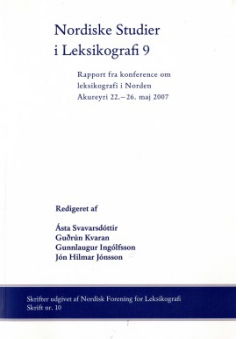


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Tilgangs- og henvisningsstruktur i digitale ordbøger - Overvejelser baseret på ordnet.dk	
Forfatter:	Lars Trap-Jensen	
Kilde:	Nordiska Studier i Lexikografi 9, 2008, s. 447-464 Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

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 leksikografi (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.

LARS TRAP-JENSEN

Tilgangs- og henvisningsstruktur i digitale ordbøger Overvejelser baseret på ordnet.dk

The concept of structure cannot be transferred from the printed book to the digital medium: the alphabetical ordering principle does not apply to a digital dictionary, hyperlinks are possible in the digital medium only, and conditions of spatial orientation, overview and browsing are different in the two media. Therefore, a digital dictionary must seek new solutions to provide user-friendly access to its contents. In this article, the task is viewed from two angles: Various solutions employed by existing online dictionaries are analyzed and evaluated as viewed by the user, and from the perspective of the dictionary-maker, special attention is devoted to the possible need to rearrange the (print-based) data structure underlying the presentation displayed on the screen.

1. Indledning

De fleste er vist enige om at traditionelle definitioner af begreberne tilgangsstruktur og henvisningsstruktur – eller mediostruktur som det også benævnes nogle steder – ikke dækker digitale ordbøger. I *Nordisk leksikografisk ordbok* lyder de korte definitioner:

Tilgangsstruktur: strukturen av de indikatorene som fører ordboksbrukeren til de opplysningene han/hun ønsker.

Henvisningsstruktur: strukturen av de eksplisitte og implisitte indikatorer som leder brukeren rundt i ordboken for å supplere eller utdype de opplysningene en har funnet.

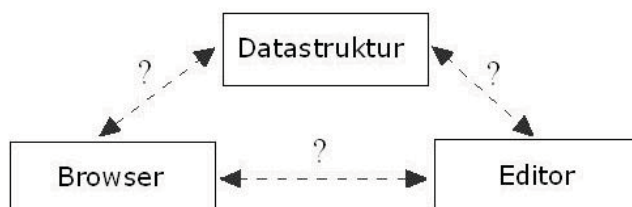
Problemet er, tror jeg, ordet struktur. Struktur handler om relationer og ordning af enheder, og det er ikke mærkeligt at en strukturel beskrivelse af den trykte ordbogs artikler og ordbogsoplysninger, ordnet sekventielt i de to dimensioner som bogsiden har, ikke kan overføres til elektroniske medier, især fordi muligheden for hyperlinks føjer en ny dimension til som gængse beskrivelser af gode grunde ikke har taget højde for. Den samme indvending kan bruges om centrale leksikografiske begreber som mikrostruktur og makrostruktur. Spørgsmålet er om begreberne kan udvides til at omfatte såvel traditionelle trykte ordbøger som digitale ordbøger. Denne artikel giver ikke det endelige og udtømmende svar på spørgsmålet, men tager fat på problemet med udgangspunkt i en række praktiske valg der har skullet træffes i forbindelse med en netordbog baseret på data

fra *Den Danske Ordbog*. I det følgende vil jeg gå ud fra et andet og meget løsere grundlag end de ovennævnte definitioner og overveje *hvad som helst der tjener til at hjælpe brugeren med at finde det han/hun ønsker*.

Jeg interesserer mig for hvordan man kan søge i en online-ordbog, hvordan resultatet af søgningen præsenteres, hvilken hjælp brugeren får, og jeg vil komme ind på de ekstra muligheder man har i en digital ordbog. På artikelniveau gælder det navigering og overblik over en artikels opbygning, henvisninger i almindelighed og derudover flerordsforbindelser og deres håndtering i den digitale ordbog. Afslutningsvis vil jeg se på hvilke leksikografiske konsekvenser det har, og hvordan det påvirker leksikografens arbejdsbetingelser.

2. Data, editor, browser

Udgangspunktet for artiklen har været *Den Danske Ordbog* (DDO) som papirordbog og brugen af data derfra til en netordbog. Interessen har især koncentreret sig om organiseringen af data og hvordan de valg der træffes, både bestemmes af og har konsekvenser for datastrukturen. Med en given struktur af data i basen kan man vælge en bestemt præsentation af dem, enten i den grænseflade som redaktørerne sidder med, i form af en editor eller et særligt redigeringssystem, eller i den brugergrænseflade som slutbrugeren anvender til at slå op i ordbogen via sin browser. Det vil sige at det er samspillet mellem database, editor og browser der er i fokus. Enhederne kan skematisk fremstilles som i figur 1, mens relationerne mellem dem indtil videre ikke er klarlagt.



Figur 1. Samspillet mellem enhederne er det der ligger bag meget af det følgende, og målet er at blive klogere på deres indbyrdes sammenhæng.

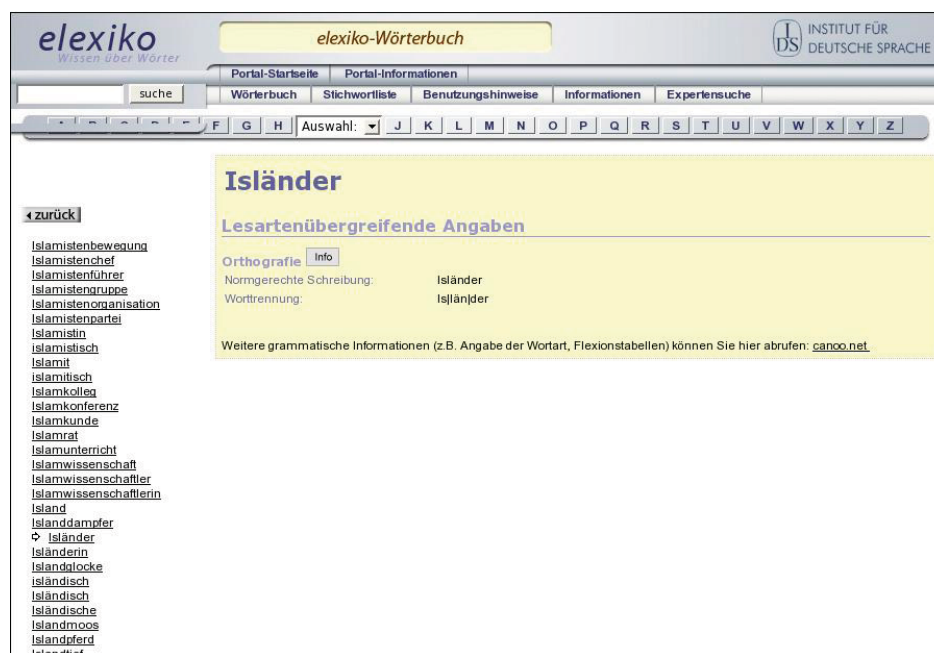
3. Opslag

3.1 Søgemåder

Den mest udbredte opslagsmulighed i online-opslagsværker er et søgefelt hvori brugeren indtaster det ønskede opslagsord. Ofte tilbydes også en trunkeret søgning, enten med jokertegn eller med valg fra en rullemenu mellem forskellige muligheder. Opslag ved hjælp af søgefelt er dog ikke den eneste mulighed. Man kan også med mus og taster navigere sig frem til det ønskede lemma gennem en systematisk eller alfabetisk ordnet opstilling og derfra klikke sig til den ønskede

