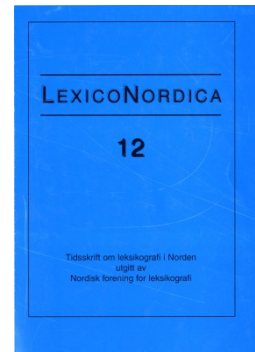


# LexicoNordica

Titel: Om elektroniske ordbøger – til brug for mennesker  
Forfatter: Anna Braasch  
Kilde: LexicoNordica 12, 2005, s. 55-69  
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.

*Anna Braasch*

## **Om elektroniske ordbøger – til brug for mennesker**

This paper deals with products which are in general terms called 'electronic dictionaries'. Firstly, a brief retrospect is given to the increasing use of computers in lexicography. Secondly, four selected types of electronic dictionaries and their basic features are discussed. Thirdly, some essential technical facilities and lexicographic design features are outlined. Hereafter the utility value of fully developed electronic dictionaries is sketched. An excursus presents briefly the use of a Danish computational lexicon, den Sprogteknologiske Ordbase, as an electronic dictionary for humans. Finally, visions for the future developments are presented.

### **0. Indledning**

På årets Schæffergård-symposium var temaet elektroniske ordbøger, et emne der i allerhøjeste grad er relevant for leksikografiens nutid og fremtid. Brugen af det elektroniske medium i udarbejdelse, præsentation og brug af leksikografiske produkter er noget nær en revolution. De nye tekniske muligheder der i høj grad er knyttet til databaseteknologien, er kun den ene side af denne udvikling. Den teknologiske udvikling stiller også voksende krav til nye veje i udarbejdelsen af det leksikografiske grundlag og design af ordbøger for elektronisk anvendelse.

Dette indlæg, der er en omarbejdet version af mit foredrag ved symposiet, beskriver leksikografiske og tekniske egenskaber ved produkter der med et overordnet udtryk kaldes elektroniske ordbøger. Beskrivelsen tager sit udgangspunkt i et meget kort tilbageblik på den historiske udvikling på området. Der afgrænses så fire hovedtyper af elektroniske ordbøger, hvorefter der fokuseres på de produkter som lever op til nogle basale leksikografiske kvalitetskriterier. I indlæggets anden halvdel ligger hovedvægten på de tekniske og leksikografiske aspekter i design, udvikling og anvendelse af genuine elektroniske ordbøger.

### **1. Historisk tilbageblik**

Det er mere end 30 år siden at computeren gjorde sit indtog i ordbøgerens verden. Erfaringerne fra elektronisk tekstbehandling og

terminologiske databanker kom hurtigt i anvendelse og videreudvikling i leksikografien. På forskningssiden bidrog den nye disciplin inden for sprogvidenskab, datalingvistikken, til den rivende udvikling, som stadig er i fuld gang.

De allerførste elektroniske 'ordbøger' blev dog udviklet i forbindelse med den spæde begyndelse af maskinoversættelse og anden maskinel behandling af naturlige sprog. Der er sket meget siden de første hulstrimmel- eller hulkortbaserede ordlister, og siden de maskinlæsbare versioner af trykte ordbøger blev lagret på magnetbånd til brug for fotosætteren frem til nutidens store internetordbøger og små elektroniske parlører på computere i lommeformat.

Computeren som værktøj vandt for alvor indpas i leksikografien med udviklingen af forskellige elektroniske hjælpemidler til behandling af tekstuelle data fra omkring 1980. Et par eksempler på relevante indsatsområder var optisk læsning og digitalisering af trykte tekster (OCR), der blev brugt til at gøre trykte ordbøger tilgængelige i maskinlæsbar form og dermed egnede til udnyttelse i senere elektroniske versioner. Disse data blev derefter analyseret, konverteret og restruktureret ved hjælp af computere.

Udarbejdelse af ordbogsmateriale, redigering og editering foregik efterhånden i tekstbehandlingsbaserede ordbogsværktøjer som fx COMPULEXIS. Senere kom det danske ordbogssystem GESTORLEX, der bygger på databaseteknologien. Computernes stadig voksende kapacitet og hastighed åbnede mulighed for udviklingen af stadig flere, bedre og større leksikografiske værktøjer og produkter.

