

NORDISKE STUDIER I LEKSIKOGRAFI

3

Rapport fra Konferanse om leksikografi i Norden
Reykjavík 7.–10. juni 1995

Redigert av

Ásta Svavarsdóttir

Guðrún Kvaran

Jón Hilmar Jónsson

Skrifter utgitt av Nordisk forening for leksikografi
Skrift nr. 3

I samarbeid med Nordisk språksekretariat og Orðabók Háskólans

Reykjavík 1995

© Nordisk forening for leksikografi. 1995.

Prentun: Steindórsprent Gutenberg

ISBN 9979-60-202-3

Rapporten kan bestilles hos:

Nordisk Språksekretariat
Postboks 8107 Dep
N-0032 Oslo

Forord

Den tredje Konferanse om leksikografi i Norden, arrangert i regi av Nordisk forening for leksikografi, ble holdt i Reykjavík dagene 7.–10. juni 1995. Det praktiske arrangementet og det forberedende arbeid til konferansen ble utført av Orðabók Háskólans (Leksikografisk institutt ved Islands universitet), Orðmennt (Islandsk forening for leksikografi) og Nordisk språksekretariat. Den lokale arrangementskomitéen bestod av Ásta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson og Sigurður Jónsson. Rikke Hauge frá Nordisk språksekretariat var komitéen behjelpelig i det forberedende arbeidet.

Følgende organisasjoner støttet konferansen økonomisk: Det islandske kultur- og utdanningsdepartementet, Den islandske nasjonalbankens kulturfond, Den islandske seddelbankens kulturfond, Letterstedtska föreningen, Clara Lachmanns fond og Nordisk språksekretariat. Dessuten ytte Orðabók Háskólans forskjellig støtte i forbindelse med forberedelsene til og arrangementet av konferansen. Instituttet har også støttet utgivelsen av denne konferanserapporten. Forlaget Mál og menning og Reykjavík kommune støttet konferansen ved å innby konferansedeltakerne til en mottakelse.

Under konferansen ble det arrangert en utstilling av ordbøker og andre leksikografiske skrifter fra nordiske forlag og institusjoner. Deres deltakelse på konferansen var til stor økonomisk hjelp for arrangørene.

Som på de tidligere konferansene var temaet for konferansen ikke avgrenset til et bestemt område. Men innholdet i rapporten vitner allikevel om at mange har etterkommet arrangementskomitéens oppfordring om å rette oppmerksomheten mot temarammen *Fra materiale til makrostruktur*. Det kommer bl.a. til uttrykk i at begreper som *lemmaseleksjon* og *leksikalisering* står sentralt i mange av artiklene.

Rapporten er redigert av Ásta Svavarsdóttir, Guðrún Kvaran og Jón Hilmar Jónsson. Bessi Aðalsteinsson har hatt ansvaret for ombrekkingen.

Artiklene er ordnet i alfabetisk rekkefølge etter forfatterens etternavn med unntak av de artiklene som er skrevet av islandske forfattere; de er ordnet etter forfatterens fornavn. Bakerst i rapporten finner vi så et forfatterregister, et register over ordbøker som er nevnt eller som det er henvist til i artiklene, og et stikkordregister.

Reykjavík, desember 1995

Ásta Svavarsdóttir
Guðrún Kvaran
Jón Hilmar Jónsson

Innhold

Henrik Andersson	
Om udvælgelse af ord til Den Danske Ordbog	1
Anna Helga Hannesdóttir	
<i>Stýft og hvatt</i> — beteckningar för öronmärken på får och deras lexikografiska relevans	11
Jens Axelsen	
Lemmaselektion i en dansk-engelsk ordbog	19
Kristin Bakken	
Leksikalisering av gammelnorske sammensetninger	27
Henning Bergenholtz	
Leksikografi. Hvad er det?	37
Frede Boje	
Hvor finder man <i>finde anvendelse</i> ?	51
Anna Braasch	
Arbejdet med „Forslag om dansk standard for lagring og udveksling af leksikalske data“	69
Anne Duekilde	
Hovedværk og supplement	83
Grete Duvå og Anna-Lise Laursen	
Tekstbaseret lemmaselektion	93
Gunhild Dyrberg og Joan Tournay	
Den fagsproglige ordbogsartikels fysiognomi	107
Eiríkur Rögnvaldsson	
A Concordance to Old Icelandic Texts and its Lexicographic Value	123
Ruth Feil	
Funktionsverber i det danske sprog	137
Ruth Vatvedt Fjeld	
Noen vanskelige adjektiver i leksikografisk perspektiv	149
Martin Gellerstam	
Brolexikon	159

Guðrún Ingólfssdóttir	
A dictionary of the Icelandic Family Sagas and semantic classification . . .	167
Guðrún Kvaran	
En historisk ordbog og dens aner	175
Helgi Haraldsson	
Plasshensyn i tospråklige ordbøker	185
Berit Heer Henriksen	
Korpusbaserede relationsoplysninger og lemmatisering af flerordsforbindelser	195
Marit Hovdenak	
Ei ordliste trekt ut av tre ordbøker/ordlister	205
Tove Jacobsen	
Elektronisk snuing av Stor norsk–fransk ordbok	213
Pia Jarvad	
Kryptosammensætninger og skabsafledninger i virkeligheden og i ordbøgerne	221
Anfinnur Johansen og Jógvan í Lon Jacobsen	
Den færøske modersmålsordbog — en midtvejsrapport	233
Jón Hilmar Jónsson	
Nøkler til ordforrådet: Om lemmafunksjon, struktur og informasjonstyper i en ny kombinatorisk ordbok over islandsk	245
Kristín Bjarnadóttir	
Lexicalization and the Selection of Compounds for a Bilingual Icelandic Dictionary Base	255
Carl-Erik Lundbladh	
Avledningar i Svenska Akademiens ordbok	265
Sven-Göran Malmgren	
Nationalencyklopedins ordbok — några karakteristiska typer av information	275
Nina Martola	
Från enspråkig till tvåspråkig ordbok	283
Magne Myhren	
Ordbøker og folkelivsgransking (etnologi)	295

Henrik Nikula	
Exemplens funktion i ordböcker	311
Kerstin Norén	
Partikelverb och lexikon: kriterier för att finna lexikaliserade partikelverb .	321
Christian-Emil Ore	
Korpus og seddelarkiv, fredelig sameksistens mellom det beste og det gode?	331
Bo Ralph	
Ordböcker som språkmodeller	339
Inger Hesslin Rider	
Hur ser en användarvänlig ordbok ut?	351
Leena Savolainen	
Finnish Sign Language	359
Rik Schutz	
Mer enn nok er for mye	365
Pål Steigan	
Multimedialeksikografi, nye utfordringer for leksikografisk teori	375
Bo Svensén	
Nordisk leksikografisk ordbok: artiklarnas struktur och innehåll	383
Lars Törnqvist	
Engelsk-svensk namnordbok	397
Lars S. Vikør	
Problem kring ordutvalet i Norsk Ordbok	403
Boye Wangenstein	
Et arkiv over forandringer i moderne norsk ordforråd og dets betydning for norsk leksikografi	413
Lotte Weilgaard Christensen	
Anvendelse af korpora i opbygningen af ordbøger til maskinoversættelsessystemet METAL	421
Petur Zachariassen og Zakaris Hansen	
Redigeringssystemet RiSt — udviklet og anvendt i et færøsk ordbogsmiljø	433

Oversikt over forfatterne	445
Refererte ordbøker	447
Stikkordregister	455

Henrik Andersson

Om udvælgelse af ord til Den Danske Ordbog

The Danish Dictionary is to contain approximately 100 000 entries selected from a text corpus of 40 million running words and four comprehensive existing dictionaries. A method of automatic selection was developed on the basis of two entry lists, which were "manually" selected from small intervals of the alphabet. Two similar entry lists were then selected computationally. The linguistic importance of a word was calculated on the basis of the number of corpus instances and the number of existing dictionaries in which the word occurs. This led to the formulation of some 30 criteria for the automatic selection of entries.

1 Indledning

Det følgende er en praktisk-leksikografisk beskrivelse af den arbejdsproces, der har ført frem til den lemmaliste, som 10 redaktører ansat på *Den Danske Ordbog* (herefter DDO) bruger i det daglige redigeringsarbejde.

Inden man giver sig i kast med at vælge ord ud, bør tre spørgsmål være besvaret:

1. Hvilken type ordbog er der tale om?
2. Hvor mange og hvilke oplysninger skal ordbogen give?
3. Hvilke kilder har redaktørerne til rådighed?

Historien har vist, at det ikke er nogen selvfølge, at disse vigtige spørgsmål er afklaret, inden ordudvælgelse og artikelskrivning går i gang. For DDO's vedkommende lyder svarene:

2 Ordbogens art

DDO er en almensproglig, monolingval betydningsordbog, der skal beskrive det danske sprog i perioden fra ca. 1950 til i dag, med hovedvægten på tiåret 1983–92 (primærperioden). Ordbogen skal dække det skrevne sprog og inddrage det talte og henvender sig i første række til brugere med dansk som modersmål, i anden instans til personer med dansk som andet- eller fremmedsprog. Den skal indeholde ca. 100 000 opslagsord. Ordbogen kommer til at fylde seks bind a ca. 100 000 linjer.

3 Oplysningstyper

Artiklerne skal give oplysning om:

- ortografi
- udtale
- bøjning
- betydning
- kombinatorik (fx kollokationer og valensmønstre)
- orddannelse (afledning og sammensætning)
- etymologi

4 Ordbogens kilder

1. Et elektronisk tekstkorpus på 40 mio. løbende ord fra primærperioden 1983–92.
2. Eksisterende ordbøger og leksikaliserede ordsamlinger, alle i nyeste udgave i maskinlæsbar form, nemlig:
 - *Retskrivningsordbogen*, udgivet af Dansk Sprognævn (RO; ca. 60 000 opslagsord).
 - Blinkenberg & Høybye: *Dansk-fransk ordbog* (B&H; over 150 000 opslagsord).
 - Vinterberg & Bodelsen: *Dansk-engelsk ordbog* (V&B; over 150 000 opslagsord).
 - Dansk Sprognævns register (DS) over især sammensætninger og afledninger i dansk fra 1950 og frem.

5 Grunde til at udvælge hele ordstoffet inden redigering

DDO tager over, hvor *Ordbog over det Danske Sprog* (ODS; i 28 bind + 5 supplementsbind, der er under udgivelse) kronologisk hører op. Men DDO er på mange måder en helt anden ordbog. For det første har den meget mindre plads til rådighed, for det andet dækker ordbogen en meget kortere periode (ca. 50 år mod ODS' ca. 250 år). Og for det tredje er DDO's arbejdsredskaber nogle helt andre end ODS'.

Det stod klart fra første færd, at redaktionen ikke kunne medtage „det hele“, som ODS + Supplement. Med et koncept, der kun stiller 600 000 ordbogslinjer til rådighed, måtte ordene udvælges med skønsomhed. Det blev hurtigt besluttet, at hele ordstoffet om muligt skulle ligge klar, inden redigeringsarbejdet for alvor gik i gang.

Herved ville der blive mulighed for at redigere i semantisk relaterede ordfelter, fx 'nedsættende betegnelser for mandspersoner': *bisse, bølle, kanalje, laban, quisling, sjuft*,

skurk osv. Det er nemmere og hurtigere at redigere den slags ordfelter på én gang end at behandle dem, efterhånden som man møder dem på sin vej gennem alfabetet. Dertil kommer, at det håndværksmæssigt er mere forsvarligt: Det bliver nemmere at opfange og beskrive hårfine betydningsnuancer mellem næsten-synonymer og chancerne for konsistent angivelse af synonymi, antonymi o.l. øges.

En anden fordel, man opnår ved at have hele ordstoffet til rådighed, er, at man kan foreskrive sig, ja måske ligefrem overholde den regel, at simpleksord, der bruges i betydningsdefinitioner, skal kunne slås op i ordbogen.

Sidst, men ikke mindst følger der en administrativ gevinst med en tidlig ordudvælgelse. Man kan løbende kontrollere, om artiklerne overholder det omfang, der er fastlagt i planen.

6 Udvikling af en metode til automatisk udvælgelse

Man kunne umiddelbart forestille sig to yderligtgående måder at vælge ord ud på. Som det ene ekstrem kunne man samle alle forskellige ord fra korpus og eksisterende ordbøger i en liste og gennemgå dem et for et. Det bedste, man kan sige om den metode, er, at den er grundig. Men med omtrent 800 000 lemmakandidater til rådighed siger det sig selv, at den ville være alt for langsom. Man kunne også gå i den modsatte grøft og udelukkende bruge korpus til at udvælge ord fra, fx ud fra reglen: „forekommer et ord x gange i korpus, skal det med“, hvor x = den værdi, der giver ca. 100 000 forskellige opslagsord. Fordelen ved denne metode ville være, at den var særdeles hurtig. Men også meget overfladisk og på anden måde uhensigtsmæssig. For det første ville der ikke blive taget højde for, at der er en indbygget diakron skævhed i korpusmaterialet, der jo kun dækker tiåret 1983–92, mens planen foreskriver, at ordbogen skal omfatte perioden helt fra 1950 til i dag. Ord fra den tidlige del af perioden ville blive mangelfuldt repræsenteret. For det andet kunne der være brist i forekomsten af fagsprog, også det mere almene. Der kan siges meget godt om den engelske *Collins Cobuild*-ordbog, hvis lemmaselektion er rent korpusbaseret. Men hvor ofte har man ikke forgæves forsøgt at slå almene fagord op i den! For det tredje har det vist sig, at korpus indeholder mange forholdsvis frekvente, men semantisk og på anden måde intetsigende ad hoc-sammensætninger og -afledninger, ord, som ikke har interesse i en betydningsordbog.

En metode til automatisk ordudvælgelse måtte manøvrere mellem *Skylla* og *Karybdis*. En lille procentdel af ordstoffet blev udvalgt efter den grundige, men langsommelige alfabetisk fremadskridende ord for ord-metode. Derefter skulle denne del af ordstoffet danne grundlag for fastlæggelsen af en automatisk selektion. Processen forløb i fire etaper:

1. Ord for ord-udvælgelse af 2% af ordstoffet (det alfabetiske interval *gal–greb*).
2. Ord for ord-udvælgelse af yderligere 4–5% af ordstoffet (hele bogstav *a*).
3. Computersimulering af de to samme alfabetiske udsnit med efterfølgende sammenligning.
4. Automatisk udvælgelse af resten af ordstoffet.

6.1 1. etape: *gal-greb*

Første etape blev afviklet i en tidlig fase af forløbet, hvor redaktionen havde brug for hurtigt at få udvalgt en del af ordstoffet til omgående redigering. Det måtte afklares, om ordbogens planlagte omfang overhovedet kunne overholdes, når det kom til stykket, eller om antallet af oplysninger om ordene måtte skæres ned. At netop udsnittet *gal-greb* blev valgt som prøveklud, skyldes, at forholdet mellem små og store artikler i dette topcentsinterval svarer ret nøje til de resterende 98% af alfabetet.

I denne tidlige fase af forløbet var kilderne endnu ikke gjort klar til automatisk udnyttelse. Metoden eller manglen på samme gik ud på, at to redaktører i fællesskab slog alfabetiske delintervaller op i korpus og talte korpusforekomster sammen, hvorefter de valgte ordene ud efter frekvens. RO og redaktørernes sprogformnemmelse supplerede med ord, som enten ikke fandtes, eller som var svagt repræsenteret i korpus.

Resultatet forekom ganske tilfredsstillende, men det havde også taget to fuldtidsansatte redaktører ca. en måned at udvælge de sølle 2% af ordstoffet. Det stod klart, at der måtte langt mere fart på lemmaselektionen; ellers ville der ved deadline i 1999 højst foreligge en ordliste, men ikke nogen ordbog.

Mens *gal-greb*-ordene blev udvalgt og redigeret, fik redaktionens to edb-eksperter, Jørg Asmussen og Ole Norling-Christensen, udviklet to vigtige redskaber til brug for arbejdet med udvælgelsen af hele bogstav *a*.

For det første blev samtlige ord i korpus forsynet med frekvensoplysninger, så man slap for det møjsommelige arbejde med at tælle korpusforekomster sammen i hovedet. Frekvensen blev angivet som to værdier: én for samlet antal forekomster og én for antal korpusdokumenter. Det er den sidste værdi, der tæller. I det følgende betyder „*x* korpusforekomster“ altså ‘forekommer i *x* forskellige tekster i korpus’. Optræder et ord fx 16 gange i den samme tekst, regnes det for én korpusforekomst.

For det andet blev samtlige ord fra ordbøgerne indlæst i en datafil, med angivelse af, hvilke(n) kilde(r) de stammede fra.

6.2 2. etape: bogstav *a*

I denne etape blev der lagt større vægt på ordbogskilderne end i *gal-greb*-udsnittet. Forekom et ord i alle ordbøgerne, altså både i RO, B&H, V&B og i Sprognævnets register, blev det udvalgt uden hensyn til, om det var repræsenteret i korpus. Ordene *adjunktur*, *agnosticisme*, *appellativ* og *attributiv* er de vistnok mest prominente eksempler på ord, der forekom i alle ordbøger, men ikke i korpus. Man kunne også nævne *akribi*, hvad der forhåbentlig ikke skal lægges noget symbolsk i!

Forekom et ord i de tre rigtige ordbøger — altså RO, B&H og V&B, men ikke i Sprognævnets register, stod det også på forhånd stærkt. Under udvælgelsen blev der skelet til korpusfrekvensen, men tommelfingerreglen lød: Er du i tvivl, skal det med.

Fandtes et ord kun i tosprogsordbøgerne B&H og V&B, stod det svagt. Tosprogsordbøger indeholder jo tit vendte fremmedsprogsækvivalenter, gamle ord, forældet fagsprog o.l., som kan have relevans i netop dén type ordbøger; men i DDO hører de ikke hjemme.

En del af ordene i Sprognævnets register var ikke overraskende ad hoc-sammensætninger og -afledninger, der aldrig har slået rod i det danske sprog. Ord, der kun var repræsenteret

i denne ordbogskilde, stod som hovedregel svagt, medmindre solid korpusfrekvens talte for det modsatte.

Efter at ordene fra de forskellige ordbogskombinationer var gennemgået og udvalgt, blev lemmalisten suppleret med højfrekvente ord fra korpus, som ikke var med i de eksisterende kilder. Hermed var der valgt lidt over 4000 bogstav *a*-ord ud, hvilket passede godt med, at dette afsnit af alfabetet erfaringsmæssigt fylder 4–5% af en ordbog.

Det samlede indtryk af de udvalgte ord var, at der var mange ord i en grå zone, ord, som måske/måske ikke skulle med. Navnlig forekom mange ord med lav korpusfrekvens problematiske. Frygten for, at korpus skulle mangle ord fra den tidlige del af perioden (før 1983) og almene fagord, havde givet de eksisterende kilder et lidt for stort ord at skulle have sagt. Ikke mindre end 20–25% af de valgte ord befandt sig i den grå zone. Her blot syv eksempler:

<i>Heterograf</i>	<i>Klasse</i>	<i>Korp. forek.</i>	<i>Lexkilder</i>
afsidning	sb	0	KUN B&H+V&B
afsikring	sb	0	KUN RO
afskedigelsesløn	sb	0	ALLE
afskedigelsesnævn	sb	0	IKKE RO
afskibningshavn	sb	0	RO,B&H,V&B
afskrue	vb	0	KUN B&H+V&B
afskrå	vb	0	KUN B&H+V&B

Ordet *afskedigelsesløn* blev automatisk udvalgt ud fra reglen om, at ord, som fandtes i alle kilder, skulle med, også selv om de ikke var repræsenteret i korpus. Eksemplet viser, at reglen ikke er skudsikker. Inden for primærperioden bruges *afskedigelsesløn* stort set ikke, det hedder i dag med en tidstypisk eufemisme *fratrædelsesgodtgørelse* (8 forekomster i korpus); men *fratrædelsesgodtgørelse* er så svagt repræsenteret i ordbøgerne, at det ikke ville komme med efter de omtalte selektionsprincipper. Eksemplet viser, at overdreven tillid til de eksisterende ordbøger kan medføre en overtrædelse af planens forskrift om, at hovedvægten i DDO's sprogbeskrivelse skal ligge på perioden 1983–92.

Ordene *afskedigelsesnævn* og *afskibningshavn* er også velrepræsenterede i ordbøgerne. Ret beset er det imidlertid diskutabelt, om de hører hjemme i DDO. Sammensætningerne *afskedigelsesnævn* og *afskibningshavn* er semantisk transparente (*afskibe* og *afskibning* er velbelagt både i korpus og i ordbøgerne), og ingen af ordene har vel været særlig udbredte på noget tidspunkt i perioden, heller ikke den tidlige (ODS-Supplementet har ét belæg på *afskibningshavn*, fra 1948, men ingen på *afskedigelsesnævn*). Korpus viser, at *-nævn* som andet sammensætningsled er produktivt i primærperioden, så det er ikke indlysende at vælge *afskedigelsesnævn* med 0 forekomster, bare fordi RO, V&B og B&H har ordet med. Sammensætninger som *adoptionsnævn*, *aftalenævn*, *forældrenævn*, *lønningnævn* og *ungdomsnævn* er alle velrepræsenteret i korpus, men står så svagt i de eksisterende ordbøger, at de ikke ville få et ben til jorden efter de her skitserede udvælgelseskriterier.

At verbalsubstantiverne *afsidning* og *afsikring* kom med trods svag kilderepresentation, skyldes, hvis sandheden skal frem, snarere en subjektiv fornemmelse end saglige forhold.

At partikelverberne *afskrue* og *afskrå* blev valgt ud trods svag ordbogsrepræsentation, er endnu dårligere begrundet. V&B og B&H følger begge princippet om, så vidt muligt at lade den sammensatte form være indgang til partikelverber. Redaktøren, der stod for

udvælgelsen, har elimineret så mange partikelverber med *af-* fra de to tosprogsordbøger, at han er blevet nervøs for, om der nu kom nok bogstav *a*-ord med. Derfor har han medtaget nogle af dem, der umiddelbart har forekommet ham mindst urimelige.

Anden fase viste, at man ikke burde lægge for megen vægt på de eksisterende ordbøger. For mange afgørelser kom til at bero på subjektive skøn. Spørgsmålet var nu, om et edb-program kunne udnytte erfaringerne og præstere et mere plausibelt resultat.

6.3 3. etape: edb-simulering af *gal-greb* og bogstav *a*

Tredje etape gik groft sagt ud på at besvare spørgsmålet: Hvordan får man et edb-program til at ramme de menneskeudvalgte ord så præcist som muligt?

Fremgangsmåden, som blev udviklet af redaktør Jørg Asmussen, var som følger:

Først blev en lille del af det oprindelige *gal-greb*-afsnit skilt ud, nemlig intervallet *gallicisme* til *gammel* (48 ord, ca. 2 promille af det samlede ordstof). For at få alle 48 ord valgt ud automatisk, viste det sig, at tre kriterier skulle være opfyldt, nemlig:

1. 4 korpusforekomster + repræsentation i mindst 3 af de 4 ordbøger. De resterende kandidater skulle
2. mindst være repræsenteret i 3 ordbøger, uanset korpusfrekvens. De sidste ord blev opfanget ved opstilling af kriteriet
3. mindst 5 korpusforekomster uanset repræsentation i ordbøgerne.

