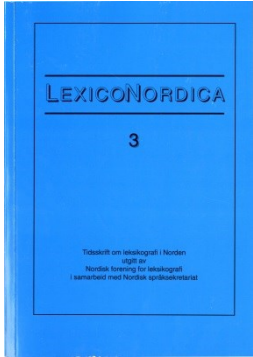


LexicoNordica

Titel:	Sammensætning af et tekstkorpus til en bilingval teknisk fagordbog og forslag til metodiske lemmaselektionskriterier	
Forfatter:	Jette Pedersen	
Kilde:	LexicoNordica 3, 1996, s. 139-152	
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.

Jette Pedersen

Sammensætning af et tekstkorpus til en bilingval teknisk fagordbog og forslag til metodiske lemmaselektionskriterier

On the basis of the inherent systematic structure of the subject field(s) in question, the criteria are laid down for the selection of texts for a monofunctional text corpus with a view to a bilingual technical dictionary. Besides, a number of recommendations are made concerning the identification of and selection criteria for associated subject fields and relevant aspects. The need for unambiguous, consistent selection criteria is emphasised, not only in connection with text selection, but also as regards the individual dictionary information categories, in the article exemplified by lemmata. Whereas the selection of certain types of lemmata which must be considered necessary for a technical dictionary to fulfil a number of encoding functions, including text production and L1–L2 translation, can only be made on the basis of a text corpus, the identification of actual LSP terms requires the participation of an expert. In either case, selection presupposes that the necessary selection criteria be established in advance.

1. Empirisk basis

At producere en fagordbog kan karakteriseres som én lang selektionsproces. For at færrest mulige valg baseres på ad hoc afgørelser og tilfældigheder, må der som et led i de prækonceptionelle overvejelser fastlægges et sæt entydige og konsekvente selektionskriterier for samtlige ordbogens informationskategorier.

Blandt fagleksikografens første valg er at afgøre, hvorfra den omtalte selektion skal foretages, eller sagt med andre ord: Hvad skal være ordbogens empiriske basis? Ser man på den fagleksikografiske praksis, finder man, at omteksten til tekniske fagordbøger sædvanligvis kun i meget vage vendinger oplyser om det materiale eller de forudsætninger, der har ligget til grund for ordbogen, jvf. fx følgende uddrag fra forteksten til Clausen da-eng hhv Chambers: "Det indsamlede materiale findes i en database" og "The software was a library database program called Inmagic which allowed all the material including additions to be indexed and" I begge citater henvises til et tilgrundliggende materiale, men uden at der redegøres nærmere for arten heraf. I andre tilfælde henvises til anonyme personer, der har ydet bistand ved ordbogens tilblivelse, jvf. fx forordet til Dictionary of Computing: "We

would like to thank many people who helped with this dictionary, by reading the text and making useful comments. The cover design is by Peter Cartwright." (sic!). I atter andre tilfælde, således fx Warrern og Skjerk Bilteknik, gives der ingen oplysninger om den empiriske basis overhovedet. Samtidig med at denne uklarhed mht eksisterende tekniske fagordbøgers empiriske basis forekommer at være reglen snarere end undtagelsen, findes der dog også eksempler på en mere reflekteret og informativ praksis. Således indeholder forteksten til Genteknologisk Ordbog en liste over den litteratur, der er blevet benyttet ved udarbejdelsen af ordbogen, mens der i forteksten til Kuçera eng-ty findes en fortegnelse over en række navngivne fageksperter, der har medvirket ved ordbogens tilblivelse, med angivelse af de nævnte personers arbejdsplads samt hvilke fagområder, de hver især repræsenterer.

Der kan på basis af heraf umiddelbart identificeres to muligheder for empirisk basis for tekniske fagordbøger: For det første det i Clausen refererede og i Genteknologisk Ordbog nærmere specificerede materiale i form af **tekstsamlinger**. For det andet kan der inddrages en eller flere **fageksperter**, således som det har været tilfældet i Kuçera, og hvortil vel også "many people" i Dictionary of Computing refererer. Herudover kan man forestille sig, at fagleksikografen har forladt sig på sin egen faglige og sproglige kompetence, dvs. **introspektion**. Endelig kan leksikografen have støttet sig til **belægsamlinger** eller **eksisterende beskrivelser**, herunder allerede foreliggende ordbøger og leksika eller hånd- og fagbøger. Disse muligheder er der dog ingen blandt de 25 undersøgte tekniske fagordbøger, der vedkender sig.

Det forekommer næppe realistisk at antage, at en teknisk fagordbog, hvis intenderede brugere uundgåeligt må omfatte tekstproducenter og oversættere i erhvervslivet, skulle blive til uden hensyntagen til de tekster, der faktisk produceres her. I modsat fald ville det fx næppe være muligt at identificere de forskellige problemtyper, som ordbogen skal medvirke til at løse. Specielt når det drejer sig om tekniske del- og enkeltfagsordbøger kommer hertil, at disse erfaringsmæssigt i stigende omfang allerede i deres udgangspunkt har en mere eller mindre eksplicit tilknytning til en konkret branche, virksomhed, organisation el. lign. – om ikke andet så af økonomiske årsager: De er dyre at producere, og fagordbøger nyder som bekendt ikke offentlig støtte. Det følger derfor, at tekster herfra må indgå som en nødvendig del af den pågældende ordbogs empiriske basis.