De første elektroniske ordbøger i Danmark udkom på disketter i 1989, og det var de digitaliserede udgaver af Gyldendals røde dansk-engelske og dansk-tyske ordbøger. De bød på nye muligheder i forhold til papirudgaven: Man kunne fx søge på dele af et udtryk på kildesproget og få oplyst den korrekte ækvivalent til hele udtrykket på målsproget, og man kunne med et enkelt tastetryk overføre det man havde slået op, til den tekst man arbejdede med på skærmen. Layoutet lignede til forveksling den trykte ordbog; fordelene ved dette var at man let kunne vænne sig til at bruge den elektroniske version, men skærbilledet var til gengæld ikke optimalt.

Disse ordbøger havde dog nogle ulemper; de krævede f.eks. ganske meget lagerplads (2,8 Mb) og 105 Kb indre hukommelse i forhold til en pc's dengang almindelige kapacitet. En af de frygtede fejlmeddelelser var "*Ikke nok internt lager*", hvilket betød at programmet (ordbogen) ikke kunne køre. Desuden kunne ordbøgerne ikke anvendes sammen med Windows o.l. programmer. De nåede vistnok ikke så stor en udbre-

delse som forventet, men de banede vejen for de senere, teknisk velfungerende udgaver.

I dag er situationen en ganske anden; størrelsen på computerens harddisk, cd-rom og internettet samt bredbåndskommunikationen udvidede grænserne betydeligt for hvad der er teknisk muligt.

## 2. Elektronisk ordbog – hvad er det?

Siden 1990'ernes hurtige teknologiske og kommercielle udvikling er udbuddet af såkaldte 'elektroniske ordbøger' vokset støt. I dag bliver ganske forskellige typer af systematiske samlinger af ord på elektronisk medium kaldt elektroniske ordbøger, men set fra en leksikografisk synsvinkel er der stor forskel på disse produkter. Der skelnes nedenfor mellem fire hovedtyper, hvoraf kun de sidste to strengt taget falder inden for kategorien ordbøger som leksikografiske produkter. De fire hovedtyper er afgrænset over for hinanden ud fra en kombination af flere kriterier, herunder anvendelsesområde, omfang, indhold, distributionsmedium og leksikografisk niveau. Formålet er ikke at præsentere en gennemarbejdet og detaljeret typologi, men blot at illustrere produkternes forskellighed ved hjælp af nogle få eksempler.

### 2.1. Lommeordbøger og -oversættere

Elektroniske lommeordbøger o.l. er i leksikografisk forstand næppe egentlige ordbøger selv om de markedsføres under betegnelsen elektronisk ordbog. De indeholder dagligdags ord og fraser, og de har en stor indholdsmæssig lighed med turistparlører. Almindeligvis er de lagret på små lommecomputere, og som regel kan der vælges mellem flere sprogkombinationer.

Oversættelsespenen kan anses som en af varianterne: Den er udstyret med en lille optisk læser der føres hen over en trykt sætning på et sprog (engelsk), og man kan aflæse en slags oversættelse til dansk på pennens display. Ordforrådet kan variere fra nogle få hundrede op til et par tusinde.

### 2.2. Online orddatabaser

Denne gruppe omfatter en række ordsamlinger (glossarier) der betegnes som elektroniske ordbøger, men set fra en leksikografisk synsvinkel

ikke opfylder alle gængse kriterier for hvad man kalder en ordbog. Der er snarere tale om forskellige typer af et-, to- eller flersprogede ord- eller termlister. Den måde disse lister opstår eller udvikler sig på, afviger også fra de almindelige leksikografiske arbejdsprocesser, eksempelvis i tilfælde af **open-source**-produkter, hvor brugerkredsen bidrager aktivt med nye ord, oversættelser mv. Det betyder at processen er dynamisk og baseres på frivillig indsats og innovative ideer, samtidig med at der er stor risiko for at en sådan ordsamling har alvorlige mangler med hensyn til systematik, præcision og pålidelighed. Modellen er eksempelvis beskrevet i Docherty (2000:67):

Open-source projects on the Internet offer Web surfers free access to a large volume of information, including dictionary resources. At the same time, individual users are invited to comment on, add to and improve existing entries, thus creating a dynamic resource which is wonderfully interactive, but, by definition, uncontrolled and perhaps even uncontrollable.

I nogle tilfælde bliver sådanne bidrag kontrolleret og redigeret af ordbasens udvikler (eller 'ejer') eller en ekspert for at tilpasse og kvalitets-sikre materialet inden det medtages i ordbasen. Dette er faktisk en ganske tidskrævende opgave.

Disse ordbasers formål er anderledes end de leksikografisk vel-funderede ordbøgers, eksempelvis bruges de især som hjælp ved surfing på internettet og læsning af internetsider på forskellige sprog. De kan også være ganske udmærkede hjælpemidler til at give en forståelse af nye termer, men lemmaselektionen kan være mere eller mindre tilfældig, og kvaliteten er uforudsigelig med hensyn til præcision og pålidelighed.