De tre selektionskriterier blev derefter overført på hele bogstav *a*. Ikke uventet viste det sig, at kriterierne var for grove; der kom alt for mange irrelevante ord med. Derfor blev der opstillet nogle finindstillingskriterier, der så vidt muligt skulle beholde de relevante, allerede udvalgte bogstav *a*-ord, men udskille de irrelevante. Det ville føre for vidt, her at anføre samtlige finindstillinger, men som eksempler kan nævnes, at ord med mindre end fire korpusforekomster og repræsentation i kun én af tosprogsordbøgerne (B&H el. V&B) blev elimineret. Ligeledes ord med mindre end to korpusforekomster og repræsentation i B&H og V&B, men ikke andre ordbøger.

Herefter blev de mere fintmærkende selektionskriterier overført på et større afsnit af *gal-greb*-ordene. En række nye kriterier blev defineret, atter overført på hele bogstav *a* og så fremdeles, indtil der var et maskinelt udvalg af *gal-greb*- og bogstav *a*-ord, der kom så tæt som muligt på ord for ordudvalgene af de samme alfabetiske intervaller.

Det viste sig umuligt at lave en helt homogen edb-simulering. Nogle få korpus/ordbogscombinationer resulterede i ordlister, der på én gang indeholdt oplagt urimelige og yderst relevante lemmakandidater. Derfor blev ordene inddelt i tre kategorier:

Kategori a): Ord fra både korpus og eksisterende ordbøger, der tilhører det centrale ordforråd og skal have selvstændig indgang, ca. 75% af ordstoffet.

Kategori b): Ord fra combinationer med uhomogene lemmakandidater. Det er op til den enkelte redaktørs skøn, om det enkelte kategori b-ord skal med som selvstændig indgang, skal degraderes til eksempel på sammensætning el. affledning eller evt. skal smides helt ud. Til denne kategori hører bl.a. alle ord med stort begyndelsesbogstav,

der jo ofte, men ikke altid er proprier, fx *A-dur*, *AIDS*, *ATP-bidrag*. Kategori b-ordene kom til at udgøre ca. 20% af det samlede ordstof.

Kategori c): Ord, der ikke findes i ordbogskilderne, men har en rimelig korpusfrekvens (fem forekomster el. mere). Kategorien repræsenterer på én gang nogle af de interessanteste ord (neologismer), og nogle af de mest uinteressante (banale ord som fx sammensætninger med *-forretning*, *-sag* og *-situation* som andetled, afledninger med *-mæssig*, blot for at nævne et par stykker).

Den samlede gennemgang af ord fra de to udvælgelsesmetoder gav et plus til den automatiserede. Gråzoneordene blev elimineret, og opdelingen af ordstoffet i de tre omtalte kategorier overlod trods alt et vist initiativ til den menneskelige dømmekraft. Som forventet var ingen af metoderne helt skudsikre. De efter bedste skøn vigtigste bogstav *a*-ord, som den automatiserede udvælgelse ikke fik med, var:

<i>Heterograf</i>	<i>Klasse</i>	<i>Korp.forek.</i>	<i>Lexkilder</i>
adjunktur	sb	0	ALLE
afhentningspris	sb	0	RO + DS
afrofrisure	sb	0	RO + DS
afspændingsmiddel	sb	1	RO + DS
agility	sb	1	INGEN
agitprop	sb	2	KUN DS
amnesi	sb	0	RO,B&H,V&B
andengenerationsindvandrer	sb	1	INGEN
antiroman	sb	0	V&B,B&H,DS
armyjakke	sb	3	KUN DS
artistnummer	sb	5	INGEN

Af disse ord er *agitprop* og *antiroman* eksempler på ord, der havde større udbredelse i sproget i den del af perioden, som ikke er dækket af korpus. En overraskende stor del af disse ord blev reddet vha. den betingelse, at ord, der forekom i RO, B&H og V&B blot skulle forekomme én gang i korpus for at blive klassificeret som kategori a-ord.

Langt flere af ordene har først fået borgerret i sproget efter primærperiodens udløb: *afhentningspris*, *afrofrisure*, *afspændingsmiddel*, *agility*, *andengenerationsindvandrer* og *armyjakke*.

At et ord som *adjunktur* savnes, kan skyldes, at ordbogsredaktører typisk er fortrolige med undervisnings- og universitetsmiljøet. Hvis der er tale om en usaglig præference, ligger den helt på linje med, at sprogvidenskabelige termer ofte prioriteres højere i almensproglige ordbøger end så mange andre nok så relevante fagområder. Det er vist først og fremmest sprogvidenskabsmænd og ordbogsredaktører, ikke den almindelige bruger, der vil savne tidligere omtalte ord som *appellativ* og *attributiv*.

Værre er det nok med et så relativt almindeligt medicinsk fagudtryk som *amnesi*, der mangler den ene sølle forekomst, der kunne have reddet ordet.

Listen antyder et problem, som det ville føre for vidt at komme ind på her, nemlig edb-støj. Ordet *artistnummer* findes ikke i de eksisterende ordbøger, hvilket medfører, at programmet ikke har kunnet lemmatisere ordet. Bøjningsformen *artistnumre* tegner sig for

fire af de fem forekomster, men er af programmet opfattet som et andet ord end *artistnummer*, der har én forekomst. Ordet er dermed blevet elimineret ud fra reglen om, at et ord, der ikke forekommer i de eksisterende ordbøger, skal forekomme mindst fem gange i korpus for at blive udpeget.

Den langsommelige ord for ord-udvælgelse af bogstav *a* lagde større vægt på kilderne; men også den havde sine smuttere, hvoraf de 11 grelleste er:

<i>Heterograf</i>	<i>Klasse</i>	<i>Korp. forek.</i>	<i>Lexkilder</i>
ABS-bremse	sb	47	KUN DS
adresseliste	sb	8	KUN V&B
advokatsalær	sb	7	KUN V&B
afbudstrejse	sb	7	IKKE B&H
aha-oplevelse	sb	6	KUN V&B
almenvæl	sb	30	RO + B&H
altmodisch	adj	9	IKKE V&B
ankelsok	sb	18	RO + B&H
appelsinsaft	sb	60	KUN B&H
artisteri	sb	16	RO + V&B
ayatollah	sb	34	RO + B&H

Som det fremgår, er der tale om ord med pæne frekvenstal (ml. 6 og 60 forekomster i korpus). *ABS-bremse* er så nyt et ord, at det af naturlige grunde ikke er kommet med i de udgaver af ordbøgerne, som fandtes på udvælgelsestidspunktet. Derimod forekommer det overraskende, at V&B er ene om at have *adresseliste*, *advokatsalær* og *aha-oplevelse* med, og at kun B&H har fundet *appelsinsaft*, listens topscorer mht. korpusfrekvens, værdigt til optagelse. I de øvrige tilfælde: *afbudstrejse*, *almenvæl*, *altmodisch*, *ankelsok*, *artisteri* og *ayatollah* er det en enkelt af de to tosprogsordbøger, der har fravalgt ordet. Det skal her understreges, at de valgte eksempler ikke er anført for at kritisere hæderkronede ordbøgers lemmaselektion. Tværtimod må man anerkende, at der er så få tilfælde af uenighed om, hvad det centrale ordforråd er.

7 Konklusion

Der er betydelige fordele ved at automatisere udvælgelsen af ord. Det er ikke alene den hurtigste (og dermed billigste) løsning, men også den bedste, for så vidt som den giver det mest konsistente udvalg. Naturligvis er sprogforneelsen en faktor, man ikke kan se bort fra, men når det kommer til konkrete afgørelser af, hvilke ord der skal med i en ordbog, og hvilke der ikke skal, er det næsten uundgåeligt, at en vis vilkårlighed — subjektive præferencer og idiosynkrasier — gør sig gældende. Den beskrevne udvælgelsesmetode giver en saglig begrundelse for optagelsen eller forkastelsen af hver eneste af de omtrent 800 000 lemmakandidater, kilderne tilsammen indeholder.

Litteratur

B&H = Andreas Blinkenberg/Poul Høybye 1991: *Dansk-fransk ordbog*. 4. udg. ved Jens Rasmussen & al. Bind 1–2. København: Nyt Nordisk Forlag Arnold Busck.

DS = Dansk Sprognævns register over sammensætninger og afledninger i dansk 1950 og frem. Ikke publiceret.

Kristensen, Kjeld 1993: Den Danske Ordbogs tekstkorpus og spORDhunde. I: Anna Garde/Pia Jarvad (udg.): *Nordiske Studier i Leksikografi II*. København: LEDA/Nordisk Forening for Leksikografi, 138–42.

ODS = *Ordbog over det Danske Sprog*. 1918–56. Bind 1–28. København: Det Danske Sprog- og Litteraturselskab.

RO = *Retskrivningsordbogen*. 1986. København: Dansk Sprognævn.

V&B = Hermann Vinterberg/C. A. Bodelsen 1990: *Dansk-engelsk Ordbog*. 3. udg. ved Viggo Hjørmager Pedersen. København: Gyldendal.

Anna Helga Hannesdóttir

Stýft og hvatt — beteckningar för öronmärken på får och deras lexikografiska relevans

In Iceland, the ownership of all sheep is indicated by a system of marks cut in the ears of the sheep. The names of these marks can be regarded as a closed lexical subsystem in the Icelandic language, a subsystem defined by its clearcut relations to phenomena in the "real world" rather than by purely linguistic features. In the present paper, the morphology of this lexical subgroup is described. Furthermore, the lexicographic relevance of subsystems established on encyclopedic grounds, primarily, is discussed with respect to the role such subsystems play in dictionaries of ordinary language.

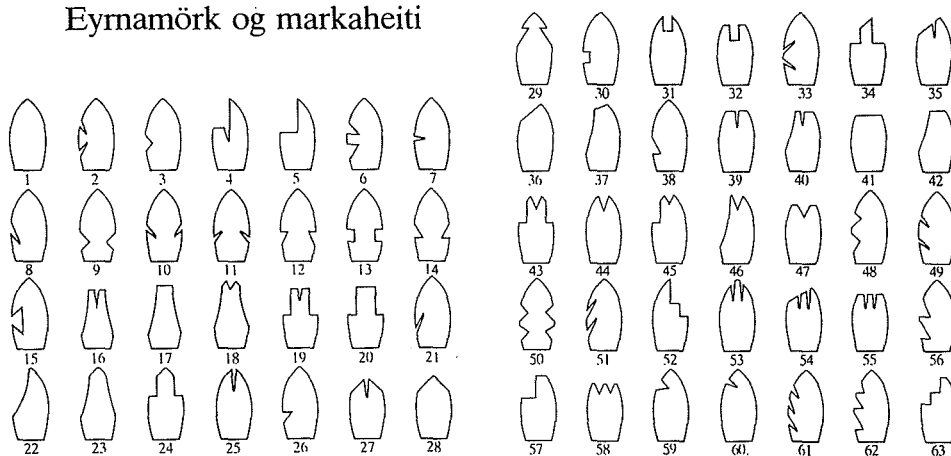
I ett språks ordförråd kan en mängd lexikaliska delsystem urskiljas. I en dimension kan det sägas vara de ingående enheternas språkliga egenskaper som ligger till grund för delsystemets avgränsning. Andra subgrupper uppkommer snarare genom en kombination av konventionaliserade användningskriterier och språkliga faktorer som återspeglas i enheternas stilistiska och pragmatiska egenskaper. Vidare finns det lexikaliska delsystem som bestäms av att de ingående enheternas referenter bildar ett system oberoende av språket, där det sålunda inte kan sägas vara språkliga kriterier över huvud taget som ligger till grund för systemets avgränsning. Sådana lexikaliska undergrupper kan till exempel utgöras av taxonomiska system, teknisk terminologi, vokabulärer som är begränsade till bestämda situationer eller användningsområden samt namn och namnliknande uttryck i allmänhet. Den senare typen av lexikaliska delsystem kan således sägas vara encyklopediskt bestämda, och de ställer lexikografen inför ett avgörande hur stort utrymme som skall beredas åt den encyklopediska informationen i en ordbok som i första hand behandlar ordförrådets språkliga aspekter. Avgörandet om huruvida den aktuella undergruppen över huvud taget skall tas med i ordboken kan bero på faktorer som områdets relevans för den allmänne språkbrukaren, om det är klart avgränsat eller inte, om det är slutet eller öppet, de enskilda enheternas frekvens etc. Och om delsystemet inkluderas i ordboken — skall då dessa enheter relateras inbördes till varandra och till det överordnade systemet eller inte, och om inte, hur skall de då behandlas?

I det följande behandlas en ordkategori som kan sägas vara encyklopediskt bestämd snarare än språkligt och som nog får betraktas som tämligen perifer i den isländska språkgemenskapens lexikon idag. Denna kategori kan närmast betecknas som ett spontant uppkommet etnosemantiskt fält som efterhand mera fått funktionen av ett slutet terminologiskt system. Kategorin innehåller de språkliga uttryck som används för att beteckna de märken som skärs in i lammets öron för att markera dess ägare. Jag beskriver i första hand det lexikaliska system som de enskilda benämningarna tillsammans utgör och försöker belysa några urskiljbara kategoriers formella och semantiska egenheter. Vidare berör jag dels hur denna ordgrupp behandlas i *Íslensk orðabók*, dels vilken lexikografisk relevans denna och liknande undergrupper kan sägas ha.

Bruket att öronmärka får antas ha varit väl etablerat i Norge redan före vikingatiden, och det lär ha överförts därifrån till atlantöarna i samband med att dessa bebygg-

des. Även benämningarna på märkena har följt med från Norge, inte bara till Island utan också till Färöarna, där det redan i Seyðabrævið (1298) anges att ägarförhållandena regleras med märkning, samt till Orkney och Shetland. Liknande benämningar återfinns i viss utsträckning också på Gotland. Vissa kontrastiva studier av märkenas utseende och benämningar i de olika områdena har genomförts, t.ex. av Kristian Østberg (1927) och Judith Chrystal (1984). Chrystal, som främst har undersökt den gotländska märkestraditionen, slår fast att benämningarna på atlantöarna och Gotland till stor del härstammar från samma märkestradition (1984:5).

Eynamörk og markaheiti



- | | | |
|--------------------|------------------------|----------------------|
| 1 Alheitt | 22 Hálfraf | 43 Sýlhamrað* |
| 2 Andfjaðrað | 23 Heilgeirað | 44 Sýlt |
| 3 Biti | 24 Heilhamrað | 45 Sýlt í blaðstýft* |
| 4 Blaðrifað* | 25 Heilrifað | 46 Sýlt í hálfraf |
| 5 Blaðstýft | 26 Hófbíti | 47 Sýlt í stúf* |
| 6 Boðbílur* | 27 Hvatrifað | 48 Tvíbitað |
| 7 Bragð | 28 Hvat | 49 Tvífjaðrað |
| 8 Fjöður | 29 Laufað* | 50 Tvígagnbitað* |
| 9 Gagnbitað | 30 Lögg | 51 Tvíhangfjaðrað |
| 10 Gagnfjaðrað | 31 Miðhlutað | 52 Tvínúmið* |
| 11 Gagnhangfjaðrað | 32 Miðhlutað í stúf | 53 Tvírifað í heilt |
| 12 Gagnhófbitað | 33 Oddfjaðrað | 54 Tvírifað í sneitt |
| 13 Gagnlaggað* | 34 Sneiðhamrað* | 55 Tvírifað í stúf |
| 14 Gagnstigað | 35 Sneiðrifað | 56 Tvístigað |
| 15 Geirskorið* | 36 Sneitt | 57 Tvístýft |
| 16 Geirstúfrifað* | 37 Sneitt í hálfraf | 58 Tvísýlt í stúf* |
| 17 Geirstýft* | 38 Stig | 59 Vaglbíti* |
| 18 Geirsýlt* | 39 Stúfrifað | 60 Vaglriðað* |
| 19 Hamarrifað* | 40 Stúfrifað í hálfraf | 61 Þrífjaðrað* |
| 20 Hamrað | 41 Stýft | 62 Þrístigað* |
| 21 Hangfjöður | 42 Stýfthálfraf | 63 Þrístýft* |

* táknar að unnið skuli gegn notkun þessara marka

Figur 1: Öronmärken och deras benämningar. Ur: *Markaskrá Eyjafjarðarsýslu, Akureyrar, Dalvíkur, Ólafsfjarðar og Siglufjarðar*. 1988

På Island reglerades örönmärkningarna redan i Grágás (ca 1250), och än idag lyder den under gällande förordningar. För varje fåravedsdistrikt finns en officiell förteckning över registrerade örönmärkeskombinationer och deras innehavare. I en sådan förteckning — *markaskrá* — finns även en översikt över de enskilda märkenas form och benämning. De äldsta källorna säger ingenting om vare sig själva örönmärkenas form eller deras benämningar, men det förefaller rimligt att anta att det redan då har varit frågan om lexikaliskt slutna system. Som underlag för min undersökning har jag använt mig av *Markaskrá Eyjafjarðarsýslu, Akureyrar, Dalvíkur, Ólafsfjarðar og Sigluffjarðar* (1988). Förteckningen tar upp totalt 63 enskilda märken och deras benämningar, som i just detta distrikt bildar 1416 olika kombinationer. De märken som används i detta distrikt och deras benämningar återfinns här i figur 1. Även *Íslensk orðabók* ger en översikt över örönmärkena fast i en något annorlunda variant (plansch F2, s. 212), och många av benämningarna behandlas också som stickord.

De 63 benämningar som utgör systemet kan i ordbildningshänseende delas in i fyra typer: benämningar i simplexform, sammansättningar, benämningar bildade med prefix samt fraser. Simplexbenämningarna kan sägas beteckna de enklaste märkena; de flesta av dessa förekommer även i särskilda konstellationer, som får ett modifierande tillägg i benämningen, och i märkeskombinationer vars benämningar ofta utgörs av en kombination av de aktuella märkenas benämningar. Om vi först ser på simplexbenämningarna (tabell 1), så är 11 av denna typ. Fem är substantiv, *biti*, *bragð*, *fjöður*, *lögg* och *stig*, medan sex har en form som i *Íslensk orðabók* i de flesta fall uppges vara particip. Jag återkommer till ordklassstillhörigheten och kallar tills vidare dessa former adjektiviska former. De märken som betecknas med en simplexbenämning kan sägas vara ett slags grundmärken, och de flesta av dem återkommer som sagt i kombinationer med en egen benämning.

biti	hamrað	lögg	stýft
bragð	hvatt	sneitt	sýlt
fjöður	laufað	stig	

Tabell 1. *Benämningar i simplexform*

Nästa grupp, sammansättningarna, innehåller 23 enheter (tabell 2). De flesta är sammansatta av ett substantiv och ett adjektiviskt led. I dessa sammansättningar förekommer några av simplexbenämningarna, antingen som för- eller efterled. Där återfinns bl.a. *biti* i *hófbiti*, *fjöður* i *hangfjöður* och *stýft* i *blaðstýft*. Men där återfinns också de grundläggande morfemen i *fjöður* och *stýft* i formen *fjaðrað* respektive *stúf-* i sammansättningarna *oddfjaðrað* och *stúfrifað*. Jämfört med de märken simplexbenämningarna betecknar är märkena bakom de sammansatta benämningarna antingen en kombination av två märken, som *sneiðhamrað*,

blaðrifað	geirsýlt	heilrifað	sneiðrifað
blaðstýft	hamarrifað	hófbiti	stúfrifað
boðbildur	hangfjöður	hvatrifað	stýfthálftaf
geirskorið	hálftaf	miðhlutað	sýlhamrað
geirstúfrifað	heilgeirað	oddfjaðrað	vaglrifað
geirstýft	heilhamrað	sneiðhamrað	

Tabell 2. *Sammansatta benämningar*

eller ett grundmärke applicerat eller upprepat på ett speciellt sätt, som *hangfjödur* och *oddfjaðrað*. Förutom att de element som kan fungera som simplexbenämningar används i sammansättningarna byggs systemet ut genom andra element som fungerar som förled och/eller efterled, t.ex. *blað-*, *geir-/geirað* och *-rifað*. Förhållandet mellan benämningen *heilrifað* och märket *rifað* är dock något annorlunda då förleden *heil-* anger att märket *rifað* skärs in i en annars hel öronspets, vilket är samma förhållande som för märket *sýlt* utan att det i det fallet framgår av benämningen.

alheilt	gagnhófbitað	tvíhangfjaðrað	tvísýlt
andfjaðrað	gagnlaggað	tvínúmið	þrífjaðrað
gagnbitað	gagnstigað	tvírifað	þristigað
gagnfjaðrað	tvíbitað	tvístigað	þrísýlt
gagnhangfjaðrað	tvífjaðrað	tvístýft	

Tabell 3. *Prefigerade benämningar*

Den tredje ordbildningskategorin, benämningar bildade med prefix, innehåller 19 enheter (tabell 3). Prefixbenämningarna bildar ett ännu mer genomfört morfologiskt mönster än sammansättningarna. I *Íslensk orðabók* klassificeras 14 av de 19 orden som particip, två som adjektiv och tre saknas. I den prefigerade kategorin återfinns vissa semantiska morfem som överensstämmer med de substantiviska benämningarna i simplexform. De kan också uppträda som sista led i sammansättningarna, där de behåller sin substantiviska form. Som prefixavledning får de emellertid adjektivisk form. Sålunda finner vi *bitað* i *gagnbitað*; jfr *biti*, *hófbiti*. Efterlederna i denna grupp är huvudsakligen systeminterna och de allra flesta känns igen från simplexgruppen. Prefixen däremot är snarare allmänspråkliga och de anger antingen var eller hur ofta ett och samma märke har placerats på örat. Ett undantag är *alheilt*, som betyder att örat är intakt.

Den fjärde och sista typen av benämningar utgörs av fraser. Denna grupp innehåller 10 enheter (tabell 4). Fraserna har alla samma form, *x í y*, där *x* har en adjektivisk form medan formen på *y* varierar.

miðhlutað í stúf	sýlt í hálf taf	tvírifað í stúf
sneitt í hálf taf	sýlt í stúf	tvísýlt í stúf
stúfrifað í hálf taf	tvírifað í heilt	
sýlt í blaðstýft	tvírifað í sneitt	

Tabell 4. *Fraser*

De olika elementen, som ensamma eller i kombination med andra utgör den totala uppsättningen benämningar, är olika frekventa inom systemet. Vi finner dels element som endast förekommer en gång, till exempel *bragð*, *boðbildur*, förleden *odd-* i *oddfjaðrað* och efterleden *-númið* i *tvínúmið*. De oftast förekommande elementen i benämningarna är formparet *fjödur/fjaðrað* samt *rifað*. Dessa förekommer i nio benämningar var.

Det som framför allt kännetecknar den morfologiska strukturen hos öronmärkenas benämningar är den frekventa förekomsten av adjektiviska former. Som redan har framgått har några av komponenterna en adjektivisk form med en substantivisk alternativform, t.ex. *hamrað* — *hamar*, andra förekommer endast som adjektiviska former, t.ex. *rifað*, medan ytterligare några förutom den adjektiviska formen även har en stamvariant i de sammansatta

benämningarna, t.ex. *stýft* — *stúf*-. Att den adjektiviska formen är så dominerande kan antas bero på en implicit syftning på ordet *eyra*. Även om de adjektiviska formerna dominerar kraftigt i systemet, förekommer i några grundläggande fall substantiviska simplexord: *biti*, *bragð*, *fjöður*, *lögg*, *stig*. I samtliga dessa fall måste märkets relativa placering på örat anges. Märkena appliceras på örats kant, antingen baktill eller framtill, och benämningarna preciseras med *framan* eller *aftan*. Märkena *hamrað*, *hvatt* etc., som har adjektivisk form, kan däremot bara placeras på ett sätt på örat och de behöver därför inte preciseras med ett platsangivande adverb. Benämningen får adjektivisk neutrumform för att kongruera med det underliggande objektet *eyra*.