artikel. Det er en løsning man især benytter i mindre værker hvor det er forholdsvis nemt at orientere sig i ordstoffet, fx privatpersoners offentligt tilgængelige ordsamlinger, eller i faglige værker hvor det er vigtigt at visualisere den systematiske opbygning, og i indscannede ældre ordbøger der kun findes som billedfiler uden indeksering på lemmaniveau. Der er sådan set ikke noget til hinder for at man kan kombinere de to søgemåder, og det finder man også i nogle netordbøger, fx det tyske projekt *elexiko*, jf. figur 2. Fordelen ved navigation fra en oversigt er at brugeren får et samlet overblik over stoffet – på samme måde som når man læser en trykt bogs indholdsfortegnelse. Dog mindskes overblikket hurtigt, jo større antallet af indgange i basen er. Hvis man skal scrolle i lang tid for at bladre forbi et stort antal poster eller klikke mange gange for at nå frem til en bestemt artikel, vendes fordelene ved overblikket nemt til en ulempe i den tid det tager at finde frem til den ønskede artikel.

For DDO's vedkommende er antallet af artikler så stort at navigationsløsningen aldrig har været overvejet seriøst – i hvert fald ikke som en primær løsning. Dog er vi opmærksomme på det tilfredsstillende ved den trykte bogs mulighed for overblik og bladring til andre artikler og vil gerne imødekomme det i et vist omfang.



Figur 2. Elexiko tillader både alfabetisk liste og søgefelt som tilgangsmulighed

Det gøres fx i netversionen af *Ordbog over det danske Sprog* (ODS) ved at tilbyde et udsnit af lemmalisten i skærmens venstreside. Her har brugeren mulighed for at bladre

frem og tilbage til artikler i alfabetisk nærhed af det søgte. En anden mulighed der minder om denne, er at vise artiklerne i rækkefølge efter hinanden på selve hovedsiden. Det bruges fx i *Macmillan English Dictionary Online* (MEDO), jf. figur 3.



Figur 3. I MEDO vises artiklerne fortløbende på skærmen med mulighed for at klikke sig videre videre frem og tilbage i den alfabetiske liste

Begge dele er selvfølgelig en illusion, nemlig om en struktur der ligner den trykte sides to dimensioner hvor artiklerne er ordnet sekventielt efter alfabetisk princip. Det har til gengæld den fordel at brugeren kan bladere rundt og læse artikler for sin fornøjelses skyld – også artikler hun ikke havde søgt efter i første omgang. Løsningen med fortløbende artikler er dog ikke så attraktiv i vores tilfælde – udelukkende af praktiske grunde. Den fungerer bedst hvis hver enkelt artikel kun er på nogle få linjer. Og løsningen med en ordliste har den fordel at man ret enkelt kan tilbyde ordbogen som baglænsordbog ved at tilføje en knap “alfabetiser baglæns” hhv. “alfabetiser forlæns”.

3.2 Søgning på hovedopslagsord

Lad os herefter se på de beslutninger der skal træffes før man får et resultat af sin søgning. Når man går ind på ordbogssiden, bør det være sådan at cursoren står i søgefeltet klar til indtastning (det er ikke altid en selvfølge). Når søgeordet er tastet ind, bør der være mulighed for enten at klikke på en søgeknapp eller trykke <retur> – eller helst begge dele. I nogle ordbøger (et eksempel er MEDO) er muligheden for at trykke <retur> ikke til stede, og det er upraktisk. Når ordet er tastet ind og der er trykket <retur>, går søgningen i gang. Men hvor søges der, og hvordan vises resultatet? Hvis vi antager at det i første omgang kun drejer sig om søgning på hovedopslagsord, kan man tænke sig tre logiske muligheder.

3.2.1 Søgeordet matcher én og kun én post i basen

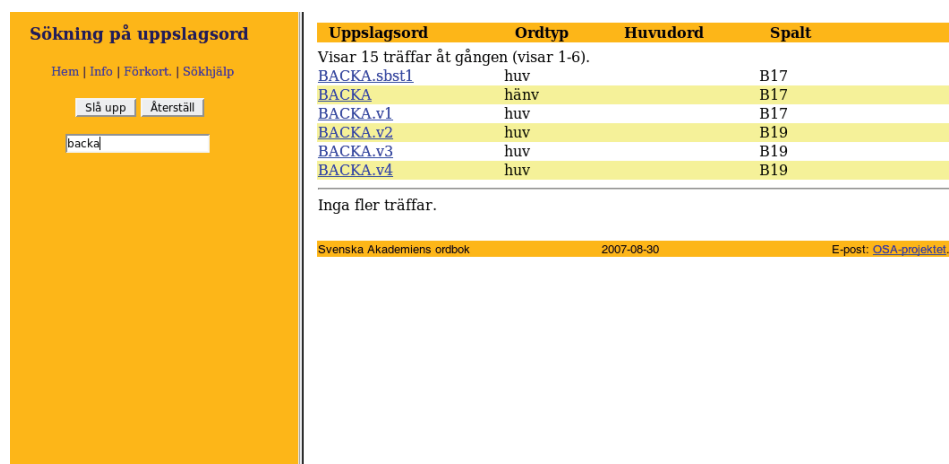
Den situation er som regel uproblematisk. Når man har trykket <retur>, kommer man straks til den pågældende artikel. Dog skal man være opmærksom på det tilfælde hvor der er et match, men bare ikke det rigtige. Et eksempel kunne være søgning efter det engelske adjektiv *savory* 'krydret, velsmagende'. Det har varianten *savory*, som er den amerikanske stavemåde, og hvis man søger efter denne form i *Oxford English Dictionary* (OED), får man ganske vist et resultat; det er bare ikke det ønskede. Det resultat der vises, er artiklen for substantivet *savory* 'sar' (plante af slægten *Satureia*), hvorimod man selv må regne ud at adjektivet skal søges under den britisk engelske stavemåde *savory*. Hvis man vil hjælpe brugeren i et tilfælde som dette, er svaret at tilbyde en udvidet søgning hvor der ikke kun søges efter opslagsordet, men også i andre elementer. Dette må man gøre som en generel mulighed da man naturligvis ikke kan vide om den viste artikel er den brugeren søgte.