### *2.3. Elektroniske udgaver af papirordbøger*

I disse tilfælde er der tale om en digitaliseret version af en trykt ordbog der ofte forhandles på cd-rom (eller dvd) til installation på pc eller netværksserver. Indholdet er identisk med den trykte ordbogs, men er bearbejdet for det elektroniske medium (eksempelvis indekseret med henblik på søgning på krydsreferencer osv.) Layoutet svarer stort set til den trykte version, tilpasset til visning på computerskærm. Desuden kan der være knyttet forskellige småprogrammer til ordbogen der kan bruges i bogstavleg.

Der er mange forskellige undertyper med hensyn til ordbogens leksi-kografiske art, indhold og omfang: Retskrivningsordbogen, Nudansk

Ordbog, forskellige skoleordbøger og store oversættelsesordbøger samt fagordbøger findes f.eks. i Danmark i elektronisk version. Fælles for dem alle er at de som regel er (vel)dokumenterede med hensyn til de tilgrundliggende leksikografiske principper og metoder samt ordbogens indhold. På denne måde kan det siges at disse ordbøgers kvalitet er forudsigelig, og brugeren med rette har bestemte forventninger til præcision og pålidelighed. Disse ordbøger har opnået en stor popularitet og bruges i mange undervisnings- og arbejdsituationer.

En videreudvikling af traditionelle ordbøger i digitaliseret form tilbydes til engelske sprogbrugere, der som bekendt udgør et meget stort aftagerpotentiale. Det drejer sig om et system af engelske ordbøger (Oxford Dictionary of English 2nd Edition, New Oxford Thesaurus of English samt Oxford Dictionary of Quotations) udbygget med stavetkontrol, krydsords- og anagramfunktion, diverse søgefinesser, variable skriftstørrelser mv. på en lommecomputer der vejer 148 gram (SHARP Electronic Dictionary PW-E500).

#### *2.4. Ordbøger udarbejdet til det elektroniske medium*

Denne hovedtype omfatter ordbøger der er designet og udarbejdet direkte til elektronisk præsentation til online (over internettet) eller offline brug (fra cd-rom eller dvd). De baserer sig på velfunderede leksikografiske principper og metodiske overvejelser der vedrører både indhold, struktur, præsentation, anvendelse og brugervenlighed.

Typen som sådan omfatter en række meget forskellige produkter, herunder kan nævnes sådanne der også indeholder en stor mængde encyklopædiske oplysninger evt. illustreret med billeder, andre er en kombination af en velstruktureret tosproget ordbog og avanceret kontekstbaseret oversætterhjælp (Prószéky 2000), eller en ordbog med integreret grammatik.

For dansk er der på nuværende tidspunkt adgang til en række ordbøger, hvoraf nogle er frit tilgængelige, f.eks. Den Danske Netordbog, Den Danske Idiomordbog og regnskabsordbøgerne. Det samme gælder den Sprogteknologiske Ordbase (STO) der er tilgængelig for online opslag (som beskrevet nedenfor i afsnit 5). Disse ordbøger er udarbejdet på offentlige institutioner.

Andre ordbøger, der er udarbejdet på kommerciel basis, kræver abonnement eller licens, eksempelvis L&H-fagordbøgerne eller Ordbogen.com (dansk-engelsk og engelsk-dansk). De først nævnte fagordbøger findes dels som enkeltordbøger, dels integreret i L&H's internetba-

serede ordbogssystem (WebWords), der kan benyttes med brugerens browser.

Hvis man interesserer sig for hvilke ordbøger der er online tilgængelige, så kan man bruge en ordbogssøgemaskine (f.eks. *Lexicool* eller *DMOZ*; sidstnævnte viser kun de tilmeldte online ordbøger) eller undersøge de lister der findes over ordbøgerne (f.eks. ved hjælp af *YourDictionary.com*). Der er dog ingen garanti for at søgeresultaterne er udtømmende. En mere udførlig liste findes på adressen <http://uniquetranslations.dk/links.htm?#ord>.

I det efterfølgende drøftes denne hovedtype af elektroniske ordbøger set fra to tæt forbundne synsvinkler: fra produktions- og anvendelsessiden, samtidig med at nogle af nutidens løsninger og fremtidens perspektiver skitseres.

