17. DEC. 1991

LEDA
LEKSIKOGRAFER I DANMARK

Nyhedsbrev nr. 9 December 1991

DANSK SPROGNÆVN
Sag B+989/2

LEDA

afholder medlemsmøde
torsdag d. 6. februar 1992 kl. 15.45–18.00
hos
IBM/European Language Services
Kongevejen 150, 3460 Birkerød

Programmet vil omfatte følgende emner:

* Velkomst: "Hvad er ELS"

* Udvikling af ordbøger til oversættelse:
  Identifikation af termer
  To-sprogede (term)ordbøger
  Opbygning af en-sprogede ordbøger
  (stavekontrol, morfologi)

* Brugen af ordbøger til oversættelse:
  Oversættelsesværktøjet i IBM (demonstration)
  Maskinoversættelse i IBM

Programmet vil tillige omfatte demonstrationer af forskellige værktøjer og processer, der indgår i oversættelses-arbejdet.

Før mødet indledes (kl. 16.00) inviteres deltagerne på kaffe, te og kager.

NB! TILMELDING til medlemsmødet på IBM er **nødvendig** og skal ske til Pia Jarvad på telefon 31 54 22 11 (lokal 2735) senest 15. januar 1992.

Man kommer til **IBM i Birkerød** enten med S-toget til Birkerød station, hvorfra der vil være ca. 15 minutters gang (ad Kajerødvej) til IBM på Kongevejen, eller med bil ad Kongevejen gennem Birkerød, hvor man vil finde IBM's bygninger lige før Texaco tankstationen.

Venligst bemærk, at mødet afholdes hos **IBM i Birkerød.**

Meddelelser

I dette nummer bringer vi som sidste side annonceringen af *InterLex 6*, Internationalt kursus i leksikografi. Kurset foregår i Exeter fra d. 5. til d. 10. april 1992. Tilmeldingsblanketter kan rekvireres hos LEDAs sekretær Pia Jarvad.

Som bilag til dette nummer af Nyhedsbrevet udsendes div. meddelelser fra ***Nordisk Forening for Leksikografi***, herunder også et *spørgeskema*. Bestyrelsen for Nordisk Forening for Leksikografi vil bede LEDAs medlemmer om at udfylde og indsende skemaet til LEDAs sekretær snarest muligt, hvorefter skemaerne vil blive ekspederet videre til Oslo.

Nyt fra medlemmerne

LEXICOGRAPHICA. SERIES MAIOR (LSM)

Som det er LEDAs læsere bekendt, har danske ordbogsforskere fået publiceret deres forskningsresultater i LEXICOGRAPHICA. SERIES MAIOR. Denne monografiserie er siden 1984 udkommet i Max Niemeyer Verlag i Tübingen og udgives af Sture Allén, P. Corbin, R.R.K. Hartmann, F.J. Hausmann, O. Reichmann, L. Zgusta og undertegnede. DANLEX-gruppen står for bind 20, *Descriptive Tools for Electronic Processing of Dictionary Data* (1987), og G. Baunebjerg Hansen for bind 35, *Artikelstruktur im zweisprachigen Wörterbuch* (1990), en bog som har været omtalt og anmeldt i LEDAs to sidste numre. Dertil kommer 4 bind fra de Københavnske *Symposium on Lexicography* ved K. Hyldgaard-Jensen og A. Zettersten. Til danske ordbogsforskere kan vi nu også regne H. Bergenholtz, der sammen med J. Mugdan har fået udgivet det usædvanlig efterspurgte og nu udsolgte bind 3 *Lexikographie und Grammatik* (1985). Hvis vi tæller bind i LSM, så står der altså danske ordbogsforskere bag 7 af de indtil nu 40 udkomne bind. Det er jo en flot indsats i international ordbogsforskning.

Men der er da også udkommet LSM-bind, der kan tilføre dansk ordbogsfremstilling og -forskning nye forskningsresultater fra udlandet. For alle ordbogsmagere og encyklopædister må f.eks. en monografi som W. Hupka *Wort und Bild. Die Illustrationen in Wörterbüchern und Enzyklopädien* (1989) være en uvurderlig kilde til inspiration, ligesom K.D. Ludwig *Markierungen im allgemeinen einsprachigen Wörterbuch des Deutschen* (1991) må anses for at være en udfordring både til leksikologer og leksikografer. Den, der er interesseret i *The Dictionary and the Language Learner*, kan studere flere bind (16, 17, 30, 31, 37, 39). Men det er ikke bare engelsk og tysk ordbogsforskning, som stær stærkt i serien, der er også udkommet bind om fransk, spansk, kinesisk, tamilsk og græsk leksikografi.

