

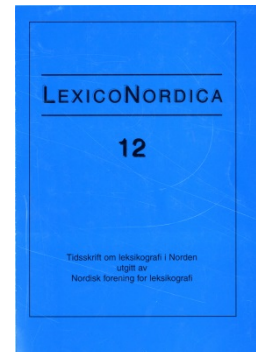
LexicoNordica

Titel: Virtuelle perspektiver for ordbogsredigering:
muligheder, strategier og virkelighedens begrænsning

Forfatter: Lars Trap-Jensen

Kilde: LexicoNordica 12, 2005, s. 109-121

URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

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 LexicoNordica (1-16) 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.

Lars Trap-Jensen

Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning

The electronic medium offers improved search facilities and allows much more information in comparison with the paper dictionary. Some of these possibilities are outlined and discussed in relation to ordnet.dk, a Danish electronic reference work that combines three existing lexicographic resources into one site available on the Internet. The project involves digitalization of the 28-volume historical dictionary *Ordbog over det danske Sprog* (ODS, Dictionary of the Danish Language) and integrating it with its 5 supplementary volumes. The ODS together with the modern, corpus-based *Den Danske Ordbog* (The Danish Dictionary) provide dictionary coverage of the Danish language from c. 1700 to the present. The integration of the two dictionaries with the corpus web site *Korpus 2000* will enable the user to access further linguistic material, and will enable lexicographers to utilize corpus material in new and more focused ways in the individual entries. Given the limits of a medium-sized project (4–5 lexicographers, 6 years) the editorial staff need to make a number of strategic choices among the options available. At present, priority is given to developing onomasiological search facilities and exploring ways of integrating corpus and dictionary material.

1. Mediebetingede forskelle på papir- og digitalt format

Elektroniske ordbøger har efterhånden været så længe på markedet at de er en naturlig del af ordbogslandskabet. Alligevel er der grund til at betragte dem som nye og forholdsvis ubeskrevne størrelser i den sammenhæng. Dels er metaleksikografisk beskrivelse og teori endnu sparsom (jf. dog *Hermes* nr. 34, 2005), dels er langt størstedelen af de elektroniske ordbøger vi har set indtil nu, ikke stort mere end skærmversioner af eksisterende papirordbøger, måske med tilføjelse af visse udvidede søgemuligheder, fx fritekstsøgning og søgning med jokertegn. Der er endnu ikke udkommet mange elektroniske ordbøger som fra begyndelsen har været konciperet udelukkende til udgivelse for det elektroniske medie. Et eksempel man kunne nævne, er det tyske *Elexiko*-projekt, der udvikles ved Institut für Deutsche Sprache i Mannheim. Der er imidlertid tale om et projekt i en indledende fase, hvilket gør en retvisende vurdering tvivlsom: Redigeringen har i skrivende stund netop rundet 1 promille af de bebudede 300.000

lemmaer. Ellers er det især learners' dictionaries på det stærkt konkurrenceprægede, engelsksprogede marked der er kommet længst med hensyn til at udnytte det elektroniske medies muligheder – selvom det i vistnok alle tilfælde er med trykte ordbøger som udgangspunkt.

Helt grundlæggende er der nogle overordnede, mediebestemte betingelser som adskiller papirordbøger fra elektroniske ordbøger:

1. Plads er ikke noget problem. Det betyder at der ikke er grænser for hvor mange lemmaer der kan bringes, hvor mange oplysningstyper, eksempler m.m. der kan gives. Ud over at pladsen ikke er begrænset, er den heller ikke bundet af bogens to dimensioner og lineære forløb. I det omfang der er en makrostruktur, er den altså illuderet – fx i form af en alfabetisk ordnet liste i den ene side af skærmen. Klassiske leksikografiske problemer som placering af flerordsforbindelser er af samme grund ophævede: Man bør kunne skrive hvad som helst i søgefeltet og komme det ønskede sted hen. Ligeledes bør den elektroniske ordbog yde hjælp ved fejlstavning.

2. Muligheden for vha. links at kalde andre resurser frem på skærmen betyder at a) man kan inddrage flere værker i ét (internt) eller pege udad mod andre (eksternt), b) man kan bruge henvisninger og teleskopløsninger (hvor man folder oplysningerne ind og ud) og c) man kan anvende andre typer filer end ordbogens tekstelementer, fx billed-, lyd- og filmsekvenser.