### **3. Produktion af elektroniske ordbøger**

Det er meget tidskrævende – og dermed også dyrt – at udarbejde store ordbøger af høj kvalitet; derfor er der på produktionssiden behov for at indføre nye leksikografiske koncepter og metoder og finde frem til teknisk mulige og sikre løsninger som i kombination kan leve op til kravene om kvalitet, fleksibilitet og rentabilitet. Nielsen (2002) redegør for overvejelserne vedrørende de leksikografiske og anvendelsesorienterede spørgsmål der gik forud for udarbejdelsen af en elektronisk bilingval regnskabsordbog. (Regnskabsordbøgerne mellem dansk og engelsk er tilgængelige over internettet fra Center for Leksikografi, HHÅ's hjemmeside.)

#### *3.1. Komplekse ordbaser – polyfunktionale ordbøger*

En farbar vej er at udarbejde komplekse ordbaser eller ordbøger hvis indhold og oplysningsstruktur kan dække behovet mht. forskellige brugertyper, funktioner og brugssituationer, hvor dataudtræk kan styres af anvendelsens formål og brugerens behov. Samtidig er det vigtigt at ordbogen har de nødvendige leksikografiske og anvendelsesmæssige kvaliteter. Duvå og Laursen (2005:93) nævner fem ”overordnede kvalitetskrav til en fagsproglig internetordbog”. Disse krav har også gyldighed i forbindelse med produktionen af ikke-fagsproglige ordbøger, og de er følgende: overskuelighed, let tilgængelighed, forudsigelighed,

pålidelighed, tilstrækkelighed. Effektiv søgning i ordbogen forudsætter at disse krav er opfyldt.

Desuden påpeges (ibid.) at der finder en brancheglidning sted, således at en og samme formidlingsopgave "...kan omfatte såvel oversættelse som produktion og reception...". Det begrundes samtidig behovet for polyfunktionale ordbøger, hvilket igen gælder både for almensprog og fagsprog. Udgangspunktet er at den tidligere opdeling i receptions- og produktionsordbøger, og ordbøger til enkeltformål som oversættelse eller undervisning er blevet mindre relevant i takt med databaseteknologiens udvikling – de aktuelle eller fremtidige rummelige ordbogskoncepter bør fokusere på at kunne omfatte indhold og datastruktur til mange formål. Denne teknologi gør det også muligt at producere dataudtræk til formålsspecifikke ordbøger, eksempelvis en ordbog der indeholder alle opslagsord og udtryk der relaterer sig til et bestemt emneområde (musik, historie osv.) eller har samme type etymologisk oprindelse mv.

### *3.2. Flere ordbøger benyttet som et kompleks*

En anden metode er at forbinde flere internetordbøger elektronisk ved hjælp af en fleksibel, fælles platform. Et vist mål af harmonisering og standardisering synes at være nødvendigt i denne sammenhæng, selv med de teknisk set fleksible løsninger. Armstrong et al. (2000:62) beskriver problemet således:

Unfortunately, the power of today's advanced search techniques is hindered by the variety of formats and proprietary interfaces of these offerings. Even among the offerings of a single publisher, not all CD-ROMs have the same interface. Users are required to deal with multiple search programs, cluttering their screens with redundant windows and toolbars.

Standardiseringen kan vedrøre beskrivelsessprog, harmoniseringen kan gælde valg af oplysningstyper, design af fordelings- og henvisningsstruktur hhv.

Behovet for ensartede eller i det mindste kompatible elektroniske leksikografiske ressourcer, herunder ordbøger, var et af hovedemnerne ved et symposium om *Lexicography in the Electronic Age* i 1981, og allerede på det tidspunkt blev koordinering og harmonisering i forbindelse med udarbejdelse af ord- og termdata formuleret som fremtidsvision uden at rundbordssamtalerne var nået frem til ideen om nogen egentlig standardisering.



Der findes forskellige tekniske løsninger i dag, eksempelvis bogskabsløsningen, som muliggør samtidig brug af flere ordbøger der er installeret på samme pc (f.eks. Gyldendals ordbøger og L&H ordbøger). En anden mulighed er rent internetbaserede ordbogssystemer, hvor ordbøgerne ligger på udbyderens server, og brugeren anvender sin standardbrowser til søgning (f.eks. det tidligere nævnt WebWord-system).

Desuden er der en række internationale initiativer i gang med henblik på at udarbejde formelle modeller og strukturer for polyfunktionale ordbaser, først og fremmest til sprogteknologiske formål (se eksempelvis Bertagna et al. 2004). Det har vist sig tidligere at ideer og forskningsresultater fra datalingvistik og sprogteknologi (eksempelvis brugen af semantiske net; difonbaseret talesyntese af transskriberet udtale) finder vej til leksikografien og bidrager til udviklingen af nye koncepter for elektroniske ordbøger. Derfor kan det også meget vel ske at de ovenfor nævnte sprogteknologisk orienterede harmoniserings- og standardiseringsbestrebelse med tiden påvirker produktionen af elektroniske ordbøger for mennesker.