Publikationssprogene er engelsk, tysk og fransk. Vi udgiver ca. 4–6 bind om året, herunder Ph.D.-afhandlinger, disputatser og kongresbind. Hvert indsendt manuskript bedømmes særskilt af 2 medlemmer og ved uenighed af 3 medlemmer af redaktionskomiteen. Når et manuskript er antaget til udgivelse, skal der ikke – i modsætning til hjemlige eller visse andre internationale forlag – betales tilskud til trykkeomkostninger.

Da det i 1993 igen bliver min tur til at være "Geschäftsführender Herausgeber", vil jeg allerede nu opmuntre til, at danske og andre skandinaviske leksikografer og ordbogsforskere begynder at overveje muligheden for at publicere forskningsresultaterne i LEXICOGRAPHICA. SERIES MAIOR. Det er en serie, som har en professionel international distribution og bliver anmeldt i førende tidsskrifter.

For yderligere oplysninger, så henvend jer til undertegnede på Handelshøjskolen i København, Dalgas Have 15, 2000 Frederiksberg, tlf. 31 19 19 19, lok. 3245 el. 3250.

Hans-Peder Kromann

Etablering af fællesnordisk dialektordbog?

Institut for Jysk Sprog- og Kulturforskning (Aarhus Universitet), der som hovedopgave har at udarbejde Jysk Ordbog, har indbudt en række dialektleksikografer fra de nordiske lande til et møde, hvor man skal diskutere mulighederne for at etablere en fællesnordisk (eller i hvert fald fællesskandinavisk) dialektordbog. Mødet finder sted d. 11.-14. maj 1992 på Aarhus Universitets kursusejendom Sandbjerg gods ved Alssund, og i mødet deltager repræsentanter fra: Ømålsordbogen, København; Norsk Ordbog og Norsk Målførearkiv Oslo; Ordbog over trøndermåla, Trondheim; Ordbok över Sveriges dialekter, Uppsala, og Ordbok över Finlands svenska folkmål, Helsingfors.

Torben Arboe Andersen

Fagnyt

Dansk-fransk ordbog

Den 4. december 1991 præsenteredes den reviderede 4. udgave af *Blinkenberg og Høybye: Dansk-fransk ordbog* ved en reception i Carlsbergfondet. Talerne var fondets formand, professor, dr.phil. Kristof Glamann, forlagsleder, cand.polyt. Lauge Stetting (HHKs Forlag), professor, dr.phil. Jens Rasmussen (Blinkenberg-Høybye Fondet), docent, cand.ling.merc. Bodil Nistrup Madsen og cand.interpret. Birgitte Lauterbach (begge fra HHK). Disse redegjorde for forlagsarbejdet og revisionsarbejdet, samt det spændende projekt at indlægge ordbogens materiale i en database. Fremtidsperspektiverne for genbrug af materialet og udvidelse af ordbasens anvendelsesmuligheder blev også skitseret. (Se også den vedlagte pressemeddelelse!)

AB

Referater

EURALEX Dictionary Use Workshop

Der er blevet afholdt en workshop d. 29.9.91 i Oxford, hvis emne var brug af ordbøger.
Der var omkring 60 deltagere, inddelt i grupper efter modersmål. Hver gruppe skulle oversætte en tekst, hvis sværhedsgrad man selv kunne vælge. Oversættelsen foregik fra modersmålet til engelsk. Der var et bredt udvalg af en- og tosprogsordbøger, man kunne benytte sig af.
Den egentlige opgave bestod i at registrere hvert ordbogsopslag og besvare et spørgeskema for hvert opslået ord. Formålet med denne workshop var at få detaljerede, ensartede (tilrettelagte) beskrivelser af opslagsprocessens enkelte trin, især for de opslag, som ikke kunne give svar på brugerens spørgsmål ved første forsøg. Disse spørgsmål samt brugerens søgestrategi kan i fremtiden få større indflydelse på ordbøgers indhold og udformning.
De indsamlede spørgeskemaer bliver analyseret med henblik på at systematisere oplysningerne og lade konklusionen indgå i arbejdet med praktisk leksikografi.