3. Det elektroniske medie byder på langt bedre muligheder for søgninger, ikke bare på tekststrengene, men også fx på specifikke elementnavne og onomasiologiske søgninger.

I det følgende vil jeg redegøre for de muligheder og valg vi står over for i forbindelse med et nyt dansk onlineprojekt, hvilke overvejelser vi har gjort os, og hvilke strategiske løsninger vi vælger at satse på. Inden da er det dog nødvendigt i korte træk at præsentere projektet.

2. *ordnet.dk* – et elektronisk opslagsværk

Der er også her tale om et opslagsværk som – i hvert fald delvis – tager udgangspunkt i eksisterende papirversioner, men ikke desto mindre har vi en ambition om at kunne være nyskabende, dels ved at sammenkoble forskellige eksisterende resurser til ét nyt opslagsværk, dels ved at udnytte det elektroniske medie til at tilbyde nye funktioner i netversionen.

Der bliver i første omgang tale om at forbinde ordbogsdata fra Den Danske Ordbog med en digital version af Ordbog over det danske Sprog og med mulighed for søgning i Det Danske Sprog- og Litteraturselskabs korpusmateriale¹.

Med digitaliseringen af Ordbog over det danske Sprog går et længe næret ønske i opfyldelse for Det Danske Sprog- og Litteraturselskab (for ikke at tale om mange af værkets trofaste brugere)². Digitaliseringsprocessen indeholder i sig selv en ikke ubetydelig udfordring, og ambitionen for denne dels vedkommende begrænser sig derfor i første række til at etablere en fuldt funktionsdygtig digital version hvor artikelstrukturen er opmærket så den lader sig konvertere til XML og til et databaseformat med de muligheder for tværgående søgninger som derved fremkommer, og hvor de fem supplementsbind flettes sammen med grundmanuskriptet. Hvor langt ned i mikrostrukturen det vil være muligt at genkende de relevante indholdselementer, kan i sagens natur ikke afgøres på forhånd – det bliver en del af projektet at afdække mulighederne.

Den Danske Ordbog foreligger allerede som digitalt manuskript, og det er derfor også disse data der kommer til at ligge til grund for eksperimenter med nye funktioner og udvidede søgemuligheder. Desuden vil manuskriptet blive redaktionelt vedligeholdt på to områder: Der skrives løbende nye artikler så lemmabestanden hele tiden udvides og opdateres, og de betydningsmæssigt gennemskuelige underopslagsord (i papirversionen anbragt i artiklernes fod som sammensætninger og afledninger) gøres til hovedopslagsord der som minimum indeholder oplysning om ordklasse og bøjning samt en definition.

Også korpusdelen bliver dynamisk ved at der løbende føjes nye tekster til de i forvejen eksisterende i Korpus 2000 og Korpus 90 (sidstnævnte er en delmængde af Den Danske Ordbogs tekstkorpus). Ud over integrationen med ordbogsdelen indgår det også i projektet at korpus-siden udvikles med en forbedret grænseflade og med udvidede søgemuligheder, bl.a. sådan at teksternes syntaktiske opmærkning udnyttes.

Projektet gennemføres i perioden 2004–2009 og finansieres af Carlsbergfondet og Kulturministeriet med støtte fra Det Elektroniske Forskningsbibliotek.

¹ En nærmere beskrivelse af projektet kan ses i Trap-Jensen (2004) og Lorentzen/Trap-Jensen (under udgivelse).

² Se Bojsen/Trap-Jensen (2005) for yderligere oplysninger om denne del af projektet.

3. Elektroniske muligheder

I det følgende gennemgås en række af de muligheder som foreligger alene som følge af overgangen fra papir til digitalt format, kombineret med koblingen mellem korpus og ordbogsdata. Fokus vil primært være på ordbogsdelen, hvilket betyder at korpusdelen kun inddrages i det omfang den er relevant for ordbogsdelen.