3.2.2 Søgeordet matcher flere poster i basen

Her kan man forestille sig (mindst) tre muligheder:

1. Der vises en søgeresultatside hvorfra brugeren kan klikke sig videre blandt de opstillede muligheder
2. Mulighederne fremgår af en sidemenu, mens den første mulighed vises centralt
3. De matchende poster vises direkte på skærmen i rækkefølge efter hinanden

Mulighed 1 er vist den mest almindelige. Den bruges i fx OED, *Svenska Akademiens Ordbog* (SAOB) og *Longman's Dictionary of Contemporary English* (LDOCE), jf. figur 4.



Sökning på uppslagsord

Hem | Info | Förkort. | Sökhjälp

Slå upp Återställ

backa

Uppslagsord	Ordtyp	Huvudord	Spalt
Visar 15 träffar åt gången (visar 1-6).			
BACKA.sbst1	huv		B17
BACKA	häv		B17
BACKA.v1	huv		B17
BACKA.v2	huv		B19
BACKA.v3	huv		B19
BACKA.v4	huv		B19

Inga fler träffar.

Svenska Akademiens ordbok 2007-08-30 E-post: [OSA-projektet](#)

Figur 4. I SAOB vises søgeresultater i hovedvinduet som en menuliste hvorfra brugeren kan klikke sig videre til den ønskede artikel

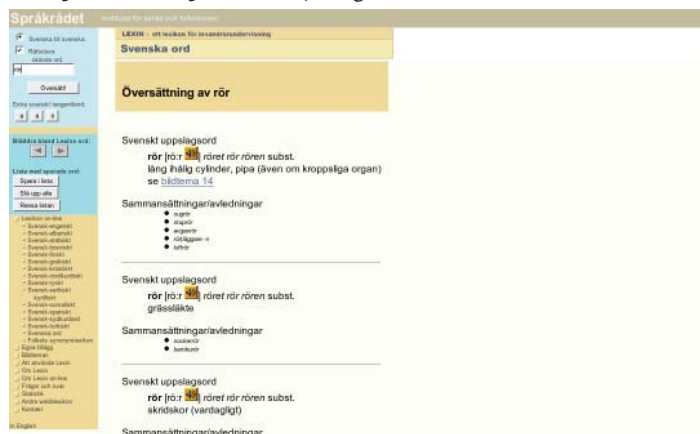
Mulighed 2 kræver at der er plads i sidemenuen til denne funktion. Til gengæld sparer den brugeren for et klik hvis den viste artikel er den søgte. I ODS er denne løsning valgt, og samtidig kombineres søgeresultatoversigten med den alfabetiske liste vha. faneblade, jf. figur 5.



Figur 5. I ODS fremgår søgeresultatet af oversigten til venstre, og det første resultat vises i hovedvinduet. Man kan veksle mellem søgeresultat og alfabetisk liste ved at vælge det ønskede faneblad.

Den løsning er ganske velfungerende og vil sandsynligvis blive videreført i den endelige grænseflade. Derimod er det muligt at listen vil blive flyttet til højremenuen fordi venstremenuen reserveres til andre formål.

Den tredje mulighed bringer brugeren direkte til de resultater der findes, men man må selv scrolle op og ned for at finde den ønskede artikel. Den mulighed finder man fx i *Lexin*-ordbøgerne, den danske *Retskrivningsordbogen* og i *Chambers 21st Century Dictionary Online*, jf. figur 6.



Figur 6. Lexin viser resultatet af en søgning på rör direkte på skærmen, og brugeren kan scrolle ned og vælge hvilken af de otte artikler der er den ønskede

3.2.3 Søgeordet matcher ingen post i basen

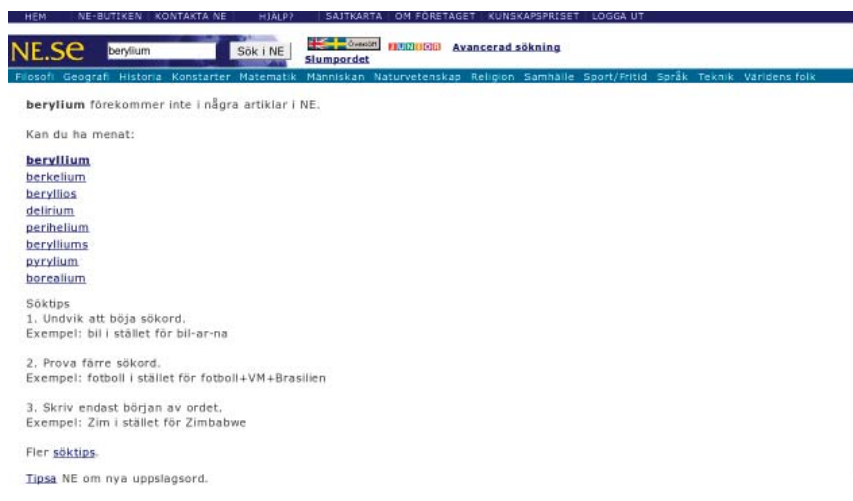
Hvis der ikke er noget match for søgningen, kan det skyldes at det pågældende ord ikke findes i basen. Der bør da komme en standardmeddelelse om det, og evt. kan der samtidig vises en relevant artikel på skærmen, fx nærmeste alfabetiske nabo. Brugeren kan også gives mulighed for at kontakte redaktionen og foreslå ordet optaget.

Den situation kan i praksis ikke skelnes fra den situation hvor det søgte ord faktisk findes i basen, men i en anden form end den der er tastet ind, hvad enten det skyldes en tastefejl eller at brugeren ikke kender den rette stavemåde.

Her bør der også komme en fejlmeddelelse der hjælper brugeren videre med søgningen. I meddelelsen kan der gives en oversigt med forslag til hvilke ord der minder om det søgte (resultatet af en såkaldt *fuzzy search*, jf. Svensén 2004: 521). I eksisterende netordbøger er det forskelligt hvor meget hjælp der tilbydes. God hjælp får man fx i online-versionen af *Nationalencyklopedins Ordbok* (NEO), hvor der ofte tilbydes brugbare alternative forslag, jf. figur 7.

Man kan se på forslagene at de er fremkommet som resultat af en algoritmisk beregning, og det er selvfølgelig ikke alle algoritmer der er lige gode.

Hvis brugeren tror at søgeordet staves anderledes, er det en fordel hvis ordbogen i forvejen kender til de mest almindelige både officielle og uofficielle varianter og indekserer dem så man kan hjælpe brugeren til at komme til det rigtige sted. I trykte ordbøger er det almindeligt at gøre det i form af henvisningsartikler: Fra officielle eller registrerede varianter henviser man til hovedopslagsordet. Det kan man også godt gøre i en netordbog, sådan som man kan se det i fx ODS eller *Grimms Deutsches Wörterbuch*.