Seventh Annual Conference of the UW Centre for the New OED and Text Research USING CORPORA, Oxford, 30.9.-1.10.91

Der var 195 tilmeldte deltagere, hvoraf ca. 180 mødte op. Konferencen bød på 14 egentlige foredrag, desuden var der et informationsoplæg om "British Corpora" og en afsluttende diskussion om "det ideelle korpus".
Foredragene gik i dybden med mange forskellige emner. De to hovedsessioner koncentrerede sig om nogle aspekter og kriterier, der er relevante for oprettelsen af korpora, samt om de forskningsmuligheder, et etableret korpus kan give.
Almenlingvistiske, sprogstatistiske og leksikologiske spørgsmål blev behandlet med så forskellige sprog som sovjet-jiddisch, italiensk, ungarsk og engelsk som udgangspunkt. Indlæg om moderne sprogteknologi, fx om konkordansværktøjer til paralleltekster og om et teksttagginssystem, bidrog til at konferencen kunne give et nuanceret billede af, hvordan man arbejder med korpora i dag.
Programmet var desværre ret stramt; der blev ikke tid til megen diskussion, og enkelte foredrag måtte endda afkortes lidt.

Der er udkommet *Proceedings of the Conference,* der kan rekvireres fra: Dictionary Dept., Oxford University Press, Walton Street, OX2 6PD Oxford, England.

Anna Braasch

KORPUSMØDE: DK87-90

På *Handelshøjskolen i Århus* blev der d. 25.10.1991 afholdet et *møde om brugen af DK87-90.* På mødet samledes fortrinsvis leksikografer, der er blevet registreret som brugere af de maskinlæsbare tekstsamlinger, som har fået navnet DK87-90. Henning Bergenholtz har stået for oprettelsen af dette korpus, og han er i det daglige arbejde blevet godt hjulpet af Richard Almind. Mødet blev indkaldt af Henning Bergenholtz og Ole Norling-Christensen med det formål at diskutere de hidtidige erfaringer med korpuset og fortælle om, i hvilke sammenhæng man har kunnet drage nytte af disse tekstsamlinger indtil nu. Der blev holdt 9 indlæg i løbet af dagen. Mødet afsluttedes med en livlig diskussion om mulighederne for oprettelsen af yderligere danske tekstkorpora. Der blev også fra flere deltagere ytret ønske om at etablere et informationsnetværk mellem brugerne.

Her bringes resumeerne af de 9 indlæg fra mødet samt et yderligere bidrag til emnet.

a) Udvælgelse af tekster til DK87-90
af Henning Bergenholtz

DK87-90 er et multifunktionelt korpus over dansk almensprog. For hvert af årene 1987 til 1990 er der udvalgt en tekstmængde på 1 mio tekstord, som består af tekster hhv deltekster fra 100 nyudkomne romaner, 50 ugeblade og 50 aviser. Fra hver roman er der udvalgt et kapitel på ca 5000 ord, fra hver avis og ugeblad er der medtaget en række større og mindre bidrag ligeledes i et omfang på ca 5000 ord. Forholdet mellem de tre tekstarter udgør således

- romaner og noveller (50% af tekstmængden)
- aviser (25% af tekstmængden)
- ugeblade (25% af tekstmængden)

Hver af disse tre tekstarter læses af en betydelig og bred del af befolkningen, og de er i modsætning til ikke-medtagne børnebøger og fagtekster ikke skrevet for en bestemt, forholdsvis begrænset del af den.

Mht romaner og noveller består hver korpustekst af mindst et kapitel eller, hvis der ingen kapitelinddeling findes, af en tekstdel, som udgør en helhed. Der var tilstræbt en tekstlængde på ca. 5.000 løbende ord. De fleste tekster ligger mellem 4.000 og 5.500 ord, men en enkelt tekst har et omfang på kun 2.000 ord og fem andre på mellem 7.000 og 9.000 ord. Valget af et kapitel i en roman hhv novelle blev foretaget efter et tilfældighedsprincip. Af de mest kendte ugeblade og et bredt udsnit af landsdækkende og lokale aviser blev tre-fem eksemplarer pr avis hhv ugeblad udtaget. Fra hver af disse aviser eller ugeblade blev der udvalgt hele tekster, der tilsammen udgør ca. 5.000 løbende ord.