3.1. Generelle søgemuligheder

Da samtlige data ligger i en database (for ODS' vedkommende dog ikke endnu), er det oplagt at tilbyde alle de muligheder for tværgående søgninger som databaseformatet tillader, dvs. søgning på hvert enkelt af basens elementnavne. Ligeledes er det en selvfølge at systemet skal kunne håndtere fritekstsøgninger, trunkeerede søgninger og søgning med forskellige booleske operatører. Om det skal ske med brug af jokertegn eller i form af valg fra en rullegardinmenu, er i den forbindelse sekundært. Endelig bør man kunne tilbyde brugerne hjælp til opslag ved at udarbejde en algoritme der kommer med relevante alternativer i tilfælde af at der indtastes en fejlstavet form i søgefeltet.

3.2. Udvidede oplysninger i ordbogsartikler

Den elektroniske ordbog giver mulighed for at arrangere allerede eksisterende oplysninger på en ny måde og tilbyde dem som en mulighed på et relevant sted i artikelstrukturen hvis brugeren måtte ønske det. Desuden giver specifikke korpussøgninger på stedet mulighed for at få supplerende oplysninger om et konkret problem. Man kan som et første skridt foretage det tankeeksperiment at opregne de søgefunktioner man kan forestille sig alene ud fra en teknisk synsvinkel, dvs. uden at bekymre sig for om det er hensigtsmæssigt set fra et brugersynspunkt. I fig. 1 illustreres nogle af mulighederne i en tilfældigt udvalgt ordbogsartikel³.

³ Denne fremstilling er en let bearbejdet version af en figur der er udviklet af min kollega Jørg Asmussen og oprindeligt brugt i anden sammenhæng.

mus sb. fk.

[1] -en, -, -ene [2][HmuZs], i de fleste sms. *mouse-* [HmusE-]

[3]

1 lille gnaver [4] med tilspidset snude, forholdsvis store ører og en ofte lang, nøgen hale [5]; SYN [6] (familien) Muridae [7]; omfatter bl.a. husmus, skovmus, rotter og dværgmus * *lille* ~ [8], ~ og rotter; *forsøg med* ~ [9] * *Mange lofter, der førhen var fulde af hø, halm og fede mus, står i dag gabende tomme* [10] BerlT90 [11] * *Det er ligegyldigt, om katten er hvid eller sort, bare den fanger mus* BT88

2 (*edb*) [12] lille redskab som bevæges med hånden på en vandret flade (ofte en lille måtte), og som sætter brugeren i stand til at flytte markøren på en edb-skærm, aktivere menuer, ikoner m.m. [13]; kendt fra 1983 [14] * *Det traditionelle greb om musen betyder ofte, at albuen tvinges ud fra kroppen i en unaturlig og anstrengende stilling* HiFi92

[1] Vis samtlige bøjningsformer i et paradigmatiske skema – og deres fordeling i korpus

[2] Udtal ordet – fra lydfiler med optagelser eller automatisk som talesyntese

[3] Vis et resume af betydninger og lemmatiserede ord for ører/haler – og deres fordeling i korpus

[4] Vis andre ord der også betegner gnavere – og deres fordeling i korpus

[5] Vis artikler med andre dyr hvor det fremhæves at de har snuder/store ører/haler – og deres fordeling i korpus