Figur 7. En søgning i NEO på *beryllium* giver ikke noget resultat. I stedet foreslås følgende muligheder: *beryllium*, *berkelium*, *beryllios*, *delirium*, *perihelium*, *berylliums*, *pyrylium* og *borealium*.

Som regel vælges denne løsning dog fordi man af resurse-mæssige grunde har været nødt til at følge den løsning der er valgt i den trykte ordbog. Mere brugervenligt er det at spare brugeren for et ekstra opslag ved at bringe ham/hende direkte til den søgte artikel. Et eksempel på det ses i NEO, som vist stort set ikke rummer henvisningsartikler. I nogle learnerordbøger foretrækkes henvisningsløsningen dog bevidst – af pædagogiske grunde (fx i MEDO). Af samme grund foretrækker vi kun at springe henvisningen over ved dobbeltformer og acceptable variantformer; derimod ikke ved stavefejl og uofficielle former. Det kan desuden blive nødvendigt at fremhæve det søgte ord i den artikel der vises som søgeresultat, så brugeren straks kan se hvorfor artiklen er et relevant resultat af den pågældende søgning.

For datastrukturen betyder denne løsning ikke andet end at henvisningsartiklerne slettes fra basen. Indekseringen af variantformerne sikrer at brugeren kan komme til den rigtige artikel, og hvis man sørger for at registrere varianterne i separate elementer, kan man også i HTML-visningen tildele dem forskellige ledsagetekster: Ved dobbeltformer og sideformer kan brugeren bringes direkte til artiklen, ved fejl og uofficielle varianter først via en meddelelsesside hvorfra de selv må klikke sig videre. I DDO's tilfælde er lavfrekvente varianter og fejl de eneste tilfælde hvor det er nødvendigt ikke bare at slette, men også at ændre i datastrukturen. I DDO-strukturen findes lavfrekvente varianter og fejl nemlig *kun* som henvisningsartikler og er ikke beskrevet i hovedartiklen. Her er det derfor nødvendigt at oprette fejlformen i det rigtige element i hovedartiklen. For brugeren gør ændringen selvfølgelig ingen forskel; det kan godt vises i form af en meddelelse på skærmen med mulighed for at klikke sig til den rette artikel. Derimod vil fejlformerne ikke længere optræde på den alfabetiske liste i siden af skærmen.

Endelig kan man tænke sig at søgeordet ikke findes som hverken hovedlemma, variant eller bøjningsform, men forekommer i et relevant element andetsteds i artiklen (fx som synonym, konstruktionseksempel eller orddannelsesmulighed). I det tilfælde er det ligeledes oplagt at vælge en løsning hvor der kommer en meddelelse om at der ikke er et match med det pågældende søgeord, men at ordet optræder i en anden artikel som brugeren så kan klikke sig hen til. Evt. kan man lade rækkefølgen af forslag være bestemt af en prioritering af artikelelementerne.

3.3 Udvidede søgemuligheder

I en netordbog har man flere muligheder end i en papirordbog. Hyperlinks gør det muligt at tilbyde opslag andre steder. Det kan være i eksterne resurser, fx Google eller *Retskrivningsordbogen*. Hvis det tilbydes som en generel mulighed, adskiller netordbogen sig ikke fra andre netresurser hvor der kan være integreret en søgemaskine i en bestemt applikation. Men muligheden kan også tilbydes i redigeret form, altså hvor en redaktør har taget stilling til at det er relevant at henvise til et websted uden for ordbogen selv. Det kan være en henvisning fra et

ord eller en bestemt betydning til en mere udførlig encyklopædisk beskrivelse, en mulighed fx *Wikipedia* benytter sig meget af. Blot skal man huske at koble en automatisk overvågning af linkene til så der ikke opstår “døde links” når en udbyder lukker sin tjeneste eller ændrer på opbygningen af sitet.

Det kan også være i form af semi-eksterne links, dvs. links til resurser som ligger uden for den aktuelle ordbog, men inden for ens egen kontrol. I vores tilfælde gælder det links til DSL’s korpusresurser og til ODS, og det er en mulighed vi ønsker at udvikle (se nærmere i fx Lorentzen og Trap-Jensen 2006, Trap-Jensen 2005). Det er fx oplagt at der samtidig med søgningen i netversionen af DDO sker en søgning i baggrunden i de andre resurser. Resultatet af baggrundssøgningen kan så tilbydes hvis en søgning i DDO ikke giver resultat, men findes i ODS (“det søgte ord findes ikke i DDO, men er registreret i ODS. Vil du se artiklen i ODS?”). Det kan også tilbydes som en generel mulighed når man står i en artikel (“Der findes også en artikel i ODS med dette ord. Vil du se artiklen i ODS?” eller “Se forekomster af dette ord i korpus”). Endelig er der de udvidede søgemuligheder som kan tilbydes internt i ordbogen. Indtil nu er det kun varianter af opslagsordet der har været behandlet, og hertil kan man også regne bøjningsformer. Disse søgemuligheder gælder de forskellige former som opslagsordet kan optræde i, men derudover er en ordbogsartikel jo opmærket i en række delelementer i mikrostrukturen, og det er oplagt at tilbyde at man også kan søge på dem, enten dem alle eller et relevant udsnit af dem. Det gør man formentlig bedst ved at tilbyde søgemuligheden “udvidet søgning” sådan som man har gjort det i fx OED, jf. figur 8.

Figur 8. I OED tilbydes udvidede søgemuligheder som valg fra et sæt af på forhånd opstillede muligheder

Her kan man vælge mellem fritekstsøgning eller søgning på bestemte elementer. Man kan se at det giver nogle helt nye søgemuligheder som ikke er til stede i papirordbogen. Fx kan man søge målrettet efter citater af bestemte forfattere eller få eksempler på ord der er lånt fra et bestemt sprog. Og man får en rimordbog forærende hvis det er muligt at søge i udtaleelementet hvor søgning på tekststregen alene kun giver meget upræcise rim (jf. Burke 2003: 244). Denne mulighed kan naturligvis også tilbydes som en selvstændig mulighed (“find ord der rimer på X”).