I det omfang de omhandler det aktuelle fagområde må alle tekster fra den pågældende branche, virksomhed m.v. betegnes som tematisk relevante. Af hensyn til oplysningernes pålidelighed bør der dog an-

lægges en teksttypologisk indfaldsvinkel, idet selektion bør begrænses til terminologisk og indholdsmæssigt præcist tekstmateriale. Dette indebærer, at reklamer i form af prospekter o.l. udelades, idet denne teksttypes primære formål netop ikke er at oplyse, men derimod at overtale. Den bevidste nedprioritering af det faglige indhold, som ligger i naturlig forlængelse af reklamers direkte kommunikationsfunktion, bevirker, at pålideligheden af denne teksttype må anses for usikker. Af samme grund udelades ligeledes tekstmateriale, der må betegnes som terminologisk tentativt. Der tænkes her specielt på de grundlæggende tekniske specifikationer, som udarbejdes på et meget tidligt stadium af produktudviklingsfasen, nemlig i forbindelse med forskning og udvikling. Som eksempel på teksttyper, der er typiske kandidater til selektion, kan nævnes de forskellige former for produktinformation, monterings-, betjenings-, og driftsvejledninger, standarder og servicemanualer.

Tekniske fagordbøger med en eksplicit tilknytning til en konkret branche eller virksomhed adskiller sig især fra rene del- og enkeltfagsordbøger ved at have en mere kompleks indgang til fagområdet. Ud over et prioriteret del- eller enkeltfag vil der typisk være behov for at medtage informationer fra andre fagområder, der har tilknytning hertil, fx ved at være omfattet af produktprogrammet. I forbindelse med et konkret virksomhedsspecifikt pumpeteknologisk ordbogsprojekt er der eksempelvis identificeret to sådanne associerede områder, idet den pågældende pumpeproducent fremstiller sine egne elmotorer og elektroniske styringer. Til forskel fra pumper produceres de pågældende elmotorer og elektroniske styringer imidlertid ikke som et mål i sig selv, men derimod alene med henblik på at indgå i det prioriterede fags produkter. Dette afspejler sig i virksomhedens tekstproduktion, hvor langt den største del af de informationer, der gives om elmotorer og elektronik, skal findes i litteratur om det prioriterede fagområde. Samtidig kan associerede fagområder give anledning til et vist omfang af selvstændig tekstproduktion. For at sikre at ordbogen vil kunne medvirke til at løse produktions- og receptionsproblemer vedr. hele produktprogrammet, brancheområdet m.v. må også disse tekster medtages. Men i modsætning til hvad der er tilfældet med ordbogens prioriterede fagområde, intenderes der for associerede områder ikke systematisk tematisk dækning, jvf. nedenfor.

Langt størstedelen af tilgængelig virksomheds- hhv. brancheintern litteratur på L2 vil typisk udgøres af relætekster, dvs. "tekster, der enten i sig selv er oversættelser, som kan være lavet af folk, der har målsproget eller kildesproget som modersmål, eller være skrevet som originaltekster af folk, der ikke har målsproget som modersmål."

(Krogh-Hansen 1994:123). De kan dermed ikke frasorteres, men må i korpus gøres identificerbare som sådanne, fx vha en særlig kodering eller ved at samles i et særligt delkorpus. For at sikre at korpus kan levere pålidelige sproglige, herunder ikke mindst grammatiske og kollokationelle, oplysninger på L2, er der altså behov for et betydeligt bredere tekstfundament end det materiale, branchen hhv virksomheden kan levere. Dette behov understøttes yderligere af, at der ved konciperingen af tekniske fagordbøger ofte forudses en bredere brugergruppe end den konkrete branche eller virksomhed.

Et muligt udgangspunkt for opstilling af bibliografi over offentligt tilgængelig litteratur omhandlende det aktuelle fagområde kan eksempelvis være bibliotekernes universelle decimalklassifikation, UDK-systemet, eller eventuelt allerede eksisterende specialbibliografier. I forbindelse med udarbejdelse af bibliografi til den danske del af et dansk-engelsk-fransk tekstkorpus inden for aftaleretten har man således taget udgangspunkt både i UDK-systemet og i en foreliggende emneinddelt dansk juridisk bibliografi (Dyrberg et al. 1991). Hvor der ikke findes sådanne bibliografier, eller hvor UDK-systemet viser sig at være uegnet, fx ved ikke at være tilstrækkeligt finmasket eller ved at være for upræcis i sin indfaldsvinkel, kan det fornødne overblik over grundmængden af tekster etableres vha on-line søgning på overordnede søgetermer i relevante fagbibliotekers databaser. I forbindelse med det her eksemplificerede fagområde er der således foretaget trunkeret on-line søgning i de tekniske bibliotekers databaser på alle sammensætninger med */pump/* som søgeterm.