H.Bergenholtz: DK87: Et korpus med dansk almensprog; in: *Hermes 1*, 1988, 229-237.
H.Bergenholtz: DK87-DK-90: Dansk korpus med almensproglige tekster; in: *3. Møde om Udforskningen af Dansk Sprog. Århus Universitet 11.-12. oktober 1990*, udg. af Mette Kunøe og Erik Vive Larsen. Århus 1991, 32-42.
H.Bergenholtz: *Dansk Frekvensordbog. Baseret på tekster fra danske romaner, ugeblade og aviser fra 1987-1990.* København: Gad 1992 (i trykken).
H.Bergenholtz/C.Hansen: Informationer om ordbogens forudsætninger og opbygning; in: H.Bergenholtz: *Dansk Frekvensordbog.* København: Gad 1992, I-XX (i trykken)
H.Bergenholtz/J.Mugdan: Korpusproblematik in der Computerlinguistik: Konstruktionsprinzipien und Repräsentativität; in: *Computational Linguistics. Computerlinguistik. An International Handbook on Computer Oriented Language Research and Applications. Ein internationales Handbuch zur computerunterstützten Sprachforschung und ihrer Anwendungen.* Hrsg. von István S. Bátori, Winfried Lenders, Wolfgang Putschke. Berlin/New York: de Gruyter 1989, 141-149.
H.Bergenholtz/B.Schaeder (Hrsg.): *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora.* Königstein/Ts.: Scriptor 1979.

b) Homogenisering af et korpus
af Richard Almind

Konverteringen af korporaerne DK87-DK90 til maskinlæsbar form blev påbegyndt i 1988. DK87 blev således scannet. DK88 og DK89 blev, grundet dårlige erfaringer med scanningen, indtastet af et indtastningsbureau i WordPerfect og gemt herfra som ASCII-tekst (menu: *Gem som DOS).* Fremskridt i scanningsmetoder gjorde, at det kunne betale sig delvist at indscanne DK90 (bøger og den bedste kvalitet af blade og aviser), mens resten blev indtastet.
Den store forskel på metoderne fra 1988 og frem til 1991, såvel som opdateringer af programmer og de mange hænder korpuset har været igennem, førte til en række mangler, der hver for sig kan virke harmløse, men i fællesskab svækker korpusets genuinitet. Følgende fejl blev derfor rettet:

1. DK87 indeholdt styrekoden ASCII 31 (en omvendt trekant). Dette tegn blev erstattet af et mellemrum.
2. WordPerfects automatiske orddeling var blevet overført til ASCII-teksten. Disse orddelinger blev fjernet under hensyntagen til bindestregskomposita, direkte tale, punktinddeling og pausestreger.
3. I DK87 fandtes ofte selvstændige ord kædet sammen med det efterfølgende med tre punktummer uden mellemrum. Flere end tre, men færre end seks punktummer i række (pauser konstrueret af forfattere for at kendetegne afbrydelser og lange tænkepauser etc.) blev til tre punktummer og har nu enten foran- eller efterstillet et mellemrum (eller er omgivet af samme).
4. Hvert afsnit og hver overskrift afsluttes nu med to linjeskift (to gange ASCII 13+10).
5. Anførselstegn blev ændret til kun at omfatte to slags: dobbelt anførselstegn (ASCII 34 = "xx") og enkelt anførselstegn (ASCII 96 = 'xx').
6. To mellemrum efter hinanden er nu blevet til et.

Dette arbejde forventes afsluttet i marts 1992. Til den tid vil alle nuværende brugere blive orienteret og et opdateringsskema vil blive udarbejdet. Derudover vil hver bruger få en folder med en kort beskrivelse af korpusets opbygning, en nøje bibliografi over de anvendte tekster og en liste over de rettelser, der er blevet foretaget siden august i år.

