



10. JUL. 1997

LEDA

LEKSIKOGRAFER  
I DANMARK

# Nyhedsbrev nr. 23 - juni 1997

ISSN 0908-3553

Redaktør af dette nr. *K.T.Thomsen*

## Indholdsfortegnelse:

Announce for Nordisk Forening for Leksikografi.....	1
Orientering om LEDA's generalforsamling 9/4-97.....	2
Svar på en anmeldelse ( <i>Jens Axelsen</i> ).....	3
Anmeldelse af Angelordbogen ( <i>Torben A. Andersen</i> ).....	10
Sprogteknologisk ordbog ( <i>Anna Braasch</i> ).....	16

## ANNOUNCE!

Efterdi der ikke er mange af LEDA's medlemmer, som tillige er medlem af Nordisk Forening for Leksikografi (NFL), skal der her bringes en liden reklame for foretagendet.

NFL har siden 1991 hvert andet år i samarbejde med Nordisk Språksekretariat/Språkråd afholdt en større konference (med 100-250 deltagere) om leksikografi i Norden. Bidragene udkommer i form af proceedings: Nordiske studier i leksikografi I, II, ..., n.

Endvidere udgiver foreningen tidsskriftet *LexicoNordica* (siden 1994), som udkommer én gang årligt i november og indeholder temabidrag fra en årlig, mindre konference (15-20 deltagere) med forud fastlagt tema: (1994: Almene énsprogede ordbøger; 1995: Grammatik i bilingvale ordbøger; 1996: Korpusbaseret leksikografi; 1997: Ordbøger for små sprogsamfund i Norden; 1998: Fagleksikografi; 1999: Indvandrerordbøger; 2000: Nordiske leksikografiske traditioner; 2001: Ordbøger mellem de nordiske sprog; 2002: Sprognormering i nordiske ordbøger). Herudover udgives et mindre medlemsblad, NFL-nytt.

NFL har et nært samarbejde med Nordisk Språkråd (de stiller bl.a. sekretariat til rådighed), er således et enestående debat- og kontaktforum, som både forlag, læreanstaltsansatte og andre med professionel interesse i og for leksikografi bør være medlem af. Erfaringen viser at det her i Norden primært er gennem NFL, at de nyeste leksikografiske landvindinger først offentliggøres og kringkastes til en bredere kreds. Så det er her det sker!

Foreningen råder i øvrigt over en liste over nulevende udøvere i leksikografi Norden.

### Adressen er:

Nordisk forening for leksikografi, Nordisk Språkråd, Postboks 8107 Dep, N-0032 Oslo

Telefon: +47-22 42 05 70, Fax: +47 22 42 76 76

Medlemskontingent er pr. 1997: 250 NOK årligt.

**DANSK SPROGNÆVN**

## Generalforsamling i LEDA den 9. april 1997

I generalforsamlingen deltog foruden bestyrelsen 5 (fem) ordinære medlemmer.

Ole Norling-Christensen blev valgt som dirigent hvorefter formanden aflagde beretning og kassereren fremlagde det reviderede regnskab som blev godkendt.

To bestyrelsesmedlemmer var på valg og blev genvalgt ligesom suppleanter, revisor og revisorsuppleant blev genvalgt.

Bestyrelsen havde ikke modtaget nogen forslag.

Under eventuelt nævnte formanden at hun havde fået en henvendelse udefra angående et medlemsmøde om Thesaurus linguae latinae. Der var enighed om at det i det hele taget kunne være interessant at afholde et halvdagsmøde om tesaurusser, så bestyrelsen vil gå videre med sagen.

Efter generalforsamlingen har bestyrelsen følgende medlemmer:

Pia Jarvad (formand), Torben Arboe Andersen (kasserer), Jens Axelsen, Anna Braasch, Anna Garde (sekretær).

Suppleanter: K.T. Thomsen og Henrik Holmboe

Revisor Boel Bøggild-Andersen, revisorsuppleant Ebba Hjorth.

Jens Axelsen

### Bemærkninger til anmeldelsen af Gyldendals Røde Dansk-Engelsk Ordbog i sidste nummer af Nyhedsbrevet.

Det er et selvfølgelig krav til en kvalificeret ordbogsanmeldelse at de oplysninger den indeholder er korrekte, og de påstande der fremsættes er underbyggede, og at den bygger på en vis indsigt i tosprogede ordbøgers funktion (jf. også Axelsen 1992, citeret i Bergenholtz/Mogensen 1993, 26-27).

Jeg skal i det følgende redegøre for hvorfor jeg ikke mener at ovennævnte anmeldelse lever op til disse krav.

#### Ækvivalenter.

De "undersættelser" anmelderen nævner s.18 er ikke undersættelser, men rigtige nok: *have you come by car* (for: "er du i bil"), *she runs a car* (for: "hun har bil"); det sidstnævnte står desuden i *Advanced Learner's Dictionary (ALD, run* tyd 16b), og i COBUILD (*run* tyd 16: *if you run a car ..., you have it and use it*); - "den korpusbaserede COBUILD 1995" (s. 21) citeres ellers flere gange af anmelderen som belæg, men er ikke anvendt her. S. 20 nævnes *great* som en dårlig oversættelse af "fed". Den er nu ikke så dårlig, jf. følgende citat fra Guardian Weekly: *No wonder .. that a younger generation .. has learned from an early age to distrust the spoken word .. confining their own utterances to such meagre limits as "great, man" and "there you go" and "that's your bag"...* (se også COBUILD *great* tyd 7).

Som eksempel på "direkte forkerte oversættelser" (s. 20) anføres oversættelsen af "bofællesskab" (*flatshare, (i hus) houseshare*), og det hedder: "... vi får helt gal besked; i den moderne danske virkelighed er her tale om en kollektiv bebyggelse, hvor hver familie har sin egen bolig og hvor man deler en række fælles faciliteter...". Følgende citater fra den moderne danske virkelighed passer ikke til denne

beskrivelse: Politiken 5/1 97, sektion 2, s. 3: "De lever i et socialistisk bofællesskab i Baldersgade på Nørrebro sammen med 14 kammerater", Politikens tillæg MAX. 9/2 97, s. 18, en annonce: "I bofællesskab i Vedbæk er andel ledig. Vi bor i et stort hus m. stor have og tæt ved skov og strand...Indskud kr. 39.000, husleje kr. 3.500..." Det gør ækvivalenterne snarere.

Videre afvises *m/f* (for: "m/k"). Det har jeg noteret fra stillingsannoncer i en engelsk avis, fx *chef m/f*, jf. *The Oxford Dictionary of Abbreviations* 1992: *male or female (in advertisements)*.

S. 21 står der: "For eksempel virker det sært at verbalidiomet 'lade i stikken' ikke gengives med det hyppige 'abandon', men med to andre, langt sjældnere udtryk". Hertil er at bemærke at disse to udtryk: *be left high and dry*, *be left in the lurch* ikke er sjældne; de findes i ALD og COBUILD (som heller ikke her nævnes); *abandon* er da også muligt, men ikke bedre, da det er et andet stillag.