Før selektion af tekster fra den opstillede bibliografi over de fremkomne titler kan påbegyndes, må der søges etableret et systematisk overblik over de(t) pågældende fagområde(r). Formålet hermed er dels at sikre indfangning af tematisk relevant tekstmateriale, dels at afgrænse det pågældende fagområde i forhold til tilgrænsende områder og dermed muliggøre frasortering af fagområde-eksternt materiale. Faget, således som det afspejler sig i opbygningen af og dermed i indholdsfortegnelsen til de fremkomne titler, kan her tjene som udgangspunkt for udarbejdelse af en egentlig fagsystematik, men det understreges, at den endelige udformning må ske i samråd med en fag-ekspert. Ved den foreløbige udformning koncentrerer man sig i første omgang om de mere generelle værker. Er ordbogens emneområde hhv ét af ordbogens emneområder som i det aktuelle tilfælde pumpe-teknologi, vil det ikke levere de fornødne oplysninger at søge hjælp i værker som *Centrifugal and Axial Flow Pumps*, der, som det fremgår af titlen, alene behandler et eller flere delfagsområder inden for enkeltfaget og derfor ikke, eller i hvert fald kun utilstrækkeligt, kan

antages at foretage den nødvendige afgrænsning i forhold til tilgrænsende områder. Dette gælder ikke blot, hvor der er behov for at afdække et fagområde i sin helhed, fx hvor der er tale om en flerfagsordbog, der blandt sine emneområder medtager et bestemt enkeltfag, eller hvor ordbogen udelukkende intenderer at dække det pågældende enkeltfag. Det gælder tillige for delfagsordbøger, også selv om det pågældende delfag netop er det, der behandles i den pågældende titel. Med henblik på selektion af tekstmateriale til en delfagsordbog er der altså i lige så høj grad som ved enkelt- og flerfagsordbøger behov for at etablere et overblik over et helt enkeltfag, idet det ellers ikke vil være muligt med tilstrækkelig stor sikkerhed at sortere delfags-eksterne værker eller dele af værker fra, samtidig med at det sikres, at det pågældende delfag dækkes og afdækkes i hele dets udstrækning. På dette stadium af arbejdsprocessen må man altså som sagt søge hjælp i bredere anlagte værker. På den anden side må der ikke være tale om for brede værker, idet bredden oftest vil være på bekostning af dybden. Værker, der indeholder en samlet behandling af et enkeltfag, må erfaringsmæssigt anses for de mest velegnede.

For at sikre at den empiriske basis også er til stede for den leksikografiske behandling af de forskellige aspekter, der falder ind under det pågældende fagområde, kan det anbefales at udarbejde en aspektfortegnelse. Potentielle aspekter kan fx identificeres med udgangspunkt i de relevante titlers indeks. I det konkrete eksempel med pumper fremkom herved så forskelligartede aspekter som funktion /arbejdsprincip; anvendelsesmuligheder, fx afvanding, vanding og varmeanlæg; komponenter; materialer, herunder bearbejdning og sammenføjning af materialer; korrosion; grundlæggende elteknik; elektrisk montage; sensorer; pumpemedier; væskestrømning; akseltætninger; koblinger; ventiler og armaturer; rørledninger; måleteknik; installation; vedligeholdelse; normer; diverse pumpebegreber, som fx trykstød i rørledninger, kavitation, kapacitet, løftehøjde og pumpetryk. I bi- og polylingval sammenhæng må de fagområderelevante dele af fagsystematikken samt de forskellige elementer i aspektfortegnelsen være repræsenteret i begge hhv alle delkorpora.

Specielt i forbindelse med selektion af offentligt tilgængeligt tekstmateriale kan det desuden blive nødvendigt at anlægge et aktualitetskriterium. Således at den empiriske basis også kommer til at omfatte den seneste udvikling på området, medtages ud over fag-, hånd- og lærebøger også tematisk relevante tidsskriftartikler, idet udgivelsestidspunktet for de øvrige værker vil være afgørende for, hvor langt det må anses for påkrævet at gå tilbage.