Trots den till synes enhetliga morfologiska strukturen i detta lexikaliska delsystem finns det vissa skillnader i hur de olika märkesbenämningarna är bildade. Som redan nämnts anges i *Íslensk orðabók* många av de former som här har kallats adjektiviska som particip, och många av dem behandlas under respektive verb. I vissa fall förefaller detta också vara riktigt. Således är *hvatt* uppenbarligen preteritumparticipet av *hvetja* och *numið* preteritumparticip av *nema*. För formparet *stig* — *stigað* kan man möjligen tänka sig verbet **stiga* = *gera stig* eller *marka stig*. Denna betydelse finns emellertid inte manifesterad som finit verb, men vilken talare som helst av modern isländska kan med hjälp av produktiva morfologiska regler skapa den mellanliggande formen **stiga*. För vissa adjektiviska former kan det vara svårt att avgöra om de formellt skall klassificeras som ett adjektiv avlett av ett substantiv som kan påvisas i språket eller en participform av ett tänkt verb. Sålunda är det inte lätt att bestämma hur ”ordbildningsparadigmet” för formparet *stúfur* — *stýft* ser ut. Skall man räkna dit också verbet *stýfa* och se *stýft* som particip av det, eller skall man se *stýft* som ett adjektiv direkt avlett av substantivet *stúfur*? Just i detta fall tycks båda analyserna kunna vara rimliga. I andra fall kan semantiska argument tala för att den adjektiviska formen skall uppfattas som adjektiv och inte particip, t.ex. *fjöður* — *fjaðrað*. Det finita verbet *fjaðra* förekommer i språket men det finns ingen överensstämmelse mellan verbets betydelse ’fjädra’ och betydelsen hos benämningens adjektiviska form ’likna en fjäder’. Dessutom är verbet *fjaðra* intransitivt. Det kan då konstateras att de former som här har kallats adjektivformer eller adjektiviska former och som i *Íslensk orðabók* för det mesta klassificeras som particip kan vara bildade på olika sätt. Dels kan de vara rena particip, som *hvatt*, och dels rena adjektiv, som *hangfjaðrað*. Behandlingen av dessa adjektivistiska former i *Íslensk orðabók* visar i viss mån bristande konsekvens.

Vid en genomgång av de enheter som utgör element i märkesbenämningarna förekommer dels enheter ur lexikonets centrala delar, som exempelvis *hálf*, *af* och *tvi-*, och dels enheter som i någon mån kan betraktas som strikt begränsade till denna lexikaliska undergrupp. Enheter av den senare typen är exempelvis *sýlt* och *boðbildur*. Även om det finns en morfologisk koppling till ord i allmänspråket har dock många av dem en så specialiserad användning att de inte längre självklart låter sig härledas semantiskt från sin allmänspråkliga motsvarighet, och de har blivit i stort sett betydelselösa, som t.ex. *sýlt*. Andra enheter har dock utvecklats i motsatt riktning och fått en distinktiv funktion som den allmänspråkliga motsvarigheten saknar. Prefixen *and-* och *gagn-* är i allmänspråket i det närmaste synonyma, ’(e)mot’, men inom märkesbenämningarna är dessa prefix ingalunda utbytbara utan klart åtskilda och sinsemellan diskriminerande. Sålunda har *gagn-* behållit den gemensamma kärnbetydelsen ’(e)mot’ och används i alla benämningar där samma märke placeras mitt emot varandra på örats båda kanter, medan *and-* endast används i benämningen *andfjaðrað*,

där märket *fjöður* har applicerats två gånger på samma kant med spetsarna ställda mot varandra. I stort sett kan de enheter som är begränsade till detta lexikaliska delsystem sägas spegla någon av variablerna hur man skär, som exempelvis *sneitt* och *skorið*, hur det märkta örat ser ut, som exempelvis *stýft*, *hvatt* och *fjaðrad*, och möjligen hur den bortskurna biten ser ut, som *biti*. En närmare beskrivning av betydelsen hos de enskilda benämningarnas allmänspråkliga motsvarigheter finns hos Østberg (1927:174 ff.), där också de färöiska och norska märkesbenämningarna redovisas.

Jag har här gett en summarisk beskrivning av några egenskaper hos ett lexikaliskt delsystem i det isländska ordförrådet — en beskrivning som naturligtvis hade kunnat göras mera ingående och precis. Detta moment lämnas emellertid därhän, och i det följande diskuteras istället denna ordgrupps lexikografiska relevans.

Det här aktuella lexikaliska delsystemet består av ett antal isländska uttryck som uttömmande ger besked om hur fåröron kan vara skurna, i syfte att entydigt ange ägare till enskilda djur. Det rör sig om en grupp uttryck som vid sidan av denna specifika funktion inte spelar någon roll i dagens isländska ordförråd. Inom ramen för denna funktion bildar de emellertid ett system, där visserligen den enskilda enheten i sig knappast heller betyder något, det är endast som en del av ett system den blir viktig. På så sätt kan detta — och liknande — delsystem tjäna som en påminnelse om vad som gäller för hela systemet. Precis som i detta system gäller generellt för hela ordförrådet att dess enheter får sin fulla betydelse endast i förhållande till andra enheter i ordförrådet. I det lexikaliska systemet låter sig enheterna egentligen inte fullt ut behandlas i isolering. Snarare är det den inbördes relationen mellan enheterna och det som skiljer dem från varandra som är av betydelse. Om lexikonet kan man som bekant tillämpa T. S. Eliots kända rader: "Where every word is at home, Taking its place to support the others."

Ett så klart avgränsat delsystem som öronmärkenas benämningar kan — trots sitt beskedliga omfång — sägas åskådliggöra den semantiska stratifiering som Weinreich (1962) ansåg vara utmärkande för lexikonet. Vissa enheter i systemet har knappast någon inherent betydelse utan får sin funktion uteslutande som en enhet i systemet, medan andra har ett mer omfattande och precist semantiskt innehåll. En mindre grupp av uttryck kan ses som mest central, då de andra kan definieras utifrån dessa. Trots de få uttryck som ingår i systemet bildar de ett komplicerat morfo-syntaktiskt och semantiskt system av ömsesidiga relationer. Det kan ge en föreställning om den enorma komplexitet som ordförrådet uppvisar i sin helhet. Vidare är det en ofta återkommande fråga i den lexikografiska debatten vilken plats som skall beredas för olika specialspråk i ett allmänspråkligt lexikon. Ett delsystem som fårmärkenbeteckningarna kan sägas reflektera ett närmast terminologiskt specialspråk. Man kan göra reflexionen att det knappast lönar sig att ta ut några få uttryck ur delsystemet och diskvalificera resten, men några långtgående slutsatser om hur termsystem skall behandlas i allmänhet kan man kanske inte dra.

På de tidigare konferenserna i denna serie har man ägnat ett visst utrymme åt frågan om den encyklopediska informationens plats i allmänspråkliga ordböcker. Att hävda att det vore motiverat att ta med just benämningarna på fårens öronmärken i en ordbok över dagens isländska vore kanske att stödja den tendens till "samlarlidenskap" som enligt Ásta Svavarsdóttir (1994) är typisk för isländsk lexikografi. Hur som helst, så aktualiserar delsystem som detta frågan om lexikografens betydelse för språkutvecklingen. En antikvariskt lagd lexikograf tar dem med om han har gott om plats i sin ordbok — eller bereder plats för dem

på andra ords bekostnad — trots att såväl systemet som dess enskilda delar blir allt mindre relevant för dagens islänningar, i takt med att jordbruket sysselsätter en allt mindre del av befolkningen, och i takt med att andra märksystem ersätter örönmärkena. Den mer deskriptivt lagde lexikografen tar kanske fasta på dessa förhållanden och utelämnar benämningarna helt — vilket i sin tur möjligen bidrar till att de försvinner fortare ur det isländska ordförrådet än de annars hade gjort. Det finns trots allt åtskilliga islänningar fortfarande som i sitt språkliga medvetande reserverar ett litet hörn åt dessa benämningar, även om detaljkunskapen kan vara begränsad. Nu när de lexikaliska databaserna i princip kan rymma ett oändligt ordförråd och den nya tekniken förser lexikografen med kraftfulla hjälpmedel som CD-ROM skulle väl 63 enheter mer eller mindre inte spela någon ekonomisk eller utrymmesmässig roll.

I den lexikografiska period som föregick dataåldern hade för övrigt några av benämningarna en alldeles egen lexikografisk relevans. *Stýft* och *hvatt*, som har fått plats i titeln på detta bidrag, var om inte en central så dock en del av den blyertsbrukande lexikografens vardag, då de något fritt kan översättas med 'trubbigt' och 'vässat'.

Litteratur

- Ásta Svavarsdóttir 1994: Sprogsamfund, sprogpolitik og ordbøger. I: *LexicoNordica* 1, 119–137.
- Chrystal, Judith 1984: *Namn på gotländska fårmärken*. En presentation. Stockholms universitet, Institutionen för nordiska språk. [Dupl.]
- Íslensk orðabók handa skólum og almenningi 1983. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Markaskrá Eyjafjarðarsýslu, Akureyrar, Dalvíkur, Ólafsfjarðar og Siglufjarðar 1988. Útg. Eiríkur Björnsson. Akureyri: Prentverk Odds Björnssonar hf.
- Weinreich, Uriel 1962: Lexicographic definition in descriptive semantics. I: *Problems in lexicography*. 2 uppl. utg. av F. W. Householder and S. Saporta. Bloomington: Indiana University, 25–44.
- Østberg, Kristian 1927: *Norsk bonderet*. Første bind. Av kreaturbrukets rets- og sedvaneregler. Oslo: J. W. Cappelens Forlag.

Jens Axelsen

Lemmaselektion i en dansk-engelsk ordbog

The article discusses possible principles for the selection of lemmas on the background of the revision of a Danish-English dictionary with 65 000 headwords. The three sources: other dictionaries, introspection and texts are briefly looked at. The problems presented by the selection are treated under the headings: time (obsolete words and neologisms), place (dialect words and foreign, particularly English, loan words), grammar (inflected forms, derivatives, compounds, proper names), style and register (slang, taboo words, technical words). The conclusion is that at the end of the day the decision depends upon the subjective judgment of the editor.

1 Indledning

At finde ud af hvad der skal tages med i en ordbog er et af ordbogsforfatterens værste dilemmaer, og samtidig er lemmaselektionen det der oftest bliver kritiseret, ikke mindst i ordbogsanmeldelser. Det fremgår af Bergenholtz/Mogensen 1993, at lemmaselektion er absolut topscorer blandt de emner der er behandlet i de anmeldelser de har undersøgt. Problemerne er nogenlunde de samme, hvadenten det drejer sig om en revision som den jeg netop har afsluttet af en dansk-engelsk (aktiv) ordbog med ca. 65 000 opslagsord, eller udarbejdelsen af en ordbog fra grunden, og de bliver ikke mindre når det drejer sig om en ordbog af begrænset omfang som den nævnte. Her må lemmaerne overvejes særlig nøje før de slippes ind eller får lov til at blive stående.

2 Mulige principper for udvælgelsen

Brugergruppen som styringsprincip for lemmaselektionen har været foreslået, men gælder mest fagordbøger. Når det som her drejer sig om en almen ordbog til det generelle publikum, foruden til gymnasiets engelskundervisning, kan *det* ikke rigtig bruges til noget.

Lemmas frekvens er et vigtigt kriterium, og her kan man jo bruge ordoptællinger på grundlag af korpus, men man kan ikke bare bestille en liste over de 65 000 almindeligste ord. I *Den Danske Ordbogs* 40 mio's korpus viser en beregning på grundlag af bogstav A at kun ca. 41 000 forekommer 10 gange eller mere. Tallet må desuden reduceres noget, da ord her betyder grafiske former, omfattende bøjningsformer m.m., og der vil også forekomme en del overflødige sammensætninger. Så et slag på tasken siger at ud over de ca. 37 000 almindeligste vil udvalget måtte bero på en personlig vurdering ud fra andre kriterier.

Jeg prøvede at formulere nogle principper i forordet til tillægget fra 1974 til V & B, 2. udgave, og det der står der vil jeg da gerne stå ved, også som rettesnor for min egen nys afsluttede ordbog, der jo kun er halvt så stor og derfor kræver endnu omhyggeligere sortering:

Hovedvægten er lagt på ord og vendinger fra det almindelige sprog, herunder også slangudtryk og vulgærisme hvis de skønnedes at have en vis udbredelse. Fra tekniske og videnskabelige fagområder er der optaget en del ord, der også er kendt uden for fagfolks kreds. Grænserne er her sat ret vidt, men det har ikke været hensigten at dække tilvæksten af termer inden for mere specielle områder som fx atomfysik, og udvalget omfatter hovedsagelig ord, der er truffet uden for rent faglig litteratur, fx i dagspressen.

Den opmærksomme læser vil bemærke en del uldne formuleringer. „Hovedvægten“ [...] hvis de *skønnedes* at have *en vis* udbredelse [...] grænserne er her sat *ret* vidt [...] *mere specielle* områder [...], *hovedsagelig* ord der er truffet uden for *rent* faglig litteratur“. Det er jo ret så generelt, og spørgsmålet er så om man kan komme det nærmere.

3 Kilder til lemmaer

De er som bekendt tre: andre ordbøger, introspektion og tekster, enten selvfundne eller indeholdt i et korpus.

Den første må bruges med megen skønsomhed. Den har sin begrænsning deri, at alle ordbøger der *er* udkommet eo ipso er forældet. Desuden vil nok ingen ordbogsforfatter med respekt for sig selv basere en ordbog eller ordbogsrevision udelukkende på, hvad han/hun har hentet fra andre ordbøger, eller afskrive en anden ordbog in extenso. Det man henter fra andre bør man i hvert tilfælde checke, og kilden bør efter god skik anføres.

Ens egen sprogviden er jo desværre ikke systematisk ordnet inde i hovedet, og der ligger også deri den fejlkilde, at man ikke ved om det man finder kun tilhører ens egen idiolekt.

Den mest tilfredsstillende kilde er derfor originale teksteksempler eller et korpus. Jeg har ikke haft korpus konstant til rådighed ved min arbejdsplads, så jeg har mest brugt *Den Danske Ordbogs* til verifikation og ikke forladt den gammeldags metode med selv at finde eksempler ved læsning. Denne metode er trods alt nyttig til supplerings af et korpus, og det går let med at finde belæg ved læsning.

4 Selektionsproblemerne

De ordnes hos Svensén (1993:46 ff.) og i forordet til ODS (s. XXII) under nogle overskrifter: tid, sted, brugsområde. Hertil kan føjes grammatik.

4.1 Det tidsmæssige:

Dels bagud: forældede ord; dels fremad: neologismer

Der er ikke noget fast kriterium for hvornår et ord er forældet, men Poul Høybyes definition (1971:313) er en rettesnor: Så længe et ord stadig bruges og forstås af nulevende, er det ikke dødt. *Trættekær* og *tilforladelig* er måske ikke kendt af den yngre generation, men er helt naturlige for min. Og arkaismer der stadig bruges spøgende, som *bole*, *bie*, *værkbruden* har jeg beholdt. Det helt forældede er fjernet, og en del var der er ingen tvivl om: *kammerpotte* (udgik efter 150 år) og *støvleknapper* (efter 60); desuden ord fra besættelsen:

spildindsamling, *særmelding*, og forældet slang: *fjong*. Men der er tilfælde som er på grænsen: *grammofonstift* og *selvbinder* bruges ganske vist ikke mere, og *futtog* findes ikke, men ordene lever nok alligevel. Skift er ofte umærkelige: nogle sammensætninger med *jernbane-* afløses af *tog-*: ingen taler mere om *jernbane|billet*, *-forbindelse*, *-kupé*, *-rejse*. Og pludselig dukker der ting op igen, som man længst troede døde, fx er *mamelukker* vendt tilbage, og *alsang*, som jeg troede tilhørte besættelsen, bruges nu om fællessang ved fodboldkampe etc. Ved neologismer er problemet, når det drejer sig om kollokvialismer, at vurdere hvad der ser ud til at holde, jf. *tjækket*, *være på skrump*, *sild* (pige). Det eneste princip man her har at holde sig til er Popes fra *An Essay on Criticism*, ll. 335–6:

Be not the first by whom the new are try'd,
Nor yet the last to lay the old aside.

Nye tekniske og andre fagord som optræder i den offentlige debat er lettere at tage stilling til: *gensplejset*, *drivhuseffekt*, *græsgrødder*, *iltsvind* må ikke savnes i ordbogen.

4.2 Det geografiske:

**Inden for landets grænser: dialektord; uden for: fremmedord,
og for en dansk-engelsk ordbogs vedkommende især anglicismer**

Dialektord er ikke det store problem, da ordbogen må basere sig på rigssproget, men de sniger sig jo ind, og man kan være nødt til at tage stilling til om de nu er accepteret: *fornæret*, *snalret*, *træls*. Fremmedord fra andre sprog end engelsk må derimod principielt medtages, hvis de er almindelige: *modus vivendi*, *berufsverbot*, *macho*, *vendetta*, da de jo kunne have en særlig engelsk form. Der må selvfølgelig sættes en grænse et sted. Jeg har fx udeladt de musikalske tempobetegnelse: *allegro*, *andante*, *largo* etc. etc. ud fra en forestilling om at de nok er kendt blandt dem der har brug for dem, men det kunne man sige om så mange andre.

Afgrænsningen ved de engelske låneord er særlig besværlig. Ækvivalenten er naturligvis i vid udstrækning identisk med lemmaet, så det kan synes overflødigt at fylde ordbogen med substantiver som *bestseller*, *carport*, *computer*, *drink*, *gentleman*, *hit*, *job*, *sandwich*, *teenager*, adjektiver som *fair*, *fifty-fifty*, *hot*, *in*, *live*, og verber som *lease*, *guide*, *line op*. Formen kan være ændret og derfor begrunde optagelse: *double*, *single* og *dart* har på engelsk „-s“, *happy end* hedder „happy ending“, til verberne er der undertiden tilføjet et *-e*: *blende*, *briefe*, *mathe*, *sniffe*, *speede*, eller de konstrueres anderledes: *kjolen matcher med jakken* „the dress matches the coat“. Det er desuden ikke sikkert at alle genkender dem som engelske eller er sikker på at formen er den samme, fx er jeg blevet spurgt om *toastmaster* hedder det samme på engelsk. Der findes endelig på dansk en del „engelske“ ord der enten ikke findes på engelsk eller har en anden betydning: *babylift* „carrycot“, *butterfly* „bow tie“, *cottoncoat* „raincoat“, *dressman* „male model“, *drop* „drip“, *shine op* „smarten up“, *smoking* „dinner jacket“, *speak* „commentary“, *speaker* „announcer“, *speeder* „accelerator“. Da en bruger aldrig kan vide om et engelskudseende ord han støder på hører til denne gruppe, er det nødvendigt også at medtage en del almindelige anglicismer som de ovennævnte, selv om deres ækvivalenter er identiske.

4.3 Det grammatiske:

Behandlingen af bøjningsformer, sammensætninger, afledninger og proprier

Verbalparticipier der dannes af verbets grundform efter de almindelige regler bliver enten ikke opført eller sat under grundformen. De kan dog have fået status som selvstændige adjektiver, især når de i mindre grad udtrykker handling. I den slags tilfælde er danske participiers engelske ækvivalent ofte et adjektiv: *bebrejdende* „reproachful“, *beskrivende* „descriptive“, *hæmmende* „inhibitory“, *underordnet* „subordinate“ og må derfor sættes som selvstændige lemmaer.

Antallet af optagne afledninger må begrænses. Mange er de forord hvor forfatteren anfører, at adverbier på „-ly“ og substantiver på „-ness“ afledt af adjektiver, af pladshensyn er blevet udeladt. I en dansk-engelsk ordbog kan man bruge det (noget tvivlsomme) princip at udelade dem når de svarer til danske afledninger på *-t* og *-hed*: *loyalt* „loyally“, *statsløshed* „statelessness“, og tage dem med hvis det ikke er tilfældet: *regelmæssighed* „regularity“, *vanskelighed* „difficulty“, men hvornår *det* er kan brugeren jo ikke forudse, og nogle ganske almindelige kan vanskeligt udelades, fx *døvhed* „deafness“. Med andre suffikser er det ikke mindre problematisk: *-løs* svarer ganske vist oftest til „-less“ (*ærmeløs*, *statsløs*) og er meget produktivt, og *-agtig* er i mange tilfælde „-like“ (*fugleagtig* „birdlike“), eller „-ish“ (*drengagtig* „boyish“), men *fløjsagtig* hedder „velvety“ og *ferskenagtig* „peachy“. Jeg mener ikke at løsningen er at sætte suffikserne som lemmaer. Den slags lemmaer er af tvivlsom værdi: Brugeren finder næppe på at slå et suffiks op, og gør han/hun det, er det ikke sikkert han/hun kan bruge den oplysning der står og nå frem til et rigtigt resultat.

Afledninger med præfiks er ligeså problematiske. Man kan ikke have alle ord på *u*-svarende til „un-“ og må følge det igen tvivlsomme princip at overlade det til brugeren at danne de ukomplicerede sammensætninger, men anføre dem, hvis de ikke dannes med „un-“: *ubevægelig* „immovable“, *ulogisk* „illogical“, *ulydig* „disobedient“, eller de er højfrekvente. Det samme gælder andre meget produktive præfikser: *af*-svarende til „de-“ (*afindustrialisere* „deindustrialize“) og *gen*-svarende til „re-“ (*genlæse* „reread“). Et præfiks kan også opføres som lemma, og her er der vel større sandsynlighed for at brugeren ser det: „af- præfiks de-“, men der er jo den fare, at nogle så tror at *afmontere* hedder „demount“ og ikke „dismantle“.

Adjektivaafledninger af personnavne er endnu et problem: hvilke personer er så betydningsfulde at de må med? Det afhænger af ordbogsforfatterens personlige smag og interesser. Yderligere dannes de på engelsk på forskellig måde: *Ibsensk* „Ibsenite“, *Kafkask* „Kafkaesque“, *Freudsk* „Freudian“, *Shakespearesk* „Shakespearean“.

Sammensætninger er nok ordbogsforfatterens største hovedpine. Man kan prøve at opstille visse kriterier for optagelse.

Længde: treleddede og derover optages kun i begrænset omfang, fx kan *generalforsamlingsbeslutning*, *arbejdskraftsbalancespørgsmål* udelades. Men *raketaffyringsrampe* og *masseødelæggelsesvåben* kan dog ikke undværes.

Leksikaliseringsgrad: ikke-leksikaliserede eller engangsord udelades, fx *lampebord* (modsat *lampeskærm*) eller *løberudruller* (brugt i en avis i forbindelse med et statsbesøg).

Gennemskuelighed: de gennemskuelige behøver ikke medtages, specielt ikke hvis de er ens på dansk og engelsk: *morgenvind* „morning wind“, *natarbejde* „night work“, *ungdomsklub* „youth club“, *dørmåtte* „doormat“.