To andre afviste ækvivalenter er ligeledes rigtige, men nok mindre almindelige end den første mulighed ordbogen giver (men de er i begge tilfælde kun alternative forslag): S. 18: *hand in one's checks* (for: "tage billetten") står i *The Oxford Dictionary of Modern Slang* 1992, s. 21, *notecase* (for: "tegnebog") har jeg noteret fra et moderne engelsk varekatalog.

Ud over kritikken af ækvivalenterne kan nævnes et par andre misvisende påstande. Anmelderen skriver s. 19: "'bildæk' (af gummi) gengives kun med 'car tyre', ikke med 'car tire'". Nej, men i artiklen står: "(*am*) automobile tire", så den amerikanske stavemåde er med. Endvidere s. 20: "Vi får heller ikke 'tus' eller - som det hed i RO 86 - 'tusse' (altså børnenes tuschfarver)..." Jo, det står som tyd 2 under "tusse", som i øvrigt nu anses for forkert, mens den rigtige form er 'tus'; det blev først offentliggjort af Sprognævnet i "Nyt fra Sprognævnet" 2, 1996, altså 1/2 år efter at ordbogen udkom.

## Lemnaer.

Anmelderen kritiserer s. 18 og 19 forekomsten af "anakronismer" eller gamle ord, som han mener demonstrerer at opdateringen af ordbogen er "langt fra tilfredsstillende udført".

Anmelderen sammenblander imidlertid i sin afvisning af disse ord og udtryk forskellige kategorier:

a) udtryk der blot er formel stil. Nogle af disse udtryk forekommer da også i Den Danske Ordbogs korpus, der omfatter perioden 1980-90, og et par også i Nudansk Ordbog 1996: "løse billet" (25 forekomster i korpus), "give sit bifald" (2) (de to sidstnævnte er med i Nudansk Ordbog 1996), "få modbydelighed for", "bifaldsmumlen" (1), "bifaldsstorm" (2). Anmelderen savner "bifaldsytring" (2) i stedet for de to sidstnævnte. Hvorfor han hellere vil have det, er ikke klart. Det er i korpus ikke mere hyppigt end de andre.

b) ældre udtryk der stadig bruges, især humoristisk: "bile" (13 forekomster).

c) fagord: "biland" (som anmelderen spøgefuldt henregner til zoologien) findes også i korpus med 5 forekomster, "studenternellike" er en almindelig haveblomst og kan ses i havebøger, "spandgardin" er med i RO96 og kendt af husmødre, som også "vandbakkelse" der har 36 forekomster i korpus, mens "bondeparti" har 16. Det sidste får følgende ironiske bemærkning med fra anmelderen: "hvis vi nu skal oversætte fra østeuropæiske sprog via dansk". Det forekommer ingeniunde kun i Østeuropa, jf. følgende citat fra korpus: "det reaktionære bondeparti Venstre".

d) Mens det ikke med rimelighed kan kritiseres at de ovennævnte kategorier er med i ordbogen, er der da også nogle helt forældede ord og udtryk med som burde være strøget, men er blevet overset. De andrager vel en halv snes (og det er da muligt at der gemmer sig flere): "bilbio", "bæreplansbåd", "Florens" (der dog står i Gyldendals Fakta, 1988: "dansk navn for Firenze"), "luftring" (der dog forekommer i fagordbøger, fx Clausens dansk-engelsk teknisk ordbog 1990, men ikke i korpus), "nytårsny", "nytårsregning", "tekstbehandlingsmaskine". Mht. "billedbånd", "billedunderskrift", "bæreplansbåd" (ingen af dem i korpus) og

"grundsprog" (2 forekomster) er der at bemærke (hvad anmelderen ikke gør) at de moderne udtryk "videobånd", "billedtekst", "hydrofoil", "originalsprog" er med i ordbogen. Om "billedunderskrift" skriver anmelderen s. 18: "det viser sig at være ... en 'caption', altså en billedtekst eller en undertekst". "Undertekst" er ikke det samme som "billedtekst", som anmelderen synes at mene, og hedder, som der står i ordbogen, *subtitle*.

Man kan også spørge om forekomsten af disse helt forældede ord er skadelig ud over at de optager plads. Anmelderen siger s. 18 at de "dels kan gøre en dansk bruger nervøs for præcisionen på den engelske side, dels være direkte vildledende for de ikke-danske brugere ordbogen har i tankerne". Dette argument er uden mening eftersom hverken danske eller udenlandske brugere af ordbogen vil støde på de forældede ord hvis ikke de slår dem op, og det gør de kun hvis de skulle træffe dem i en tekst.

Til gengæld finder anmelderen at der mangler en del moderne ord. Nu er diskussion om lemmaselektion, trods sin popularitet blandt anmeldere (se Bergenholtz/Mogensen 1993), en lidet relevant øvelse, da udvalget af lemmaer altid vil være subjektivt (jf. Axelsen 1995). Naturligvis er der medtaget et betragteligt antal nye lemmaer - anmelderen nævner da også et par stykker i slutningen af artiklen. Af de ca. 30 han her nævner som savnede, er nogle oplagte kandidater til optagelse, andre diskutabile og atter andre efter mit skøn absolut ikke optagelsesværdige. Det kan man drøfte i det uendelige. At ord vedrørende Internettet ikke er med skyldes at det endnu ikke var så omtalt den gang ordbogen blev redigeret. At "bidt" ikke er sat som lemma skyldes at verbalformer normalt står under verbets grundform. Kun participier der optræder som adjektiver står som lemmaer: "fortabt", "flydende".

Man kan så spørge om anmelderen ud fra de ovennævnte ca. 50, delvis fejlagtige lemma- og ækvivalenteksempler har godtgjort at opdateringen af ordbogen, der er på over 1.000 sider og har over 100.000 oplysninger (65.000 opslagsord + adskillige delbetydninger, og 30.000 udtryk), er utilfredsstillende. Det mener jeg ikke, og bemærkningen s. 20: "Til gengæld for de nye som faldt er der gamle

overalt" (min fremhævelse) holder ikke stik, eftersom den halve snes virkelig gamle udtryk anført under punkt d ovenfor giver 1 pr. 100 sider eller 1 pr. 10.000 oplysninger.

### **Principielle forhold.**

Mere relevant end bemærkninger om mere eller mindre tilfældige enkeltheder er diskussion af principielle ting. Opstilling, tyddifferentiering, eksemplificering er også væsentlige i en ordbog, men det siges der ikke meget om.

Opstillingen i artiklernes "hoved" er ændret på den måde at der er lavet indrykning ved hver ny tyd, især i de lange artikler, som derved bliver mere overskuelige fordi brugeren, i stedet for at være nødt til at læse en blok på en hel eller halv spalte igennem for at finde den relevante tyd, kan nøjes med at lade blikket glide ned langs venstre forkant. Det er naturligvis mindre vigtigt i artikler med kun to eller tre tyd, som ikke er svære at få overblik over. Det eneste anmelderen har at sige herom angår igen enkeltheder: Han finder to artikler hvor der ikke er indrykning. De har kun to tyd og det derfor er ret ligegyldigt, anføres ikke.