### *3.3. Teknologiens rolle i leksikografisk praksis*

Computerbaserede processer har ændret den leksikografiske praksis ganske betydeligt. Det leksikografiske arbejdsgrundlag udgøres i høj grad af digitale tekster; dataindsamling sker ved hel- eller halvautomatiske processer. Computeren kan håndtere meget store tekstmængder med stor sikkerhed og fremstille konkordanser, frekvenslister, morfologiske analyser, strukturelle opmærkninger – blot for at nævne de mest almindelige operationer. Leksikografen har redigeringsystemer til sin rådighed, der samtidig sørger for fuld formel konsistens i hele ordbogen, indholdsmæssigt mellem forskellige medarbejders kodninger. Desuden er det altid muligt at få overblik over ordbogens, artikeltypers eller enkelte artiklers status, udføre systematiske omstruktureringer og opdateringer (med de farer det indebærer for utilsigtede ændringer ...).

Alle disse teknikbaserede processer sammen med computerens stadig voksende kapacitet fører til at leksikografien har helt nye muligheder med hensyn til ordbogens nærmest ubegrænsede omfang inkl. antallet af opslagsord og artikellængde (der igen bestemmes af mængden af oplysninger til de enkelte opslagsord, antal og længde af eksempler, citater osv.). Et yderligere aspekt er at begreberne nyt oplag og revideret udgave erstattes med **version** – opdatering kan ideelt set ske både med fastlagte mellemrum og efter behov.

Det er et interessant aspekt at en sådan behovsbaseret opdatering kan udløses af brugernes mislykkede eller resultatløse søgninger der registreres i en logbog. Disse registreringer analyseres automatisk med henblik på at finde grunden til at brugerne ikke fik det ønskede resultat. Analyseresultaterne systematiseres ud fra fejlkilden som kriterium (fejlstavninger, søgninger på nydannelser og nye fremmedord er hyppige grunde til 'ikke fundet'-svar). Metoden beskrives nærmere i de Schryver & Joffe (2004).

De nye leksikografiske muligheder sprænger helt de fysiske rammer der kendes fra trykte ordbøger, omfang/papirforbrug spiller nemlig ingen rolle i denne sammenhæng. Det er fristende at ønske at kunne udnytte alle tekniske muligheder og faciliteter i produktionen af en og samme ordbog. Der er dog stadig de givne økonomiske rammer at tage hensyn til – og den væsentligste udfordring ligger derfor i at optimere forholdet mellem teknik og leksikografisk produkt.

#### **4. Elektroniske ordbøger som produkt i anvendelse**

Den elektroniske ordbog som produkt har nogle generelle fortrin i forhold til den trykte ordbog. Brugeren kan forvente nye, opdaterede versioner med jævne mellemrum hvor nye ord og/eller betydninger er registreret og beskrevet. En banal, men ikke uvæsentlig detalje er at en elektronisk ordbog ikke kræver hyldeplads til opbevaring.

##### *4.1. Leksikografisk kvalitet*

Det er indlysende at en elektronisk ordbog har nogle særlige leksikografiske kvaliteter hvis de tekniske muligheder er optimalt udnyttet. Dette vil i princippet sikre at ordbogen har høj brugsværdi fordi den eksempelvis indeholder mange opslagsord der er udførligt beskrevet og er forsynet med flere eksempler, citater og indeholder mange fraseologiske enheder, eftersom en relationel database kan rumme meget store datamængder i komplekse datastrukturer. Det bør bemærkes at tekstkondensering inkl. pakkede informationer er uhensigtsmæssig fordi sådanne forhold komplicerer eller forhindrer korrekt indeksering, som er en forudsætning for at finde de søgte informationer.

Det er velkendt at en elektronisk ordbog er meget konsekvent i behandlingen af ord både med hensyn til oplysningsmængde og henvisningssystemet, og den kan med fordel bruges sammen med et tekstbehandlingssystem; desuden kan brugeren indsætte egne noter. Lignende

egenskaber kan også opnås ved digitaliserede ordbøger med passende implementering.