Men brugeren kan aldrig vide om en given sammensætning, transparent eller ej, også dannes ved simpel sammensætning på engelsk, som de nævnte. Engelsk har især tre andre almindelige modsvarigheder:

1. adjektiv + substantiv: *ungdomsforbryder* „young offender, juvenile delinquent“, *kernefamilie* „nuclear family“.
2. Præpositionsforbindelse: *ungdomsideal* „ideal of one's youth“, *skyldfølelse* „sense of guilt“, *interessekonflikt* „conflict of interests“.
3. Genitiv + substantiv: *fuglerede* „bird's nest“, *dukkehus* „doll's house“.
4. Og endelig kan det være et helt andet ord: *narkoselæge* „anaesthetist“, *perrontunnel* „subway“.

Ordbogsforfatteren er derfor alligevel nødt til at foretage et udvalg, da alle ikke kan komme med. Som der står i forordet til ODS (s. XXVII) kan man lige så lidt forvente at en ordbog har alle sammensætninger med som at en regnebog har hele talrækken (i forvejen er $\frac{3}{4}$ af alle opslagsord i ODS sammensætninger, jf. Egholm-Pedersen 1994:72). Jeg har valgt den udvej at udelade dem der kan dannes som på dansk og tage så mange som muligt af dem der ikke gør. Det princip er naturligvis utilfredsstillende: den bruger der søger en bestemt sammensætning og ikke finder den, kan ikke vide om den er udeladt fordi den dannes ved simpel sammenstilling, eller den bare er oversat, og en række af de ukomplicerede sammensætninger bør tages med alligevel da de er så almindelige: *natklub*, *bordben*, *stoleryg*.

Der er nogle substantivsammensætninger hvor førsteleddet kan variere, om ikke uendeligt så dog ret vidt, fx dem med andetled som: *-branche*, *-regning* (*varme-*, *el-*, *mælke-* etc.) og *-time*: hvis man tager *sprogtime* med, melder alle andre fag sig: *fysik-*, *matematik-* etc. etc., listen kan forlænges. Det er et spørgsmål om man også i disse tilfælde skal nøjes med at opføre andetled som lemma, men man bør vel under alle omstændigheder tilstræbe en vis konsekvens. Her opviser praksis, inklusive min egen tidligere, et broget billede: har man *byghøst*, *rughøst*, *hvedehøst* bør man nok også have *havrehøst*, *rishøst*, *æblehøst*, *roehøst*, har man *damebekendtskab*, *herredouble* bør man også have *herrebekendtskab*, *damedouble* (og *-single*).

Et særligt problem er talordene som jo er ord i sproget og hver for sig kunne være lemmer, herunder sammensætninger med talord, fx *-alder*: *i tiårsalderen*, *tyveårsalderen*, *tredivårsalderen* etc. Praksis opviser også her et broget billede: nogle har slet ingen, andre en enkelt og ikke andre. Jeg har fulgt det princip at nøjes med tiårene, da de mellemliggende vel er mindre almindelige, bortset fra dem under ti. Ved adjektiver afledte af talord + *-årig* kunne man måske tage dem allesammen, da mennesker jo sjældent bliver over 100, eller indføre en artikel med *-årig* som lemma, men vil brugeren tænke på at slå op der?

Ved ordenstallene og heraf afledte brøker endende på *-del* er praksis i eksisterende ordbøger igen ret forskellig og uforklarlig. V & B tager: (1) Ordenstallene + brøkerne op til

nittende, nittededel, minus femtendedel, syttendedel, attendedel. (2) Ordenstallene til tierne fra *tyvende* til og med *halvfemsindstyvende*, men af brøker kun *tyvendedel, tredivtedel, tresindstyvendedel*. Gyldendals Røde 1984 har ordenstallene til og med *nittende*, men af brøkerne kun *ellevtedel, tolvtedel, sekstendedel, nittededel*, mens ordenstallene og brøkerne til tierne er behandlet som i V & B. Gad 1993 har: (1) Ordenstallene op til *nittende*, dog minus *sekstende, syttende*, samt brøkerne kun op til *tolvtedel*. (2) Ordenstallene til tierne kun fra *tyvende* til *fyrretyvende*, og af brøkerne kun *tyvendedel*. Ingen ordbøger (undtagen *Retskrivningsordbogen*) har også *enogtyvendedel, toogtyvendedel* etc.

Adjektivsammensætninger hvis første led kan varieres med mange ord som er beslægtede frembyder et lignende problem og lige så forskellig praksis, fx dem med *-sproget, -talende* som andetled. Forleddet kan være *fransk, tysk, swahili* etc. etc. (en nøjes med *dansk- og engelsk-*). Eller *-venlig (børne-, bruger-. . .), -farvet, -malet (rød-, blå-, grøn-, orange- etc.)*.

Proprier vil man næppe vente at finde i en dansk-engelsk ordbog, bortset fra geografiske navne som har en særlig engelsk form: *Køln „Cologne“, København „Copenhagen“, Kärnten „Carinthia“, Schlesien „Silesia“*. Personnavne slet ikke, ud over historiske: *Henrik „Henry“, Ludvig „Louis“, Vilhelm „William“, Knud (den Store) „Canute“, Jakob* (også som apostelnavn) *„James“,* og andre bibelske navne: *Matthæus „Matthew“, Johannes „John“* (også som pavenavn). At Brynildsen 1927 har *Jens „Jack“, Jeppe „James“, Hans „Hans, Jack, John, skotsk Jock“, Ole „Auley“, Olsen „Macaulay“* hører til kuriositeterne.

Tegneseriefigurer spiller en så vigtig rolle i tidens mytologi, at de også bør med: *Knold og Tot „the Katzenjammer Kids“, Gyldenspjæt „Jiggs“* i serien „Educating Father“, og de nu om dage fra *Radiserne „the Peanuts“* mere kendte *Søren Brun „Charlie Brown“, Trine „Lucy“* m.v., foruden hele Disneys galleri: *Anders And „Donald Duck“, Fedtmule „Goofy“, Onkel Joakim „Scrooge McDuck“*. Praksis er også her højst forskellig: man finder *Gyldenspjæt*, men ikke *Rasmine, Nuser* men ikke *Trine*, og grænsen er unægtelig svær at trække.

Med hensyn til forkortelser har jeg i ordbogen begrænset mig til de helt almindelige: *s.u., f. eks. samt bl. a.* (hvor man ikke har noget lige så almindeligt alternativ). Enkelte velkendte internationale forkortelser som *Nato* fortjener optagelse, det kunne jo være de dækkede over noget dansk, som *FN „UN“,* men hvor langt skal man gå? Skal *WHO „World Health Organization“* med? *IBRD „International Bank for Reconstruction and Development“* kan man finde optaget.

4.4 Det „stilistiske“:

Slang, tabuord og fagord

Det er en velkendt sag at slangord ofte har en kort levetid. De er af begrænset værdi for en der vil udtrykke sig på engelsk, da forkert anvendte slangord vil virke pinlige på en indfødt, og derfor skal de behandles med forsigtighed. Helt udelade dem bør man nok ikke, de kan være nødvendige i en oversættelse, men man gør sig umage med at prøve at vurdere hvilke der har en fremtid for sig. Tabuordene har ligeledes en meget høj frekvensgrad i talesproget, og da de er så få bør de vel optages, selvom der for dem gælder det samme forbehold som ved slang.

Fagordene er til gengæld en meget stor gruppe. Spang-Hanssen (1988) siger at deres antal er langt større end det almene ordforråd, så her er en kraftig sortering nødvendig. Det oven for citerede kriterium fra forordet til tillægget til V & B er det nærmeste jeg kan

komme til et princip. Det er nemmere at se hvad man i hvert tilfælde ikke behøver. Ord som i denne tekst (fra en stillingsannonce i Magisterbladet 18/11 93): *Isoformen af glutamat decarboxylase i de Langerhanske øer* falder klart udenfor. Det er den slags ord som man kun vil støde på i videnskabelige sammenhænge, og som de der har brug for dem kender i forvejen.

Men ligesom det kan ske at forældede ord pludselig kommer til live igen, kan det ske at ret specielle fagord pludselig dukker op i den offentlige debat og bliver almindeligt kendte. Det gælder, fra atomkraftdebatten *formeringsreaktor*, fra rumfarten *bremseraketter*, *varmeskjold*, fra foreningsdebatten *algeopblomstring*, *stråforkortningsmiddel*, og fornylig *vækstfremmer*, og et eksempel er også den uhyggelige aktualitet et maritimt fagord som *bovport* fik fornylig.

Det kan næppe undgås at netop udvalget af fagord vil blive præget af redaktørens eller redaktørernes personlige interesser. F. eks. kan man tydeligt se at mange ordbøger laves af cand.mag.'er, for hvem ord fra fonetikken: *triftong*, *stemmeridse*, *lukkelyd*, *stavelseaccent*, og grammatikken: *hensigtsbisætning*, *sætningskløvning*, *verbaladjektiv*, *nomen*, er ganske dagligdags og har fået indpas i ordbøger af begrænset omfang som Gyldendals Røde 1984, hvor ord af en lignende specialiseringsgrad fra andre fagområder ikke er blevet optaget. Jeg har udryddet en del sådanne i den Røde 1995.

Som det vil fremgå af det forudgående er praksis på flere områder stærkt varierende, hvilket illustrerer de problemer lemmaselektionen volder. Det bliver i den sidste ende redaktørens personlige skøn og præference der kommer til at afgøre tvivlstilfælde. I forordet til ODS er dette udtrykt flere steder, fx „Sluttelig maa det fremhæves, at valget af de ord, der medtages, og af dem, der udskydes, i mange tilfælde maa blive vilkaarligt“ (s. XXVII; jf. også s. XXIV og XXIX). Man kunne i forordet til enhver ordbog skrive de ord som undertiden bliver sagt ved afslutningen af tv-avisen i Danmark: „Dette var hvad vi valgte at bringe“.

Bibliografi

Ordbøger

Brynildsen 1927 = John Brynildsen: *Norsk-Engelsk Ordbok*. Tredie omarbejdede Udgave. Oslo: H. Aschehoug & Co.

Gad 1993 = *Gads store ordbøger engelsk-dansk dansk-engelsk*. Redaktion Anna Garde. København: G. E. C. Gads Forlag.

Gyldendals Røde 1984 = Jens Axelsen: *Dansk-engelsk Ordbog*. 9. udgave. [1995 10. udgave.] København: Gyldendal.

ODS = *Ordbog over det danske Sprog* 1919–54. Udgivet af Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

Retskrivningsordbogen = *Retskrivningsordbogen* 1986. København: Dansk Sprognævn/Gyldendal.

V & B 1990 = Hermann Vinterberg/C. A. Bodelsen: *Dansk-engelsk Ordbog*. 3. udgave v. Viggo Højmager Petersen. [2. udgave ved C. A. Bodelsen, medredaktører: Jens Axelsen, B. Kjærulff Nielsen, Edith Frey; tillæg 1974 udarbejdet af Jens Axelsen.] København: Gyldendal.

Anden litteratur

- Bergenholtz, Henning/Jens Erik Mogensen 1993: Wörterbuchkritik in Dänemark. I: *Lexicographica*, International Annual for Lexicography 9. Tübingen: Max Niemayer Verlag, 8–35.
- Egholm-Petersen, Sv. 1994: ODS-Supplementet. Præsentation af bind 1: A(aa)–Bh 1992. I: Anna Garde/Pia Jarvad (red.) *Nordiske Studier i Leksikografi* II. København: Nordisk forening for leksikografi, 69–77.
- Høybye, Poul 1971: En oversættelsesordbog. I: Axel Andersen (red.): *Danske opslagsværker* 9. København: G. E. C. Gad, 299–324.
- Pope, Alexander 1711: *An Essay on Criticism*. London.
- Spang-Hanssen, Henning 1988: Det faglige ordforråd. I: Erik Hansen/Jørn Lund (red.): *Sproget her og nu*. København: Gyldendal, 92–102.
- Svensén, Bo 1993: *Practical Lexicography*. Oxford/New York: Oxford University Press.

Kristin Bakken

Leksikalisering av gammelnorske sammensetninger

In linguistic theory compounds are recognized as occupying an intermediate position between the syntax and the lexicon. They share features both with typical syntactic expressions and with typical lexical expressions. The process by which compounds gradually develop from largely syntactic items to typical lexical items is in this paper termed the lexicalization process, and a hypothetical model of this process is presented. This hypothesis is then tested at one specific point by the use of an Old Norse material of noun + noun compounds. If the lexicalization model is a true reflection of the way compounds actually behave, one can predict that compounds which are lexicalized to a large extent, probably will show a larger tendency toward being written as one graphic unit (i.e. without a space), than those which are lexicalized to a lesser degree. The Old Norse material confirms this prediction, and the lexicalization model is thereby strengthened as a hypothesis.

1 Disposisjon

I dette foredraget skal jeg kort innlede om sammensetningers stilling innenfor lingvistisk teori. Deretter skal jeg presentere i hypoteseform en modell for hvordan leksikaliseringprosessen tenkes foregå, og til slutt skal jeg teste denne leksikaliseringshypotesen mot et gammelnorsk materiale på ett punkt, nemlig prediksjonen om at sterkt leksikaliserede sammensetninger trolig vil sammenskrives i materialet.

2 Innledning

Komplekse ord som sammensetninger og avledninger, er klassiske eksempler på språkstørrelser som det er vanskelig å kategorisere absolutt. De kan sies å befinne seg i grenseland mellom syntaks og leksikon. Avhengig av det fokuset og den teori språkforskeren legger an, kan f eks sammensetninger beskrives ved hjelp av syntaktiske regler eller som typiske ord lagret i et leksikon.

Eksempler på språkforskere som har fokusert på det syntaktiske aspektet ved komplekse ord, er Lees (1960) som beskriver sammensetninger innenfor tidlig transformasjonell syntaktisk teori, Aronoff (1976) som beskriver avledninger innenfor rammene for generativ morfologi, nærmere bestemt innenfor leksikalistisk morfologi, Selkirk (1982) som vil beskrive komplekse ord ved hjelp av frasestrukturregler og bruker til det en variant av \bar{X} -syntaks, og Lieber (1992) som også beskriver ordintern struktur ved hjelp av syntaktiske modeller.¹

¹Felles for disse lingvistene er at de alle arbeider innenfor rammene av generativ teori, og trolig skyldes ikke dette bare tilfeldigheter. Hvis man i utgangspunktet har som ideal å unngå redundans i grammatikkene, vil

På den annen side er det klart at det er lang tradisjon for å behandle og beskrive sammensetninger som **ord**, dette gjelder ikke minst blant leksikografene. Når de fleste intuitivt vil kategorisere sammensetninger som ord, er dette et uttrykk for at sammensetninger deler en del vesentlige karakteristika med andre ord. Allerede Bloomfield (1933:180ff) vier sammensetningene plass fordi de representerer et kritisk punkt i forsøket på å definere ordet. De har karakteristika som er typiske både for ord og syntaks. Spencer gir en mer oppdatert versjon av den samme diskusjonen (Spencer 1991:310ff), og Bybee (1985:106) ender opp med å plassere sammensetninger i en mellomstilling mellom ord og syntaks på følgende grunnlag:

Compounding is distinct from all other combinatory processes of language in that it has the characteristics of both *syntactic* and *lexical* expression. Compounding resembles syntactic expression in that the units combined also always exist independently as words — that is, they are complete both phonologically and semantically. Compounding resembles lexical expression in that the resulting unit is a *word*, and the meaning of this word is not predictable from a summation of the meaning of its parts. Even though compounding may be productive in the sense that new compounds are freely created, the results of the compounding process are lexicalized, and they tend gradually to lose their semantic and phonological transparency.

Det er klart at sammensetninger for en del kan beskrives ved hjelp av syntaktiske regler i selve lagingsøyeblikket. I germanske språk er sammensetning en av de aller mest produktive prosessene, og en ny sammensetning kan lages når som helst for å dekke et kodingsproblem. Forholdet mellom første- og andreledd i lagingsøyeblikket kan f.eks. beskrives ved hjelp av syntaktiske begreper som underordning, sideordning eller objektsforhold og genitivsforhold.

Etter at selve den formelle ordlagingen er fullført, er det nylagete ordets videre skjebne i språket usikker. Noen ord blir aldri gjentatt, mens andre vinner allmenn aksept. Her vil jeg presentere en modell for sammensetningens utvikling fra nylaget sammensetning med intern syntaktisk struktur, til etablert sammensetning uten slik intern struktur. Denne prosessen kaller jeg i tråd med f.eks. Bybee i sitatet over, leksikaliseringprosessen, og den kan defineres som vegen fra syntaks til leksikon (jf Bybee 1985:106).

Det blei tidlig klart for meg i arbeidet med gammelnorske sammensetninger i hvilken grad ordforrådet er relikvisk, og dermed hvor begrenset verdi det har å forsøke å formalisere ordlagingsreglene på bakgrunn av et synkront snitt gjennom ordforrådet. Aronoff (1976:35ff) sier eksplisitt at ordlagingsregler bare kan settes opp for produktive ordlagingsprosesser. Dette kan synes opplagt, men innenfor f.eks. leksikalistisk generativ morfologi er dette slett ikke alltid en forutsetning når ordlagingsregler formuleres. Koblingen mellom regler og produktivitet er et vilkår for at de reglene en setter opp har hypotesestatus. Hvis denne koblingen ikke gjøres vil reglene kun være generaliseringer over finite domener (jf Dyvik 1986:152f).

man regelgenerere mest mulig og lagre minst mulig. Det er derfor ikke helt uventet at det er innenfor generativ grammatikk at man har vektlagt de syntaktiske aspektene av sammensetningene så sterkt. På den annen side har man alltid sett det syntaktiske aspektet ved sammensetningene. Bak tradisjonelle termer som 'bahuvrihi', 'tatpuruṣa' og 'dvandva', opprinnelig hentet fra de gamle indiske grammatikerne, ligger det jo også syntaktiske klassifiseringer (henholdsvis underordning i eksosentriske sammensetninger, underordning i endosentriske sammensetninger og sideordning).

Det er en grunntanke for meg at hver enkelt sammensetning starter å leve sitt eget liv fra det øyeblikket den konvensjonaliseres i ordforrådet. Den utviklingen som hver enkelt sammensetning følger er i grove trekk lik, men poenget blir at den enkelte sammensetning står på forskjellig sted i prosessen hvis en sammenligner ordforrådet på et synkront tidspunkt. I min terminologi kan det uttrykkes som at sammensetningene er leksikalisert i ulik grad.

Før jeg fortsetter å gjøre nærmere rede for leksikaliseringprosessen, må det nevnes at materialet som jeg bruker som testbakgrunn, er et ekserpert materiale av substantiviske hypotagmer fra den første halvdel av 1300-tallet. Jeg har ekserpert alle diplomene fra denne perioden som kan stedfestes til Vestfold fylke. Jeg omtaler dem bevisst som substantiviske hypotagmer fordi sammenskrivings- og særskrivingskonvensjonene i denne perioden fortøner seg som svært vakkende, og det finnes få sikre kriterier for å utskille bare de prosodiske sammensetningene (dvs sammensetning med én ordaksent) fra tekstene. Dette kommer jeg tilbake til under punkt 3 under. Materialet består av 1296 eksempler.

3 Leksikaliseringshypotesen

Formuleringa av leksikaliseringshypotesen trekker vekslers på framstillingene i Bauer (1983), Lipka (1977 og 1992), og i Grimm (1991).

Idet to ord settes sammen til ett, regner jeg med at det eksisterer et semantisk forhold mellom ledda som kan beskrives ved hjelp av syntaktiske begreper. I norske sammensetninger er det vanligvis et førsteledd som nærmere spesifiserer andreleddet. Førsteleddet har aldri spesifikk referanse, men kun generell referanse (jf Diderichsen 1941:102, Dyvik 1979 og Spencer 1991:312). Det vil si at førsteleddet ikke peker på én spesiell referent i verden, men kun på førsteleddets klasse av denotata, eller mer spesielt, førsteleddet har ikke referanse som førsteledd, kun betydning.

Her støtter jeg meg til den semantiske modellen som presenteres i Lyons (1977:174ff). I denne modellen er meningsbegrepet tredelt. Lyons skiller mellom bruksavhengig referanse, en betydning eller et begrep (sense) som er abstrahert vekk fra alle konkrete representasjoner, og en klasse av mulige denotata som utgjør den kategorien i verden som faller inn under begrepet. Så vidt det går fram av Lyons' framstilling (jf Lyons 1977:209f) er klassen av denotata ikke fast avgrenset, men åpen for nye ekstensjoner.

I lagingsøyeblikket har altså sammensetningen komposisjonell betydning, dvs at helhetens betydning er et produkt av begge sammensetningselementenes betydning. Samtidig vil sammensetningen i lagingsøyeblikket lages i en bestemt kontekst, dvs at sammensetningen i lagingsøyeblikket brukes med spesifikk referanse. Den nye sammensetningens denotatum er sekundær til den første referansen sammensetningen gis (jf Lyons 1977:224ff).

Mens man i bruken av et simpleks som *hund* er regulert av det settet av denotata som betydningen muliggjør, er bruken av en nylaget sammensetning regulert av de to ulike leddas betydning og av forholdet mellom disse to leddas betydning. Så lenge det er dette forholdet mellom to leddas betydning som regulerer hvilken referanse det nye uttrykket kan gis, vil jeg regne sammensetningens betydning som komposisjonell.

Leksikaliseringprosessen tar altså en sammensetning med komposisjonell betydning som input. Denne spontanlagingen kan eventuelt konvensjonaliseres. Dette regner jeg som leksikaliseringssens første fase. Konvensjonalisering defineres her som at forbindelsen mellom et gitt innhold og et gitt uttrykk blir automatisert. Dette innebærer at et gitt uttrykk

entydig utløser forestillingen om et gitt innhold. Til forskjell fra spontanlagingene avhenger denne assosiasjonen ikke av konteksten. Forskjellen i kontekstavhengighet kan illustreres ved et par eksempler. At en spontanlaging som *nøkkelbord* skal forstås avhenger av den konteksten det forekommer i, f eks:

I resepsjonen oppbevares ikke nøklene i en nøkkelhylle, men på et nøkkelbord.

Ved konvensjonaliserte sammensetninger er betydningen entydig også uten kontekst, jf sammensetninger som *nøkkelknippe* eller *nøkkelhull*. Med konvensjonalisering siktes det her til noe som skjer i forholdet mellom ordets uttrykks- og innholdsside, og tanken er at dette først og fremst skjer i det enkelte individs psyke. En sammensetning kan altså i ekstreme tilfeller være konvensjonalisert for meg, men ikke for andre. Fra produksjonssida betyr dette at en konvensjonalisert sammensetning ikke konstrueres i produksjonsøyeblikket, men at den som helhet hentes fram. Fra mottagersynspunkt betyr dette at sammensetningen er forståelig uten et tillegg av forklarende kontekst.

Individets språk kan likevel ikke ses uavhengig av det språklige fellesskapet, og jeg regner med at det til enhver tid er stor grad av overensstemmelse mellom individets språk og fellesskapets. Enkeltords betydning er et godt eksempel på dette, språket fungerer ikke om individer konvensjonaliserer språktegn uavhengig av andre språkbrukere. Jeg regner derfor med at sammensetninger konvensjonaliseres for den enkelte språkbruker, men at dette skjer i kontinuerlig interaksjon med fellesskapets språkbruk. I de aller fleste tilfeller vil derfor et ord som framstår som konvensjonalisert for meg, også oppleves slik av andre språkbrukere i min språkbrukerkrets. Selv om jeg her prinsipielt vil regne en sammensetnings leksikaliseringsgrad som to distinkte størrelser i en kommunikasjonssituasjon, nemlig henholdsvis sammensetningens leksikaliseringsgrad hos sender og hos mottaker, betyr ikke dette at det i praksis vil være store forskjeller. Oftest vil konvensjonaliseringsfasen også innebære en konvensjonalisering mellom individer.