Anmelderen strejfer også andre principielle spørgsmål, men uden at se at der kunne ligge mange problemer deri. Et af dem er behandlingen af kulturspecifikke ord. Der er forskellige, lige utilfredsstillende løsninger: man kan udelade ("kravlenisse", "sangskjuler" er ikke med), man kan skrive: "[findes ikke i Engl. el. USA]", for dog at vise at ordet ikke er oversat, man kan give en forklaring, som i de af anmelderen citerede tilfælde. Man kan også, som han foreslår, give en omtrentlig ækvivalent, fx *anniversary* for "rund fødselsdag". Det giver dog kun den ene side af indholdet, som hvis man oversatte "kravlenisse" ved *pixie*. Endelig kan man konstruere en ækvivalent og skrive "(kan gengives)", som den af anmelderen foreslåede gengivelse af "klippekort": *multitrip ticket*, der er en god idé. Disse løsninger er brugt forskellige steder i ordbogen, men kan ikke bruges i alle tilfælde. En redaktør foreslog en gang i desperation over problemets uløselighed at indføre en ny forkortelse *fpns* = find på noget selv.

Et andet principielt problem er placeringen af stilbetegnelser. Spørgsmålet har været diskuteret i litteraturen (fx Grindsted 1988), men de foreslåede løsninger har jeg ikke fundet anvendelige. Jeg har eksperimenteret med anbringelsen i tydspecifikationen: forrest, når det refererer til det danske: "(T: ur)", og sidst når det refererer til det engelske: "(gå, T)" men er endnu ikke sikker på at det er en god løsning. Anmelderen skriver s. 21: "Forkortelsen "T" i eksemplet her står i øvrigt for 'uformelt sprog', og det forklares intetsteds hvorfor man ikke fx bruger et 'U' til denne funktion". Det fremgår af forkortelseslisten at T står for "uformelt sprog; uformelt skriftsprog og daglig tale" (min understregning), derfor T. Det kan dårligt kan erstattes af U, fordi U leder tanken hen på "utælleligt", hvad det også betegner i flere *Learner Dictionaries*, og det skal det måske senere i den Røde.

Spørgsmålet om amerikanismer er også principielt. Det er vel rimeligt at tage stavemåder og afvigende ækvivalenter med - måske kun i de tilfælde, hvor det britisk-engelske udtryk ikke forstås i USA. Vælger man at tage stavemåder med, skal de selvfølgelig med overalt, og under "radioprogram" burde *radio program* også være med. Man kan diskutere om man derudover også skal have syntaktiske afvigelser med, som den af anmelderen anførte *I don't have a clue*, og som *out the window*, *in the door*, *protest* brugt transitivt, foruden de nævnte verbalformer *dove* og *snuck*. (*I haven't a clue*, som anmelderen kritiserer som mindre almindeligt på britisk-engelsk, er den form både COBUILD og ALD har).

Temmelig overraskende og komplet irrelevant får man også oplysning om de ting som bogen ikke har lovet at tage med og derfor heller ikke har, fx etymologi. (Det nævnes ikke om det skulle være til lemmaer eller ækvivalenter der kunne tænkes at være oplysning om etymologi.) Man kunne lige så godt have oplyst at bogen heller ikke lover at bringe billeder af kongehuset, hvad den så heller ikke gør. Helt grotesk er følgende bemærkning: "At forvente engelske udtaleangivelser i en tosproget ordbog af denne type er måske for meget forlangt" (min fremhævelse). Det er ikke "måske" for meget forlangt, det er ganske urealistisk. Skulle alle ækvivalenter have udtalebetegnelse? Også eksemplerne? Også idiomerne i foden af artiklerne? Hvor mange bind ville en sådan ordbog mon skulle omfatte? Jeg me-



ner heller ikke der skal være udtaleoplysning til lemmaerne. Anmelderens begrundelse for at ønske det: "Så kunne selv en usikker bruger skelne mellem nabo-opslagsordene 'teist' og 'tejst'" viser igen en bagvendt opfattelse af hvad en dansk-engelsk ordbog bruges til. Hvis brugeren støder på ordet "teist" slår han det op, hvis "tejst" slår han det op, men han søger ikke oplysning om forskellen på to danske ord i en dansk-engelsk ordbog.

### **Balance.**

Anmeldelsen kan ikke siges at repræsentere en balanceret vurdering af ordbogens gode og dårlige sider. Kritikken optager ca. 4 1/2 side og et afsnit kaldet "Dansk-engelsk ordbogs stærke sider" godt 1/2. I dette afsnit vil anmelderen "give nogle få eksempler på de kvaliteter denne bog besidder" (min understregning; prioriteringen er klar nok: hvad anmelderen anser for fejl er vigtigere). Han nævner som positive træk 8 ækvivalenter, tyddifferentieringen under "overhøre" og "godt", og 2 pseudoanglicismer (men der bliver dog også plads til lidt kritik af en formulering under "forfordele").

### **Empiri.**

Endelig skriver anmelderen s. 22: "Men den største ros en sådan ordbog kunne få, ville være at der var korpusbelæg for både valg og betydningsdifferentiering af opslagsordene, og - ikke mindst - at der også var empirisk dækning for de fremmedsproglige ækvivalenter." Det er uklart hvor anmelderen forestiller sig at de nævnte oplysninger her i ordbogen kommer fra hvis ikke fra empiriske belæg, både for opslagsord, betydningsdifferentiering og ækvivalenter. De stammer ganske vist fra eksempler der enkeltvis er fundet i tekster (jf. ordbogens forord og Axelsen 1995; dog især for de engelskes vedkommende også fra ordbøger) og ikke fra et elektronisk korpus. Den Danske Ordbogs korpus har jeg kun brugt til at verificere ting jeg har fundet, hvis almindelighed jeg var i tvivl om, og ikke til at

kontrollere hele den danske side, hvad anmelderen s. 18 synes at lægge ind i bemærkningen i forordet. Det ville ganske rigtigt, som han skriver, være "umuligt at gennemføre i praksis" (selvom han s. 22 - noget ulogisk - efterlyser korpusbelæg for valg af opslagsordene). Det er jo nemlig ikke uden problemer at bruge korpus, ikke mindst tidsmæssige, hvad anmelderen ikke synes at overveje s. 22.

### **Bibliografi:**

**ALD = Oxford Advanced Learner's Dictionary**, 1995, 5<sup>th</sup> edition, Oxford University Press.

**Axelsen, Jens** 1992: En anmeldelse af en anmeldelse. *SPRINT*, sproginstitutternes tidsskrift (fra Handelshøjskolen i København), 1992, 2, s. 1-5.

**Axelsen, Jens** 1995: Lemmasektion i en dansk-engelsk ordbog. I Asta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson (red.) *Nordiske Studier i Leksikografi* 3. Reykjavík: Nordisk Forening for Leksikografi, 19-26.

**Bergenholtz, Henning/Jens Erik Mogensen** 1993: Wörterbuchkritik in Dänemark. I *Lexicographica*, International Annual for Lexicography 9. Tübingen: Max Niemeyer Verlag, 8-35.

**Clausens tekniske ordbøger dansk-engelsk** 1990, udarbejdet i samarbejde med Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København, København: Grafisk.