c) Forudsætninger for optimal brug af et korpus
af Ole Norling-Christensen

Hovedtesen i mit indlæg var, at et korpus skal være en veldokumenteret én-én-tydig afbildning *af de træk ved* teksterne, som ønskes studeret. Jo flere træk, der ønskes repræsenteret, desto vanskeligere bliver korpus at bruge.
En lang række valg må træffes, hvis korpuset skal blive konsistent; og konsistens er en forudsætning for at elektroniske søge- og bearbejdningsmetoder kan give pålidelige resultater.
Eksempler på nødvendige valg: Ved hvilke "ASCII-tegn" skal *æ, ø, å* og *é* repræsenteres; skal der skelnes mellem forskellige slags citationstegn; forskellige slags afsnitsinddeling; forskellige slags fremhævelser? - og hvor mange og efter hvilke kriterier?
Som et egnet værktøj til systematisk, generisk kodning omtalte jeg den internationale standard SGML og det internationale *Text Encoding Initiative's* arbejde med at opstille *Guidelines for the Encoding and Interchange of Machine-Readable Texts.*

d) FICTION - En fritekst database
af Søren Brandt og Jens Ørnfeldt Thomsen

De fleste "almindelige" databaser adresserer behovet for at strukturere store mængder af ensartede oplysninger på en sådan måde, at det er let og hurtigt at finde frem til disse igen på et senere tidspunkt. Derimod vil en typisk database have problemer med data, som ikke har en tilsvarende klar struktur.

Et korpus er et eksempel på en sådan datamængde. Har man behov for at søge efter specielle ord, ordstammer, endelser, sætninger og lignende, er der ikke megen hjælp at hente.

FICTION er udviklet med henblik på hurtig og fleksibel søgning i store tekstmængder, og sigter altså imod at gøre det let at finde og udtage relevante dele af tekster til videreforarbejdning i f.eks. et tekstbehandlingsprogram. Giver en søgning resultat, vil forekomster blive vist i deres kontekst, hvorefter der kan læses, bladres og evt. noteres på en indbygget notesblok.

Eksempler på ting der kan søges efter v.h.a. FICTION:
- Forekomster af enkeltord
- Ord med bestemte stammer/endelser/længder
- Samtidige forekomster af flere ord
- Alle kombinationer af ovenstående.

Nærmere oplysninger om systemkrav, specifikationer, pris mm. kan fås hos de to forfattere/programmører (LEDA har adressen).

e) Erfaringer med DK-Korporaerne
af Carsten Hansen

I mit indlæg redegjorde jeg for Dansk Sprognævns erfaringer med DK-Korpusset, hvor det er lagt ind i tekstsøgningsværktøjet WordCruncher. Med mange eksempler fra korpusset gav jeg en vurdering af dets brugsværdi til ortografiske, orddannelsesmæssige, syntaktiske, semantiske, kollokationelle og tekstlingvistiske undersøgelser. Sprognævnets erfaringer er med undtagelse af ortografien meget positive, idet det fungerer som et fint supplement til Sprognævnets manuelt excerperede seddelsamling. Årsagen til den ortografiske utroværdighed lokaliseredes til fejl og mangler ved selve processen omkring etableringen af korpusset. Jeg pointerede supplerende den store værdi korpusset har i mit arbejde som redaktør af en dansk-tysk-dansk ordbog.

f) "Fejl" og fejlfinding i DK
af Peter Widell

På *Politikens Forlag* er arbejdet med DK-korpus så småt ved at komme igang. Indtil videre har vi fortrinsvis arbejdet med opbygningen af forskellige edb-værktøjer til behandling af korpuset. Værktøjerne omfatter bl.a.: (A) et *konkordansprogram* med sofistikerede søgemuligheder; (B) et *program til registrering af forskellige former for kollokationer* ved hjælp af statistiske beregningsmetoder; (C) et *program til generering af fuldformsordlister* på baggrund af bøjningsoplysninger i en ordbog - in casu bøjningsoplysninger i Nudansk Ordbog; (D) et *morfologisk analyseprogram.* Programmet har bl.a. været anvendt til at finde slåfejl og optiske læsefejl i DK-korpus (dvs. fejl af typen *advokternes, afdelngsleder, udgangspukt, søvningt).*

g) Anlæg til Syntaktisk Analyse
af Hans Arndt

Baggrunden for mit indlæg er den opfattelse, at de søgninger man kan foretage på et 'råt' korpus alene ved definition af alfabetiske strenge, ikke er tilstrækkelige til en fuldstændig udnyttelse af korpus' empiriske værdi; og at der derfor er behov for udarbejdelse af et *taggingssystem,* der muliggør grammatisk definerede søgninger.