Den andre fasen i leksikaliseringsprosessen kan sies å karakteriseres ved begrepssammensmelting. Igjen er det i det individuelle systemet forandringen tenkes skje, og igjen er det meningssida til uttrykket som forandres. Fra å ha komposisjonell betydning, går nå uttrykket over til å ha ikke-komposisjonell betydning. Betydningen framstår ikke lenger som summen av delene. Dette kommer best til uttrykk ved at man ikke lenger kan bytte ut en del av uttrykket med et synonym uten at helhetens betydning forandres. Noen eksempler kan illustrere dette. Mens man i *nøkkelhull* kan erstatte *hull* med *åpning*, og *knippe* med *bunt* i *nøkkelknippe* uten at betydningen blir endret, er ikke dette mulig i en sammensetning der det har skjedd begrepssammensmelting. (Sammensetningene *nøkkelåpning* og *nøkkelbunt* er selvfølgelig spontanlaginger og som sådan ikke konvensjonalisert, men de er i den tilsvarende konteksten synonyme med de konvensjonaliserte orda.)

Ser man på uttrykk der begrepssammensmelting har skjedd, slik som i *brannmann*, *barnehage* og *visittkort*, lar delene seg ikke skifte ut på samme måte.

En brannmann hentet katten ned fra treet.

? En ildmann/brannkar hentet katten ned fra treet.

Det er en barnehage preget av foreldet pedagogikk.

? Det er en ungehage preget av foreldet pedagogikk.

Han la igjen et visittkort.

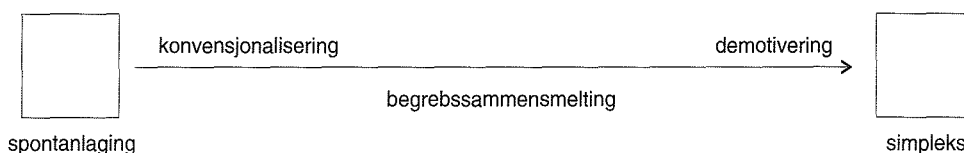
? Han la igjen et besøkskort/visittark.

Av disse eksemplene går det fram at betydningen til disse sammensetningene ikke lenger er komposisjonell i den forstand at helhetens betydning er en funksjon av betydningen til de to ledda som inngår. Det har skjedd en begrepssammensmelting, slik at meningselementer ikke lenger kan utskilles fra helheten. Ser en på sammensetningsleddas opprinnelige betydning og helhetens betydning, er det likevel klart at den nye sammensetningens betydning fremdeles fortøner seg som motivert utifra de ledda som inngår.

Fase tre i leksikaliseringsprosessen endrer på dette, idet denne kjennetegnes ved demotivering. Her er sammensetningens betydning så avvikende i forhold til det utgangspunktet leddas betydning gir, at uttrykket ser umotivert ut i forhold til betydningen. Det har ofte vært anført at tegnet i Saussures mening er arbitrært, men allerede Saussure selv modifiserte dette ved å innføre termene absolutt og relativt arbitrære tegn (jf de Saussure 1983:130ff). Sammensetninger er blant eksemplene på relativt arbitrære eller relativt motiverte tegn. Sammensetningene er motiverte uttrykk i den grad uttrykkselementene som inngår i dem er semantisk motivert. I demotiveringsfasen nærmer sammensetningen seg det prototypiske arbitrære tegnet, idet uttrykket ikke lenger kan sies å være motivert ut ifra innholdet. Gitt dette definisjonskriteriet slutter de å være sammensetninger.

Det er en forutsetning for demotiveringsfasen at sammensetningen allerede har fått en betydning av ikke-komposisjonell type. Så lenge delene bidrar til helhetens betydning, sperer dette for de ulike demotiveringsprosessene. Eksempler på demotiverte sammensetninger er *jordmor*, *nøkkelost*, *brudgom*, *snippkjole*, *beksømstøvel*. I alle disse har ett eller flere av ledda en betydning som virker umotivert i forhold til helhetens betydning.

For å oppsummere kan følgende figur tegnes.



Figur 1. *Leksikaliseringsskalaen*

Leksikaliseringsprosessen blir med vilje her kalt en skala, idet det er klart at disse fasene glir over i hverandre. Det er vanskelig å tenke seg denne prosessen som annet enn en gradvis prosess uten faste grenser. Som input tar leksikaliseringsprosessen spontanlagninger, eller førstegangsord uten hevd i språket. Endepunktet for leksikaliseringsprosessen er sammensetninger som er demotivert, hvor den indre semantiske og morfologiske strukturen er irrelevant, og som derfor ligner på simplekser. De er som simpleksene blitt arbitrære tegn uten intern struktur.

En kommentar kunne knyttes til påstanden om manglende intern struktur. Resonnementet er som følger: Når betydningen til en sammensetning ikke lenger er komposisjonell, har den morfologiske strukturen ikke noe semantisk korrelat, og den har slik sett ingen relevans for språkbrukerne. Den kan selvfølgelig rekonstrueres ved aktiv analyse fra språkbrukernes side, men i daglig bruk av ordet aktiviseres denne strukturen ikke.

Når den interne strukturen i sammensetningen er gjort irrelevant, åpner dette for både morfologisk reanalyse, folkeetymologiske omdanninger og for fonologiske endringer, som

ellers ville ha vært forhindret ved forestillingen om morfologisk struktur. Jeg skal ikke komme inn på disse fenomenene her, men nevne eksempler som *barsel*, *fføs* og *stakkar* som viser sammensetninger der det har foregått lydendringer som har hatt som følge at komposisjonsgrensene i disse orda er utydeliggjort. Disse fonologiske endringene er uttrykk for sterk leksikaliseringsgrad, og de bringer sammensetningene enda mer i harmoni med simpleksene. Leksikaliseringsprosessen er her likevel gitt en semantisk definisjon. Det vil si at de fonologiske endringene bare er bieffekter av leksikaliseringen, selv om de kan tenkes virke tilbake på leksikaliseringsprosessen og forsterke den.

4 Leksikaliseringsgrad og skrivekonvensjoner i middelalderdiplomer

Leksikaliseringsskalaen presentert over har her status som en hypotese om hvordan sammensetningens veg fra syntaks til leksikon kan tenkes foregå. I avhandlingsarbeidet har jeg forsøkt teste denne hypotesen ut på et materiale av gammelnorske substantivfølger (dvs substantiviske hypotagmer). Jeg har merket et materiale på 1296 substantivfølger for leksikaliseringsgrad, henholdsvis svak, sterk eller usikker leksikaliseringsgrad. I dette ligger det metodiske problemet som jeg ikke skal drøfte nærmere her; men det første er velkjent for alle som driver med historisk lingvistik. Hvilke muligheter har vi i dag for å få tak på historiske språkstegs leksikalske semantikk? Jeg har brukt den enkelte kontekst, kunnskap om ordets betydning i moderne språk og ordets bruk i andre historiske kontekster slik det dokumenteres først og fremst i Fritzner (1883–1896 og 1972).

Et annet problem er hvordan en gradert skala som leksikaliseringsskalaen kan testes ut empirisk. Jeg har løst det ved å merke materialet for sterk, svak og usikker leksikaliseringsgrad, og slik sett valgt å måle en gradert virkelighet ved hjelp av en diskret målestokk. For nærmere drøfting av disse problemene henvises til Bakken (1995*).

På bakgrunn av leksikaliseringsskalaen slik den her er skissert, kan en gjøre noen statistiske prediksjoner. Hvis disse prediksjonene slår til i materialet, vil jeg regne leksikaliseringshypotesen som styrket. Her vil jeg i kortform presentere én slik prediksjon og påfølgende test.

Prediksjonen kan gjøres slik:

Hvis det er slik at sterk leksikaliseringsgrad betyr at sammensetningens morfologiske struktur ikke lenger har psykologisk relevans, er det mer sannsynlig at sterkt leksikaliserede substantivfølger blir sammenskrevet enn at svakt leksikaliserede substantivfølger blir det.

Det som her gjøres er å krysse et semantisk kriterium (leksikaliseringsgrad) med et grafisk kriterium (sammenskriving). Hvis disse to størrelsene samvarierer, anses prediksjonen som innfridd og leksikaliseringsskalaen styrket som teori.

Én forutsetning for at denne testen er meningsfull, er at det grafiske kriteriet faktisk er variabelt. I et moderne skriftspråkssamfunn er skriftnormen så sterk at det i praksis må regnes med liten individuell grafisk variasjon.² I perioden det her gjelder, norsk høymiddelalder, er

²Den moderne tendensen til å særskrive sammensetninger selv om skriftnormen tilsier noe annet, er slik sett et svært interessant og svært relevant unntak.

skriftnormen mye løsere, og det gis rom for en mye større grad av individuell variasjon innenfor de mer generelle konvensjonene som rådde. Et av de punkta hvor filologer tradisjonelt har vurdert normen som spesielt vaklende, er nettopp ved sammenskriving av sammensetninger (jf f.eks Diderichsen 1941:102, Lagman 1964). Det er på denne bakgrunnen at det gir mening å krysse leksikaliseringsgrad med variablene sammen- og særskriving. Fra materialet kan jeg nevne et par eksempler. Ordet *vitnisburðar* i formelen *till sanns vitnisburðar* sammenskrives 35 ganger og særskrives 8 ganger. I ett av eksemplene som sammenskrives, står *sanns* også sammenskrevet med *vitnisburðar*. Måleenheten *eyrisból* sammenskrives 3 ganger, og særskrives 10 ganger, mens *auraból* sammenskrives 8 ganger og særskrives 24 ganger. Det burde altså være klart at entydige fordelinger på dette punktet er vanskelige å finne, og det er da også dette som har frustrert filologer i beskrivelsen av emnet. Alle forsøk på å fange opp variasjonen som fonologisk eller prosodisk betinget møter selvmotsigelser som eksemplene over.

Prediksjonen over regner med at det statistisk sett vil være større tendens til sammenskriving ved de sterkt leksikalisererte substantivfølgene enn ved de svakt leksikalisererte substantivfølgene. Siden leksikaliseringsgrad i siste instans vurderes som en individuell størrelse, og siden ulike individ trolig vil analysere eget språk i ulik grad, er unntak til de statistiske fordelingene å vente.

Materialet som helhet fordeler seg slik:

	Sammenskriving		Særskriving		Sum	
	antall	andel	antall	andel	antall	andel
Sterkt leksikalisererte	426	73%	155	27%	581	100%
Svakt leksikalisererte	26	6%	425	94%	451	100%
Usikkert leksikalisererte	61	23%	203	77%	264	100%
Sum	513	40%	783	60%	1296	100%

Tabell 1. *Substantivfølgenes leksikaliseringsgrad koblet til variabelen sammen- eller særskriving.*

Tendensen her er klar. De sterkt leksikalisererte substantivfølgene viser en klar tendens til sammenskriving i forhold til de svakt leksikalisererte substantivfølgene. 73% av de sterkt leksikalisererte sammenskrives mot 6% av de svakt leksikalisererte.

Disse tallene kan kommenteres på flere ulike måter, og jeg vil bare skissere hvordan dette blir gjort i avhandlingen. For det første kan det innvendes at det blant de svakt leksikalisererte substantivfølgene nok er en overvekt ikke-sammensetninger (etter prosodiske kriterier), og særskrivinga her er derfor trolig heller motivert ut ifra prosodiske forhold enn ut ifra semantiske forhold.

Det er klart at prosodi er en vesentlig faktor i reguleringa av sammen- og særskriving i denne perioden. Når det først er understreket at sterk leksikaliseringsgrad og status som prosodisk sammensetning ofte følges at, kan leksikaliseringsgrad brukes som en forklarende tilleggsfaktor. Tilfeller som i lys av prosodiske forhold ser anomale ut, kan i lys av faktoren leksikaliseringsgrad bli lettere å harmonisere med mønstrene i materialet. I materialet har jeg plukket ut alle de substantivfølgene som etter formelle kriterier må regnes som prosodiske sammensetninger (stammesammensetninger som *lǫgbók*, og substantivfølger med etterhengt

bestemt artikkel som f eks *laugardaginn*). Disse viser en klar overvekt av sammenskriving, men det er også et faktum at de fleste av disse er sterkt leksikalisert, og følgelig peker både prosodi og semantikk her i samme retning. Det er også et faktum at 37 mot 158 av disse prosodiske sammensetningene særskrives. Som forklaring på denne typen avvik kunne det være aktuelt å se på leksikaliseringsgraden til disse avvika, men materialet mitt er for lite til å trekke sikre slutninger på grunnlag av fordelingene her.

Det kunne også spørres hvor vesentlig individuell variasjon er i denne sammenhengen. Etter å ha vurdert åtte individuelle skrivere opp mot hverandre, kan syv skrivere sies å stå på linje med materialet som helhet. Selv om enkeltteksemplene som knyttes til den enkelte skriver er relativt lite, kan det også skisseres en tendens der de kongelige skriversne, altså de som presumptivt er best skolert, er de som best innfrir prediksjonen over. De kongelige klerkene Håkon Ivarsson, Ivar Audunsson og Pål Styrkársson sammenskriver 90% eller flere av de sterkt leksikaliserte substantivfølgene, mens de sammenskriver 5% eller færre av de svakt leksikaliserte substantivfølgene.

En nærmere undersøkelse av hyppige enkeltord i materialet viser at enkeltteksempler med sterk leksikaliseringsgrad som regel sammenskrives, mens de svakt leksikaliserte som regel særskrives. Tendensen er mindre klar ved substantivfølger som jeg har markert som usikre, jf *auraból* og *eyrisból* over.

5 Konklusjon

Etter å ha testet et materiale av substantivfølger fra 1300-tallet mht samvariasjon mellom leksikaliseringsgrad og sammenskriving evt særskriving, ser det ut til at sterk leksikaliseringsgrad virkelig er en faktor som hemmer særskrivingstendensen. Selv om mange forbehold bør tas — det viktigste er usikkerhetsmomentet som hefter til hvor stor andelen av prosodiske sammensetninger i materialet er — gir denne testen likevel støtte til leksikaliseringsskalaen slik den er formulert. Siden særskriving jo i siste instans er uttrykk for en morfologisk analyse, er det mindre sannsynlig at skriversne i sin skriving gjenfinner en ordintern struktur som fra betydningssynspunkt er gjort irrelevant.

Litteratur

- Aronoff, Mark 1976: *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs 1. Cambridge MA: MIT Press.
- Bakken, Kristin 1995*: *Leksikalisering av sammensetninger. En studie av leksikaliseringsprosessen belyst ved et diplommateriale fra 1300-tallet*. Upublisert manus til doktoravhandling.
- Bauer, Laurie 1983: *English Word-Formation*. Cambridge: Cambridge University Press.
- Bloomfield, Leonard 1933: *Language*. Chicago/London: The University of Chicago Press.
- Bybee, Joan 1985: *Morphology. A Study of the Relation between Meaning and Form*. Typological Studies in Language 9. Amsterdam: John Benjamins.
- Grimm, Ursula 1991: *Lexikalisierung im heutigen Englisch am Beispiel der Ableitungen*. Tübinger Beiträge zur Linguistik 360. Tübingen: Gunter Narr Verlag.
- Diderichsen, Paul 1941: *Sætningsbygningen i Skaanske Lov*. København: Ejnar Munksgaard.

- Dyvik, Helge J. J. 1979: Omkring fremveksten av artiklene i norsk. I: *Maal og Minne* 1/2, 40–78.
- Dyvik, Helge J. J. 1986: *Grammatikk og empiri. En syntaktisk modell og dens forutsetninger*. Bd I. Skriftserie. Nr 24. Serie B. Institutt for fonetikk og lingvistikk. Universitetet i Bergen. Bergen: Universitetet i Bergen.
- Fritzner, Johan 1883–96: *Ordbog over Det gamle norske Sprog. Omarbeidet, forøget og forbedret Udgave I–III*. Kristiania: Den norske Forlagsforening.
- Fritzner, Johan 1972: *Ordbog over Det gamle norske Sprog IV. Rettelser og tillegg ved Finn Hødnebo*. Oslo/Bergen/Tromsø: Universitetsforlaget.
- Lagman, Edvin 1964: Om nominala sammansätningar i fornsvenskan och i nutida svenska dialekter. I: *Arkiv för Nordisk Filologi* 79, 172–194.
- Lees, Robert B. 1960: *The Grammar of English Nominalizations*. Indiana University Research Center in Anthropology, Folklore, and Linguistics 12. Bloomington: Indiana University Press.
- Lieber, Rochelle 1992: *Deconstructing Morphology. Word-Formation in Syntactic Theory*. Chicago: The University of Chicago Press.
- Lipka, Leonhard 1977: Lexikalisierung, Idiomatisierung und Hypostasierung als Probleme einer synchronischen Wortbildungslehre. I: Herbert E. Brekle/Dieter Kastovsky (red.): *Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.–10. Juli 1976. Anlässlich des 70. Geburtstags von Hans Marchand am 1. Oktober 1977*. Bonn: Bouvier Verlag Herbert Grundmann, 155–163.
- Lipka, Leonhard 1992: [1. utg. 1989] *An Outline of English Lexicology*. Forschung & Studium Anglistik 3. Tübingen: Max Niemeyer Verlag.
- Lyons, John 1977: *Semantics*. Volume 1. Cambridge: Cambridge University Press.
- Saussure, Ferdinand de 1983: *Course in General Linguistics*. [Engelsk oversettelse ved Roy Harris.] London: Duckworth.
- Selkirk, Elizabeth 1982: *The Syntax of Words*. Cambridge MA: MIT Press.
- Spencer, Andrew 1991: *Morphological Theory. An Introduction to Word Structure in Generative Grammar*. Oxford: Blackwell.

Leksikografi. Hvad er det?

Lexicography is not a linguistic discipline and specially not a genuine part of lexicology. Lexicography is not only the compiling of dictionaries, but theory and practice for dictionaries, lexica and encyclopaedias.

1 Sproglige og faglige kategorier

I 1930'ernes fonologiske debat og 70'ernes ordklassediskussion blev der skelnet mellem to grupper: (1) tilhængere af guds sandhed og (2) hokus-pokus-tilhængere. Principielt mener jeg, at den anden gruppe havde de bedste argumenter. Dette betyder dog ikke, at en hvilken som helst opdeling er acceptabel, men at forskellige opdelinger er hensigtsmæssige til forskellige sammenhæng. I forbindelse med fastlæggelse af indhold og antal af ordklasser betyder det, at en henvisning til en angivelig guddommelig eller ideologisk inspiration kan betegnes som en sådan hensigtsmæssig sammenhæng, men må betegnes som en immuniseringsstrategi:

Multi plures, multi vero pauciores partes esse dixerunt. Modo autem octo universalis tenet ecclesia. Quod divinitus inspiratum esse non dubito. [Nogle har hævdet, at der er flere, andre, at der i sandhed er færre ordklasser. Hele kirken mener, at der kun er otte. Jeg er overbevist om, at dette synspunkt beror på guddommelig inspiration.]

(Smaragdus, 9. årh., cit. efter Michael 1970:51)

Die marxistisch-leninistische Auffassung vom Wesen und zur Einteilung der Wortarten ist wiederholt dargestellt worden. Dabei ist man sich einig, daß die Wortarten Ausdruck der begrifflich-kategorialen Widerspiegelung der objektiven Realität sind.

(Sommerfeldt 1977:150, som skelner mellem ti ordklasser)

Tværtimod vil vi ideelt få præcist lige så mange ordklasser, som vi i en bestemt situation har brug for. Men det er på den anden side upraktisk, hvis hver enkelt grammatiker laver sit eget system eller måske flere forskellige systemer, som fx Bergenholtz (1976) med fem klasser, Bergenholtz/Schaeder (1977) med 51 klasser og Bergenholtz (1984) med ti klasser.

Det er denne upraktiske situation, kategorien eller rettere disciplinen leksikografi finder sig i. Der foreligger så mange forskellige opfattelser, at det er forstyrrende for den leksikografiske diskussion. Nogle opfattelser kan ganske vist afvises med henvisning til, at de hverken begrundes eller svarer til en forhenværende tradition. Det gælder fx for Jacoby (1990) i sin monografi om nordisk leksikografi, når han betegner "die Lehre von den Lexika" som leksikologi.

For en lang række andre opfattelser gælder det, at hver enkelt leksikograf, leksikolog, glossolog, encyklopædist, terminograf tilsyneladende har sin egen opfattelse, som han eller

hun anser for den eneste rigtige. Mange opdelinger er så specielle, at de næppe kan eller bør overføres til generelle opfattelser af leksikografi. Det er fx tilfældet, når Lindemann (1994:2) i sin behandling af franske ordbøger før 1600 skelner mellem glossografi som lære om ordlister og leksikografi som læren om ordbøger og leksika.

Nu er det ikke sådan, at jeg forestiller mig, at alle tilhørerne vil kunne være enige i alle dele af min opfattelse. Det vil jeg ganske vist ikke have noget imod. Mere væsentligt er det at give oplæg til en debat, som vi i Norden har brug for, så vi får bedre kendskab til de forskellige opfattelser, hvilket igen er forudsætningen for at nå frem til større entydighed og evt. til enighed.

2 Leksikografiske definitioner af *leksikograf* og *leksikografi*

Almensproglige ordbøger kan og skal heller ikke være fagligt helt nøjagtige, men de gengiver ofte en forståelse, som har stor udbredelse. Det gælder også for udtrykkene *leksikograf* og *leksikografi*, hvor funktionen (skrivning) og objektet (ordbog hhv. leksikon) udgør elementerne i den leksikografiske definition:

leksikograf person, som giver sig af med leksikonarbejde (Nudansk ordbog 1992)

Leksikograf Ordbogsskriver (Dansk Ordbog for Folket 1907)

Leksikografi Ordbogsskrivning (Dansk Ordbog for Folket 1907)

lexicography [. . .] the compiling of dictionaries (Longman 1992)

Meget sigende er indordningen af leksikografers virke som noget, man ”giver sig af med”. Forskellen til beskrivelsen af ingeniør i samme ordbog er påfaldende:

ingeniør videnskabeligt uddannet tekniker
(Nudansk ordbog 1992)

En tilsvarende beskrivelse af leksikograf er i øvrigt ikke mulig, da enhver, der udarbejder ordbøger eller skriver om ordbøger kan kalde sig og bliver kaldt leksikograf. Men hvis vi i stedet går fra den personrelaterede angivelse til disciplinen, kan man i nogle ordbøger finde angivelse af leksikografiens videnskabelighed:

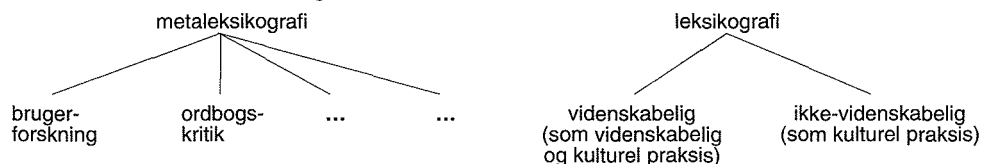
Lexikographie [Wissenschaft von der] Aufzeichnung u. Erklärung des Wortschatzes in Form eines Wörterbuchs (Duden GWB 1978)

Lexikographie Wissenschaft von der Zusammenstellung und Erklärung des Wortschatzes in Wörterbüchern und Lexika (Handwörterbuch 1984)

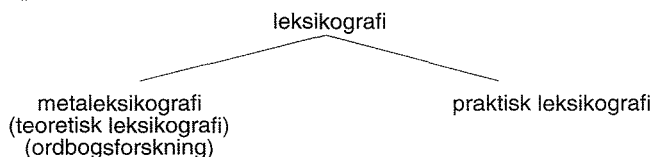
I den leksikografiske definition fra Duden bør man bemærke den kantede parentes, som angiver delinformationens fakultativitet, altså enten er den videnskabelig, eller den er det ikke. Da var man mere entydig i ordbogen fra det tidligere DDR, hvor man i Akademie der Wissenschaften havde ansat op mod 100 videnskabsfolk, der udarbejdede ordbøger. Da ordbogen blev udarbejdet i dette videnskabelige akademi, hvor de ansatte videnskabsfolk både udarbejdede ordbøger og leksika, kan det selvfølgelig ikke undre, at de betegnede deres

arbejde som videnskab. Den hermed antagede altomfattende videnskabelighed kan man ikke have ment deskriptivt, det er snarere en præskriptiv angivelse, som kan være forbundet med et ønske om at overføre akademiets praksis til en alment gældende praksis. Begge ordbøger tager ikke højde for, at man dels kan skelne mellem teori og praksis, dels mellem forskellige former for leksikografisk praksis.