I forbindelse med oprettelsen af et dansk og et tysk genteknologisk tekstkorpus er der tilstræbt kvantitativt ligelig repræsentation af tekstord inden for hver af de fire kommunikationsretninger fagekspert -> fagekspert / fagekspert -> lægmand / lægmand -> fagekspert / lægmand -> lægmand (Lauridsen et al. 1989:10). Samtidig har man intenderet et polyfunktionalt tekstkorpus, idet korpus er sammensat på en sådan måde, at det skal kunne tjene som udgangspunkt for "diskursanalyser, såvel som grammatisk-syntaktiske, leksikografiske samt terminologiske undersøgelser" (Lauridsen et al. 1989:4). Repræsentation af tekster med lægfolk på afsendersiden har efterfølgende vist sig problematisk i forbindelse med den leksikografiske anvendelse af det danske genteknologiske korpus, nemlig som empirisk basis for udarbejdelsen af Genteknologisk Ordbog, idet den fagkompetente del af det leksikografiske team har konstateret, at denne teksttype indeholdt en stor risiko for fejlinformation. Det kan ud fra denne erfaring konkluderes, at hvor et korpus skal tjene som værktøj i en faglig sammenhæng, må det også opbygges på et korrekt fagligt grundlag (Kaufmann 1992:67). I modsat fald vil tekster af typen lægmand -> fagekspert og lægmand -> lægmand, kræve en omfattende selektiv indsats, hvis korpus også, således som det var intenderet, skal kunne anvendes til leksikografiske formål. "Med mindre en ekspert (molekylærbiolog) tages med på råd ved en sådan nødvendig frasortering af materiale, der i faglig henseende er mindre lødigt, vil der kunne drages mange forkerte og modstridende konklusioner ud fra det foreliggende materiale." (Stummann 1992:70-71).

Ovenstående tematiske og teksttypologiske overvejelser resulterede for det her eksemplificerede fagområde, inklusive aspekter og associerede områder, i et korpus i størrelsesordenen 1.390.000 løbende tekstord, fordelt med 900.000 ord på L2 og 490.000 på L1. Den skæve fordeling mellem L1 og L2 skyldes dels, at anvendeligheden af langt størstedelen af de virksomhedsinterne L2 tekster er begrænset, ikke mindst i forbindelse med ækvivalentkontrol og selektion af sproglige oplysninger, idet der er tale om relætekster, dels at tekstproduktion på og oversættelse til modersmålet ikke er intenderede brugsfunktioner.

2. Lemmaselektion

For at afdække hvilke lemmaselektionskriterier, der har været anvendt i eksisterende tekniske ordbøger, har jeg undersøgt omteksterne til 25 mono- og bilingvale tekniske fagordbøger. Heraf giver tolv ingen oplysninger overhovedet, mens lemmaselektion i de resterende tretten

kun omtales i meget løse vendinger. I en tilsvarende undersøgelse af ni juridiske fagordbøger når Nielsen 1989:59 frem til et lignende resultat: "Det er karakteristisk for alle de undersøgte ordbøger, at de enten ingen oplysninger indeholder om selektionen af lemmata eller, at de kun indeholder korte, vage oplysninger herom." De sparsomme oplysninger står i stærk kontrast til den interesse, emnet vises i ordbogsanmeldelser, jvf. følgende uddrag af anmeldelser af fagordbøger i tidsskriftet *Hermes*, hvor de to første omhandler flerfagsordbøger, mens de sidste fire er anmeldelser af enkeltfagsordbøger:

Møller (1990:148): "Selv om en række centrale termer er medtaget, mangler mindst lige så mange, ikke mindst af moderne karakter."

Riiber/Thomsen (1991:141): "Den konklusion, man må drage på grundlag af disse og andre tilsvarende stikprøver, må være, at DTTO nok har taget munden lidt for fuld ved at påstå, at 'det centrale ordforråd' er medtaget. Dertil er der for store mangler i systematikken. Når hullerne i systematikken snarere synes at være reglen end undtagelsen, kunne man let få den tanke, at udvælgelsen af termer er noget tilfældig, eller måske styret af et hyppighedskriterium, hvilket ikke bør være tilfældet for en fagordbog."

Duvå (1991:148 og 152): "Hvad laver ganske almindelige ord som fx ... *fedte for, mulighed, på grund af* i en ordbog for speciel fagterminologi? ... Er de nyeste fagudtryk medtaget? Forgæves leder jeg efter: ... *børsetik, deregulering* Den manglende lemmatisering gælder dog ikke kun nye udtryk, men desværre gode, gamle, velkendte fagudtryk: *bygge- og anlægsarbejde* ...".

Laursen/Tarp (1991:156): "Men hvorfor kan man f.eks. finde ord som *afladedokumenter, eksport* ..., når helt relevante og hyppigt brugte ord som *anløbshavn, bortfragter, fortolde* ... ikke er medtaget?"

Duvå/Nielsen (1992:117): "Vi kan finde termer, der er relateret til organisation, fx *stabschef* ..., men hvorfor ikke termer som fx *åben-dørpolitik, quality management* ..., som er nogle af nøgleterminerne indenfor moderne management og ledelsesstrategi? Til gengæld kan vi finde adskillige termer inden for regnskabsområdet ... Det kan være svært at se, hvorfor de hører hjemme i en marketingordbog."

Barlach et al. (1992:196): "I denne sammenhæng skal opremses følgende manglende udtryk i de to stikordslister: *økonomisk kriminalitet, criminal code, criminal law* ...".