[6] Skriv alle forkortelser helt ud

[7] Vis alle medlemmer af denne familie – og deres fordeling i korpus

[8] Vis flere typiske ordgrupper med mus, hentet fra korpus

[9] Erstat tilde med opslagsordet, evt. fremhævet på anden måde

[10] Vis flere/alle eksempler fra korpus eller med mere kontekst

[11] Vis alle bibliografiske oplysninger

[12] Vis alle it-ord – og deres fordeling i korpus

[13] Vis alle ord der betegner (små) redskaber – og deres fordeling i korpus

[14] Vis alle betydninger som var nye i 1980'erne

- 3** (*med.*) = 'ledmus [15] * *Hvis brusk slås løs kan det bevæge sig rundt i knæet. Det kaldes en mus i knæet* [16] *hj.side, medicin 2000* [17]
- grå mus** [18] (*ofø*) kedelig, tilbageholdende el. uanseelig person [19]; JF kontormus [20] * *Den politiker, der er opdraget til at være en lidt grå, anonym mus, bliver også en lidt grå og anonym politiker* BerlT90
- hvid mus** [21] albino af husmus • holdes især som forsøgsdyr og kæledyr [22] * *det varede lidt, før*
hun fik øje på en lille hvid mus, der halvvejs gemte sig i noget halm i bunden FamJour86
- med mand og mus** [23] se 'mand
- når katten er ude, spiller musene på bordet** [24] se 'kat¹
- våd som en druknet mus** [25] meget våd; gennemblødt; SYN drivvåd [*varierer i form*] * *pastor Maclaren [røg] på enden i baljen og sjoskede op på stranden våd som en druknet mus* BerlT91
- AFL til bet. 1** SUFF muse *vb.* [26]
SMS til bet. 1 musefælde, -hale, -rede, -unge; hunmus; *til bet. 2* musearbejde, -knap, -måtte [27]
- HIST** oldn. *mús*, ty. *Maus*, gr. *mys* [28]
- [15] Giv en fuld definition uden henvisning til anden artikel
- [16] Vis flere belæg fra korpus med denne betydning
- [17] Gør link til hjemmeside aktivt
- [18] Vis andre faste forbindelser hvori der indgår dyr
- [19] Vis andre ord/udtryk der betegner en kedelig person
- [20] Vis andre ord/udtryk der bruger et dyr til at betegne en person
- [21] Vis eksempler fra korpus på adj. der betegner en farve, efterfulgt af sb. der betegner et dyr
- [22] eksempler på andre dyr som holdes som kæledyr
- [23] Vis eksempler på parallelle konstruktioner efter skabelonen *med [sb.] og [sb.]*
- [24] Vis andre eksempler på ordsprogs-lignende vendinger hvori dyrebetegnelser indgår
- [25] Vis andre eksempler på sammenligningskonstruktioner bestående af adj. og et nominalsyntaxme
- [26] Vis andre verber som er suffiksafledninger af sb. (der betegner et dyr)
- [27] Vis flere/andre sammenhænge hvori mus/gnaver/dyr indgår
- [28] Vis andre ord der er indlånt fra græsk. Vis andre dyrebetegnelser som har en beslægtet form på græsk

Fig. 1. Mulige søgefunktioner i en digital ordbog

Der er to ting der springer i øjnene når man ser ned over figuren: For det første hvor mange steder der er mulighed for at udvide oplysningen med supplerende materiale fra ordbog eller korpus, og for det andet at det langt fra er alt hvad der er muligt, som vil opfylde et reelt brugerbehov. Men uanset om man hæfter sig ved det ene eller det andet, er det væsentligt at notere sig at der i langt de fleste tilfælde er tale om oplysninger som principielt er indeholdt i ordbogen eller korpus i forvejen – udfordringen består i at finde de relevante data frem og præsentere dem for brugeren. Når man skal beslutte sig for hvilke muligheder der skal prioriteres i det elektroniske værk, bør man i hvert enkelt tilfælde stille sig spørgsmål som de følgende:

- Kan det lade sig gøre at generere de pågældende data fra ordbasen på en forholdsvis enkel måde?
- I hvor høj grad kræver det manuel (efter)redigering af materialet?
- Er materialet af en sådan beskaffenhed at det principielt mulige også er praktisk gennemførligt?
- Modsvares muligheden af et reelt behov hos brugeren?

Ideelt ville det være at analysere brugerbehovene, dernæst at rangordne søgefunktionerne så de matcher behovene, og så ellers gå i gang fra en ende af. Desværre sætter virkelighedens resurseknaphed hurtigt en grænse for denne fremgangsmåde. I et projekt af middelstørrelse som det indeværende, med 3–5 redaktører tilknyttet over en projektperiode på 6 år (eller i alt ca 13 mio. kr.), er det klart at nytten af en søgefunktion nøje må sammenholdes med de resurser der er forbundet med at gøre den tilgængelig. Hvis fx en funktion kræver at hele ordbogsmanuskriptet gennemgås systematisk og redaktionelle ændringer udføres manuelt, bliver resurserne hurtigt forbrugt på denne funktion alene. Et eksempel fra fig. 1 kunne være udtale. Om denne funktion hedder det lidt bombastisk hos Landau (2001:97): ”To my mind, hearing the pronunciation is the only unarguable improvement of a CD over a book”. Selvom man ikke behøver være enig i denne anskuelse, har udtale indiskutabelt stor værdi for de mange brugere der har svært ved at afkode en trykt lydskrifts mange specialtegn. Men hvis løsningen er at oplysningen indtales og tilføjes databasen som separate lydfiler, er det dels en temmelig kostbar investering, dels en omstændelig procedure at bruge i et dynamisk projekt der løbende suppleres med nye artikler. Umiddelbart mere attraktivt forekommer det derfor at bruge et talesynteseprogram der kan downloades fra hjemmesiden. Heller ikke denne løsning er dog uden problemer, men det er til dels et spørgsmål om den rette teknologi på det rette tidspunkt.