Den manglende dimension i de nævnte definitioner er i den nyere leksikografiske diskussion inddraget gennem en opdeling i **praktisk leksikografi** og **metaleksikografi**. Men det er ikke sådan, at udtrykket **metaleksikografi** overtages af alle ordbogsforskere. Fx er Schaefer (1992) yderst kritisk. Han skelner mellem praktisk leksikografi og teoretisk leksikografi. Udtrykket *meta-* i forbindelse med leksikografi anser han for overflødig, ja direkte forvirrende. Så mener han, skulle man også tale om **metametaleksikografi**, når man forsker i **metaleksikografiske** resultater. Man har jo heller ikke **geografi** og **metageografi** eller **ortografi** og **metaortografi**, siger han. Dette argument er dog svagt, mener jeg. Der findes fx **metaetik** (som ikke regnes som en del af etikken) eller **metamatematik** (som regnes som en del af matematikken). Problemet er snarere, hvordan den praktiske leksikografi forholder sig til **metaleksikografien** og hvordan **leksikografien** forholder sig til andre discipliner. I mit svar går jeg i stor udstrækning ud fra Wiegand (1989), dog uden at følge den helt, da Wiegand ikke forudsiger en fælles betegnelse for leksikografi og **metaleksikografi**. Der deles ind i en videnskabelig og en ikke-videnskabelig leksikografi svarende til den leksikografiske virkelighed, hvor en lang række ordbøger ikke kan siges at have nogen relation til overvejelser, som kan kaldes videnskabelige.



Jeg vil foretrække at beholde leksikografi som fælles betegnelse for både konception af og udarbejdelse af ordbøger, dvs. praktisk leksikografi, og for den ikke direkte praksis-orienterede disciplin, dvs. **metaleksikografi**:



Den foreløbige skelnen mellem praktisk og teoretisk leksikografi siger intet eller kun lidt om leksikografiens relation til andre discipliner. Her vil jeg i det følgende diskutere nogle forslag til forståelse af sådanne relationer på en sådan måde, at nogle bliver beskrevet som for omklamrende, andre som for afgrænsende.

3 Leksikografi ctr. anvendt lingvistik

En meget udbredt opfattelse ser leksikografien som en del af lingvistikken, som anvendt lingvistik. Hvis leksikografien ikke kun har (sprog)ordbøger som objekt, men også sagleksika,

encyklopædier og såkaldte alordbøger hhv. encyklopædiske ordbøger som objekt, ville det medføre, at lingvistikken måtte anses for det overordnede område for alle videnskaber, i det mindste i forbindelse med kodificering af viden i form af leksikografiske opslagsværker. Der kan ikke være tvivl om, at indsigt i bestemte sprogvidenskabelige områder vil være fordelagtigt, ja påkrævet, hvis en række ordbøger, leksika og encyklopædier skal konciperes og udarbejdes på optimal vis. Men ved at anvende lingvistiske teorier bliver hverken udarbejdet metaleksikografiske teorier eller konkrete ordbøger. Lingvistiske kundskaber er ikke tilstrækkelige. Der kræves ud over sagviden i særdeleshed indsigt i ordbogsstrukturer og andre ikke-lingvistiske, men centrale leksikografiske problemstillinger. Af samme grund kan heller ikke den såkaldte sprogleksikografi siges at være en genuin del af lingvistikken. Her vil det være naturligt og i de fleste tilfælde nødvendigt, at sprogleksikografen har en lingvistisk uddannelse. Af samme grund vil megen leksikografisk forskning finde sted på sproginstitutioner, men også her er udtrykket anvendt lingvistisk misvisende i forhold til praktisk og teoretisk leksikografi.

4 Leksikografi ctr. leksikologi

En anden udbredt opfattelse ser leksikografi som en del af leksikologien hhv. som anvendt leksikologi. Det er uden tvivl også baggrunden for følgende forståelse af leksikografi under brug af det synonyme udtryk "ordbogsvidenskab":

lexikologi vetenskapen om ordförrådets uppbyggnad och struktur □
 äv. inkluderend्रे ordbogsvidenskab (Svensk Ordbok 1986)

Hertil svarer den leksikografiske definition i mange lingvistiske leksika og almene encyklopædier:

Lexicology is the study of the origin and meaning of words; lexicography is practice of dictionary making. Lexicology is a theoretical science; lexicography, an applied science. Good dictionary makers, however, inevitably are both lexicologists and lexicographers. (Grolier 1993)

En sådan opfattelse bygger på to forudsætninger, som ikke er hensigtsmæssige i relation til objektet "leksikografisk opslagsværk":

1. Leksikografi forstås som sprogleksikografi — hvorved objekterne leksikon og encyklopædi udelukkes og albogsleksikografi negligeres.
2. Leksikografi forstås alene som praktisk leksikografi som udarbejdelse af ordbøger — hvorved metaleksikografi udelukkes eller negligeres.

Der kan stadig ikke være tvivl om, at det vil være fordelagtigt for al leksikografi, at der bygges på kendskab til leksikologiske fremstillinger. Men det er ikke sådan, som Thomsen (1995) citerer et mundtligt udsagn fra H. P. Kromann: "Leksikografien stiller spørgsmålene, leksikologien forsøger at besvare dem." Leksikologien besvarer nogle af leksikografiens overordnede spørgsmål, men ikke alle detaljespørgsmål, og beskæftiger sig frem for alt ikke med leksikografisk så vigtige områder som brugerbehov, ordbogsstrukturer, layout, fonologi eller ortografi.

5 Leksikografi ctr. encyklopædologi

Hvor jeg i de første to tilfælde har prøvet at påvise, at en omklamring fører til en u hensigtsmæssig forståelse af leksikografien, vil jeg i to andre tilfælde argumentere for det omvendte forhold. En indskrænkning af leksikografi til sprogleksikografi, fører til, at vigtige leksikografiske problemstillinger udelukkes eller negligeres. Den afgørende fællesnævner for leksikografien er teori og praksis for leksikografiske opslagsværker, både for papirbøger og andre medier. De i kap. 2 citerede leksikografiske definitioner af leksikografi angiver dels ordbøger dels både ordbøger og leksika som objekt, men samtidig anser selvsamme bøger disse betegnelser for at være enten polyseme eller homonyme, fx:

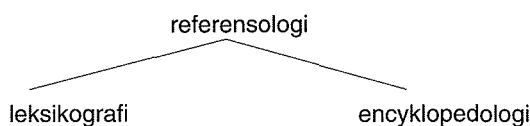
leksikon [...] dels om værk med alfabetisk ordnede artikler indeholdende faglige oplysninger [...] dels om ordbog (Nudansk Ordbog 1992)

Andre skelner meget skarpt mellem ordbog og leksikon. Hertil svarer en skarp opdeling mellem sprog og sag og en dermed forbunden opfattelse, som indskrænker leksikografi til kodificering af sproglige enheder:

Ifølge hevdvunne forestillinger beskjeftiger både leksikografien og encyklopedologien seg med *alfabetiske oppslagsverk*. I ordbøger er emnet (gjenstanden for artikkelskrivning) *ord* (språklige enheder), i encyklopedier ting og foreteelser, *verdens tilskikkelser*, realoplysninger.

(Henriksen 1992:219)

Herudfra foretager Henriksen følgende opdeling:



At denne opdeling ikke tager hensyn til alordbøger hhv. encyklopædiske ordbøger og af uforståelige grunde kun indregner alfabetiske opslagsværk som objekt for "referensologi", er ikke hovedproblemet. Det grundlæggende problem er et andet: Kan der virkelig skelnes klart og entydigt mellem "real-leksikografi" og "ord-leksikografi"? Jeg er mere end skeptisk og vil samtidig tage afstand fra egne tidligere fremstillinger, hvor jeg for klart har skelnet mellem sprogleksikografi, sagleksikografi og albogsleksikografi. Hvis vi går ind på at diskutere tilfældige definitioner i monolingvale ordbøger, kan denne skelnen ikke opretholdes, se hertil Bergenholtz (1994) og Bergenholtz/Kaufmann (1995), som i denne sammenhæng bygger på Haiman (1980). I stedet for at gentage de der bragte argumenter vil jeg henvise til Møller (1959) og Pedersen (1994), som begge — uden at tage den fulde konsekvens — nærmer sig de synspunkter, som Haiman (1980) har fremlagt:

I det hele taget kan man da kun modtage et klart indtryk af, at de programagtige, teoretiske udtalelser om den (næsten) absolutte forskel mellem ordbog og leksikon hviler på et meget løst grundlag, og i praksis har man, så vidt det kan ses, kun i ringe grad haft en sådan vejledning for øje, når en forklaring til de forskellige ord blev udformet. (Møller 1959:55)

Når så både Pedersen (1994) og Møller (1959) alligevel i praksis mener at kunne trække en praktisk grænse, kan det forklares mere entydigt, end det bliver gjort:

Til trods for at der ikke kan drages en blot nogenlunde sikker grænse mellem rent sproglige oplysninger til forståelse af ordene og realoplysninger om disse, har man således i praksis en tydelig opfattelse af, at en eller anden omstændelig saglig redegørelse i hvert fald hører hjemme i et såkaldt realleksikon.

(Møller 1959:55)

Den ikke nævnte, men sandsynlige grund for en sådan praktisk grænsedragning, ligger ikke i en genskabelse af grænsen mellem ord og sag, men er en hensyntagen til den her ikke nævnte brugergruppe: Det drejer sig hverken om en fagekspert eller en semiekspert, men en lægmand, for hvem der forudsættes et vist kendskab til den fælles kultur. Samtidig kan man gå ud fra, at begge monolingvale almensproglige leksikografer i særlig grad vil tage hensyn til det første og i en vis grad også den anden af de følgende grundlæggende brugssituationer, hvorimod et monolingvalt leksikon eller en encyklopædi normalt både tager hensyn til den første og den sidste brugssituation.

1. reception af L1-tekster
2. produktion af L1-tekster
3. reception af L2-tekster
4. produktion af L2-tekster
5. oversættelse fra L1 → L2
6. oversættelse fra L2 → L1
7. tekstuafhængig oplysning om sprog
8. tekstuafhængig oplysning om en sag

Hvor der på "leksikografiske" kongresser ofte holdes entydigt leksikologiske foredrag, som jeg som leksikograf finder fejlplacerede, hører man sjældent bidrag fra og om monolingvale leksika og encyklopædier. Jeg har det håb, at denne tilstand skal kunne ændre sig — til fordel for både teori og praksis i ordbøger, leksika og encyklopædier.

6 Leksikografi ctr. terminologi

Hvis man ved leksikografi forstår sprogleksikografi, bliver fagsprog tilbage som objekt for en anden disciplin, det vil i praksis sige terminologi eller rettere terminografi. En sådan opdeling svarer på sin vis til store dele af leksikografiens historie, hvor al væsentlig metaleksikografi har behandlet almensproglige ordbøger samt leksika og encyklopædier. Samtidig har kvaliteten af de fleste fagordbøger været yderst ringe, særligt fordi, de har været kendetegnet af en manglende fagsystematisk tilgang til det fagsproglige ordforråd. At påpege denne mangel og udarbejde modeller til en anden tilgang har været terminologiens styrke, som den siden 1930'erne er blevet initieret og præget af Wüster (1936).

En forsonlig holdning har set en art symbiose mellem leksikografi og terminologi:

Die Arbeit an Übersetzungsfachwörterbüchern umfaßt Terminologie und Lexikographie. Als Lexikograph muß der Autor den bestehenden Fachsprachschatz in alphabetischer Reihenfolge möglichst vollständig erfassen. Als Terminologe muß er Begriffsfelder erarbeiten und vergleichen, inwieweit die Teile

eines Fachwortes zum Begriffsinhalt oder Begriffsumfang passen oder nicht passen. (Bucksch 1973:7)

Man skal således være både leksikograf og terminolog, hvis man vil koncipere og udarbejde en fagordbog. Andre ser en art arbejdsopdeling, hvor leksikograferne bygger på terminologernes resultater, idet terminologi opfattes som den systematiske indsamling af fagtermer, mens der ved leksikografi forstås udarbejdelsen af ordbøger og leksika (Andersen 1979:7). Grundholdningen hos de fleste leksikografer og terminologer er dog en anden. Alment kan man sige, at leksikografer indtil nu kun i ringe grad har forholdt sig til den grænsedragning, hvorimod terminologer i høj grad har forsøgt at fremhæve formodede og virkelige forskelle. Alt i alt kan man finde følgende grundholdninger om relationen mellem fagleksikografi og terminologi:

- (A) Terminologi hhv. terminografi har fagsprog som objekt, hvorimod leksikografi har almensprog som objekt. Denne opfattelse fører i praksis til en afskaffelse af udtrykket fagleksikografi.
- (B) Terminografi er en del af terminologien, mens fagleksikografi er en del af leksikografien. Terminografi og fagleksikografi har ordforrådet i fagsprog som objekt, men er klart adskilte discipliner, idet de har forskellige opgaver, forskellige teoretiske udgangspunkter og forskellige metodiske fremgangsmåder.
- (C) Fagleksikografi, hvortil der findes det synonyme udtryk terminografi, er en del af leksikografien. Derudover findes der terminologien, som i firmainterne, brancheinterne, nationale og internationale gremier søger at udarbejde fælles saglige og sproglige normer.

Løsning (A) er entydig: Ethvert deskriptiv eller præskriptiv arbejde med fagtermer er en opgave for terminologien hhv. dets delområde terminografi. Også løsning (C) er på sin vis klar. Den svarer til opdelingen i HSK 13, den planlagte håndbog i fagsprogsforskning, hvor tre af de 28 kapitler har overskriften Fagleksikografi I, II og III. Sidstnævnte har tilføjelsen "die Terminographie im 20. Jahrhundert". Derudover findes der et eget kapitel med overskriften Terminologi. Denne løsning er faktisk udarbejdet under deltagelse af både leksikografer og terminologer, men før vi evt. vil kunne tilslutte os den, må vi tage stilling til løsning (B): Foreligger der virkelig klart forskellige opgaver, teorier og metoder?

Efter 1975 er det blevet almindeligt at antage et terminologisk delområde, som kaldes **terminografi**. I overensstemmelse med ISO 1087 skal dette udtryk erstatte "terminological lexicography" og "special lexicography" (Picht 1985:353). Adskillelsen mellem terminologi hhv. terminografi og leksikografi bliver begrundet med en række argumenter, som påviser en forskel eller en modsætning til leksikografien:

- (1) Terminologien har udelukkende fagsprog som objekt (Dubois 1979:57f., Rey 1979:239, Picht 1985:353, Thomas 1993:44, Grinsted/Toft 1994:71)
- (2) I terminologien arbejdes der kun med fageksperter som informanter, hvorimod man i leksikografien kan bruge enhver native speaker (Frandsen 1979:24f)
- (3) Terminologien er computeriseret, dens resultater fremlægges ikke bare i form af papirordbøger, men også på disketter eller CD-ROM (Riggs 1989:107)

- (4) Terminologien udarbejder hjælpemidler til fagekspertes
- (5) Terminologien udarbejder hjælpemidler til tekstproduktion (Riggs 1989:90, Roald 1987:11)
- (6) Terminologien er rent synkron (Wüster 1985:102, Roald 1987:13, Svensén 1992a:107, 1992b:25)
- (7) Terminologien er primært præskriptiv, mens leksikografien primært har et deskriptivt sigte (Roald 1987:10f., Riggs 1989:90, Maidahl Christiansen/Duvå/Laursen 1994:272)
- (8) I terminologien arbejdes der onomasiologisk (Riggs 1989:89, Maidahl Christiansen/Duvå/Laursen 1994:276)
- (9) I terminologien findes ingen polysemi, men der er nogle gange flere termer til et begreb (Riggs 1989:107, Roald 1987:9)
- (10) I terminologien går man ud fra en fagsystematik (Roald 1987:6f., Felber/Budin 1989:139)
- (11) Terminologer beskriver begrebsammenhæng og -tilordninger, hvorimod leksikografer bruger lingvistiske metoder (Felber/Budin 1989:139)
- (12) I modsætning til leksikografien foretrækker man i terminologien en systematisk makrostruktur fremfor en alfabetisk makrostruktur (Roald 1987:6, Riggs 1989:91, Felber/Budin 1989:139, Picht 1985:366ff.)
- (13) I terminologien arbejder man med begreber og termer og ikke med sproglige tegn (Felber/Budin 1989:139)

De fleste af disse argumenter gælder dog ikke kun for terminologien hhv. terminografien, men også for den del af leksikografien, der kaldes fagleksikografi. Der kan i (1), (2), (3), (6), (7), (9) og (10) indsættes fagleksikografi i stedet for eller sammen med terminologi. Betydningen af (8) er ikke helt entydig, men hvis der menes ca. det samme som (10) gælder dette argument også for fagleksikografi. Om punkt (11) kan der siges, at der her foreligger en misforståelse; om punkt (12), at denne modsætning ikke er reel. Kun under punkt (13) kan der findes en virkelig forskel i valg af semiotisk teori.

Om de enkelte argumenter kan der siges meget, det har jeg gjort i et selvstændigt bidrag (Bergenholtz 1995). Jeg vil dog ikke undlade nogle få kommentarer.

Ad (2): Selvfølgelig kan man i fagleksikografien kun arbejde med fagekspertes som informanter. Men også i den almensproglige leksikografi vil man, når man behandler bestemte fagrelaterede ord, inddrage fagekspertes. Det fremgår af de fleste store fagordbøgers forord (se hertil Bergenholtz 1994).

Ad (3): Det er historisk rigtigt, at terminologer meget tidligere end mange leksikografer, herunder også fagleksikografer, har gjort brug af computeren som hjælpemiddel. Men i dag er også næsten al leksikografi "computeriseret".

Ad (7): I nogle lande er sprogleksikografien først og fremmest deskriptiv (fx i Danmark). I andre lande er den langt mere præskriptiv (fx på Island). Men tendentielt er fagleksikografien

i alle lande stærkere præskriptiv end den almensproglige leksikografi. Jeg har dog været ude for et forlag, som ønskede præskriptive udsagn i en fagordbog fjernet med bemærkningen, at det ikke kan være en ordbogs opgave at oplyse, om en bestemt brug af et ord er forkert. Det kan evt. diskuteres i almensproglig sammenhæng, ikke i fagsproglig.

Ad (9): Hertil kan der siges, at der faktisk findes en del almensproglige ordbøger, som principielt ikke giver mulighed for polysemi, men lemmatiserer hver betydning i forskellige semantiske felter, fx Verben in Feldern 1986. Argumentet i fagsproglig sammenhæng er dog svagt, sml. Roald (1987:9):

Polysemi er derfor ikke noe relevant begrep i TG [= terminologien, HB]. Tilfeller hvor ett identisk uttrykk representerer to eller flere begreper, betraktes som homonymi. Således vil f.eks. "morfologi" som leksem i LG [= leksikografi, HB] være flertydig og fordeles på ev. tre betydninger, med fagområdehenvisning, innen samme ordartikel: (a) (lingv.), (b) (medisin), (c) (geologi).

I en medicinsk eller geologisk eller lingvistisk fagordbog vil denne homonymi ikke være relevant, idet andre fag ikke inddrages. I en flerfagsordbog eller i en almensproglig ordbog, vil der efter normal lingvistisk opfattelse ikke være polysemt, men homonymt. Om homonymer lemmatiseres under et indgangslemma eller under flere forskellige, er dog et andet spørgsmål. I øvrigt forekommer eksempler med klare tilfælde af polysemi i mange fagsprog ikke særligt hyppigt, fx har vi i Genteknologisk Ordbog 1992 blandt ca. 3000 ordbogsartikler kun tre med polyseme angivelser.

Ad (10): Mange fagordbøger kan med rette kritiseres for at behandle de enkelte fagtermer isoleret. At gå ud fra en fagsystematik er ikke blevet patenteret af terminologer, men er tværtimod blevet krævet og også brugt af fagleksikografer (fx Pedersen 1994a og 1994b).

Ad (11): Nu kan man diskutere, hvori lingvistiske metoder i leksikografisk arbejde kan bestå, men at beskrive begrebsammenhænge og -tilordninger, er nu ikke uden videre udelukket fra lingvistikken. I øvrigt kan der henvises til, at Wüster (1936:165) ud fra sit kendskab til lingvistiske metoder netop bygger på ordfeldmetoden, som den blev grundlagt af Jost Trier: "Jeder Begriff und seine Benennung sind also auf zweifache Weise in den Begriffsschatz der Gesamtsprache eingebettet: durch das Begriffssystem und durch das Bedeutungsnetz."

På sin vis er det noget overraskende, at (12) fremhæves så kraftigt af terminologer, når de samtidig med rette peger på andre medier end papirordbøger. På en CD-ROM er tilgangsstrukturen altafgørende for brugeren, hvorimod makrostrukturen ikke er væsentlig. Men selvfølgelig kan man fx give tilgang til oversigter over opslagsord indenfor samme knude i fagsystematikken, eller også give mulighed for at få alle eller nogle af de direkte fagrelaterede ordbogsartikler samtidig med den søgte artikel. Men også for papirordbøger er modsætningen ikke reel. Fx findes der en lang række begrebsordbøger og systematisk opbyggede encyklopædier. I praksis er der både fordele og ulemper ved en alfabetisk og også en systematisk makrostruktur (se hertil Bergenholtz/Tarp 1995). Bl.a. kan det være fordelagtigere at vælge en delvis systematisk makrostruktur, hvis en større fagordbog offentliggøres i mindre dele over en længere årrække. Man kan så fordelagtigt have en systematisk opdeling og ordne de enkelte afsnit alfabetisk (Wüster 1936:167). Denne blandede makrostruktur kan også være fordelagtig for etbindsværker. Det var den mest udbredte struktur i middelalderens fagordbøger, den findes også i nye etbindsværker, fx i den førnævnte tyske monolingvale

valensordbog (Verben in Feldern 1986). Det helt afgørende er efter min mening ikke i første omgang præsentationsformen, men den metodiske fremgangsmåde, hvor en fagsystematisk tilgang og metode er en væsentlig forudsætning for fagligt relevante oplysninger: "bei der Sammlung der in ein Fachwörterbuch aufzunehmenden Wörter kann man eigentlich gar nicht anders vorgehen als systematisch" (Wüster 1936:167).