#### 4.2. *Brugergrænseflade*

En god brugergrænseflade gør tilgangen til oplysningerne i ordbogen hurtig, effektiv, præcis og fleksibel. Design af informationsindholdets fordelingsstruktur og udvikling af brugergrænseflade er to uadskillelige processer. Fordelingsstrukturen vedrører opdeling af informationsindholdet i oplysningstyper, fordeler disse på en hensigtsmæssig måde i tabeller og indekserer dem på passende vis. Jo mere eksplicit og finmasket fordelingen er, desto mere præcist kan oplysningerne indekseres, hvilket i sidste ende er grundlaget for vellykkede søgninger og overskuelige præsentationer af søgeresultater. Det forhindrer dog ikke at søgeresultatet vises i kondenseret form på skærmen.

En elektronisk ordbog kan have mange funktionaliteter, herunder forskellige søgeretninger, præsentationsformer af en ordbogsartikel og præsentation af andre oplysninger end rent leksikalske. Den mest almindelige søgeretning i eksisterende produkter er **semasiologisk**-baseret, dvs. den tager udgangspunkt i det sproglige tegns forside og finder bl.a. dets betydninger. Den semasiologiske søgning kan også foretages med ordet i trunkeret form og med jokertegn.

Den **onomasiologiske** søgeretning tager sit udgangspunkt i det sproglige tegns indholdsside, dvs. begrebet, og finder de(t) relevante ord (hvis der er flere, så er de hinandens synonyme). Dette er langt mere krævende end den formbaserede (semasiologiske) søgning, idet onomasiologisk søgning forudsætter implementering af et underliggende begrebssystem og eksplicit og konsekvent multiindeksering af ordbasens fulde indhold. I denne søgningsform benyttes som regel de logiske operatører OG, ELLER og IKKE, som er grundlaget for komplekse søgninger. De særlige leksikografiske og sprogteknologiske udfordringer samt forskellige søgningsmuligheder er skitseret i Geeraerts (2000:80 ff.).

Der er to vigtige præsentationsformer som er tæt beslægtede med hinanden; begge er baseret på oplysningernes hierarki og fordeling og beregnet til at vælge/fravælge oplysninger der indgår i søgeordets beskrivelse.

Den første kaldes **lagdelt visning**, som især er relevant i tilfælde af lange, komplekse artikler, hvor artiklen kan deles op i hoved, krop og fod, hvorefter både kroppen og foden igen kan deles op i flere dele. I denne visning bestemmer brugeren hvilken del af basisvisningen der er

relevant for ham, eksempelvis artikelhovedet med grammatiske og udtalemæssige oplysninger.

Den anden metode er **teleskopvisningen**, hvor brugeren vælger en bestemt oplysningstype som han gerne vil se i alle tilgængelige detaljer fra en hel oversigtsartikel. Her drejer det sig om at folde artiklen ud på et bestemt punkt, f.eks. hvis brugeren gerne vil se flere kollokationer hvor søgeordet optræder i sin tredje registrerede betydning.

En række søgefaciliteter understøtter forskellige underholdningsmæssige anvendelser, f.eks. med henblik på løsning af krydsord og sproglige gåder eller til forskellige bogstavspil. En speciel søgemulighed returnerer et sæt af ord med enslydende sidste stavelse, søgeresultatet er baseret på udtaleoplysningerne i ordbogen. Sådanne faciliteter kalder man ofte **edutainment** (*education + entertainment*); det er ikke usandsynligt at netop disse faciliteter er medvirkende til at en ordbogs brugerkreds vokser.

#### 4.3. Andre understøttende funktioner

Grænsen mellem ordbog, encyklopædi, grammatik og tekstkorpus udviskes efterhånden fordi man kan søge i flere ressourcer, eller fordi der i samme ressource er integreret forskellige oplysningskomplekser.

Mange ordbøger har en grammatik i form af bøjningsparadigmer og mere eller mindre simple syntaktiske regler. Disse kan slås op i forbindelse med et lemma, og den relevante oplysning bliver vist i et pop op-vindue. Ordbogen kan desuden indeholde hyperlinks, multimediaelementer – eksempelvis billede eller lyd, idet sidstnævnte kan stamme fra en udtalefil eller være genereret ved talesyntese.

Der er desuden nogle standardfunktionaliteter som online teknisk og leksikografisk brugervejledning og adgang til forskellige omtækster.

## 5. Ekskurs: den Sprogteknologiske Ordbase (STO) for dansk

STO er, som navnet siger, primært udviklet til anvendelse i sprogteknologiske produkter og til datalingvistisk forskning. Den indeholder i lighed med en elektronisk ordbog et nærmere afgrænset ordforråd og dets lingvistiske beskrivelse som er udarbejdet i overensstemmelse med faste leksikografiske retningslinjer, de såkaldte lingvistiske specifikationer.