Derfor er det nødvendigt at udvælge nogle strategiske satsningsområder. Det kan betyde at visse søgefunktioner må fravælges fordi de er for omkostningstunge, også selvom de ideelt set er nyttige og efterspurgt. Og omvendt kan det betyde at visse søgefunktioner tilbydes selvom de ikke er indlysende påkrævede, fordi det er enkelt at tilbyde dem. Man kan fx forestille sig sådanne funktioner gemt bag en knap ”flere oplysninger”, som brugerne efter eget valg kan benytte sig af eller lade være hvis de ikke er interesserede i informationen.

Det kan være nyttigt at sammenfatte mulighederne fra fig. 1 i forskellige typer efter både art og arbejdsindsats. For hver type kan mulighederne være indbyrdes sammenhængende eller uafhængige. Og de kan være forholdsvis simple, eller de kan være omkostningstunge.

Type a) omfatter muligheder der går på at præsentere en eksisterende oplysning på en anden måde, typisk ved at udfolde hvor papirordbogen sammentrænger, beskærer og forkorter af pladsøkonomiske grunde. Mulighederne [1], [6], [9], [15] og [17] hører til denne type.

Type b) indebærer at der indføres en ny mulighed som ikke eksisterer i papirforlægget. Her bør skelnes imellem muligheder der kræver manuel redigering ([2]), og muligheder der evt. kan genereres automatisk ([3], [11]).

Type c) udnytter det digitale mediums mulighed for at linke til andre ord i basen som ligner opslagsordet i en bestemt henseende. I sin simpleste form dækker funktionen over en SQL-forespørgsel til basen om at generere et søgeresultat med samme eller lignende indhold som opslagsordet for det pågældende element (fx [14]); i andre tilfælde kræves yderligere en statistisk analyse af elementindholdet. Til denne type hører [4], [5], [7], [12], [13], [14], [18], [19], [20], [22], [23], [24], [25], [26], [27] og [28].

Type d) kombinerer ordbogsopslag med en fokuseret korpussøgning, dvs. undersøger forekomster eller fordelinger i korpus på netop det sted i ordbogsartiklen hvor brugeren befinder sig. Koblingen til korpus kan også ske i kombination med generering af søgeresultater nævnt under type c), hvorfor der er et vist overlap mellem de to typer. Til denne type hører [4], [5], [7], [8], [10], [12], [13], [16], [21] og til dels [1] og [3].

4. Strategiske satsningsområder

De muligheder der hører under type a), udmærker sig ved at være forholdsvis enkle at udføre da de kan udføres globalt og automatisk, og bør

derfor prioriteres. Det er også en logisk følge af betingelsen om at plads ikke er noget problem: I en elektronisk ordbog bør redaktionelt indførte forkortelser, tilder, bindestreger osv. så vidt muligt erstattes af fuldt udskrevne former. I nogle tilfælde kan man dog forestille sig en ”teleskopløsning” hvor brugeren først ser en kondenseret opstilling, fx papirversionens, og derefter har mulighed for at klikke sig til en fuldt udfoldet beskrivelse. Det kunne fx være en attraktiv løsning når det gælder fleksionsangivelser, fordi det fuldt udfoldede paradigme med tilhørende officielle og uofficielle varianter undertiden fylder så meget at det kan blive en kilde til irritation for den bruger der leder efter noget helt andet. Valget mellem forskellige præsentationsmuligheder accentuerer i øvrigt vigtigheden af at adskille indhold og præsentation allerede i redigeringsfasen.