**COBUILD = Collins Cobuild English Dictionary**, 1995, new edition. Editor-in-Chief John Sinclair, HarperCollins.

**Fakta Gyldendals etbinds leksikon** 1988, København: Gyldendal.

**Grindsted, Annette** 1988: Principper for præsentation af ækvivalenter i oversættelsesordbøger, *Odense University Studies in Linguistics* Vol. 8, Odense Universitetsforlag, s. 65 ff.

**Retskrivningsordbogen** 2. udgave 1996, Dansk Sprognævn, København: Aschehoug.

**The Oxford Dictionary of Modern Slang** 1992, ed. John Ayto & John Simpson, Oxford, New York

Torben Arboe Andersen

Bent Jul Nielsen og Magda Nyberg:

**Ordbog over den danske dialekt i Angel**

med en indledning om dialektens udtrykssystem.

C.A. Reitzels forlag. 1995. – 602 sider. Kr. 531,25. – ISBN 87-7421-947-2.

Institut for Dansk Dialektforsknings publikationer, serie A, nr. 33.

Universitets-Jubilæets danske Samfunds skrifter nr. 532.

Indhold: Indledning s. 5-10, Lydsystem s. 11-53, Indledning til ordbogen s. 54-62, Ordbog s. 63-588, Tekstbilag s. 589-602.

Tidligere anmeldelse: Marie Bjerrum: Angel-ordbogen – en imponerende præstation. – i: *Ord & Sag*. 15 (1995) s. 46-51.

**1. Oversigt.** Angel-ordbogen dækker det sønderjyske sprog, som det tidligere er blevet talt i den nordlige del af Angel i en radius af 20-25 km øst og syd for Flensborg. I dette område holdt det sønderjyske sprog stand mod nedertysk længere end i det øvrige Angel: det var levende indtil midt i 1800-tallet, blev derefter trængt tilbage af nedertysk og uddøde stort set i løbet af 1930-erne (jf. s. 5). Der foreligger omfattende optegnelser af dialekten; de danner grundlaget for både ordbogen over Angelmålet og opstillingen af dets lydsystem.

I indledningen redegøres nærmere for disse kilder; de anføres endvidere i oversigtsform s. 59f. Fra 1800-tallet er det et hidtil uudgivet manuskript af J.J. Kappel (1837ff.), men vigtigst E.H. Hagerup: *Om det danske Sprog i Angel*. 1867 (204 s.). Hagerup ville vise, at der i Angel fandtes et "levende dansk Folkesprog"; Jul Nielsen og Nyberg betegner bogen som "en sprogbeskrivelse af høj kvalitet, især på det leksikalske – og herunder det morfologiske – område. Det er virkelig *hørte* ord og sætninger, der præsenteres" (s. 7). Hagerups udtaleangivelse er, skønt der på det tidspunkt ikke forelå nogen entydig lydskrift, "meget informativ og forbavsende konsekvent" (smst.). Fra 1900-tallet er ordbogens kilder først og fremmest optegnelser i 1930-erne ved filologerne Anders Bjerrum og Kristian Hald m.fl. med anvendelse af Danias lydskrift (s. 5 ff.). Deres meddelere var født 1847-72 og talte mest nedertysk, men huskede fra deres barndom tydeligt den sønderjyske dialekt, som blev talt af deres forældres generation (s. 5). Ekscerperingen af de nævnte kilder omfatter: Hagerup ca. 6.700 ordsedler, Kappel ca. 1.800, Bjerrum m.fl. ca. 9.700, øvrige ca. 800 (jf. s. 59f.), dvs. i alt ca. 19.000 ordsedler.

I afsnittet Lydsystem opstilles dialektens foneminventar ud fra en omhyggelig analyse af de nævnte filologoptegnelser – med henblik på, at ordbogens udtaleformer så vidt muligt

skal anføres i fonematisk form, betegnet "systematisk notation" (s. 12ff., oversigt s. 50f.). Dette afsnit er især af interesse for dialektologer (og andre filologer) og skal ikke omtales nærmere her; dog bør et par af dialektens lydlige karakteristika lige nævnes: fællesdansk stød mangler, og dialekten har tonal accent ("musikalsk accent"), jf. afsnittet "Accentuation" s. 18-27.

I Indledning til ordbogen gives (s. 54-59) en kort og i hovedtræk instruktiv beskrivelse af de benyttede redaktionsregler efterfulgt af en forkortelsesoversigt. I Tekstbilag gengives fra 1930-erne dels korte lydskrevne optegnelser, dels udskrift af gram mofonoptagelser af dialekten (s. 589-601), desuden 2 ældre optegnelser, bl.a. fra Kappels manuskript (s. 601f.), alt med oversættelse til rigsmål.

## 2. Ordbog A-Å s. 63-588

Efter års arbejde med materialet finder forfatterne det berettiget at behandle det som ét korpus, skønt "der er hundede år mellem de ældste og de yngste kilder" (s. 9). De ytrer en smule tvivl om, hvorvidt materialet er tilstrækkeligt for en egentlig ordbog, "dvs. en nogenlunde fyldig og alsidig afspejling af Angelboernes danske talesprog", idet det "frembyder ... mange tilfældige "huller"" (smst.). Efter gennemsyn af bogen er jeg dog ikke i tvivl om, at man har fået langt den væsentligste del af ordforrådet med.

### Artikelstruktur (jf. s. 54-59)

En fuldt udbygget artikel består af: opslagsform og ordklasseangivelse – lydformer – betydningsangivelse(r), herunder faste forbindelser og talemåder, med citater fra kilderne – evt. saglig-terminologisk afsnit – evt. henvisning – evt. etymologisk oplysning. Et eksempel, artiklen *kilde* s., ses i afsnit 3 ndf.

Som antydnet er ikke alle kategorier repræsenteret i hver enkelt artikel. I "oplagte tilfælde", dvs. hvor betydningen er helt den samme som i rigsmålet, kan endog betydningsangivelsen udelades (jf. s. 57). Overgang til betydningsdelen hhv. den etymologiske del markeres med linieskift og tankestreg. Hvor betydningsdelen opdeles yderligere, får hver hovedbetydning ligeledes linieskift (og tankestreg), hvilket bestemt medvirker til at gøre artiklerne overskuelige. Til adskillelse af afsnit vedr. mindre betydningsforskelle bruges planke. Henvisninger til saglig-terminologiske artikler (jf. afsnit 3) er placeret i parentes først i betydningsdelen.