3.4 Artikelnavigering

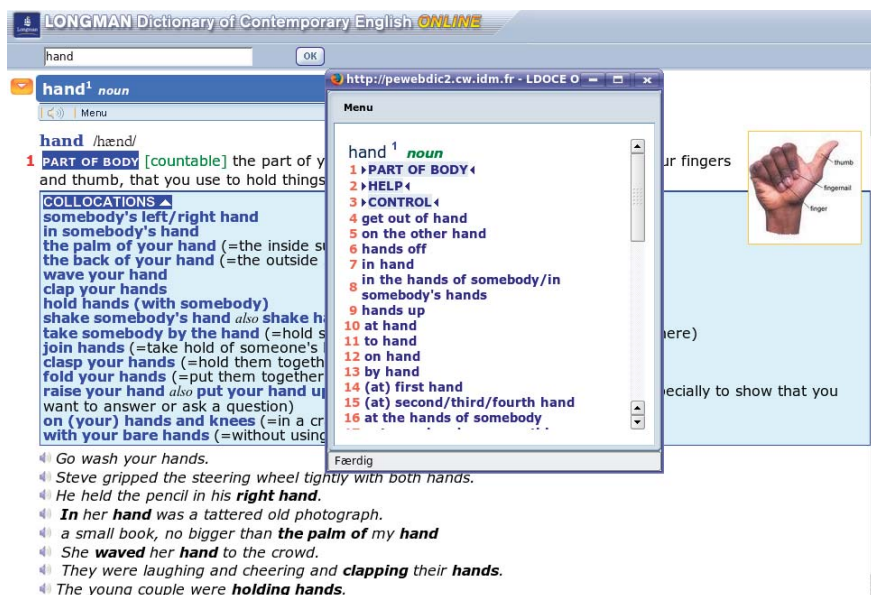
Når man er nået frem til den ønskede artikel, er spørgsmålet hvordan den ønskede information findes inden for artiklen. Det gælder at et skærmbillede på én gang er mere begrænset og rummer flere muligheder end papiret. Skærmen er begrænset i forhold til bogens sider ved at der står betydelig mindre på et skærmbillede end der gør på end papirsider. Og da undersøgelser samstemmende viser at læsbarheden nedsættes på skærmen, er det forståeligt nok sværere at skaffe sig et overblik over den samlede artikel på en skærm. En løsning på dette problem kan være at vise et resumé af artiklens hovedinddeling foroven. Det skaber overblik, og hvis oversigten er klikbar, har brugeren mulighed for straks at klikke sig frem til det ønskede sted frem for at skulle scrolle sig gennem et måske betragteligt antal skærmbilleder. For den der ikke vil læse artiklen i dens helhed, er det en god måde at finde den ønskede information hurtigt og præcist. Det er mærkelig nok ikke så udbredt som man kunne tro, men kendes dog i fx *Grimms* ordbog, i LDOCE og i *Wikipedia*. Alene af denne grund er *Grimms* ordbog væsentlig nemmere at bruge i netversionen end papirordbogen. Se figur 9.

Af figur 9 kan man også se at betydningsresuméet i *Grimm* er fremkommet ved simpelthen at tage de første ca. 100 karakterer af definitionen. Der er ikke foretaget nogen manuel redigering fra redaktørernes side. Det er gjort anderledes i LDOCE, hvor resuméet er holdt i korte, knappe overskrifter der dækker definitionen. Denne facilitet er med andre ord tænkt ind i konceptualiseringen af ordbogen (også af den trykte ordbog), jf. figur 10.

Det er klart at det gør en stor forskel om man vælger den ene eller den anden løsning. Hvis hver enkelt betydning skal forsynes med et særligt element der indeholder en glosse eller en sammentrængt definition til brug i resuméet, kræver det en betydelig redaktionel indsats, og det ligger uden for mulighederne i dette projekt. Der kan dog godt tænkes en automatisk eller semi-automatisk løsning som er anderledes end *Grimms* simple løsning. Fx fungerer fag- og stilmarkører ofte som måder at orientere sig i et ords betydninger på, og redaktionsinterne eksperimenter tyder på at en kombination af markører og uddrag af definitioner kan føre til et brugbart resultat (jf. Lorentzen og Trap-Jensen 2006, Trap-Jensen 2005).



Figur 9. Grimms Deutsches Wörterbuch. Bemærk at hele skærmens højrespalte er forbeholdt artikelnavigation.



Figur 10. Betydningsresumé i LDOCE. Et klik på "Menu" i øverste venstre hjørne kalder det centrale vindue frem.

I OED, som ellers har gjort meget for at fremme overskueligheden, gives der ikke artikelresuméer. En artikel som *go* har fx 96 nummererede hovedbetydninger og endnu flere underbetydninger (*go out* har alene underbetydninger op til bogstav u, dvs. 21 betydninger). Til gengæld er det en stor hjælp at man har mulighed for

at slå visse oplysningstyper til og fra i en såkaldt teleskopvisning. Navnlig citater fylder meget i en artikel, og det giver et væsentlig bedre overblik over artiklen hvis de ikke vises på skærmen. Af samme grund bør man overveje om artikelresuméet også skal være klikbart så brugeren selv kan slå det til og fra. Navnlig ved korte, overskuelige artikler vil mange nok finde det generende.

Den mulighed at brugeren selv bestemmer om oplysninger skal vises eller skjules, vil formentlig kunne udnyttes endnu bedre og derved få større betydning i fremtidens netordbøger. På den måde vil nøgleord som brugergrupper og ordbogsfunktioner komme til at få en anden rolle, idet den samme resurse kan rumme flere forskellige funktioner og henvende sig til forskellige brugergrupper. Det kommer nemlig helt an på hvilke oplysningstyper brugeren har slået til i sin opsætning, og tilsvarende kan den samme type oplysning formidles på forskellige måder. Man kan forestille sig at brugeren definerer en brugerprofil som svarer til netop de forudsætninger og behov han eller hun har, og i overensstemmelse hermed bliver de relevante oplysningstyper vist på skærmen. Idéen med at have fx både en kort og præcis ordbogsdefinition og en længere uddybende forklaring er ikke ny, men kendes også fra papirordbøger, ikke mindst fagordbøger (et eksempel er *Nordisk leksikografisk ordbok*). Der er ikke langt fra denne idé til at overlade det til brugeren at vælge hvilken udgave der skal vises i en netordbog. Et eksempel er *Musikordbogen*, der giver mulighed for at vælge kort eller lang visning.

4. Henvisninger

4.1 Henvisninger i trykte ordbøger

Hvis man tager udgangspunkt i den trykte ordbogs mere klart definerede henvisningsstruktur, er der i hvert fald tre forskellige argumenter for de henvisninger der optræder i en trykt ordbog som DDO. For det første et pladsøkonomisk argument. Det tager for meget plads op at gentage en oplysning to steder. Derfor henvises der fra ét sted til et andet. Eksempler er synonyme betydninger og dobbeltformer, altså officielt godkendte alternative skrivemåder. Ved redigeringen af DDO blev der i samarbejde med Dansk Sprognævn og Den Store Danske Encyklopædi udarbejdet en liste over hvilke dobbeltformer der skulle foretrækkes som hovedopslagsord og i redaktionel tekst.