Af mulighederne under type b) er det klart at de funktioner som kan genereres automatisk, umiddelbart får højere prioritet. At give adgang til alle bibliografiske oplysninger er overkommeligt da alle korpustekster i forvejen er forsynet med headeroplysninger, og det er ikke noget problem at videregive samtlige disse oplysninger. Mest sandsynligt vil det blive som teleskopløsning, og vi vil i samme forbindelse overveje om det kan lade sig gøre at udfolde papirordbogens kildeforkortelser mere hensigtsmæssigt. Vi har også eksperimenteret med muligheden for at give automatiske betydningsresumeeer, og det ser lovende ud. En ulempe ved den digitale ordbog er at brugeren ikke kan overskue hele artiklen så godt som i papirordbogen. Derfor er det vigtigt, ikke mindst i de længere artikler, at yde hjælp så brugeren kan orientere sig i artiklen og få mulighed for at navigere mellem de forskellige dele.

Udtale er et område som principielt har høj prioritet fordi behovet i visse brugergrupper er stort. Til gengæld er det, som nævnt ovenfor, afhængigt af at den rette tekniske løsning er til stede. Det er en mulighed vi undersøger, og meget gerne ser indført.

De muligheder der er sammenfattet i type c), har alle at gøre med at præsentere andre artikler eller træk fra andre artikler som kan være relevante for det pågældende opslagsord. Det er oplysninger som findes i ordbogen i forvejen, og som det kan være nyttigt at give brugeren adgang til i visse sammenhænge, især i forbindelse med vidensrelaterede funktioner, fx læring. Nogle af oplysningerne kan stilles til rådighed uden de store problemer. Fx løses [12] ”vis andre it-ord” ved at formulere en forespørgsel i basen hvor indholdet af elementet Fag = edb, eller [14] ”vis andre ord/betydninger som var nye i 1980’erne” med en forespørgsel til basen hvor indholdet i elementet Kron eller Dat >= 1980 og <1990. Andre af mulighederne kræver

derimod en betydelig indsats. En mulighed som [4] ("vis andre ord der også betegner gnavere") forekommer umiddelbart simpel, men for at den skal fungere tilfredsstillende, kræver det en meget stor grad af konsistens i betydningsdefinitionerne – i dette eksempel således for det første at ordet *gnaver* er benyttet som genus proximum i samtlige definitioner af ord der betegner gnavere, for det andet at genus proximum kan udpeges entydigt i hver enkelt definition. Det sidste problem er heldigvis ikke så stort idet det under redigeringen af Den Danske Ordbog har været et eksplicit krav at redaktørerne anførte genus proximum for hver betydning i et separat felt, netop med henblik på en senere elektronisk udnyttelse af det. En analyse af Den Danske Ordbogs definitioner for udvalgte betydningsområder har imidlertid vist at den systematiske konsistens er svingende, fra næsten acceptabel til noget ufuldstændig. Det skal understreges at kravet om systematik alene drejer sig om definitionernes egnethed til automatisk sprogteknologisk processering og ikke må forveksles med en kvalitativ vurdering af definitionerne sådan som de fungerer for ordbogsbrugeren. Mest overbevisende virker systematikken for konkrete substantiver, mindre godt for abstrakte og overførte betydninger. I eksemplet 'gnaver' er det således kun artiklen *hare* der falder uden for systematikken ('gråbrunt pattedyr med hvid bug, to store fortænder i over- og undermund, lange ører og kraftige bagben' – hvor redaktøren altså har valgt at anføre *pattedyr* frem for *gnaver* som genus proximum). Men også inden for gruppen af konkrete substantiver er der variation. Når man forlader det som nogle semantikere kalder 'basic level concepts' (se fx Lakoff 1987:31ff.), bliver afstanden mellem almensproget og den faglige systematik hurtigt for stor, og princippet om at vælge nærmeste systematiske overbegreb fungerer derfor mindre godt her. Det gælder fx mere primitive niveauer i den biologiske systematik – eksempler er stilksporesvampe, polypper eller rødalger. Det afspejler sig i definitionerne ved at der typisk anføres et mere basalt niveau som genus proximum: ordet *støvbald* defineres fx som en svamp og ikke som en stilksporesvamp.