**Opslagsform.** Fastlæggelse af opslagsformer for ord, som ikke findes i rigsmålet, udgør altid en særlig opgave for redaktører af dialektordbøger. Opgaven er i Angelordbogen løst på følgende måde: ord, der er fælles med rigsmålet, får opslagsform efter Retskrivningsordbogen og Ordbog over det Danske Sprog (ODS), mens man for de øvrige ord følger dels H.F. Feilberg: *Ordbog over jyske folkemål* (I-IV, 1896-1914), dels Otto Mensing: *Schleswig-Holsteinisches Wörterbuch* (I-V, 1927-35). Det sidste skyldes, at dialekten som

venteligt har et anseeligt islæt af nedertyske ord, fx. numeralier som *veertig*, *tagentig* (= 40, 80). Opslagsformer hos Feilberg og Mensing tillempes dog undertiden, fx *hyffel* (= kristtjørn) i stedet for *høffel* hos Feilberg, *fryntlig* (= venlig) i stedet for nty. *fründlich* hos Mensing. For de ord, der findes i både ODS og Feilberg, følges ODS; det gælder fx subst. *kåd* = mindre ejendom, hus med jordtilliggende, der hos Feilberg placeres som *kod*. Feilbergs ortografi følges ved fx *volfejl*, der (som i andre syd-sønderjyske dialekter) er det almindelige ord for adj. billig, jf. tysk *wohlfeil*. Ud over de nævnte ordbøger har forfatterne (jf. s. 54) især konsulteret M. og A. Bjerrum: *Ordbog over Fjoldemålet* (I-II, 1974), der er en ordbog over den danske dialekt i området lidt sydvest for Angelordbogens område. Også ordbøger over flere nedertyske dialekter er anvendt.

Homografer markeres med foransat romertal. Sammensatte ord anføres med bindestreg mellem første- og sidsteled, fx *skudspole-pind* (= del af vævespole).

I enkelte tilfælde er der oprettet artikler på ord, der ikke er optegnet i dialekten; her er opslagsform og ordklasseangivelse placeret i kantet parentes, jf. fx: [ringridning s], der følges af et rigsmålsцитат fra Hagerup. På tilsvarende måde anføres også enkelte henvisninger fra rigsmålsord til dialektens ord (jf. om henvisninger ndf.).

**Lydformer.** Hvor det har været muligt at fremanalysere en fonematiske form, anføres den (mellem skråstreger) som den første, efterfulgt af lydformer citeret fra kilderne, jf. eksemplet i afsnit 3 (det hævdede to-tal forrest i den fonematiske form markerer tonal accent 2 ligesom "vinkeltegnet" forrest i to af de fonetiske former). De fonetiske former, hovedsagelig fra 1930-optegnelserne, er optegnet i Dania-lydskrift. Også bøjningsformers udtale angives.

**Definitioner og betydningsopdelinger.** Definitionerne er, som det tilstræbes (jf. p. 57), klare og præcise. Betydningsopdelingerne er meget omhyggeligt gennemført, også i artikler vedr. fællesdanske ord, fx præpositioner som *II i* med betydningerne 1-12 og verber som *tage* med betydningerne 1-10.6.

**Citater.** Anføres i kursiv, med oversættelse af enkeltord eller ordgrupper i antikva og parentes. Der oversættes kun så meget, som skønnes nødvendigt for forståelsen. I kildeartiklen i afsnit 3 ndf. oversættes fx ikke citat nr. 2, fordi sidste bemærkning i sag-afsnittet gengiver citatet næsten ordret. Det vigtigste for forfatterne har været at bringe citaterne i en forståelig form, og den valgte praksis fungerer udmærket, blot man er indstillet på at bruge lidt tid på selv at oversætte dele af et givet citat. I hvert fald er det brugervenligt, at de fleste dialektord straks oversættes eller forklares, så læseren ikke skal hen i en anden artikel i ordbogen at få de oplysninger. En ulempe er det måske, at de parentetiske oversættelser af og til bryder citatet vel meget op. Det mest brugervenlige ville være at oversætte hvert citat fuldt ud, men det er naturligvis mere plads- og tidkrævende. De mange citater er med til at give et fyldigt og varieret indblik i egnens talesprog.

**Flerleddede udtryk** (faste forbindelser) anføres spatieret, ofte efter planke.

**Talemåder** (inkl. ordsprog) citeres direkte, jf. fx i artiklen i afsnit 3 ndf.: *det er for silde at smide kilden efter, når barnet er druknet* – dialektens variant af et også i rigsmålet kendt ordsprog.

**Henvisninger.** Angelordbogen er forsynet med et meget omfattende henvisningsapparat til stor hjælp for brugeren. Fra mange artikler henvises til synonymer hhv. til saglig-terminologiske artikler (jf. afsnit 3); der henvises også til de ord, det givne ord indgår i som sidsteled, fx: *-bræt s*, se *byge-*, *mangle- ...vrikkebræt*. Desuden findes der rene henvisningsartikler af flere typer, især følgende. 1) Til uventede stavemåder, fx *fertig* se *veertig* (= 40), opslagsform her i overensstemmelse med Mensing. Også fra tyske eller tysklignende ord til det angel-danske ord: *fuhr-værk* se *foerværk* (= vogn), og fra tilsvarende danske stavemåder: *for* se *foer* (hvor *foer* = kørsel mv., jf. ty. *Fuhre*). 2) Fra suffikser: alle ordbogens eksempler på et givet suffiks opregnes, fx *-else* se *begravelse*, *betydelse*, ... *vådtelse*; jf. også de mange eksempler under *-er*, *-et* og *-ig*, der kunne bruges i en undersøgelse af adjektiv- og adverbiumdannelse. 3) Participialadjektiver og sammensatte infinitiver – et udbredt fænomen i sønderjysk – henvises til behandling under hovedordet, fx: *ihjel-slagen part* se *slå I* og *op-bedt part* se *II bede 4* (= blive konfirmeret).

En særlig type er henvisninger fra rigsmålsord, der ikke findes i dialekten, til dialektens ord: rigsmålsordet sættes da i kantet parentes, fx [*bryllup*] se *II kost 2*. Denne brugervenlige henvisningstype anvendes også i enkelte andre tilfælde, fx fra [*bindingsværk*] til dialektens *tavlværk*, eller rettere: til *ege-tavlværk*, da *tavlværk* ikke foreligger som simpleks. En variant ses ved nogle af numeralierne, fx: *firs* – *ikke opt*(egnet), se *tagentig*, tilsvarende for *halvfems* (m.fl.) ...*negentig* (m.fl.), men anvendes også ved ord fra andre ordklasser: *hin-anden* ... findes ikke, jf. *enanden*, *enandre*, *hverandre*.

Henvisninger savnes kun i enkelte sammenhænge. Et generelt ønske kunne være henvisninger fra stærke bøjningsformer til grundform, fx fra (præsens) *stær* til *stå*, fra *æche* til *age*, men det ville være temmelig arbejdskrævende at gennemføre. Ellers har jeg kun savnet meget få henvisninger, fx *fesper* se *vesper*, idet der under *vesper* (= (eftermiddags)måltid) kun anføres udtaleform med *f-*; tilsvarende fra *skæhsvun* til *chaise-vogn*, da *sk*-formen er eneste udtale af ordet. Men det er småting. Som nævnt er der virkelig gjort meget ud af at sikre et righoldigt og brugervenligt henvisningsapparat.