Ad (13): Når det drejer sig om den semiotiske teori og det hermed forbundne teorisprog, har vi det eneste punkt med en klar forskel mellem fagleksikografi og den del af terminologien, der nu kaldes terminografi. Mens man i leksikografien taler om lemma, ordbogsartikel, sprogligt tegn med indhold og udtryk, bruger terminologer termer som begreb, begrebsindhold, benævnelse. I hvor høj grad disse teorier og termer gør et samarbejde eller i det mindste en gensidig diskussion umulig, er uklart. Jeg anser selv forskellene for at være noget mindre, end det nogle gange er blevet antaget (Bergenholtz 1995).

Alt i alt vil jeg — trods (13) — holde fast ved opfattelsen af den næsten fuldstændige synonymi mellem fagleksikografi og den del af terminologien, som beskæftiger sig med ordbøger, altså terminografien. Jeg kan her tilslutte mig Joly (1986:12):

Dois-je avouer que je n'ai jamais beaucoup cru à l'autonomie de la terminologie par rapport à la lexicographie? Je n'y vois guère que le prestige de la mode et la manie de parer de nouvelles dénominations une activité bien connue et fort ancienne. La terminologie est-elle vraiment autre chose que la redécouverte de la lexicographie par les techniciens, avec tout le côté systématique, théoricien et l'utilisation d'un arsenal de pensée et de concepts que les hommes de technique et d'organisation du travail ont l'habitude d'apporter à leurs diverses tâches de fabrication?

7 Afslutning

Leksikografien relaterer sig direkte og indirekte til den type håndbøger, som kan kaldes leksikografiske opslagsværker og som adskiller sig fra andre håndbøger ved at indeholde ordbogsartikler. Nogle af de vigtigste betegnelser for leksikografiske opslagsværker er ordbog, leksikon og encyklopædi. Disse og andre betegnelser bruges dog ikke alle i samme betydning og vil desuden alt efter opslagsværkstype blive forsynet med indskrænkende eller informative tilføjelser, som tager højde for den forudsete genuine brug. Et leksikografisk opslagsværk er kendetegnet ved at indeholde en række strukturer, det drejer sig om mindst 30 forskellige ordbogsstrukturer. De vigtigste er makrostruktur, mikrostruktur og tilgangsstruktur. Med leksikografisk opslagsværk menes ikke kun trykte bøger, men også andre medier som CD-ROM.

Litteratur

- Andersen, Eyvind 1979: *Ordbøger. Oversættelse. Terminologi*. Lyngby: Danmarks tekniske bibliotek.
- Bergenholtz, Henning 1976: *Zur Morphologie deutscher Substantive, Verben und Adjektive*. Bonn: Dümmler.
- Bergenholtz, Henning 1984: *Grammatik im Wörterbuch: Wortarten*. I: Herbert Ernst Wiegand (hrsg.): *Studien zur neuhochdeutschen Lexikographie IV*. Hildesheim/Zürich/New York: Olms, 19–72.

- Bergenholtz, Henning 1994: Faglige oplysninger i monolingvale betydningsordbøger. I: *LexicoNordica* 1, 13–26.
- Bergenholtz, Henning 1995 (i trykken): Wodurch unterscheidet sich Fachlexikographie von Terminographie? I: *Lexicographica* 11, 37–46.
- Bergenholtz, Henning/Uwe Kaufmann 1995 (i trykken): Enzyklopædische Informationen in Wörterbüchern. I: Nico Weber (hrsg.): *Theorie der Semantik und Theorie der Lexikographie*. Tübingen: Niemeyer.
- Bergenholtz, Henning/Burkhard Schaefer 1977: *Die Wortarten des Deutschen. Versuch einer syntaktisch orientierten Klassifikation*. Stuttgart: Klett.
- Bergenholtz, Henning/Sven Tarp 1995 (i trykken): Makro- og Mikrostrukturen i Fachwörterbüchern. I: *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Hrsg. von Lothar Hoffmann, Hartwig Kalverkämper, Herbert Ernst Wiegand in Verbindung mit Christian Galinski und Werner Hüllen. Berlin: de Gruyter.
- Bucksch, Herbert 1973: Das technische Übersetzungsfachwörterbuch. I: *Mitteilungsblatt für Dolmetscher und Übersetzer* 19.5, 6–8.
- Dansk Ordbog for Folket 1907 = *Dansk Ordbog for Folket*. Udarbejdet af B. T. Dahl/H. Hammer/Hans Dahl. Første Bind A–L. Andet Bind M–Ø. København/Kristiania: Gyldendalske Boghandel/Nordisk Forlag.
- Dubois, Claude 1979: La spécificité de la définition en terminologie (par comparaison avec la définition en lexicographie). I: *Actes du sixième colloque international de terminologie. Pointe-au-Pic (Québec) du 2 au 7 octobre 1977*. Québec, 45–59.
- Duden GWB 1978 = *Duden*. Das große Wörterbuch der deutschen Sprache in sechs Bänden. Hrsg. u. bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von Günther Drosdowski. Bd. 4 Kam–N. Mannheim/Wien/Zürich: Bibliographisches Institut.
- Felber, Helmut/Gerhard Budin 1989: *Terminologie in Theorie und Praxis*. Tübingen: Narr.
- Frandsen, Lene 1979: *Terminologi og Leksikologi*. Forskelle og Lighed. København: Handelshøjskolen i København.
- Genteknologisk Ordbog 1992 = Uwe Kaufmann/Henning Bergenholtz: *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi*. København: Gad.
- Grinsted, Annelise/Bertha Toft 1994: Leksikografi og terminologi. I: Viggo Hjørnager Pedersen/Niels Krogh-Hansen (red.): *Oversættelses Håndbogen*. København: Munksgaard, 61–75.
- Grolier 1993 = *Grolier Multimedia Encyclopedia*. Grolier.
- Haiman, John 1980: Dictionaries and Encyclopedias. I: *Lingua* 50, 329–357.
- Handwörterbuch 1984 = *Handwörterbuch der deutschen Gegenwartssprache*. In zwei Bänden. Von einem Autorenkollektiv unter der Leitung von Günter Kempcke. Berlin: Akademie-Verlag.
- Henriksen, Petter 1992: Forholdet mellem encyklopedologi og leksikografi. I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991*. Oslo: Nordisk Forening for Leksikografi, 219–227.
- Jacoby, Michael 1990: *Historische Lexikologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin*. Wiesbaden: Otto Harrassowitz.
- Joly, Hubert 1986: Itinéraire d'un lexicographe de fortune. I: *Language Monthly*. September '86, 12–14.
- Lindemann, Margarete 1994: *Die französischen Wörterbücher von den Anfängen bis 1600. Entstehung und typologische Beschreibung*. Tübingen: Niemeyer.

- Longman 1992 = Jack C. Richards/John Platt/Heidi Platt: *Longman Dictionary of language teaching and applied linguistics*. Second edition. Essex: Longman.
- Maidahl Christiansen, Lisbet/Grete Duvå/Anna-Lise Laursen 1994: Das Translationswörterbuch für Fachsprachen. Ein integriertes Konzept. I: Burkhard Schaefer/Henning Bergenholtz (Hrsg.): *Das Fachwörterbuch*. Fachwissen und seine Repräsentation in Wörterbüchern. Tübingen: Narr, 269–284.
- Michael, Ian 1970: *English Grammatical Categories and the Tradition to 1800*. Cambridge: University Press.
- Møller, Kristen 1959: *Leksikologi og leksikografi*. København: Schultz.
- Nudansk Ordbog 1992 = *Politikens Nudansk Ordbog*. Bd. 1 A–K. Bd. 2 L–Å. København: Politikens Forlag.
- Pedersen, Jette 1994a: Fagsystematik. I: Henning Bergenholtz/Sven Tarp (red.): *Manual i fagleksikografi. Udarbejdelse af fagordbøger — problemer og løsningsforslag*. Herning: Systime, 84–90.
- Pedersen, Jette 1994b: Anskueliggørelse af faglige relationer i brancheordbøger. I: *LexicoNordica* 1, 203–228.
- Pedersen, Karen Margrethe 1994: Ordbog og encyklopædi. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konferance om leksikografi i Norden 11.–14. maj 1993*. København: Nordisk Forening for Leksikografi, 248–253.
- Picht, Heribert 1985: Leksikografi — Terminografi. I: *Nordisk terminologikursus* 2. Bind 1, 351–372.
- Picht, Heribert 1992: Grænseflader mellem terminologi og vidensteknik. I: Bertha Toft (red.): *Proceedings af 2. nordisk symposium om Terminologi EDB & Vidensteknik*, [Kolding], 22–33.
- Rey, Alain 1979: Définition de la terminologie en tant que discipline linguistique autonome état: de la question. I: *Actes du sixième colloque international de terminologie. Pointe-au-Pic (Québec) du 2 au 7 octobre 1977*. Québec, 229–257.
- Riggs, Fred W. 1989: Terminology and Lexicography: Their Complementarity. I: *International Journal of Lexicography* 2, 89–110.
- Roald, Jan 1987: Leksikografi og terminologi: en sammenligning. I: Tove Jacobsen/Jan Roald (Red.): *Flerspråklig leksikografi og terminologi. Seminarrapport Bergen 22.–23. mai 1987*. Bergen: Universitetet i Bergen, 3–13.
- Schaefer, Burkhard 1992: Probleme einer Darstellung der Geschichte der deutschen Lexikographie. Oder: ein ordentliche erzählung und erklärungen waarhafter, grundtlicher und geschächner dingen (Maaler). I: Gregor Meder/Andreas Dörner (hrsg.): *Worte, Wörter, Wörterbücher. Lexikographische Beiträge zum Essener Linguistischen Kolloquium*. Tübingen: Niemeyer, 7–24.
- Sommerfeldt, Karl-Ernst 1977: Zur Klassifizierung der Wortarten. I: *Deutschunterricht* 30, 150–152.
- Störig, Hans Joachim 1986: Zur Abgrenzung der Lexikographie. I: Herbert Ernst Wiegand (hrsg.): *Studien zur neuhochdeutschen Lexikographie* VI. 1. Hildesheim/Zürich/New York: Olms, 183–195.
- Svensén, Bo 1992a: Terminologi och leksikografi. I: *Terminologiläran och dess relationer till andra områden*. Nordisk forskarkurs i Mariehamn, Åland, september 1990. Stockholm: NordTerm, 97–122.
- Svensén, Bo 1992b: Leksikografi och leksikografiska produkter — några grundbegrepp. I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991*. Oslo: Nordisk Forening for Leksikografi, 9–27.
- Svensk Ordbok 1986 = *Svensk Ordbok*. Utarbetad vid Språkdata Göteborgs universitet. Stockholm: Esselte Studium.

- Thomas, Patricia 1993: Choosing headwords from language-for-special-purposes (SLP) collocations for entry into a terminology data bank (term bank). I: Helmi B. Sonneveld/Kurt L. Loening (eds.): *Terminology. Applications in interdisciplinary communication*. Amsterdam: Benjamins, 43–68.
- Thomsen, Knud Troels 1995 (i trykken): [Anmeldelse af Sven Göran Malmgren: Svensk lexikologi]. I: *LexicoNordica* 2.
- Verben in Feldern 1986 = Helmut Schumacher (Hrsg.): *Verben in Feldern, Valenzwörterbuch zur Syntax und Semantik deutscher Verben*. Schriften des Instituts für deutsche Sprache 1. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst 1989: Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. I: Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta (hrsg.): *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*. Erster Teilband. Berlin/New York: de Gruyter, 246–280.
- Wüster, Eugen 1936: Wörterbücher der Technik. I: *Wörter und Sachen* 17, 164–173.
- Wüster, Eugen 1985: *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 2. Aufl. Kopenhagen. (1. Aufl. Wien/New York: Springer 1979)

Hvor finder man *finde anvendelse*?¹

Function verb constructions (FVC, Funktionsverbgefüge) / Support verb constructions are very frequently used in Danish and in German, but since the translations are rather unpredictable they should be described precisely in active dictionaries. In the introduction different types of constructions are briefly described. Then the treatment of FVCs with Danish *anvendelse* in dictionaries relevant for Danes translating into German is examined, revealing lacunas and inconsistencies. Results of a contrastive research based on parallel corpora in German and Danish are described, and finally a list of desiderata for a thorough contrastive lexicographic description is proposed.

1 Indledning

1.1 Afgrænsningsproblemer

Funktionsverber (forkortet FV) og **funktionsverbumbonstruktioner** (forkortet FVK), er problematiske på flere måder, specielt med hensyn til afgrænsning. For det første hører de til grænseområdet mellem grammatik og leksikografi. Man kan diskutere, hvad der hører hjemme i en grammatisk beskrivelse, og hvad der hører til i en ordbog. For det andet er de problematiske, fordi der er stærkt delte meninger om, hvad der forstås ved et FV og en FVK. Hvor går grænsen til andre klasser af verber respektive til frie forbindelser, og hvor hører de hjemme i et fraseologisk begrebssystem? En semasiologisk beskrivelse er altså vanskelig. Men onomasiologisk er de lige så problematiske, fordi der bruges forskellige betegnelser for begreber, som er mere eller mindre identiske, fx **funktionsverbum**, **støtteverbum** og „**Streckverbum**“.

I denne artikel tages der ikke stilling til, hvilken afgrænsning der bør foretrækkes. I de undersøgelser, der omtales, er inddraget konstruktioner, som af alle eller mange anses for FVK'er, og forhåbentlig kan de bidrage til at give et bedre grundlag for at foretage en afgrænsning. Tilsvarende gælder det for definitioner, at det endnu ikke er lykkedes nogen at give en alment accepteret definition af, hvad en FVK er, og det skal heller ikke forsøges her. Det er næppe muligt at give én enkelt definition, som dækker alt det, som man i almindelighed anser for FVK'er. Derfor gives i det følgende ikke definitioner, men blot beskrivelser af karakteristiske egenskaber ved forskellige typer af FVK'er.

1.2 Formål med arbejdet

Udgangspunktet for mit arbejde er et praktisk behov, nemlig behovet for at skaffe oplysninger, som er tilstrækkelige til at opfylde kravene i to specifikke oversættelsessituationer:

¹Dette arbejde har været muliggjort ved en forskningsbevilling fra Statens Humanistiske Forskningsråd til undersøgelser af de genteknologiske korpora.

(1) maskinoversættelse. (2) „Hinübersetzung“, altså oversættelse *til* fremmedsprog, in casu fra dansk til tysk.

De oplysninger, der skal bruges i en ordbog til maskinoversættelse, skal som bekendt være overordentlig omfattende og eksplicitte, da man ikke kan overlade noget til en ikke-eksisterende sprogfornemmelse.

I ordbøger til „Hinübersetzung“, altså aktive ordbøger, er det bl.a. et problem, hvor stor en del af den nødvendige information man kan give implicit, fx ved henvisning til generelle regler eller ved at lade brugeren udlede informationen af eksempler. Ordbogsforfattere stoler i mange tilfælde i høj grad på brugerens evne til at danne analogier og bruge sin common sense. Min erfaring som lærer i oversættelse siger mig, at hvis ordbogens formuleringer åbner mulighed for falske analogier, så bliver muligheden udnyttet! Jeg vil derfor plædere for, at man gør aktive ordbøger mere eksplicitte, og jeg vil til sidst komme ind på, hvad det indebærer, men i øvrigt ligger hovedvægten på indsamling og bearbejdelse af de relevante oplysninger.

2 Funktionsverber og funktionsverbumbonstruktioner

2.1 Definition / beskrivelse

En FVK består af et verbum og en størrelse, som enten *er* et substantiv eller som *indeholder* et substantiv. Dette substantiv kaldes ofte for **funktionsnomen** (FN).

Det lyder ikke som nogen særlig præcis beskrivelse af fænomenet. Ikke desto mindre er det det eneste, som de fleste, men ikke engang alle, er enige om.² Når man skal karakterisere FVK'er nærmere, begynder uenigheden for alvor. Mange forskere forsøger slet ikke at give en egentlig definition af fænomenet, men opregner blot en række egenskaber, der typisk gælder for konstruktionerne, og indfører samtidig en række forbehold som „in der Regel“, „weitgehend“, „in seltenen Fällen“. Da afgrænsningen ikke er mit hovedanliggende, vil jeg her gøre det samme.

2.1.1 Konstruktionstyper

Når det drejer sig om sprogene dansk-tysk, hører den helt overvejende del af de relevante konstruktioner til to hovedtyper, som her blot kaldes type 1 og 2.³ Type 1 omfatter konstruktioner som *træffe aftale med nogen om noget*, type 2 konstruktioner som *bringe noget i anvendelse*. Desuden er der en mindre gruppe, type 3, som af nogle regnes med til en af hovedtyperne. Som typeeksempel har vi *finde anvendelse*. Herudover findes der et større eller mindre antal yderligere konstruktioner, som man eventuelt kan betragte som FVK, afhængigt af, hvilke egenskaber man anser for at være afgørende kriterier.⁴

²Weinrich (1993:45) regner fx FN for en del af verbet.

³Type 1 og 2 er hovedtyper i den forstand, at de rummer de fleste og de almindeligste konstruktioner og derfor har størst praktisk betydning. Da de imidlertid i forhold til en lingvistisk beskrivelse ikke er overordnet, men sideordnet type 3 og eventuelle yderligere typer, er jeg gået bort fra at kalde dem „hovedtype 1–2“. Når afgrænsningsproblemerne er mere afklaret, bør der selvfølgelig findes en mere „sigende“ terminologi.

⁴Dyhr (1980) opererer fx med fem „Subklassen“ af tyske FVK'er på grundlag af morfosyntaktiske forskelle.

2.1.2 Fællestræk

Fælles for de tre typer er:

1. FN er et abstrakt substantiv, i de fleste tilfælde deverbalt, undertiden deadjektivisk.
2. Verbet har — i konstruktionen — ingen leksikalsk betydning i sædvanlig forstand.
3. Konstruktionens betydning er kompositionel. Dvs. den kan forstås ud fra enkelt-elementernes betydning. Herved adskiller FVK'er sig fra idiomatiske udtryk.
4. Det enkelte FN kan — i en FVK — kun forbindes med et enkelt eller nogle få verber (fra en lukket liste), mens det for de fleste FV's vedkommende er umuligt at opstille en endelig liste over mulige FN'er.

2.1.3 De tre typer

Herudover gælder for den første type (med typeeks.: *træffe aftale med nogen om noget*):

1. FN er direkte objekt på dansk, på tysk akkusativobjekt.
2. Det logiske subjekt for FN er samtidig grammatisk subjekt for verbet. Dette logiske subjekt kan uden for konstruktionen udtrykkes med en genitiv eller et ækvivalerende udtryk (poss. pron. eller præp. forb. med *af*, *von* eller *durch*).
3. Hvis FN har mulighed for at tilknytte andre valensbundne led, er mulighederne de samme i FVK'en og i andre anvendelser.
4. FN kan ofte modificeres med fx adjektiviske attributter. Modifikationsmulighederne kan ikke beskrives generelt, men er leksikalsk betingede, altså afhængige af det enkelte FN.

For den anden type (typeeks.: *bringe noget i anvendelse*) gælder i reglen følgende:

1. FN er styrelse (regimen) for en præposition.
2. Artikelbrugen ligger fast, i de fleste tilfælde er der nul-artikel, eller — på tysk — artiklen er smeltet sammen med præpositionen, typisk til *zum*.
3. FN kan i de fleste tilfælde ikke modificeres, hverken ved foranstillede eller ved efterstillede attributter.⁵
4. Verbet indgår i en kommutationsrække med et eller flere andre verber. Der udtrykkes herved normalt en forskel i aktionsart og/eller kausativitet, men ingen anden betydningsforskel.
5. Konstruktionen er leksikaliseret.

⁵Konstruktioner med *have* eller *være* som FV har normalt ikke denne begrænsning.

2.1.3.1 Type 1

I den franske tradition har man interesseret sig for **constructions à verbe support**, på dansk **støtteverbumbonstruktioner**. Det svarer stort set til det, jeg her har kaldt type 1. Man fokuserer især på substantivernes valensforhold, som indgår i definitionskriterierne.

Man kan altså fx *træffe aftale med nogen om noget*. I FVK'en i (1) har vi *banken* som grammatisk subjekt. Uden for FVK'er udtrykkes denne relation som en subjektiv genitiv som i (2):

- (1) Banken traf aftale med kunden om et lån.
- (2) Man diskuterede, hvilke konsekvenser bankens aftale med kunden om et lån ville få for hans skatteregnskab.

I en hel del tilfælde findes der to eller flere verber, der kan forbindes med det samme FN. Undertiden — men langt sjældnere end ved type 2 — udtrykkes herved en forskel i aktionsart, i øvrigt bruges verberne ofte mere eller mindre synonymt som i nedenstående autentiske eksempler fra korpus med henholdsvis *træffe* og *indgå aftale*:

- (3) AMI er nu kommet så langt i udvikling af en målemetode, at der er *truffet aftale med* en virksomhed *om* at prøve at måle gensplejsede organismer.
- (4) Ifølge forsøgsleder dr. agro B. Bech Andersen er stoffet totalt ufarligt for mennesker, og det er umuligt at spore hormonet i mælken. Al- ligevel har Foulum *indgået en aftale med* Kødbranchens Fællesråd og Danske Mejeriers Fællesorganisation *om* at destruere al mælk og kød fra forsøgsdyrene for at forbrugerne ikke skal frygte, det havner i fri handel.

2.1.3.2 Type 2

I modsætning til den franske tradition er der i den germanistiske tradition nogle forskere, som foretrækker at definere FVK'er på en måde, så type 1 slet ikke hører med til FVK'erne. Deres definition dækker stort set kun det, jeg her kalder type 2. Det gælder især Ingemar Persson (fx 1992) og von Polenz (fx 1987). Som fællesbetegnelse for de forskellige typer foreslår von Polenz (1987:170) **nominaliseringsverbumbonstruktion**.

Ved beskrivelsen af type 2 interesserer man sig især for det, der i den tyske tradition i reglen betegnes som „Aktionsart“. De, der er mere fortrolige med fransk eller engelsk tradition, ville nok tale om „aspekt“. Der er ikke enighed om hverken antal af undertyper eller om terminologien, men fx von Polenz inddeler type 2 i tre undertyper. De to udtrykker henholdsvis **inchoativ** og **durativ** aktionsart, den tredje udtrykker den **kausative** betydning, som altså ikke er en aktionsart, selv om den fx hos Helbig/Buscha (1984:96) regnes med blandt aktionsarterne.

På tysk vil det kausative typisk udtrykkes ved verber som *bringen* og *setzen*, det inchoative fx med *kommen*, det durative med *sein* eller *stehen*, og for fx von Polenz er det en forudsætning for, at man kan tale om en FVK, at den verbale del indgår i en kommutationsrække bestående af mindst to verber, der udtrykker denne forskel i betydning.

Man har således fx:

- (5) in Gang setzen (kausativt) — in Gang kommen (inchoativt) — in Gang / im Gang sein (durativt)
- (6) in Verdacht bringen (kausativt) — in Verdacht kommen (inchoativt) — im Verdacht stehen (durativt)

2.1.3.3 Type 3

Denne type kalder von Polenz (184:174) **Passiv-FVK**. Det drejer sig om udtryk med *finden* og *erfahren* m.m. som fx *Anerkennung finden* og *Erweiterung erfahren*, hvor konstruktionen svarer til en passivkonstruktion af det tilgrundliggende verbum, altså *anerkennen* og *erweitern*.

Jeg medtager her også konstruktioner med *geschehen* eller *erfolgen*, der ligeledes semantisk svarer til en passivkonstruktion.

