det viser seg altså ved sammensetninger at dansk og norsk har forskjellige begrepsstrukturer, som man ikke vil få fram ved å oversette det danske ordnettet til norsk. Man må belage seg på mye manuelt ekstraarbeid også om man velger å oversette ordnett. Basert på observasjonene vi selv har gjort fra arbeidet med NorNet og hva andre som har laget ordnett har rapportert, mener vi at behovet for etterarbeid er høyt uansett hvilken metode man velger. Vi har derfor valgt å fokusere på metodeutvikling, da vi mener at dette vil være verdifullt ikke bare for utviklingen av NorNet, men også for utvikling av relaterte leksikalske ressurser.

8. Videreutvikling av NorNet

Planen fremover er å inkludere flere ordklasser og relasjonstyper, både mellom delbetydninger internt innenfor hver ordklasse og der avledninger og sammensetninger under samme begrepsfelt koples sammen på tvers av ordklassene. Vi vil eksperimentere med forskjellige teknikker for uttrekking av semantisk informasjon, både fra ordboksdata og naturlig tekst, og dokumentere resultatene. Eksperimentene vil basere seg på regelbaserte metoder, statistiske modeller og kombinasjoner av disse.

Definisjonsspråket i tradisjonelle ordbøker er dels formelt, dels naturlig, og det er derfor interessant å analysere – spesielt med regelbaserte metoder. Ved å forsøke å utvikle en generell metode som i stor grad kan automatiseres, vil det åpne for nye måter å hente ut informasjon fra liknende, strengt strukturerte teksttyper på, også om de i utgangspunktet ikke ble laget med maskinlesbarhet i tankene. På bakgrunn av disse observasjonene er analyse og klassifisering av definisjonene i Bokmålsordboka under utvikling. Denne analysen baserer seg på ordklassemønstrene for definisjonene og vil kunne si noe om konsistens ved hjelp av statistikk over unike eller sjeldne definisjonsmønstre, avvik fra de vanligste mønstrene mm. Resultatet av denne analysen danner også grunnlaget for utvikling av flere regelbaserte metoder for uttrekking av semantisk informasjon som vil bli brukt i NorNet.

LITTERATUR

- Andre ord, ved Anders Johanssen. <http://andreord.dk/>
- BOB = Wangensteen, Boye, 2005: Bokmålsordboka, Kunnskapsforlaget, Oslo
- Fjeld, Ruth Vatvedt & Lars Nygaard (2009): NorNet – a monolingual wordnet of modern Norwegian. I: B.S Pedersen, A. Braasch, S. Nimb & R. V. Fjeld (red.): Proceedings of the NODALIDA 2009 workshop, NEALT, s. 13–16
- LBK=Leksikografisk Bokmålskorpus (<http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html>)
- Lorentzen, Henrik & Sanni Nimb, 2010: Fra ordbog til wordnet. Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse? I: H. Lönnroth og K. Nikula (red.): Nordiske studier i leksikografi 10, s. 329–344
- Miller, George A. 1998: Nouns in WordNet. I: C. Fellbaum (red.): WordNet – An electronic database, MIT Press, Cambridge, s. 23–46
- Nygaard, Lars, 2006: Frå ordbok til ordnett, upublisert hovedoppgave, UiO
- Pedersen, Bolette Sandford & Anna Brasch, 2009: What do we need to know about humans? Viby, Mnemo, s.158–165
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen & Henrik Lorentzen, 2009: DanNet: the challenge of compiling a

wordnet for Danish by reusing a monolingual dictionary. I: Language Resources and Evaluation 43, s. 269–299.

Ruth E. Varvedt Fjeld

Institutt for lingvistiske og nordiske studier, Oslo universitet
r.e.v.fjeld@iln.uio.no

Rune Lain Knutsen

runelk@ifi.uio.no

Julie M. Torjusen

julieimt@student.ilos.uio.no