Konklusionen på de stikprøveanalyser vi har foretaget i Den Danske Ordbog, er at selvom ordbogens definitioner forekommer velstrukturerede og konsistente i sammenligning med mange andre ordbøger, ville det kræve en meget stor indsats, herunder ikke mindst en betragtelig portion manuel redigering af materialet, at få blot genus proximum-delen til at fungere tilfredsstillende. Umiddelbart synes der derfor at være to mulige veje at gå med hensyn til *ordnet.dk*: enten at forkaste muligheden eller at begrænse den til de semantiske områder hvor pålideligheden er størst og resurseforbruget mindst. Ikke desto mindre

har vi valgt at satse på en helt tredje udvej: at indgå i et samarbejde om etablering af en selvstændig digital 'ordbog' hvor ordforrådets betydningsstruktur og interne relationer beskrives på en formaliseret måde så de kan bruges af it-systemer, et såkaldt leksikalsk ordnet. Dette projekt, DanNet, skal ses som en dansk udgave af de internationale EuroWordNet og Princetons WordNet og gennemføres over en treårig periode i samarbejde med Center for Sprogteknologi ved Københavns Universitet. Projektet finansieres af Statens Humanistiske Forskningsråd. Udarbejdelsen sker med udgangspunkt i dels Den Danske Ordbogs definitioner, dels den sprogteknologiske ordbase Sto/Simple. Ud over at et sådant leksikalsk ordnet har en selvstændig anvendelse i forbindelse med forskellige former for informationsnavigering, fx ved søgning på internettet, vil det kunne indgå i *ordnet.dk* som en vigtig onomasiologisk komponent. Dvs. det vil kunne udnyttes ikke kun til at håndtere supplerende oplysninger af den art jeg har diskuteret her, men vil gøre det muligt at bruge *ordnet.dk* som en begrebsordbog hvor man også kan få besked om synonymer, antonymer, over- og underbegreber, del-helheds-relationer og andre tesauruslignende opslag. Det vil blive det første større satsningsområde for projektet.

Det andet større satsningsområde vil blive de muligheder der opregnes under type d), altså udnyttelsen af korpusmateriale til at supplere ordbogsoplysninger med. I den forbindelse bør det understreges at eksempelartiklen i fig. 1 først og fremmest tjener til at vise nogle af de mange muligheder der findes, men netop ikke er udtømmende. Artikler med ord fra andre ordklasser ville indebære nogle andre muligheder der ikke er beskrevet her. Et af de måske mest oplagte områder er valens og konstruktionsoplysninger, som jo især optræder i artikler med verber. Her kunne det være en stor hjælp for mange brugere hvis den trykte ordbogs temmelig formaliserede beskrivelse med et enkelt klik kunne suppleres med autentiske korpuseksempler på faktisk brug. Den mulighed er til stede, idet korpusteksterne er syntaktisk opmærkede; opmærkningen bliver blot ikke udnyttet i den aktuelle netversion, men det vil ændre sig i en senere version. Tilsvarende er det oplagt at supplere kollokationer med yderligere materiale fra korpus, og det samme gælder til en vis grad for citater. Her må vi dog understrege at vi, og for den sags skyld sprogteknologien, ikke er kommet så langt at det kan lade sig gøre at generere korpuseksempler over et givet ord i en bestemt betydning.

5. Sammenfatning

Jeg har forsøgt at antyde nogle af de muligheder som elektronisk publicering frembyder. Gennemgangen skulle gerne have vist at mulighederne er langt flere end resurserne rækker til – i hvert fald i et projekt som det her beskrevne. Men omvendt håber jeg også at den har vist at der er mange spændende udfordringer at gå i gang med, og at det elektroniske medie er særdeles velegnet til netop den komprimerede type af information som leksikografer beskæftiger sig med. I projektet *ordnet.dk* vil vi gerne udvikle bare nogle af de muligheder som er til stede, og jeg har redegjort for nogle af de strategiske overvejelser vi har gjort os: Ud over nogle oplagte enkeltstående muligheder kommer satsningsområderne især til at ligge inden for onomasiologiske søgninger samt integrationen mellem ordbog og korpus.