Ordbasens materiale er korpusbaseret og omfatter i alt mere end 81.000 lemmaer, hvoraf ca. 13.500 stammer fra fagsproglige tekster, men ikke er egentlige eksperttermer. Den lingvistiske beskrivelse omfatter en udtømmende bøjningsmorfologi og stavevarianter for alle

opslagsord, desuden meget detaljeret beskrivelse af syntaktiske konstruktioner for ca. 45.000 opslagsord. De semantiske oplysninger som ordbasen indeholder, er ikke tilgængelige vha. grænsefladen.

Leksikografer, lingvister og andre interesserede kan søge i basen over internettet fra Center for Sprogteknologis hjemmeside (<http://cst.dk>, vælg STO ordbase fra forsiden); her findes brugervejledning og et par tips med interessante søgeeksempler og en beskrivelse af ordbasen.

Grænsefladen har fire søgemuligheder:

- **Ordsøgning** (med lemma eller en bøjet ordform, f.eks. *bog/bøger*).  
Søgeresultat: Lemmaets/lemmaernes bøjningsformer, stavning mv. samt syntaktiske konstruktioner inkl. styrede præpositioner (med eksempler).
- **Substantiviske sammensætninger** (hvis man ønsker at vide hvilke sammensætninger et substantiv indgår i, f.eks. *barn*).  
Søgeresultat: En liste med alle substantiver som søgeordet indgår i (f.eks. *barnepige, børnehave, blebarn*) inkl. eventuelle kandidater i kursiv.
- **Korpussøgning** (med op til tre lemmaer og/eller bøjede ordformer i kombination).  
Søgeresultat: konkordanslinjer fra et aviskorpus (Berlingske Tidende 1990–92).
- **Parameterbaseret søgning** med kombination af forskellige lingvistiske kriterier, f.eks. ordklasse, valens, styrede præpositioner.  
Søgeresultat: En liste af tilfældigt udvalgte ord (op til 30) der opfylder den valgte kombination af kriterier.

For at brugeren skal kunne sammenligne STOs oplysninger med andre internetordbøger og se ordets forekomster i tekster, er der desuden etableret links fra grænsefladen til Retskrivningsordbogen, Den Danske Netordbog, Korpus2000 og Google (søgning på danske sider).

Takket være brugergrænsefladen, der viser de sprogteknologiske koder i læsbar form, er STO i mange henseender en virkelig anvendelig elektronisk ordbog for mennesker, eksempelvis i sprogindlæring/undervisning på højt niveau.

## 6. Opsummering og visioner

Dette indlæg var tænkt som en almen præsentation og har givet nogle eksempler på hvilke produkter der populært betegnes som 'elektronisk ordbog', og beskrevet disse produkter ganske kort. Formålet med dette var at udpege nogle fundamentale forskelle og enkelte ligheder mellem de forskellige produkttyper. Herefter fokuseredes på de elektroniske ordbøger der opfylder grundlæggende leksikografiske og anvendelsesmæssige kriterier. Der er valgt nogle overordnede træk og kun som belysende eksempel er der fremhævet enkelte detaljer selv om det ville have været spændende at fordybe sig i enkelthederne vedrørende de forskellige ordbøgers leksikografiske egenskaber og brugervenlighed.

Konklusionen er, ikke uventet, at kombinationen af leksikografisk kvalitet og teknisk brugervenlighed bør være det overordnede mål i den fremtidige udvikling på området. Visionerne for det danske sprog omfatter først og fremmest ønsket om at få flere elektroniske ordbøger, både en rigtig stor almensproglig nutidsordbog og flere forskellige fagsproglige to- eller flersprogede ordbøger. Visionerne omfatter udstrakt brug af nye præsentationsmåder i form af lyd, billeder og billedsekvenser i forbindelse med sådanne ordbogsartikler der giver naturlig basis for dette. Udviklingen af nye elektroniske ordbøger vil automatisk medføre udarbejdelse af passende ordbogskoncepter, oplysningsstrukturer (f.eks. implementeringen af semantiske net for dansk).

På den mere tekniske side er der betydelige perspektiver i sammenkobling af forskellige ordbøger både online og offline, eksempelvis til at producere materiale for nye sprogkombinationer; dertil skal der udvikles fleksible platforme. Dette kræver at leksikografien støtter den internationale og nationale standardisering der vedrører strukturering af leksikalske data, harmonisering af metadata og beskrivelsessprog.