Der foreligger to veldokumenterede forsøg med tagging, som beskrevet dels i Garside et al. 1987, dels af Fred Karlsson, Helsinki Universitet (f.eks. 1990). Jeg fremlagde på mødet nogle *principper for tagging,* baseret hovedsagelig på sidstnævnte. Min hovedpointe var at der skal foretages dels en primærtagging, baseret på morfologi og leksis, dels en sekundærtagging, baseret på en empirisk verificeret syntaktisk analyse. Det arbejde Karlsson har udført i Helsinki viser, at et sådant tagginssystem fungerer for finsk, engelsk og svensk, således at nye tekster kan tagges med over 97% disambiguering, og uden uoverkommelige behov for datakraft.

Garside, Roger, Geoffrey Leech & Geoffrey Sampson (eds.) (1978) *The Computational Analysis of English.* London, Longman.
Karlsson, Fred (1990) *Constraint Grammar as a Framework for Parsing Running Text* in Proceedings from the XIIIth International Conference on Computational Linguistics, Helsinki.

h) Statistiske resultater af arbejdet med DK87-90
af Henrik Holmboe

På Afdeling for Datalingvistik, HHÅ, er der til hvert af de fire textcorpora på ca. 1 mill. danske ord udarbejdet en progressivt alfabetisk, en rangordnet og en retrograd alfabetisk ordbog med en række statistiske oplysninger til hver enkelt i texterne forekommende ord. Disse oplysninger omfatter for hvert ord:
- absolutfrekvens i hver af de tre genrer AVIS - BLAD - BOG
- den samlede absolutfrekvens i de tre genrer
- relativ frekvens
- rang
- rangniveau

I den rangordnede ordbog er der tillige oplysning om den akkumulerede relativfrekvens samt om, hvor mange ord der optræder på hvert rangniveau. I hver ordbog er ordene forsynet med et løbenummer, og for hvert ord i de enkelte ordbøger er der henvisninger til det pågældende ords løbenummer i de to andre ordbøger.
Formålet med ordbøgerne er at bidrage til vor viden om dansk sprogbrug og sprogbygning, specielt hvad angår morfologi og orddannelse.
Under forberedelse er en fyldigere publikation om de anvendte metoder og med angivelse af en række statistiske iagttagelser, som ikke fremgår af de hér omtalte ordbøger. Publikationen vil udkomme i Afd. f. Datalingvistiks skriftserie *"Skriften på skærmen".*

i) Erfaringer med excerperinger af sprogbrugseksempler fra DK87-90
af Anna Braasch og Frede Boje

Vi har i vores daglige arbejde behov for at kunne foretage en systematisk leksikografisk beskrivelse af flerordsenheders egenskaber, men har savnet en brugbar model til dette formål. Både humanoversættelse (især *til* fremmedsprog) og automatisk behandling af naturlige sprog stiller store krav til denne beskrivelse, der skal være entydig, eksplicit og udtømmende.
Vi har taget udgangspunkt i materialet og typologien i "Dansk Sprogbrug" af Erik Bruun og sammenlignet opslag heri med korpuseksempler for samme nøgleord (hånd, hul, hjælp...). Arbejdet indebærer en systematisk bearbejdelse af belæggene med henblik på målsætningen: opstilling af en fintypologi og udarbejdelse af en videre beskrivelsesmodel for leksikalske og morfosyntaktiske egenskaber. Fordelen ved i dette arbejde at bruge et maskinlæsbart materiale af denne type er en (lang) række autentiske eksempler på søgeordene i nutidig sprogbrug og hermed et bredere sprogligt materiale ("corpus evidence") der kan danne grundlag for justering af leksikografisk beskrivelse af flerordsenheder. Materialet er desuden umiddelbart egnet til videre datamatisk behandling (lemmatisering, sortering, sammenligning...).

Til sidst bringes herunder yderligere et bidrag til emnet, der er indkommet efter mødet.

j) Korpus i praksis
af Christian Michelsen, forfatter til Dansk Idiomatisk Ordbog for Udlændinge og
Inge-Lise Dalager, ordbogsredaktør, Gyldendal

Vi har brugt DK87-90 i forbindelse med arbejdet på *Dansk Idiomatisk Ordbog for Udlændinge* (arbejdstitel) med det formål at:

a) finde illustrerende og mundrette eksempler til idiomerne,
b) indkredse definitionen af udtrykket så den bliver helt dækkende og præcis,
c) registrere konstruktionsvarianter,
d) vurdere bredden i vores udtryk.