Set fra et leksikografisk synspunkt ligger der en stor fordel i mediets dynamiske natur. Det indebærer nemlig at det er muligt til enhver tid at udvide og indføre ændringer, fx løbende at tilføje nye ord og betydninger til det elektroniske opslagsværk i takt med at de opstår i sproget. For en trykt ordbog kræver den slags nye og reviderede udgaver – og ofte alvorlige økonomiske overvejelser inden de kan udføres. Også på et andet område ligger det i mediets natur at materialet kan udvides: Det aktuelle projekt omfatter to ordbøger knyttet sammen med et korpus over moderne tekster, men på længere sigt kan man nemt forestille sig *ordnet.dk* udvidet på både ordbogssiden og tekstsiden så det efterhånden udvikler sig til en portal med leksikalske beskrivelser af dansk ordforråd fra den tidligste runeperiode og helt op til dagen i dag, og med mulighed for at brugeren kan foretage egne undersøgelser i materialet og få autentiske eksempler i tekster fra den aktuelle periode.

Litteratur

- Bojsen, Else/Trap-Jensen, Lars 2005: ODS, ODS-S og fremtiden. I: Mette Kunøe og Peter Widell (udg.): *10. Møde om Udforskningen af Dansk Sprog*. Århus, 58–67.
- DanNet = Pedersen, Bolette S. m.fl. (red.): *DanNet*. København: Center for Sprogteknologi/Det Danske Sprog- og Litteraturselskab 2005–2008. www.wordnet.dk.
- Den Danske Ordbog = Hjorth, Ebba/Kristensen, Kjeld m.fl. (red.): *Den Danske Ordbog 1–6*. København: Det Danske Sprog- og Litteraturselskab/Gyldendal 2003–2005.
- Elexiko = Klosa, Annette m.fl. (red.): *elexiko*. Mannheim: Institut für Deutsche Sprache. www.elexiko.de.

- EuroWordNet = Vossen, Piet m.fl. (red.): *EuroWordNet*. Amsterdam 1996–1999. www.illc.uva.nl/EuroWordNet.
- Hermes 34*. Bergholtz, Henning m.fl. (red.). Århus 2005.
- Korpus 2000 = Asmussen, Jørg m.fl. (red.): *Korpus 2000*. København: Det Danske Sprog- og Litteraturselskab 2000–2002. <http://korpus.dsl.dk/korpus2000>.
- Lakoff, George 1987: *Women, Fire, and Dangerous Things*. Chicago & London: The Chicago University Press.
- Landau, Sidney I. 2001: *Dictionaries. The Art and Craft of Lexicography*, 2. udgave. Cambridge: Cambridge University Press.
- Lorentzen, Henrik/Trap-Jensen, Lars (under udgivelse): *ordnet.dk* – et nyt sprogligt opslagsværk på internettet. I: Carsten Hansen, Henrik Lorentzen, Lars Trap-Jensen (udg.): *Nordiske Studier i Leksikografi 8: Rapport fra Konference om Leksikografi i Norden, Sønderborg 24.–28. maj 2005*. København.
- Ordbog over det danske Sprog 1–28. København: Det Danske Sprog- og Litteraturselskab/Gyldendal 1918–1956. Supplementsbind 1–5, 1992–2005.
- ordnet.dk* = Andersson, Henrik/Asmussen, Jørg/Lorentzen, Henrik/Sørensen, Nicolai H./Trap-Jensen, Lars (red.): *ordnet.dk*. København: Det Danske Sprog- og Litteraturselskab 2004–2009. www.ordnet.dk
- Simple = Norling-Christensen, Ole/Pedersen, Bolette S. (red.): *SIMPLE*. København: Det Danske Sprog- og Litteraturselskab/Center for Sprogteknologi 1998–2000.
- Sto = Braasch, Anna m.fl. (red.): *Sprogteknologisk Ordbase*. København: Center for Sprogteknologi 2001–2004. www.cst.dk/cgi-bin/defisto.
- Trap-Jensen, Lars 2004: Et net af ord – *ordnet.dk*. I: *Mål & Mæle nr. 4*. København, 20–27.
- WordNet = Miller, George A. m.fl. (red.): *WordNet*. Princeton 1985–. <http://wordnet.princeton.edu>.

Lars Trap-Jensen
 ledende redaktør
 Det Danske Sprog- og Litteraturselskab
 Christians Brygge 1
 DK-1219 København K
ltj@dsl.dk