**Etymologiske oplysninger** gives kun, hvor forfatterne har nyt at tilføje til "lettilgængelige" oplysninger i ODS og Feilberg (dvs. oplysninger placeret under samme opslagsord dér som i Angelordbogen, jf. s. 58). Ved lån fra nedertysk er oplysningerne hentet fra Mensing (jf. ovf.), fx under *pratle* (= våse): nt *praddeln*. Også tilsvarende højtyske former kan være anført, fx under *presse-kop* (= sylte, s.): jf nt *Presskopp* (hty *Preßkopf*). I mange tilfælde er der lagt et stort arbejde i eftersøgningen af ords oprindelse. Fx i artiklen *rydde-*

*stage* (= stang på gammeldags plov), udtale *Røstag*, hos Feilberg placeret under *rostage*. Etymologien betegnes som uvis, men forfatterne anfører parallelformer som *Röllstock*, *Rossstöcker* fundet under *Ploog* (= plov) hos Mensing, desuden nævnes *rüststol* fra Ordbog over Fjoldemålet som "de mest nærliggende ord for samme genstand". Ved indlån fra (neder)tyisk savnes i enkelte artikler den långivende tyske form, fx ved *kursvil* s. (= spøg, tidsfordriv); til gengæld gives der henvisning fra den tyske form: *kurzweil* se *kursvil*.

### 3. Saglig-terminologiske artikler

En utraditionel mellemkategori i ordbogen udgøres af de såkaldte saglig-terminologiske oversigtsartikler, der har karakter af leksikonartikler snarere end af ordbogsartikler. I adskillige kulturhistorisk set centrale artikler som *bage*, *barsel*, *dørnsk* (= stue), *høst*, *II kost* (= bryllup), *spindehjul* (= rok), *vogn* gives således en kortere eller længere rigsmålsbeskrivelse af den omhandlede ting, proces eller begivenhed med brug af dialektens ord for del-proceser, tilhørende genstande mv. Beskrivelserne varierer fra 5 linier i artiklen *fremgulv* (= forstue, entré) til 4-6 spalter i artikler som *hus*, *hør*, *væve* og *II kost* (= bryllup).

Denne artikeltype tages i ordbogen som en selvfølge og får ikke nogen egentlig introduktion. Første gang den nævnes sker det blot i forbiarten i omtalen af henvisninger til synonymer og antonymer mv. (kap. 3.2, Definitioner, s. 57): "Specielt gives der henvisning til de saglig-terminologiske oversigtsartikler, der måtte være relevante. For at angive, at der henvises til en sådan sagartikel, forsynes henvisningsordet med en \* (fx "jf \**æde*")." Få linier senere (kap. 3.3, Andre redaktionelle konventioner), omtales disse artikler igen meget kortfattet: "De saglig-terminologiske artikler anbringes foran citaterne; de indledes og afsluttes med en stiplede linie. Hvis de kun hører til en af underbetydningerne, anbringes de der" (min understregning). Forfatterne synes dermed at betragte de saglig-terminologiske redegørelser som "artikler i artiklerne", indplaceret i de pågældende ordbogsartiklers struktur. Anmelderen ville foretrække blot at tale om saglig-terminologiske afsnit i artiklerne. De i alt 36 artikler af denne type opregnes lige efter (s. 57), men beskrives ikke yderligere, skønt de fleste brugere af ordbogen nok kunne ønske lidt flere oplysninger om denne "nye" artikeltype.

Til illustration har jeg valgt artiklen for subst. *kilde* = brønd (se fig. 1 på omstående side).

kilde s fx /<sup>2</sup>ke:l/ <sup>1</sup>ke:l, <sup>1</sup>ke:l, ke:l H, ke:t  
 MbrJH; if Hag 45, 119, 176 ofl *Kehl*,  
*Kjehl*, *Kjèhl*, *Kjèhld*, if JKapp *Keel*; pl  
 if Hag 45 *Kjiller*.  
 – brønd.

-----  
*Kilden* (brønden) var omgivet af en  
*kildekasten* (brøndkarm; den kunne  
 også kaldes *karm*), som nåede op i  
 brysthøjde. Man kunne enten *drage*  
*vandet op* i et reb vha en *vrang*  
 (håndsving), eller man kunne benytte

en *svænke*, (if Hag) *kildesvange* (brønd-  
 vippe). Den lodrette stang på brønd-  
 vippet benævntes (if Hag) *kildesvøbe*.  
 Ved storvask stod man ved kilden og  
*byggede* (dvs bankede) tøjet.

-----  
 æ <sup>1</sup>ke:l æ 'dy.f H. di 'sdo. u. ve <sup>1</sup>ke:l ɔ  
 'byzd H. hvor (vor) *Keel hær got Vand*  
 JKapp. (talem:) *De è fò sild' á smi æ*  
*Kjèhld etter* (dvs kaste brønden til),  
*ven æ Baen è drunknt StSoltHag 176.* –  
 jf *vældkilde*.

Fig. 1. Artiklen *kilde* i Angelordbogen.

De kursiverede ord i sag-afsnittet er opslagsord i ordbogen, og der henvises ofte fra disse til sag-afsnittet med brug af \*, fx under *karm* s.: (jf *\*kilde*). Som det ses, er formuleringerne koncise og alligevel letlæselige.

Inspirationen til de saglig-terminologiske artikler kommer fra **Ømålsordbogen**, hvor denne artikeltype udgør en meget central kategori. Angelordbogen har ligesom Ømålsordbogen hjemsted på Institut for Dansk Dialektforskning, Københavns universitet. De hidtil udgivne 3 bind af Ømålsordbogen, der dækker afsnittet *a – fald*, håber jeg at få lejlighed til at anmelde i et kommende LEDA-nyhedsbrev; en grundigere vurdering af princippet med brug af saglig-terminologiske artikler udskydes dertil. Men som bekendt for flere af LEDA-Nyhedsbrevets læsere har Karen Margrethe Pedersen, der er en af redaktørerne på Ømålsordbogen, beskrevet princippet i flere foredrag og artikler, jf. fx artiklen "Ordbog og encyklopædi", i *Nordiske Studier i Leksikografi. II.* (1994). 248-53 – og senest artiklen "Ømålsordbogen" i *Mål & Mæle*, 1997 nr. 1 s. 12-18.

I Angelordbogen giver de saglig-terminologiske artikler gode fremstillinger af bestemte sagsområder med de tilhørende dialektale ord indføjnet i deres steds- eller proces-sammenhæng mv. Enkelte gange overraskes man dog lidt af sag-afsnittenes indplacering. Fx er hele redegørelsen for tærskning mv. placeret under *plejl*: her beskrives ud over plejl og tærskproces også, hvordan man bandt neg til stråtag, og hvordan det aftærskede korn blev rensat og sigtet. Andre ordbøger ville formentlig placere beskrivelserne af delprocesserne opdelt på flere opslagsord. Men hovedindtrykket er, at leksikon-afsnittene er harmonisk indplaceret i ordbogsartiklerne, og at de fremmer overblikket over et givet sagsområde. Det er således en fornøjelse at læse den meget informative beskrivelse under *II kost* af et gammel-



dags bryllup med *lovensøl* (trolovelsesgilde) 4 uger før, indbydelse (ved *bydesvend*), forberedelser i dagevis (*sysseldag* og *brudesengaften* sidste dag før), selve bryllupsdagen/-gildet (*første kostdag*) og det afsluttende gilde med naboer og *sysselfolk* (på *anden kostdag*).