For det andet er der et pædagogisk argument. Hvis man ikke ved hvordan et ord staves, kan man jo ikke slå det op. Derfor er henvisninger nødvendige hvis brugeren skal have reel hjælp til at komme fra en fejlagtig form til den rigtige. I tillæg er der en, omend indirekte, vejledende funktion i at henvise fra en lavfrekvent til den mere almindelige form eller betydning.

For det tredje bruges henvisninger til at gøre brugeren opmærksom på yderligere oplysninger i artikler andetsteds i ordbogen som kan være relevant for det ord eller den betydning man har slået op, og som man måske slet ikke var op-

mærksom på i udgangssituationen. Den slags henvisninger kaldes her plushenvisninger.

4.2 Henvisninger i digitale ordbøger

Set fra netordbogens synspunkt ser de samme argumenter – ikke overraskende – lidt anderledes ud: Det pladsøkonomiske argument gælder ikke som det efterhånden nok vil være mange bekendt. Og det er en grund til at nedlægge de variantartikler der har været nævnt. Ud fra samme argument kan man i princippet også nedlægge alle henvisninger mellem betydninger. Men inden den beslutning tages, må de øvrige argumenter lige overvejes.

Det pædagogiske argument falder i to dele. Teknisk set er der ikke noget til hinder for at lede brugeren direkte til det rette opslagsord uden et ekstra klik. Men der kan godt være en pædagogisk pointe i at vælge omvejen alligevel: Hvis man indtaster en anden form end opslagsformen og derefter kommer til en resultatside hvor det oplyses at det indtastede ikke er den rette form, er det muligt at brugerens opmærksomhed rettes mod forholdet og at han/hun måske lærer af det. Hvis man derimod kommer til den rigtige artikel med det samme, er der risiko for at den pædagogiske effekt udebliver. Derfor ser man i mange ordbøger at der stadig opereres med henvisninger, ikke mindst i learnerordbøger, fx i LDOCE og MEDO hvor der henvises fra fx *children* til *child*. Andre ordbøger vælger at bringe brugeren direkte til den rette opslagsform, fx *Oxford Advanced Learner's Dictionary Online* i eksemplet *children*.

Personlig finder jeg at man bør skelne mellem officielle og ikke-officielle varianter sådan at man kun føres direkte til opslagsordet hvis der er tale om en autoriseret form. Men man skal bemærke at det standpunkt er pædagogisk begrundet, og valget – uanset hvilken løsning man vælger – påvirker ikke beslutningen om at nedlægge henvisningsartikler som selvstændige poster i basen. Så længe tilladte, uofficielle og fejlagtige varianter er opmærket med separate tags i hver deres felt, kan man knytte forskellige visninger til dem i HTML-versionen. Med andre ord gør det ikke nogen forskel for redaktøren i det daglige arbejde. Men præsentationen af datastrukturen for brugeren bliver forskellig. Hvis der er gode argumenter for det, kan beslutningen nemt gøres om uden at det kræver anden redaktionel indsats end at ændre HTML-visningen af de involverede elementer globalt.

Det andet tilfælde, hvor der henvises fra en sjælden til en hyppig betydning, er lidt anderledes. I langt de fleste tilfælde vil brugeren være bedst tjent med at få besked på stedet, frem for at blive sendt til en anden artikel. Men undertiden kan man af pædagogiske grunde overveje at bevare henvisningen, mest oplagt hvis der er tale om en ekstrem fordeling, eller hvis betydningen er forkættet, fx ved kontamination af to betydninger.

Man bør dog være opmærksom på at det er nødvendigt at ændre datastrukturen en del for at den fungerer hensigtsmæssigt. Hvis man ønsker at vise den

samme oplysning to gange i forskellige artikler, bør det ikke ske ved at gentage samme tekst i identiske elementer i flere artikler. Det fører til uønsket redundans, som kan skabe problemer hvis man foretager en revision i den ene artikel, men glemmer den anden. Derfor bør betydningsbeskrivelsen kun stå i én artikel, mens der i eventuelle synonymartikler blot optræder en "pegepind" der viser videre til den fulde beskrivelse. I HTML-visningen ser brugeren selvfølgelig kun målartiklens fulde ordlyd. Derved sikres at en ændring af definitionen automatisk afspejler sig i artikler med synonyme betydninger.

En forudsætning for dette princip er at det er muligt at pege entydigt på en bestemt betydning. Det var ikke muligt under redigeringen af den trykte DDO fordi lemmaet som grundenhed var det eneste element der kunne bestemmes entydigt ved hjælp af en ID-kode. I forbindelse med projektet DanNet (se mere herom i fx Nimb, Pedersen og Trap-Jensen i denne publikation) er der imidlertid oprettet ID-numre på betydningsniveau, og det kan nu udnyttes til dette formål. Det indebærer en vis redaktionel indsats at ændre datastrukturen for eksisterende artikler så den konsekvent følger det beskrevne princip. For redaktørerne betyder det en ændret praksis fremover, på den måde at der kun skal skrives én definition, mens der ved synonyme betydninger skrives en reference til betydnings-ID'et i målartiklen. For ordbogsbrugeren gør det som nævnt ingen forskel: På skærmen ser man blot to artikler med identiske betydningsangivelser.

Endelig er der plushenvisningerne. De bør selvsagt stadig have en plads i ordbogen, for brugeren kan ikke på forhånd vide at der findes relevante oplysninger i andre artikler. Det kan være oplysninger af encyklopædisk art, eller det kan være oplysninger om grammatiske former eller beslægtede ord.

Svaret på spørgsmålet om alle henvisninger skal nedlægges, er selvfølgelig et nej, det skal de ikke. Men der er god grund til at nedlægge rigtig mange af dem, frem for alt dem der kun er begrundet i pladsøkonomi. Dog skal man være opmærksom på at det kræver en vis omorganisering af data.

Det omvendte spørgsmål, om alt skal være henvisninger, kan ligeledes besvares negativt, i hvert fald i den forstand der her er talt om henvisninger. Derimod kan man godt operere med to forskellige slags henvisninger, eksplicitte hyperlinks der er fremhævet som man er vant til at se dem, fx understreget eller i en bestemt farve, og derudover en generel mulighed for at dobbeltklikke på et ord, hvorefter det slås op i ordbogen. Det er en facilitet som man også ser i andre ordbøger (fx *Cambridge Advanced Learner's Dictionary Online* eller LDOCE), og en mulighed som vi også overvejer. For at fungere hensigtsmæssigt kræves det dog at alle forekomster i ordbogen entydigt kan tilordnes den rigtige opslagsform, dvs. at bøjnings- og homografproblemer er løst.