Ser man på de tretten tekniske ordbøger, hvis omtekst kommer ind på lemmaselektion, omend det altså kun er summarisk, hedder det eksempelvis i *Elektroteknisk Ordbog*: "... indeholder i alfabetisk orden samtlige de i følgende danske standarder indeholdte termer: ..." og *Webster's New World Computer Ordbog*: "Formålet med denne bog har

været at samle så mange som muligt af disse [= computer-terminologiske] ord." I modsætning hertil intenderes i Skibsted en mindre dybtgående dækning: "Bogen omfatter den væsentligste terminologi inden for M, E, P og B retningerne." Det samme er tilfældet med rækken af bilingvale tekniske flerfagsordbøger fra Clausen: "Der er ved udvælgelsen af opslagsordene lagt vægt på at medtage det centrale ordforråd inden for de behandlede emneområder." og L&H: "[Ordbogen] repræsenterer et systematiseret ekstrakt af en translatør-udviklet database." Til sammenligning hedder det i McGraw-Hill: "... the vocabulary of science and technology was surveyed by a select group of consulting editors. The terms ... included in the Dictionary are those considered important in each consultant's particular area of specialisation."

Om den tilgrundliggende empiriske basis for selektion af lemmata kan det af ovenstående citater fra tekniske fagordbøger udledes, at denne har bestået dels af tekstsamlinger (Clausen og L&H), dels af introspektion, her i form af en fagkompetent vurdering (McGraw-Hill). Bortset fra belægsamlinger, som i denne sammenhæng må anses for uegnede, eksisterer for selektion af lemmata principielt de samme muligheder for empirisk basis som for selektion af ordbogens øvrige informationskategorier. Specielt for ordbøger med tilknytning til en konkret branche eller virksomhed gælder det, at selektion må tage særlige hensyn til de tekster, der produceres i den pågældende branche eller virksomhed. Når det drejer sig om del- og enkeltfagsordbøger, vil eksisterende fagordbøger alene kunne have en supplerende funktion, og kun i det omfang, der allerede måtte foreligge en del- eller enkeltfagsordbog for det pågældende fagområde, idet flerfagsordbøger ikke kan anses for tilstrækkeligt dybtgående til at sikre dækning af et fags termer.

Yderligere kan der på basis af citaterne identificeres to grundholdninger med hensyn til dækningen af et konkret fagområde, således som den afspejler sig i ordbogens lemmabestand. Det drejer sig om maksimerende lemmaselektion, som fx i Elektroteknisk Ordbog og Webster's New World Computer Ordbog, henholdsvis minimerende lemmaselektion, jvf. fx Skibsted og Clausen. En maksimerende ordbog er altså en ordbog, der intenderer at dække den størst mulige del af et bestemt fags ordforråd. Der er her tale om et yderst arbejds- og pladskrævende ordbogsprojekt, og som det var tilfældet med de to ovenfor citerede ordbøger, der i henhold til deres omtækt intenderede maksimerende lemmaselektion, vil maksimerende ordbøger da også typisk begrænse sig til et enkelt fagområde. Når bortses fra enkelte, meget omfattende flerbindsværker, er flerfagsordbøger i langt de fleste til-

fælde minimerende, idet de kun intenderer at medtage det centrale ordforråd inden for hvert fagområde, jvf. ovenstående citater fra omteksterne til Skibsted, Clausen og McGraw-Hill.

Som det fremgik af ovenstående uddrag fra fagordbogsanmeldelser, kan en ordbogs begrænsning til et enkelt fagområde omvendt ikke ses som nogen garanti for, at den pågældende ordbog også er maksimerende, dvs. dækkende i forhold til fagets ordforråd. Samtidig konstateredes i samtlige anmeldelser en række eksempler på lemmata, som måtte anses for irrelevante for det pågældende fagområde. Endelig fandtes der eksempler på, at det ikke var lykkedes minimerende fagordbøger at identificere det centrale ordforråd, således som det fx var intenderet i flerfagsordbogen fra *Clausen*, hvor anmelderne antyder, at lemmaselektion har været styret af et frekvenskriterium. Med et konkordansprogram lagt ind over et maskinlæsbart tekstkorpus er det en simpel sag at generere et alfabetisk indeks over samtlige de i tekstkorpus forekommende ortografiske ord med angivelse af den frekvens, hvormed de forekommer i korpus. Med udgangspunkt i en sådan alfabetisk frekvensliste kan der herefter anlægges et rent objektivt baseret frekvenskriterium for selektion af lemmata. I den polylingvale Grundwortschatz Naturwissenschaften, der intenderer at dække fagområderne kemi, matematik og fysik, har man koblet frekvens med et tekstspredningskriterium, hvilket i ordbogens engelsk-tyske del har resulteret i følgende komplette liste over lemmata under bogstavet *w*: *wall, want, wash, water, wave, wavelength, way, we, weak, weigh, weight, well, well-known, what, whatever, wheel, when, whence, whenever, where, whereas, whereby, whether, which, whichever, while, white, who, whole, whose, wide, widely, width, will, wire, wish, with, withdraw, within, without, word, work, worker, world, worth, would*. Det er her om muligt næsten mere end umuligt at forestille sig, hvilke brugere der i forbindelse med hvilke brugssituationer vil kunne drage nytte af denne type lemmata – de intenderede fagområder taget i betragtning.