På alle punkter har korpus kunnet hjælpe os, i nogen udstrækning. Vi kan fx tage et idiom som *at ligge på den lade side.* Her gav korpus os anledning til at udvide vores forklaring. Herudover fik vi oplysninger der tyder på, at udtrykket overvejende (måske kun) bliver brugt i benægtende form, og at det mest (måske

kun) bliver brugt i fornutid. Endelig leverede korpus stof til vores eksempel.

DK87-90 har dog nogle *begrænsninger,* både med hensyn til ***bredden*** og ***dybden.*** Til vores overraskelse har vi således ikke kunnet finde eksempler på ganske almindelige udtryk som *at snuse noget op, at sætte sig mellem to stole* og *at have mord i øjnene.* Hermed er også sagt, at korpus med den nuværende størrelse ikke er i stand til at fortælle noget om et udtryks hyppighed.

Vi synes, at de fiktive tekster med deres 50% er overrepræsenteret. Fiktion er en egen, indforstået verden hvor den digteriske frihed gør, at sproget bliver vendt og drejet efter forfatterens forgodtbefindende. En god ting for en underholdende roman, men ikke for en leksikograf der søger eksempler på en mere generel, dagligdags, sprogbrug.

Efter vores mening er der for få eksempler på den generelle, dagligdags sprogbrug. En udvidelse med det vi vil kalde "god journalistik" ville være en væsentlig styrkelse af korpus. Vi tænker på tekster som kronikker, anmeldelser, temaartikler og kulturjournalistik fra nogle af de største dagblade. Det kunne tilføre korpus et værdifuldt eksempelmateriale vedrørende moderne dansk som det bruges af en stor del af befolkningen.

Anmeldelser

Clausens Tekniske Ordbøger, Dansk-Engelsk, udarbejdet i samarbejde med Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København. Grafisk Forlag 1990. 379 s.

Bogen er led i en serie der omfatter dansk-engelsk, dansk-fransk og dansk-tysk samt den modsatte vej. Den er behagelig at bruge, typografien er klar og opstillingen overskuelig. Kun synes jeg godt man kunne have ofret ny linie ved begyndelsen af sammensætningsafsnittene, som fx ved **solfanger,** hvor den halvfede skrift i udtrykkene gør det lidt svært at finde hvor sammensætningerne begynder. Artiklerne er ligeledes klart opbygget med opdeling efter anvendelsesområder og i rækkefølge efter hyppighed: først dem med flere anvendelsesområder (med de helt generelle allerførst), derefter i alfabetisk rækkefølge enkeltanvendelser. Så vidt jeg som ikke-fagmand kan bedømme får man god besked, men jeg savner i nogle tilfælde hvor der gives en række ækvivalenter, som under **snæver** (der har 7), **rubank** (4), **regnmåler** (4+1 der er forklaret) besked om anvendelsen af de forskellige muligheder. Der anføres også amerikanske ækvivalenter, dog ikke de almindelige stavevarianter. Ækvivalenter som er pluralis hvor det danske er singularis er markeret, men der gives ellers ikke grammatiske oplysninger.

Bogen gør et solidt og tiltalende indtryk, og jeg har ved forskellige prøveopslag kun fundet enkelte ting som jeg vil sætte spørgsmålstegn ved: fx findes oversættelsen *shutter* både under **lukker** og **blænde** (her sammen med *diaphragm* og *stop),* **scenetekniker** oversættes med *property man,* som jeg mener snarere er en rekvisitør, oversættelsen til **lægtehammer** tror jeg ikke på (så vidt jeg har kunnet konstatere har de slet ikke sådan nogle i England) og under **rense** kunne *treat* også være medtaget om spildevand, da **spildevandsrensning** oversættes *sewage treatment;* en enkelt trykfejl er jeg faldet over: under **højttaler** står der *elektrodymanisk;* men det er kun småting som man vil kunne finde i enhver ordbog.

Om udvalget af opslagsord er rimeligt er det svært at afgøre for en ikke-fagmand. Det må under alle omstændigheder være en uhyre vanskelig opgave at bestemme hvad der skal med og hvad ikke i en teknisk ordbog af begrænset omfang som den foreliggende, der har ca. 30.000 opslagsord. Mulighederne er jo fuldstændig uoverskuelige. Forordet anfører det fornuftige princip at der er lagt vægt på at medtage det centrale ordforråd inden for de behandlede områder. En mere detaljeret og sagkyndig bedømmelse af udvalget findes i **HERMES**, 5-1990, ved Bernt C. Møller i en anmeldelse af den dansk-franske, og da det danske ordforråd i alt væsentligt er det samme (det anføres på klapteksten at ordbøgerne bliver til ud fra en fælles database) må den også have gyldighed for den dansk-engelske, så jeg henviser derfor til denne. Hans konklusion er, trods nogle indvendinger, at det er en gennemgående gedigen ordbog.