Andre eksempler på udvidelsen af den enkelte ordbogs anvendelsesområde er inddragelse af forskellige sprogressourcer ved en sammenkobling. Dette kræver en forudgående analyse af formater og indhold for at kunne gøre ressourcer som ordbog, termsamling, grammatik, tekstkorpus mv. kompatible.

Udviklingen på disse områder vil forhåbentlig blive båret frem af et frugtbart samarbejde mellem flere discipliner: leksikografi, sprogteknologi, videnshåndtering, sprogdidaktik, datalogi m.fl. – Det kan godt lade sig gøre!

## Referenceliste

### *Ordbøger*

*Den Danske Idiomordbog*, Center for Leksikografi, Handelshøjskolen i Århus; <http://netdob.asb.dk/idiom/index.html>

*Den Danske Netordbog*, Center for Leksikografi, Handelshøjskolen i Århus; <http://www.dendanskenetordbog.dk/netdob/index.html>

Gyldendals elektroniske ordbøger, 1989, Nordisk Forlag A/S, Copenhagen; *Dansk-engelsk*. Programmel: TEXTware A/S, Copenhagen

*Ordbogen.com*, Cool Systems ApS. <http://www.ordbogen.com>

*Regnskabsordbogen*, Center for Leksikografi, Handelshøjskolen i Århus; <http://netdob.asb.dk/iasdk/index.html>

*Retskrivningsordbogen*, Dansk Sprognævn; <http://www.dsn.dk>

*Sprogteknologisk Ordbase STO*, Center for Sprogteknologi, Københavns Universitet; <http://cst.dk>

*L&H Ordbøger*; <http://www.lhpublishing.dk/>

*L&H WebWords*; <http://www.lhpublishing.dk/webwords/index.html>

#### *Anden litteratur*

- Armstrong Susan, Colin Brace, Dominique Petitpierre, Gilbert Robert & Derek Walker 2000: DicoPro: An Online Dictionary Consultation Tool for Language Professionals. I: *Proceedings of the 9th EURALEX International Congress*. Eds.: U. Heid, S. Evert, E. Lehmann, Chr. Rohrer. Stuttgart. IMS, 61–66.
- Bertagna, Francesca, Alessandro Lenci, Monica Monachini & Nicoletta Calzolari 2004: Content Interoperability of Lexical Resources: Open Issues and “MILE” Perspectives. I: *Proceedings of the 4th International Conference on Language Resources and Evaluation*, ELRA, Lisboa, 131–134.
- de Schryver, Gilles-Maurice & David Joffe 2004: On How Electronic Dictionaries are Really Used. I: *Proceedings of the 11th EURALEX International Congress*. Eds.: G. Williams and Sandra Vessier. Université de Bretagne-Sud, Lorient, 187–194.
- Docherty, Vincent J. 2000: Dictionaries on the Internet. I: *Proceedings of the 9th EURALEX International Congress*. Eds.: U. Heid, S. Evert, E. Lehmann, Chr. Rohrer. Stuttgart. IMS, 67–74.
- Duvå, Grete og Anne Lise Laursen 2005: Fordeling af informationer og selektion af henvisninger i en fagsproglig internetordbog. I:

- Nordiske Studiar i leksikografi 7. Rapport frå Konferanse om leksikografi i Norden Volda 20.–24. mai 2003.* Eds.: R.V. Fjeld og D. Worren. Oslo, 93–105.
- Geeraerts, Dirk 2000: Adding Electronic Value. The electronic version of the Grote Van Dale. I: *Proceedings of the 9th EURALEX International Congress*. Eds.: U. Heid, S. Evert, E. Lehmann, Chr. Rohrer. Stuttgart. IMS, 75–84.
- Nielsen, Sandro 2002: Overvejelser angående det leksikografiske grundlag for en elektronisk bilingval regnskabsordbog. I: *Lexico-Nordica 9*, 173–192.
- Prószéky, Gábor og Balázs Kis 2000: Development of a Context-Sensitive Electronic Dictionary. I: *Proceedings of the 11th EURALEX International Congress*. Eds.: A. Braasch & C. Povlsen. Copenhagen. Center for Sprogteknologi, 281–290.
- Lexicography in the Electronic Age 1982: Proceedings*. Eds.: J. Goetschalckx and L. Rolling. North-Holland Publishing Company, Amsterdam, New York, Oxford.

Anna Braasch  
seniorforsker  
Center for Sprogteknologi  
Københavns Universitet  
Njalsgade 80  
DK-2300 København S  
[anna@cst.dk](mailto:anna@cst.dk)