En løsning på de ovenfor eksemplificerede lemmatiseringsproblemer forudsætter en teoretisk afklaring af de for lemmaselektion tilgrundliggende kriterier. Dette indebærer konkret, at der som første skridt opstilles et sæt kriterier for identifikation af det prioriterede fagområdes termer. Dette gælder, uanset om der er tale om en maksimerende del- eller enkeltfagsordbog eller en minimerende flerfagsordbog. I førstnævnte tilfælde vil nemlig alle fagtermer skulle identificeres med henblik på selektion, mens det i sidstnævnte tilfælde ikke vil være muligt at selekttere det centrale ordforråd, medmindre man først har etableret et overblik over hele ordforrådet, jvf. fx diskussionen af

Clausen ovenfor. Der foreslås med henblik herpå udarbejdet systematiserede oversigter over det pågældende områdes fagterminologi. Bortset fra at dette arbejde må involvere et tæt samarbejde mellem leksikograf og fagekspert (hvor der ikke er personsammenfald mellem de to), er det næppe muligt at opstille et sæt universelle retningslinier for en terminologisk strukturering af samtlige teknikkens emneområder. Der kan hentes inspiration i lære- og andre fagbøger samt i virksomhedsinternt materiale, hvor tekster af typen eksploderede billeder og snitmodeller med ledsagende tekst må anses for særligt velegnede, når det drejer sig om etablering af et overblik. På baggrund heraf foreslås der udarbejdet såkaldte termsystematikker med udgangspunkt i de forskellige produkttyper, som igen kan tilordnes en systemiseret, funktionelt baseret fagsystematik (Pedersen 1995a:83–90 og Pedersen 1994:203–227).

Mens en fagordbog, i hvilken lemmabestanden er begrænset til rene fagtermer, ofte vil være tilstrækkelig til receptionsformål, anses en sådan rent terminologisk ordbog ikke at kunne yde den fornødne hjælp i forbindelse med L2 tekstproduktion og L1-L2 oversættelse. I modsætning til hvad der gør sig gældende for terminologiske ordbøger, forudses der altså hvor det drejer sig om andre fagordbøger end rene receptionsordbøger at være behov for en betydeligt bredere lemmaforståelse end den, der begrænser sig til egentlige fagtermer. Branche- og virksomhedsspecifikke tekniske fagordbøger medtager som nævnt også emneområder, som ligger uden for det prioriterede fagområde, men som fx er karakteriseret ved at være omfattet af produktprogrammet. Dette indebærer, at der også må selekteres fagudtryk herfra i det omfang, disse indgår i branchens hhv virksomhedens tekstproduktion. I modsætning til hvad der er tilfældet med det prioriterede fagområdes termer er der altså for associerede områder ikke behov for systemiseret terminologisk dækning. Yderligere kan der være behov for at selektere fagudtryk fra en række mere perifere emneområder, som ikke direkte er omfattet af produktprogrammet, men fx relaterer sig til de for fremstillingen heraf tilgrundliggende materialer, processer m.v. I lighed med hvad der var tilfældet for associerede områder, vil der også for perifere fagområder alene være behov for tekstuel betinget selektion. Det forudsættes desuden, at de pågældende udtryk optræder i de tilgrundliggende tekster med en vis fastlagt minimumsfrekvens og tekstspredning. Tallene herfor ansættes i forhold til den samlede størrelse af tekstkorpus. Der er dermed for associerede hhv perifere områder tale om varierende grader af minimerende, tekstbaseret lemmaselektion, idet der for selektion af fagudtryk fra perifere områder

er indlagt yderligere to kriterier. De ovenfor foreslåede selektionskriterier kan herefter opsummeres som følger:

- (1) fagtermer i henhold til termsystematikker for det prioriterede fagområde,
- (2) fagudtryk fra associerede fagområder, der indgår i branchens hhv virksomhedens tekstproduktion,
- (3) fagudtryk fra perifere fagområder, der optræder i tekster om det prioriterede fagområde med en vis minimumsfrekvens og med en vis tekstspreddning.

Ovenstående kriterium (1), der vedrører maksimerende selektion af termer inden for ordbogens prioriterede område, tager alene udgangspunkt i selve faget, nærmere betegnet i viden herom. Mens metoden således sikrer en terminologisk dækkende ordbog, er den ingen garanti for, at de pågældende fagtermer faktisk optræder i de tekster, der produceres inden for faget. Omvendt er der i forbindelse med de minimerende, tekstbaserede fremgangsmåder (2) og (3) ingen garanti for, at der overhovedet vil blive selekteret deciderede fagtermer fra de pågældende områder.