5. Flerordsforbindelser

Det er ikke fordi flerordsforbindelser adskiller sig principielt fra resten af ar-

tikelelementerne, men traditionelt er de et problembarn for papirordbøger på i hvert fald to områder. Det første er problemet med at finde ud af under hvilket af de indgående ord forbindelsen skal anbringes, det andet er udtrykkenes formmæssige variation. Et par eksempler fra den trykte DDO: Hvis man vil vide hvad *anabolske steroider* er, står forklaringen i artiklen *steroid*, mens der i artiklen *anabolsk* er en henvisning til udtrykket i artiklen *steroid*. Og hvis man vil vide hvad *som bare pokker* betyder, bliver man i artiklen *pokker* henvist til *som bare fanden*, som står i artiklen *fanden*, hvor man kan finde forklaringen. Det er fremgået at hvis pladsøkonomi er det eneste argument for at sende brugeren rundt ved hjælp af henvisninger, er der ingen grund til at bevare dem, og det gælder også i disse tilfælde. Det betyder at langt de fleste af henvisningerne kan nedlægges, og brugeren kan få forklaringen på stedet.

Med hensyn til datastrukturen kan man spørge sig om flerordsforbindelser egentlig behøver være knyttet til et bestemt af de indgående ord. Det kan forekomme mere hensigtsmæssigt med et databasedesign hvor flerordsforbindelserne er registreret i en selvstændig base, og hvor hvert enkelt af forbindelsernes indgående (betydende) ord peger på det tilsvarende opslagsord via ID-nummeret. Ved opslag på et bestemt ord er det derefter disse pegepinde der bestemmer hvilke ordforbindelser der vises under det pågældende ord. På den måde kan samme flerordsforbindelse vises i sin helhed under hvert enkelt af de indgående ord uden at man tager stilling til dens placering. Og der opstår ikke overflødig redundans fordi forbindelserne kun findes ét centralt sted. Ud over det redaktionelle arbejde med at omstrukturere data efterlader denne løsning blot et enkelt problem at tage stilling til: Hvordan håndterer man en søgning direkte på en flerordsforbindelse? Skal man bevare illusionen om artikeltilknytning og udnævne én artikel til "vært" for den pågældende forbindelse (som i fx NEO), eller skal man tillade et søgeresultat der består af andet end artikler (som i fx MEDO)? Det er indtil videre et uafklaret spørgsmål.

Endelig forudsætter løsningen ligeledes (jf. afsnit 4 nederst) at bøjnings- og homografiproblemer er løst.

6. Konklusion

Jeg har gennemgået nogle af de muligheder man har for at hjælpe brugeren med at finde frem til den ønskede information. Jeg har især interesseret mig for om de kræver ændringer i den eksisterende datastruktur. De muligheder og valg jeg har diskuteret, kræver forskellige indsatser fra redaktionens side.

En første redaktionel indsats vedrører konvertering af eksisterende data til en passende struktur. Den vil i en lang række tilfælde kunne foretages automatisk eller semi-automatisk for restgruppen af problematiske tilfælde, og gennemgangen har peget på tre typer af data. Den første er nedlæggelse af et antal artikler hvor der henvises fra én formvariant til en hovedartikel – dobbeltform, uofficiel

variant og sproglig eller ortografisk fejlform er de typer der blev opereret med i DDO. Nedlæggelse af henvisningsartiklerne vil kunne ske automatisk i ca. 80 % af tilfældene, mens de resterende 20 % kræver en manuel redaktionel indsats. Det skyldes at lavfrekvente fejlformer ikke har fået en beskrivelse i hovedartiklen i DDO, og derfor må oplysning herom manuelt overføres til hovedartiklen når henvisningen nedlægges.

Endvidere gælder det nedlæggelse af mange henvisninger på betydningsniveau, nemlig dem der ikke er hverken pædagogiske henvisninger eller plushenvisninger i DDO, men udelukkende har sin begrundelse i pladsøkonomi. Her er udfordringen at udskille de pædagogiske fra de ikke-pædagogiske. Det er ikke så svært i de tilfælde hvor der er tale om kontroversiel brug (eksempelvis kontamination), for dér er betydningen forsynet med en kommentar der gør opmærksom på det. Dog kan vi ikke isolere de tilfælde hvor en henvisning kunne være ønskelig pga. en ekstrem frekvensfordeling, til forskel fra de pladsøkonomisk begrundede henvisninger, og derfor bliver vi nødt til at træffe et valg. Alternativet er at gennemgå ca. 10.000 henvisninger manuelt for at tage stilling til om der skal bevares en henvisning af pædagogiske grunde, og det er usandsynligt at resurserne rækker til at prioritere denne opgave.

Endelig er der flerordsforbindelserne, som bør konverteres til en separat base og forsynes med pegepinde der henviser til de indgående opslagord via ID-numre. Også dette kan gøres automatisk for manges vedkommende, men der vil være en restgruppe af problematiske tilfælde (typisk homografiproblemer) som kræver manuel efterbehandling.

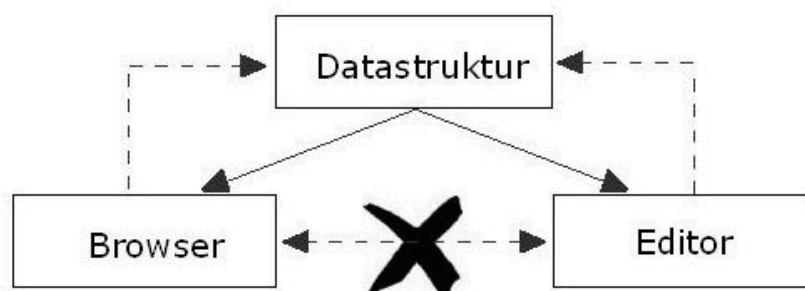
Den anden type af redaktionel indsats er redigering. Jeg har allerede nævnt at der vil være en række problematiske tilfælde som kræver en manuel bearbejdning. Så vidt jeg kan bedømme, er opgaven dog overkommelig så længe det er muligt at isolere disse tilfælde.

Den sidste type redaktionel indsats er programmeringsopgaver, og her drejer det sig især om visning af data i brugergrænsefladen. Blandt de opgaver jeg har nævnt, er hjælp til udvidet søgning, søgehjælp i form af trunkeret søgning og en algoritme der giver meningsfulde bud på mulige alternativer. Derudover er det artikelresumé, teleskopvisninger og endelig en generel mulighed for at kunne slå op vha. dobbeltklik.

Hvis vi til sidst vender tilbage til diagrammet fra artiklens indledning, er vi måske blevet lidt klogere på sammenhængen mellem data, browser og editor. Jeg er hverken programmør, it-ekspert eller datalingvist, men vil gerne være med til at bygge bro mellem det der undertiden opfattes som to adskilte verdener, leksikografens og teknikerens.