#### 4. Afslutning

I indledningen udtrykker Jul Nielsen og Nyberg beskedent håbet om, at "publikationen i nogen grad /vil/ kunne gøre krav på betegnelsen "ordbog" (s.9). Forbeholdet kan man roligt slette: Angelordbogen giver med sine velstrukturerede artikler, de saglig terminologisk afsnit, det fyldige citatmateriale, det velordnede henvisningsapparat mv. et gedigent indtryk at det tidligere talte sprog i en væsentlig del af Sydslesvig. Der er i høj grad grund til at takke forfatterne for, at de gennem "års arbejde" fik samlet og udnyttet Angel-materialet til den vægtige ordbog (og den omhyggelige beskrivelse af lydsystemet), der nu foreligger. Leksikografisk, men også kulturhistorisk set er det væsentligt, at der udarbejdes ordbøger som Angelordbogen, der registrerer dialekttalende samfunds ord og udtryksmåder samt beskrivelser af gamle arbejdsprocesser og traditioner. Jeg vil derfor slutte med gøre første anmelders ord til mine: Angelordbogen er en imponerende præstation.

Angelordbogen giver sammen med Ordbog over Fjoldemålet et godt overblik over ordforrådet i de gamle danske dialekter syd for den nuværende rigsgrænse. Den ene af forfatterne har tidligere deltaget i udgivelsen af en ordbog over en dialekt lidt nord for grænsen, nemlig M.B. Ottsen: *Hostrup-Dansk* II bd. 1-5 (1961-69), udg. af Ella Jensen og Magda Nyberg. Ligeledes fra egnen nord for grænsen er der udgivet en mindre omfattende ordbog: *Alsisk Ordsamling* (1988). Ordforrådet i disse 4 værker indgår som et meget væsentligt materiale i udarbejdelsen af Jysk Ordbog.

*Anna Braasch*

#### En stor dansk sprogteknologisk ordbog - et nationalt projekt

Center for Sprogteknologi (CST) er i gang med planlægningen af en stor dansk sprogteknologisk ordbog (*STO*). Da projektet givetvis kan påregne interesse i det leksikografiske miljø, vil vi nedenfor fremlægge de væsentligste punkter i disse planer og berette om de forberedende skridt der er taget indtil nu. Indlægget er tænkt som en første, bred orientering om initiativet og vil blive fulgt op ved senere lejlighed med en mere detaljeret redegørelse.

### **Hvad er en sprogteknologisk ordbog?**

En sprogteknologisk ordbog er en samling organiserede leksikalske data lagret på et elektronisk medium, som regel i en databasestruktur, og beregnet til sprogteknologisk (maskinel) anvendelse. Maskinel anvendelse stiller særlige krav til den leksikografiske beskrivelse mht. indhold og format og især mht. detaljeringsgrad.

Den planlagte sprogteknologiske ordbog vil være et basisprodukt der indeholder systematisk beskrivelse af leksikalske enheders morfologiske og syntaktiske egenskaber og forsynes med en detaljeret dokumentation af ordbogens lingvistiske og leksikografiske principper. Derved sikres at ordbogens materiale også kan udnyttes i andre end sprogteknologiske sammenhænge, eksempelvis som basis for andre leksikografiske produkter.

### **Projektets relevans, baggrund og perspektiver**

Udarbejdelsen af en stor sprogteknologisk ordbog har to hovedformål. For det første er den nødvendig for udviklingen af de allerfleste avancerede sprogteknologiske anvendelsesprogrammer; for det andet vil den også være en uvurderlig ressource i forskningssammenhæng.

### **Forskningsrelaterede aspekter**

Udviklingsarbejdet indebærer i forbindelse med etableringen af strukturen for ordbogen nogen leksikografisk forskning bl.a. inden for syntaks og semantik som i hovedtræk vil være afsluttet inden det egentlige kodningsarbejde påbegyndes.

Der vil i et vist omfang også være behov for forskning afledt af selve ordbogsarbejdet især fordi dette er det første store danske ordbogsprojekt hvor datalingvistiske aspekter og leksikografiske metoder søges forenet i skabelsen af et sprogteknologisk basisprodukt. Disse områder vil givetvis også skabe en udmærket baggrund for specialer og studenterprojekter.

På et mere fremskredent stadium vil ordbogen imidlertid også have forskningsmæssig betydning som en ressource i sig selv. Ordbogen vil ikke blot omfatte leksemer og deres beskrivelse, men i meget høj grad også deres kombinatorik. Der er altså tale om et leksikografisk projekt der i sig selv er syntaktisk orienteret og som giver mulighed for datamatiske syntaktiske analyser i stort omfang, når det er færdigt.

### **Anvendelsesorienterede aspekter**

Sprogteknologiske produkter omfatter alle former for programmer, der tjener til forbedring og effektivisering af menneskers arbejde med tekster i naturligt sprog: stave- og grammatikkontrol, oversættelse, undervisningsprogrammer osv. I alle sådanne produkter

indgår en ordbog. Fraværet af en stor ordbog gør derfor udviklingsopgaver af denne type næsten uoverkommelige, både rent arbejdsmæssigt og økonomisk.

Det er vigtigt at udvikle forskellige slags sprogteknologiske værktøjer af høj kvalitet også for dansk, da dette vil bidrage til at sproget bliver brugt i alle dets funktioner i både erhvervs- og kulturlivet samt i den offentlige sektor. Det er således indlysende, at udover den forskningsmæssige interesse har udvikling af sprogteknologien også stor samfundsmæssig og kulturel betydning.

Denne erkendelse var baggrunden for, at Teknologinævnet i 1994 lod udarbejde en rapport om dansk sprogteknologi. En af konklusionerne var at udviklingen af en stor sprogteknologisk ordbog er en meget stor og kompleks opgave, som markedet på ingen måde selv kan løfte. Derfor foreslog rapporten, at Center for Sprogteknologi skulle tage initiativ til at igangsætte et ordbogsprojekt.

Projektet har fået positiv modtagelse og vurdering hos både Forskningsrådene, Forskningsministeriet, Det Danske Sprog- og Litteraturselskab og hos forlag, samt ikke mindst hos andre relevante institutioner.

### **Projektets organisering og løbetid**

Projektet bør være et *samarbejdsprojekt* med mange partnere. Center for Sprogteknologi har taget initiativ til projektet og forventer at udføre en betydelig del af arbejdet, herunder at styre projektet. Andre medvirkende vil være datalingvistiske og leksikografiske miljøer, forlag mv.

Fordelene ved et samarbejdsprojekt er dels, set fra projektets side, at det giver adgang til at drage nytte af flere menneskers og miljøers ekspertise, og dels, set fra miljøernes side, at den inspiration der udgår fra projektet vil kunne bidrage til flere institutioners forskning.

CST er gået i gang med de indledende dele af projektet, idet opsparede reserver fra tidligere år anvendes hertil. Løbetiden anslås til 6-7 år; men dette kan først vurderes helt konkret, når de igangværende undersøgelser vedr. samarbejde, genbrug af eksisterende ordbogsdata, programmell mv. er gennemført. Projektplanen lægger op til at brugbare dele af ordbogen kan frigives efterhånden. Her tænkes på mindre udsnit af ordbogen, fx en ordklasse, ordforrådet fra et afgrænset fagområde eller et udsnit af ordbogen med kun basislingvistiske oplysninger e.l.