Herudover må der tages højde for de aspekter, der optræder i den i forbindelse med oprettelse af tekstkorpus udarbejdede aspektfortegnelse, idet der selekteres

- (4) fagudtryk, der må anses for særligt vigtige eller relevante for det prioriterede delfagsområde.

For at en fagordbog skal kunne gøre fyldest i forbindelse med brugsfunktionerne tekstproduktion på og oversættelse til fremmedsproget foreslås tillige selekteret

- (5) øvrige leksikalske enheder, der i forbindelse med tekstproduktion på og oversættelse til fremmedsproget frembyder særlige grammatiske eller kollokationelle problemer.

Også i forbindelse med kriterium (5) er der tale om tekstbaseret selektion, idet identifikation af denne problemtype kræver medvirken af en fagsprogskyndig lingvist. Vedrørende selektion og præsentation af grammatiske og kollokationelle oplysninger i tekniske fagordbøger henvises til Pedersen (1995b:121–142 hhv 1995c:60–73).

3. Afsluttende bemærkninger

Rene terminologiske ordbøger er som nævnt bl.a. karakteriseret ved at begrænse sig til egentlige fagtermer, foruden ved en mikrostruktur, som er fattig på sproglige, men rig på faglige oplysninger. Denne ordbogstypes primære brugsværdi er dermed i forbindelse med tekstreception og som et supplement til egentlige fagbøger. Til forskel herfra intenderer specielt bi- og polylingvale fagleksikografiske produkter typisk at yde hjælp i forbindelse med tekstproduktion på og oversættelse til fremmedsproget. For at kunne opfylde disse funktioner, er det bl.a. påkrævet med en lemmapforståelse, der er ganske betydeligt bredere end den, der begrænser sig til vidensbaseret selektion af egentlige fagtermer. Dette gælder ikke mindst, når vi taler om den hastigt voksende del af fagleksikografien, som tager udgangspunkt i eller orienterer sig mod en konkret branche eller virksomhed. Alene relevant tekstmateriale vil kunne levere de fornødne oplysninger til hjælp for selektion af lemmata efter kriterierne (2), (3) og (5), samt i vid udstrækning også (4), ovenfor, kontrol af ækvivalenter til disse lemmata samt de for L2 sprogproduktion og L1-L2 oversættelse nødvendige grammatiske og kollokationelle angivelser. Det er altså her af største vigtighed, at fagleksikografen har et monofunktionalt tekstkorpus til sin rådighed, dvs. et korpus som er sammensat med dette specielle formål for øje og efter nøje fastlagte tematiske, teksttypologiske og øvrige kriterier, således som det er beskrevet ovenfor. For at sikre at korpus opbygges på et korrekt fagligt grundlag, skal igen fremhæves nødvendigheden af en fagkompetent vurdering allerede på tekstselektionsstadiet af den leksikografiske proces.

Lige så uomgængeligt nødvendigt det er med kriterier for selektion af tekster **til** korpus, lige så påkrævet er det at fastlægge kriterier for selektion af informationer **fra** korpus. Sidstnævnte er her eksemplificeret ved selektion af lemmata men må i videst muligt omfang forhåndsfastlægges for samtlige ordbogens informationskategorier, ikke blot under hensyntagen til den pågældende ordbogs fagområde(r) men også med henblik på ordbogens forudsete brugere og brugsfunktioner samt de problemtyper, ordbogen skal medvirke til at løse.

Litteratur

- Barlach, Richard/Henning Bergenholtz/Wilhelm Gubba 1992: [Anmeldelse af] Steen Hjelmblink: Retsplejeordbog. København: Munksgaard 1991. I: *Hermes* 9, 193–196.
- Stepanoff, A. J. 1957: *Centrifugal and Axial Flow Pumps. Theory, Design, and Application*. New York etc.: John Wiley & Sons.