I arbejdet med netordbogen går det i stigende grad op for mig hvor vigtig organiseringen af data er. Datastrukturen er det der bestemmer hvad det er muligt at vise både for redaktøren i redigeringssystemet og for brugeren i en browser. Derfor går påvirkningsretningen fra strukturen til grænsefladerne. Til gengæld

kan vi som brugere kun forholde os til data via det interface der afspejler data, altså gennem browserens grænseflade hvis vi er brugere af ordbogen, eller gennem editorens grænseflade hvis vi redigerer ordbogen. Editoren er også det sted hvor redaktørerne kan kode de data som lagres i basen. Så der sker en vis form for feedback fra det program vi ser data i. Derimod, og det er måske det vigtigste budskab i denne sammenhæng, er der ingen direkte forbindelse mellem det redaktøren ser når han/hun redigerer en artikel, og det slutbrugeren ser i sin browser. Redaktøren kan redigere data via et databasesystem, en simpel XML-editor, en webbaseret grænseflade eller et særligt redigeringsystem udviklet til formålet. Det kan alt sammen være mere eller mindre hensigtsmæssigt set fra redaktørens side, men det har i bund og grund ikke noget at gøre med det skærbillede som slutbrugeren kommer til at se på. Et bud på sammenhængen kunne derfor være som i figur 11.



Figur 11.

Det er i stigende grad vigtigt at vi som leksikografer også interesserer os for emner som databasedesign, XML-struktur og elementær programmering, ikke fordi vi skal overtage teknikernes arbejde, men fordi det er nødvendigt for at forstå hvilke muligheder de digitale medier stiller til rådighed, og ikke mindst er det nødvendigt i projektstyringen for at kunne vurdere hvor meget tid og hvor mange resurser en bestemt beslutning kræver for at blive ført ud i livet.

Litteratur

Ordbøger:

Cambridge Advanced Learner's Dictionary Online: <http://dictionary.cambridge.org>

Chambers 21st Century Dictionary Online: <http://www.chambersharrap.co.uk>

DDO = Ebba Hjorth, Kjeld Kristensen m.fl. (red.). *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab og Gyldendal, København 2003–2005.

Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm: <http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/dwb/wbgui?lemid=GA00001>

ellexiko: http://hypermedia.ids-mannheim.de/pls/ellexiko/p4_start.portal

LDOCE = *Longman Dictionary of Contemporary English Online*: <http://pewebdic2.cw.idm.fr>

Lexin: <http://lexin.nada.kth.se/lexin.html>

MEDO = *Macmillan English Dictionary Online*: <http://online.macmillandictionary.com>

Musikordbogen: <http://musikordbogen.dk>

NEO = *Nationalencyklopedins Ordbok*: <http://www.ne.se>

Nordisk leksikografisk ordbok, Henning Bergenholtz, Ilse Cantell, Ruth V. Fjeld, Dag Gundersen, Jón Hilmar Jónsson, Bo Svensén. Oslo: Universitetsforlaget 1997.

ODS = *Ordbog over det danske Sprog*: <http://ordnet.dk/ods>

OED = *Oxford English Dictionary*: <http://dictionary.oed.com/>

Oxford Advanced Learner's Dictionary Online: <http://www.oup.com/elt/catalogue/teachersites/oald7/?cc=dk>

Retskrivningsordbogen: <http://www.retskrivningsordbogen.dk>

SAOB = *Svenska Akademiens Ordbok*: <http://g3.spraakdata.gu.se/saob/>

Wikipedia: <http://da.wikipedia.org>

Anden litteratur

Burke, Sean Michael 2003: The design of online dictionaries. I: Piet van Sterkenburg (ed.): *A Practical Guide to Lexicography*. Amsterdam: John Benjamins Publishing Company, 240-249.

Lorentzen, Henrik og Lars Trap-Jensen 2006: *ordnet.dk* – et nyt sprogligt opslagsværk på internettet. I: Henrik Lorentzen og Lars Trap-Jensen (red.). *Nordiske Studier i Leksikografi, NFL-skrift nr. 9*. København, 253–264.

Svensén, Bo 2004: *Handbok i lexicografi*. Stockholm: Norstedts Akademiska Förlag.

Trap-Jensen, Lars 2005: Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: *LexicoNordica 12*, 109–122.

Lars Trap-Jensen
cand.mag, MPhil, f. 1960
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
ltj@dsl.dk

LARS S. VIKØR

Behandling av fraseologi i Norsk Ordbok

Treatment of phraseology is an important part of lexicography, which has been given extensive attention during the most recent decades. In *Norsk Ordbok* ('Norwegian Dictionary'), as in many sister projects, this has been done less systematically than desirable in the past. At present, when the project is being thoroughly revised, a systematization has been undertaken. This is not yet finished, but here I introduce a typology of Norwegian phrases related to three possible ways of presenting them in the dictionary: as separate lemmas (only recommended for foreign phrases imported integrally into Norwegian), as sublemmas (roughly for the most idiomatic phrases) and as editorial examples (mostly collocations or looser word combinations, and expressions with a normal sentence structure). The typology is based on grammatical criteria, and for each category examples are given, followed by an indication of the most suitable lexicographic treatment according to the view of the author.

1 Bakgrunn

Fraseologi er ikkje berre eit spesialområde for ein bestemt ordbokssjanger, men eit kjerneområde for alle større ordbøker, og aller mest for store nasjonale definisjonsordbøker. I tradisjonell leksikografi har dette språknivået ofte vorte behandla ganske usystematisk, og mitt prosjekt, *Norsk Ordbok* er ikkje noko unntak i så måte. Dei siste åra har vi arbeidd med ei systematisering som blant anna har bestått i ei klassifisering av frasetypar. Det kan gjerast på ulike måtar og ut frå ulike kriterium, men eg legg her fram eit forsøk eg har gjort i *Norsk Ordbok*-samanheng. Eg har bygd på tidlegare forsøk og grunnlagsmateriale frå kollega Oddrun Grønvik i det same prosjektet, og på fraseologi-kapitlet i Valerij Berkovs *Norsk ordlære* (Berkov 1997: 93–107). Men slik denne typologien blir lagd fram her, er eg eineansvarleg for han. Han er ikkje "vedteken" i *Norsk Ordbok*-samanheng, men skal munne ut i eit omforeint sett retningslinjer for den leksikografiske behandlinga av frasar, som vi i *Norsk Ordbok* kallar *fleirordige uttrykk*. Og sidan dei nordgermanske språka vel er nokså like når det gjeld grunnmønstra i fraseologien, håpar eg at dette kan ha ein viss overføringsverdi til andre, i alle fall som eit diskusjonsgrunnlag.

2 Tre behandlingsmåtar

Fleirordige uttrykk kan i *Norsk Ordbok* i prinsippet behandlast på tre måtar: Dei kan vere *eigne oppslag* (fleirordslemma), dei kan vere *underoppslag* (sublemma),