## Ordbogens indhold og format

### Størrelse og ordforråd

Ordbogens størrelse er planlagt til ca. 100.000 opslagsenheder, således at der i denne optælling medtages lemmata og sublemmata. Ordforrådet i *STO* bliver sammensat af almensproglige ord og ord der optræder både i almensproglige og fagsproglige tekster, samt ord der optræder i tekster fra flere fagområder. De sidstnævnte typer, dvs. de såkaldte gråzoneord, fællesord eller almene fagord, relaterer sig i større eller mindre grad til fagsprog. Der er udbredt enighed om at sådanne kategoriseringer af ord ikke kan anses for at være absolutte men at de altid er afhængige af et bestemt praktisk formål. Det valgte ordforråd bliver beskrevet i leksikografisk-lingvistisk henseende på en ensartet måde uanset hvilken type af ord det drejer sig om.

Ved almensproglige ord forstår vi det ordstof der anvendes i nutidige dagligdags kommunikationssituationer. Ved udvælgelsen af lemmata går vi ud fra en kombination af nogle grundlæggende kriterier mht. kilde, frekvens og relevans som vil blive forfinet trinvist. På et meget basalt plan kan det f.eks. fastlægges at det almensproglige ordstof skal tilhøre standardsprogligt skriftsprog og den centrale del af ordforrådet (baseret på frekvensundersøgelser i korpus mm.) og at det skal have relevans i forhold til *STO*'s applikationsområder.

En foreløbig dækningsmæssig sammenligning mellem Retskrivningsordbogens ordforråd og det ordforråd der skønnes relevant for *STO* viser at der er sandsynlighed for at ca. 40% (40.000) af *STO*'s opslagsord vil falde inden for og ca. 60% uden for Retskrivningsordbogens dækning.

Der kan ikke drages en skarp grænse mellem de almensproglige og fagsproglige dele af sproget da al fagsproglig kommunikation også indeholder almensproglige ord og udtryk samtidig med at almensproget udvides løbende med ord og udtryk fra forskellige fagsprog fx med almen udbredelse af kendskabet til bestemte fagområder. Et eksempel er informationsteknologiens indtog i dagspressens spalter der medfører at fagord som *computer*, *tekstbehandling*, *printer* ikke mere er forbeholdt fagsproglig kommunikation; de opfattes snarere som tekniske eller fagrelaterede udtryk brugt i dagligdags kommunikation.

Det fagsproglige ordstof skal selekteres fra nutidigt, fagrelateret skriftsprog og fra flere udvalgte fagområder med sprogteknologisk relevans og under hensyntagen til forskellige faktorer såsom frekvens i almen- og i fagsproglige tekster, tekstens type og dens fagsproglighedsgrad osv. Der skal indsamles relevante fagsproglige tekster og eksisterende, genbrugelige leksikografiske beskrivelser.

### Leksikalske oplysninger

Udvælgelsen og struktureringen af oplysningstyper vil i høj grad ske på grundlag af eksisterende praksisser inden for det sprogteknologiske område. I denne forbindelse har EU-projektet PAROLE en særlig betydning, da der inden for rammerne af dette projekt produceres en sprogteknologisk ordbog med leksikalske beskrivelser af 20.000 lemmata som CST er ansvarlig for. Det er tanken at der skal bygges videre på denne ordbog (herom også senere).

Den traditionelle måde er at opdele leksikalske oplysninger efter deres indhold i morfologi (inkl. morfosyntaks), syntaks og semantik svarende til lingvistikens discipliner. I *STO* vil oplysningerne desuden blive klassificeret og struktureret efter lingvistisk niveau i basisoplysninger (fx ordklasse, bøjning, syntaktisk valens, semantiske roller) og avancerede oplysninger (fx derivation, kontroltype, 'raising', læsningsopdelinger).

### Arbejdsmetode - et par nøgleord

#### Genbrug

En af hjørnestenene i projektet er genbrug af eksisterende ordbogsdata og korpora. Det er klart, at allerede eksisterende datamatiske ordbogsdata skal genbruges i videst muligt omfang. Disse er imidlertid ganske begrænsede og vil i vidt omfang allerede være indgået i PAROLE-ordbogen. Hertil kommer muligheden for at genbruge traditionelle ordbøger. Selv om der er store forskelle på, hvordan ordbøger formuleres for mennesker og for maskiner, så er der dog også en del væsentlige fællestræk. Som eksempler på ordbogsarbejde som kan være af interesse for *STO* kan nævnes Den Danske Ordbog, der er under udarbejdelse under Det Danske Sprog- og Litteraturselskab, et delprojekt om valens under det SHF-støttede projekt om Udforskning af Dansk Ordforråd og Grammatik (UDOG) på Odense Universitet, samt de store ordbogsarbejder der udføres i det danske leksikografiske miljø og på de danske forlag. Det skal tilføjes at ikke kun leksikalske beskrivelser, men også korpussamlinger mv. vil være af interesse.

Vi er gået i gang med henvendelser til en del datalingvistiske og leksikografiske miljøer og til forlag og vil fortsætte hermed. Henvendelserne er i det store og hele blevet besvaret positivt; men der er behov for en lidt større afklaring af projektet før man kan indgå faste aftaler.

Selv med en større afklaring omkring projektets status må det dog påregnes at forhandlinger

om betingelser for deltagelse, herunder rettigheder i forhold til den færdige ordbog, vil være ret tidskrævende. Rettighedsproblematikken er vanskelig, da der ikke findes ret mange love eller regler på dette område, der direkte dækker vore behov.

### **Model**

Ordbogsarbejdet bliver udført vha. elektroniske værktøjer der er bygget til at håndtere en bestemt beskrivelsesmodel. I denne model struktureres de klassificerede oplysninger på tre niveauer mht. deres lingvistiske indhold (morfologi, syntaks og semantik). Modellen muliggør desuden opdeling af oplysningerne mht. lingvistisk niveau, som omtalt ovenfor, i basis og avanceret beskrivelse.

Der er en række indlysende fordele ved at strukturere ordbogen vha. denne model. Ordbogen kan udarbejdes og tages i brug etapevist, fx kan basisniveauet afsluttes før det avancerede niveau; morfologien kan færdiggøres før syntaksen; der kan laves en prioriteringsrækkefølge for fag/emneområder osv. En anden fordel er at det bliver enklere at uddrage netop den delmængde af oplysninger som kræves for bestemte formål eller anvendelser.

På denne måde bliver ordbogen fleksibel og dermed anvendelig til flere forskellige formål.

### **Afsluttende bemærkninger**

Et så omfattende projekt som vi har skitseret ovenfor kræver naturligvis nøje planlægning. I dette oplæg har vi beskrevet nogle udvalgte aspekter medens andre, ligeså relevante, ikke kunne medtages her. Vi vil derfor gerne fremlægge projektet for LEDAs medlemmer på et medlemsmøde i løbet af efteråret hvor vi vil fortælle om flere aspekter og gå mere i dybden med enkelte, leksikografisk interessante punkter. Samtidig bliver der lejlighed til at få besvaret spørgsmål om projektet, selve ordbogen og samarbejdsmulighederne.

*Anna Braasch, Bente Maegaard, Bolette Sandford Pedersen*