- Chambers 1988 = Walker, Peter M.B. (ed.): *Chambers Science and Technology Dictionary*. Edinburgh/Cambridge: W. & R. Chambers and Cambridge University Press.
- Clausen da-eng 1990 = *Dansk-engelsk teknisk ordbog*. København: Grafisk Forlag.
- Dictionary of Computing 1988: *Dictionary of Computing*. Middlesex: Peter Collin Publishing.
- Duvå, Grete 1991: [Anmeldelse af] Orlando Gorlindo: Erhvervs- og økonomiordbog – Dictionnaire économique et de l'entreprise. Dansk-Fransk. Danois-Français. GTO Forlag 1990. I: *Hermes* 7, 147–152.
- Duvå, Grete/Anne Ellerup Nielsen 1992: [Anmeldelse af] Anne Møllmann: Dansk-fransk marketing ordbog. Foreningen til Unge Handelsmænds Uddannelse 1990. I: Handelshøjskolen i Århus: *Hermes* 8, 111–126.
- Dyrberg, Gunhild/Dorrit Faber/Steffen Leo Hansen/Joan Tournay 1991: *Oprettelse af fagsproglige tekstkorpora. Engelsk-fransk-dansk juridisk sprog aftaleret*. Handelshøjskolen i København: (= ARK 60)
- Elektroteknisk Ordbog 1983 = *Elektroteknisk Ordbog. Engelsk-dansk. Dansk-engelsk*. Udarbejdet og vedtaget af Dansk Elektroteknisk Komité. København: Dansk Standardiseringsråd.
- Genteknologisk Ordbog 1992 = Kaufmann, Uwe og Bergenholtz, Henning: *Genteknologisk Ordbog. Dansk-engelsk/engelsk-dansk*. København: Gad.
- Grundwortschatz Naturwissenschaften 1988 = Hoffmann, Lothar: *Grundwortschatz Naturwissenschaften. Russisch-Deutsch Englisch-Deutsch Französisch-Deutsch*. Leipzig: VEB Verlag Enzyklopädie.
- Kaufmann, Uwe 1992: Anvendelse af det danske genteknologiske tekstkorpus ved udarbejdelsen af Genteknologisk ordbog, med specielt henblik på udvælgelsen af eksempler. I: Gert Engel (ed): *Korpuslingvistik i Fagsprogsforskningen*. Hindsgavl Slot 26. og 27. november. [Kolding: Handelshøjskole Syd], 56–68.
- Krogh-Hansen, Niels 1994: Oversættelse af fagsproglige tekster. I: Hjørnager Pedersen, Viggo/Niels Krogh-Hansen (eds): *Oversættelsehåndbogen*. København: Munksgaard.
- Kucera eng-ty 1989 = Kucera, A.: *The compact dictionary of exact science and technology*. Volume I: English-German. Wiesbaden: Brandstetter Verlag.
- L & H 1990 = Langkilde, Charlotte et al: *L & H Teknisk Ordbog dansk-engelsk /engelsk-dansk*. L & H Ordbøger.

- Lauridsen, Ole/Ingrid Skytte Andersen 1989: *Rapport over oprettelsen af et dansk og et tysk maskinlæsbart tekstkorpus inden for området genteknik*. København: Handelshøjskolen i København (= ARK 66)
- Laursen, Anna-Lise/Sven Tarp 1991: [Anmeldelse af] Orlando Galindo: Dansk-spansk fagordbog (teknisk forslag a/s 1979) og Dansk-spansk erhvervs- og økonomiordbog (GTO Forlag 1989). I: *Hermes* 7, 153–162.
- McGraw-Hill 1974 = *McGraw-Hill Dictionary of Scientific and Technical Terms*. Editor in Chief: Lapedes, Daniel N. New York etc. : Mc-Graw-Hill.
- Møller, Bernt 1990: [Anmeldelse af] Dansk-fransk teknisk ordbog. Fonden for Fagsproglig Leksikografi. Handelshøjskolen i København og Aschehoug Dansk Forlag A/S 1989. I: *Hermes* 5, 143–152.
- Nielsen, Sandro 1989: Kritisk oversigt over engelske og danske juridiske ordbøger. I: Handelshøjskolen i Århus: *Hermes* 2, 55–75.
- Pedersen, Jette 1994: Anskueliggørelse af faglige relationer i brancheordbøger. I: *LexicoNordica* 1, 203–228.
- Pedersen, Jette 1995a: Systematic classification. I: Henning Bergenholtz/Sven Tarp (eds): *Manual of Specialised Lexicography*. Amsterdam: Benjamins, 83–90.
- Pedersen, Jette: Grammatiske oplysninger i tekniske fagordbøger. I: *LexicoNordica* 2, 1995b:121–142.
- Pedersen, Jette 1995c: The Identification and Selection of Collocations in Technical Dictionaries. I: *Lexicographica* 11, 60–73.
- Riiber, Theis/K.T. Thomsen 1991: [Anmeldelse af] Dansk-tysk teknisk ordbog. Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København og Grafisk Forlag A/S, København 1990. I: *Hermes* 7, 139–145.
- Skibsted 1971 = Skibsted, Sigurd: *Teknisk engelsk-dansk ordbog*. København: Polyteknisk Forlag.
- Skjerk Bilteknik 1991 = Skjerk, Ebbe: *Bilteknisk Ordbog, engelsk-dansk*. København: Teknisk Forlag.
- Stummann, Bjarne 1992: Anvendelsesmuligheder og fagligt indhold af det danske genteknologiske tekstkorpus. I: Gert Engel (ed): *Korpuslingvistik i Fagsprogsforskningen*. Hindsgavl Slot 26. og 27. november [Kolding: Handelshøjskole Syd], 69–74.
- Warrern da-eng 1988 = Warrern, Allan: *Dansk-engelsk teknisk ordbog*. København: Clausen.
- Webster 1984 = *Webster's New World Computer Ordbog*. Frederiksund: Thorsgaard.