Et par mere generelle betragtninger, som jeg har gjort mig, skulle dog måske tages under overvejelse til en ny udgave: Der savnes i bogen en nærmere redegørelse for, hvad redaktørerne forstår ved "teknisk sprog", og også for, hvilke områder der er lagt særlig vægt på. Listen over emneforkortelser omfatter 129 emner, herunder nogle der må være marginale, som *musikinstrumenter, postvæsen, billedhuggerkunst, spil, sport, tourisme* (sic) og *teater,* og det ville være nyttigt at vide i hvilken udstrækning de er dækket. En anden ting der er problematisk er de opslagsord "af mere generel karakter" som ifølge forordet også er medtaget "i et vist omfang". Her må det ligeledes være uhyre svært at trække grænsen, fx er **lighed, ligne, lunken, papirspose, ødegård** med, og man spørger sig selv om det er rimeligt, eller om bogen hellere skulle be-

grænse sig til tekniske ord man ikke så let kan finde andre steder. Endelig ville oplysning om det kildemateriale bogen bygger på være nyttig. Forordet siger at "en del af ordbogsmaterialet er gennemset af fageksperter", men det røber intet om hvilken del eller hvilke fageksperter.

Bogen foreligger også i en elektronisk udgave der ligner den som også mine egne ordbøger findes i, og som har stort set de samme faciliteter: opslag, forekomstsøgning, kopiering til tekstbehandling etc. Jeg har prøvet den og den fungerer udmærket. Når man har den installeret er det jo faktisk hurtigere at slå op i den på skærmen end hvis man skal have fat i bogen. Det eneste jeg savner er muligheden for at søge på emneforkortelserne. Jeg ville godt have set hvad der fx er af ord mærket TEAT, men det kunne ikke lade sig gøre.

Jens Axelsen

LEDAs modtagne bøger pr. 1.12.1991

International Journal of Lexicography. Vol. 4, no. 2, 1991, Oxford University Press. Oxford, 1991.

International Journal of Lexicography. Vol. 4, no. 3, 1991, Oxford University Press. Oxford, 1991.

Gads store ordbøger, engelsk-dansk dansk-engelsk, G.E.C. Gad, København, 1991.

Gads ordbøger, spansk-dansk dansk-spansk, G.E.C. Gad, København, 1991.

Gads stribede ordbøger, spansk-dansk dansk-spansk, G.E.C. Gad, København, 1991.

Engelsk-dansk teknisk ordbog, København, 1991, udarbejdet i samarbejde med Fonden for Fagsproglig Leksikografi og udgivet af Grafisk Forlag A/S i samarbejde med Fonden for Fagsproglig Leksikografi. I serien Clausens tekniske ordbøger. 400 s.

Tysk-dansk teknisk ordbog, København, 1991, udarbejdet i samarbejde med Fonden for Fagsproglig Leksikografi og udgivet af Grafisk Forlag A/S i samarbejde med Fonden for Fagsproglig Leksikografi. I serien Clausens tekniske ordbøger. 432 s.

Henriksen, Torben: Juridisk ordbog spansk-dansk, G.E.C. Gad, København, 1991. 440 s.

Becker-Christensen, Christian: Politikens Forkortelsesordbog, Politikens Forlag, København 1991. 307 s.

Jarvad, Pia: Politikens Lille Danske Ordbog, Politikens Forlag, København 1991. 303 s.

Karker, Allan: Politikens Synonymordbog, 10. reviderede udgave, Politikens Forlag, København 1991. 355 s.

Jacobsen, Henrik Galberg og Jørgensen, Peter Stray: Håndbog i Nudansk, 2. reviderede udgave, Politikens Forlag, København 1991. 538 s.

Mohr, Ingeborg m.fl.: Danske bogstavpar, Munksgaards Forlag, København 1991. 495 s.

Meddelelser fra redaktionen

Dette nummer af Nyhedsbrevet er redigeret af Anna Braasch med venlig bistand af Frede Boje. En speciel tak til Anna Garde for den afsluttende opsætning og trykning af bladet.