Hverken *finden* og *erfahren* eller *geschehen* og *erfolgen* passer helt på beskrivelsen af de to hovedtyper. På den anden side spiller de en så stor rolle i oversættelsessammenhæng, at det — uanset hvordan man vil klassificere dem — er vigtigt at give dem en lige så grundlig leksikografisk beskrivelse som hovedtyperne.

3 Leksikografiske beskrivelser

(Stikprøver i ordbøgerne: *finde anvendelse*)

Efter denne meget summariske oversigt over FVK-typerne er det naturligt at undersøge ordbøgerne for at konstatere: Hvor — i ordbøgerne — finder man *finde anvendelse* og de andre FVK'er? Når der blandt grammatikerne er så divergerende opfattelse af, hvorledes man beskriver FV'er og FVK'er, vil det ikke være overraskende, hvis beskrivelserne i ordbøgerne er lige så inkonsekvente og fyldt med huller.

Her skal blot vises resultatet af nogle opslag, som kunne være relevante, hvis man — som dansker — skal oversætte *finde anvendelse* til tysk. Ved samme lejlighed kan vi se på, hvad vi får at vide om andre FVK'er med *anvendelse*, dvs.

bringe i/til anvendelse
komme i/til anvendelse
tage i anvendelse
være i anvendelse

Ved oversættelse af det danske verbum *anvende* kan man næsten altid bruge enten *anwenden* eller *verwenden*, men det er vigtigt, at man vælger det rigtige af dem, da de ikke er synonyme. Tilsvarende gælder for substantiveringerne *Anwendung* og *Verwendung*. Problemet for ordbogsbrugeren er altså dels at blive klar over, at det danske udtryk er polysemt, og så blive sat i stand til at monosemere det polyseme danske udtryk rigtigt, dels at få bekræftet — eller afkræftet — en formodning om, at man på tysk kan bruge en konstruktion med *finden*.

Stikprøverne er taget i tre dansk-tyske, en norsk-tysk og to tysk-tyske ordbøger. De relevante opslag er vist i deres helhed i figur 1, og i det følgende refereres der hertil.

1. Norsk-tysk:

fin'ne^s *Verb* ^{°fann/°fant, °funne/°funnet}
 (1) finden (a-u), entdecken; f- døden, sin død den Tod f-; f- olje Öl f-; let, så skal I f- suchet, so werdet ihr f- (Matth 7,7); jøeg f-er ikke det uttrykket i ordboka ich finde den Ausdruck nicht im Wörterbuch; funne saker Fundsachen (2) f-. meinen, dafürhalten

2. Bork:

finde 1. finden* (etw passend; etw Verlorenes; die Lösung; Trost; Ruhe; Aufnahme; keine Gnade; keine Worte; sich A darein f.; Vergnügen/Geschmack an+D f.; Trost in+D f.; e-n Freund an/in e-m f.); 2. auffinden* (e-n Toten/Vermissten); *ikke til at* ~ nicht auffindbar || die Leiche wurde im Kanal geborgen || 3. ausfindig machen (nach langem Suchen e-e Wohnung); 4. (*kende, skønne, anse for*) befinden* (e-n schuldig b.); 5. (= søge) suchen; 6. + *adv/præp*: ~ *for godt* für gut befinden*; *som De -r for godt* wie es Ihnen beliebt; ~ *frem* hervorsuchen; ~ *frem til* aus-

3. Hjorth:

anvendelse (*se anvende*) 1. die Verwendung -/en; 2. die Anwendung -/en;
 □ *bringe i* ~ in Anwendung bringen*; *finde* (*el. få*) ~ *for* für +A Verwendung finden*; *rummet finder* ~ *som bibliotek* der Raum wird als Bibliothek verwendet (*el. gebraucht*).

4. Bork:

anvendelse Verwendung, Anwendung f (A./V. finden*; zur A./V. kommen*); Gebrauch m; (*af tropper*) Einsatz m.

5. Munksgaard:

anvendelse (*if anvende*) Gebrauch m (-e)s; +pl; Verwendung f -; -en; Nutzung f -; (-en); Anwendung f -; -en; Verwertung [-'ver-] f -; -en
 □ *fredelig* ~ *af atomenergi* friedliche Nutzung der Kernenergie; *bringe n. til* ~ *etw. (A) in/zur Anwendung bringen**; *finde* ~ (*på en bestemt situation*) (auf eine bestimmte Situation) Anwendung finden*; *finde* ~ *for n.* für *etw. Verwendung finden**; *have* ~ *for en, n.* für jn, *etw. Verwendung haben**; *det har jeg ingen* ~ *for* dafür habe* ich keine Verwendung; *komme til* ~ zur Anwendung kommen*/gelangen s; *tage n. i* ~ *etw. (A) in Verwendung nehmen**.

6. Norsk-tysk:

***an**'ven'delse -n Anwendung f (på auf); Verwendung, Gebrauch m* (av von *el. gen.*): loven finner ingen a- her das Gesetz findet hier keine A-, trifft auf diesen Fall nicht zu; få a- for noe(n) für *etw. (j-n) V-* finden; finne a- angewandt werden; verwendet werden, V- finden

7. Wahrig:

bilag 1

'fin-den 1 (V. t. 134) 1 zufällig, durch Suchen od. durch Nachdenken auf etwas stoßen, etwas entdecken, erlangen; erhalten, erwerben; der Meinung, Ansicht sein, erachten, halten für, erkennen als 2 ich finde, daß ...; das finde ich auch; ich kann ihn nirgends ~; gute Aufnahme ~ gut aufgenommen werden; wir müssen einen Ausweg, Mittel und Wege ~, damit ...; keine Beachtung ~ nicht beachtet werden; ich finde meine Brille nicht; seine Freude, sein Vergnügen an etwas ~; in ihm habe ich einen zuverlässigen Freund gefunden; (keinen) Gefallen, Geschmack an etwas ~; man findet immer wieder Leute, die ...; ich habe die Lösung gefunden; wir konnten keinen Platz mehr ~; ich kann das Richtige nicht ~; sie kann keine Ruhe ~; den Tod (auf dem Schlachtfeld, bei einem Unfall, in den Wellen) ~ (fig.) umkommen; er versucht, Trost in der Kunst zu ~; er konnte vor Überraschung keine Worte ~ er war sprachlos; die richtigen, passenden Worte ~ 3 jmdn. bereit ~, etwas zu tun; etwas falsch, gut, prak-

8. DUW:

6. einer Sache teilhaftig werden: Hilfe, Beifall, Befriedigung f.; (oft als Funktionsverb.) Beachtung, Berücksichtigung, Verwendung f. (*beachtet, berücksichtigt, verwendet werden*).

9. Wahrig:

'An-wen-dung (f. 20) Verwendung, Gebrauch, Umsetzung in die Tat; ~ finden angewendet werden; zur ~ bringen anwenden

Ver-wen-dung (f. 20; unz.) das Verwenden, Gebrauch, Anwendung, Benutzung; dafür werden wir schon noch eine ~ finden; ich habe keine ~ dafür ich kann es nicht gebrauchen

10. DUW:

Anwen-dung, die; -, -en: 1. das Anwenden (1, 2); die A. eines Verfahrens; die A. einer Bestimmung auf einen Fall; *etw. in/zur A. bringen* (Papierdt.: *anwenden*); zur A. kommen/gelangen/A. finden (Papierdt.: *angewendet werden*). 2. (Med.) *therapeutische Maßnahme, bes. im Rahmen einer Kur*;

Verwen-dung, die; -, -en: 1. das Verwenden: keine V. für *etw., jmdn. haben* (*etw., jmdn. nicht gebrauchen können*); V. finden (*verwendet werden*); in V. stehen (österreich.; *in Gebrauch sein*); in V. nehmen (österreich.; *in Gebrauch nehmen*). 2. (o. Pl.) (geh.) das Sichverwenden (2) für *jmdn., etw.*;

Men skal man nu slå op under *finde* eller under *anvendelse*? Selv hærdede ordbogsbrugere er tit i tvivl, men lad os begynde med at se under *finde* i oversættelsesordbøgerne. Det viser sig, at ingen af de undersøgte ordbøger har medtaget kombinationen med *anvendelse*. Vi ved, at forholdet mellem verbum og FN typisk er 'få til mange', derfor er det rimeligt, at man ikke forsøger en opremsning af de mulige FN, men der er heller ingen af dem, der overhovedet nævner muligheden af at bruge *finde* som FV. Det nærmeste vi kommer, er, at Bork (figur 1: 2) som den eneste blandt de mange ting, man kan *finden*, nævner *Aufnahme*. Ved analogislutning kan brugeren her få bestyrket sin formodning om, at en tilsvarende konstruktion med *finden* + *Anwendung* eller *Verwendung* er mulig, men vished får han ikke.

Vi prøver at slå op under *anvendelse*. Hjorth (figur 1: 3) har

finde (el. få) *anvendelse* for: für +A Verwendung finden.

Samtidig kan man se, at *bringe i anvendelse* hedder in *Anwendung bringen*. Den utrænede bruger vil måske heraf slutte, at det er verberne *finde* / *bringe*, der afgør, om man bruger *Verwendung* eller *Anwendung*. Han vil sikkert slet ikke tænke over, at konstruktionen med *for* / *für* ikke er passivsynonym, men har en helt anden betydning. Sammenlign:

(7) Den nye teknik har fundet anvendelse inden for flere forskningsområder

(8) Syersken har fundet anvendelse for alle resterne

Vi får ingen oplysninger om kombinationerne med *komme*, *tage*, *være*.

Hos Bork (figur 1: 4) finder vi:

anvendelse: Verwendung, Anwendung f (A./V. finden; zur A./V. kommen) [. . .]

Her ser vi begge mulighederne med *finden* anført uden forklaringer, men hvis man ser eksemplerne ved verbet, har man en rimelig chance for at ramme rigtigt. Samtidig får vi *kommen* med de samme to muligheder. Her mangler til gengæld oplysning om kombinationerne med *bringe*, *tage*, *være*.

Munksgaards ordbog er den mest omfattende. Det viser sig også på dette punkt, for her finder vi begge betydninger af *finde anvendelse* (figur 1: 5). Eksemplerne giver en vis hjælp til at vælge rigtigt, men specielt brugen af parentes eller ikke omkring den følgende PP stiller store krav til brugerens evne til at tolke implicit information. Munksgaard har udover *finde* både *bringe*, *komme* og *tage*; her mangler altså kun *være*. Men der er flere problemer, specielt i ækvivalentoplysningerne. Ved *bringe* har man kun medtaget præpositionen *til* og ikke det almindeligere *i*, men som ækvivalent angives *etw. (A) in/zur Anwendung bringen*. Man kan altså ikke her se, at man også kan — og i givet fald skal — bruge *Verwendung*, og man får intet at vide om forholdet mellem *in* og *zur*.

Ved *komme* er situationen næsten parallel: Kun én præposition på dansk, kun *Anwendung* er nævnt som muligt substantiv, og man får ingen hjælp til at vælge mellem *kommen* og *gelangen*. Endelig har vi at *tage n. i anvendelse*. Det er gengivet som *etw. (A) in Verwendung nehmen*, hvilket lyder helt rimeligt for en dansker. Problemet er blot, at det ifølge flere andre ordbøger, bl.a. Duden, er en „österreichische Besonderheit“.

De følgende ordbøger vil jeg ikke kommentere lige så grundigt.

Om den norsk-tyske ordbog (figur 1: 6) vil jeg uden at gå i detaljer blot sige, at den er lige så mangelfuld og inkonsekvent som de danske. Af tysk-tyske ordbøger har jeg her begrænset mig til de to mest udbredte etbindsværker, *Duden Universalwörterbuch* (DUW) og Wahrig.

Når man slår op under *finden* i Wahrig (figur 1:7), finder man, i overensstemmelse med Wahrigs systematik, under punkt 2 en række eksempler med *finden* + substantiv, i alt 15 forskellige, heraf to FVK'er: med *Aufnahme* og *Beachtung*, men hverken *An-* eller *Verwendung*. I DUW finder man under *finden* pkt. 6 (figur 1:8): <oft als Funktionsverb:> *Beachtung, Berücksichtigung, Verwendung f.* — men altså ikke *Anwendung*.

Vi fortsætter med *Anwendung* og *Verwendung*.

I Wahrig (figur 1:9) finder man under *Anwendung* udtrykkene *Anwendung finden* og *zur Anwendung bringen* med parafraiser, men ellers ukommenteret. Under *Verwendung* har Wahrig af relevant kun *Verwendung finden* — uden parafrase. I DUW (figur 1:10) finder man *etw. in/zur Anwendung bringen; zur Anwendung kommen/gelangen/Anwendung finden* — med parafrase, men i stedet for at tale om Funktionsverbgefüge hæfter man her etiketten „Papierdeutsch“ på udtrykkene. *Verwendung finden* står med parafrase, men uden værdiladet etikette, desuden har vi *in Verwendung stehen* og *in Verwendung nehmen* med etiketten *österr.*

Sammenfattende kan vi altså konstatere, at de undersøgte ordbøger i hvert fald ikke har afkræftet vores formodning om, at behandlingen ville være mangelfuld og inkonsekvent. Vi kan desuden konstatere, at det for en rutineret ordbogsbruger er muligt at finde frem til en korrekt oversættelse af *finde anvendelse* ved hjælp af de nævnte ordbøger i passende kombination, men at han kommer i store vanskeligheder, hvis han alene på grundlag af disse værker skal sikre sig, at han får oversat de andre konstruktioner med *anvendelse* korrekt.

Nu er der næppe andre end ordbogsredaktørerne selv, der kan sige, i hvor høj grad den mangelfulde behandling af denne type konstruktioner skyldes, at det materiale, man har haft som grundlag, har været præget af tilfældige huller, eller det skyldes, at redaktøren har ønsket eller været tvunget til at begrænse omfanget og så har foretaget begrænsningen subjektivt efter kriterier, som ikke er umiddelbart gennemskuelige. De forlagstekniske begrænsninger kan man som udenforstående ikke tage stilling til, men risikoen for tilfældige huller og inkonsekvenser kan i hvert fald mindskes, hvis man kan basere arbejdet på undersøgelser af elektroniske korpora.

4 Korpusundersøgelserne

4.1 Gentek-korpus

Gentek-korpus rummer tekster på tre sprog: dansk, tysk og spansk, og er, som en række andre korpora fra de senere år, på rundt regnet en million løbende ord pr. sprog. I mine undersøgelser er inddraget det danske og det tyske korpus. Gentek-korpus har til formål inden for et mere eller mindre tilfældigt valgt emneområde at give materiale til at belyse sproglige forskelle, som hænger sammen med nogle bestemte forskelle i kommunikationssituationen, nemlig forskelle i graden af faglighed. Man taler om kommunikationsniveau og opererer med fire niveauer:⁶

- niv. 1 fagmand til fagmand
- niv. 2 A fagmand til lægmand
- niv. 2 B lægmand til fagmand
- niv. 3 lægmand til lægmand

⁶Gentek-korpus er beskrevet nærmere i Lauridsen/Andersen (1993).

„Fagmand“ skal forstås som én, der har genteknologi som fag, „lægmand“ som én, der ikke har det som fag, uanset hvilket uddannelsesniveau man i øvrigt har. Denne skelnen, som på sin vis er logisk, har nogle konsekvenser, som jeg vil vende tilbage til nedenfor.

4.2 Mine undersøgelser

4.2.1 Inventaret og metoden

Hvilke FV'er og FVK'er har man på tysk og på dansk?

Eftersom der er uenighed om afgrænsningen, kan det ikke undre, at der er stor forskel på, hvilke verber og hvilke konstruktioner de enkelte forskere omtaler, men ud over de teoretisk begrundede forskelle er der også forskelle, der bærer præg af at skyldes tilfældigheder. Man finder ingen steder en oversigt, som gør krav på at være udtømmende — endsige *er* udtømmende. Det gælder også for resultatet af mine undersøgelser.

Min oprindelige hensigt og mit håb var at kunne lave en udtømmende beskrivelse af FVK'er i det foreliggende korpus, men selv med denne kraftige begrænsning er det ikke tilfældet. Eftersom der ikke fandtes en samlet liste over FV'er, har jeg arbejdet eklektisk. Ved at sammenholde en række beskrivelser har jeg forsøgt at opstille en liste, der så at sige udgør foreningsmængden af de verber og konstruktioner, som nævnes af andre. Listen er så suppleret med, hvad jeg stødte på under arbejdet med korpus eller ved læsning i øvrigt.

Kort fortalt har arbejdets første fase derefter hovedsagelig bestået i ved hjælp af et konkordansprogram at undersøge de potentielle FV'ers kontekst for at fastslå, om de rent faktisk forekom som FV og i givet fald i hvilke konstruktioner. Derefter har jeg vendt søgeretningen og taget udgangspunkt i FN'erne, dels dem, der i forvejen stod på min liste, dels dem, der blev fundet med udgangspunkt i verberne. For den første gruppes vedkommende drejede det sig — ligesom ved verberne — om at konstatere, om de overhovedet forekom i FVK'er og i givet fald i hvilke konstruktioner. For den anden gruppes vedkommende var første spørgsmål jo allerede besvaret positivt, så her drejede det sig om at finde ud af, om de indgik i konstruktioner med andre FV'er end det først fundne.

For det danske korpus' vedkommende har metoden været nøjagtig den samme, men udgangspunktet har været ret forskelligt. Der har hidtil været skrevet meget lidt om danske funktionsverber.⁷ Den startliste, jeg kunne opstille på det grundlag, var derfor væsentligt mere sparsom end den tyske, så den måtte i højere grad bygge på, hvad jeg selv kunne komme i tanker om, selvfølgelig stærkt inspireret af den tyske liste.

4.2.2 Undersøgelsesresultater

4.2.2.1 Generelt

Det har tidsmæssigt ikke været muligt at gennemføre en undersøgelse af samtlige de verber og substantiver, som efter de nævnte principper skulle undersøges. For nogle af verberne og for de fleste af substantiverne gælder det, at undersøgelsen er ret tidkrævende, og at udbyttet i forhold hertil er pauvert. At finde et nyt relevant eksempel er næsten som at finde en nål i en høstak. Mine forventninger om at finde en række nye FV'er og konstruktionstyper er ikke blevet indfriet. Jeg har selvfølgelig i korpus fundet mange eksempler på FVK'er, som jeg

⁷Se i litteraturlisten under Boje, Dyhr, Feil, Hansen og Nøhr Pedersen.

ikke har fundet i litteraturen om FVK'er, men det har hovedsagelig været nye kombinationer med de mest produktive FV'er som *bringen* og *erfolgen* samt *foretage*.

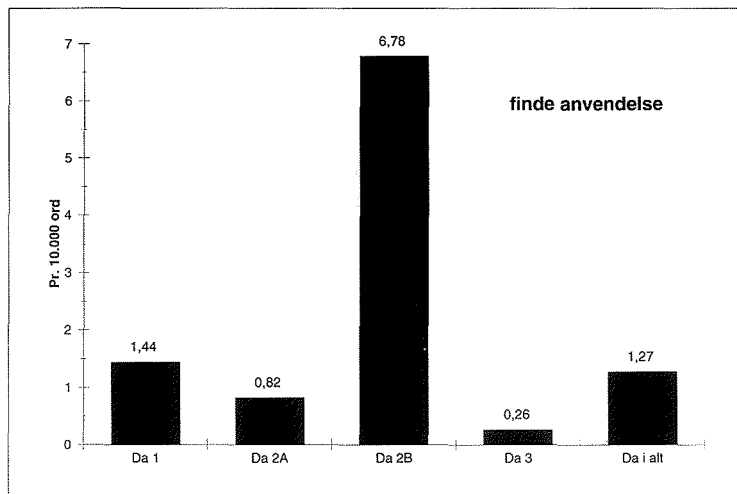
Jeg har derfor valgt at koncentrere mig om to aspekter. Dels har jeg især beskæftiget mig med de konstruktioner, der ser ud til at være af særlig interesse i kontrastiv sammenhæng, og dels har jeg gjort nogle statistiske iagttagelser, som Gentek-korpusets særlige sammensætning giver mulighed for.

4.2.2.2 Statistiske undersøgelser

Hermed kommer jeg ind på den anden betydning af artiklens titel: Hvor — i korpus — finder man *finde anvendelse*?

Det er vist en almindelig opfattelse, at nominalstil generelt og dermed også FVK'er specielt er stiltræk, som er særlig udpræget i fagsprog, og det giver Gentek-korpus en udmærket mulighed for at undersøge. Nedenfor vises nogle diagrammer med resultater fra denne type undersøgelser. Det drejer sig om udtrykket *finde anvendelse* og dermed beslægtede udtryk. Da de fire kommunikationsniveauer ikke er ligeligt repræsenteret, er en fordeling af forekomsterne i absolutte tal misvisende. Fordelingen er derfor i det følgende omregnet til antal forekomster pr. 10.000 ord, altså en procentfordeling.⁸

I Gentek-korpus er der i alt 127 forekomster af udtrykket *finde anvendelse*, og deres fordeling ses i fig. 2.



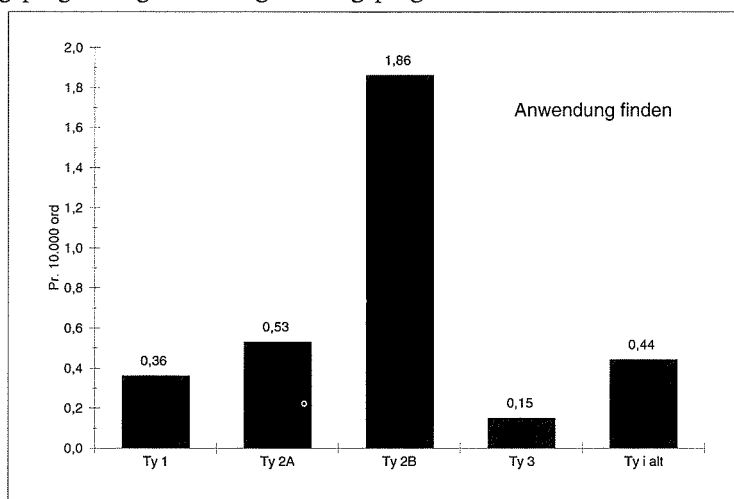
Figur 2.

Hvis vi et øjeblik „glemmer“ niv. 2B og ser på de tre andre, ser fordelingen meget fornuftig ud. På niv. 2A, som man lidt forenklet kunne kalde halvfagligt, bruges konstruktionen godt tre gange så hyppigt som mellem lægfolk, og på det helt faglige niveau næsten seks gange så hyppigt. Det er i smukkeste overensstemmelse med, hvad man ville forvente. Men niv. 2B ser umiddelbart helt forkert ud. Hvorfor skulle lægfolk bruge konstruktionen 26 gange

⁸En oversigt i tabelform over de i det følgende behandlede forekomster, dels i absolutte tal, dels i procent, ses i bilaget.

så tit, når de skriver til fagfolk, som når de skriver til hinanden? Det er selvfølgelig ikke sådan, det hænger sammen. Grunden er, at de tekster, der hører til niv. 2B, alle er skrevet af fagfolk, blot en anden slags, nemlig af jurister. Det ville føre for vidt at redegøre for grunden til det her, men det betyder, at vi i niv. 2B har juridisk fagsprog, og det vil igen sige, at Gentek-korpus bl.a. giver mulighed for at sammenligne noget, som ikke er fagsprog (niv. 3), med to forskellige fagsprog (niv. 1 og niv. 2B).

Vi kan altså se, at det danske udtryk *finde anvendelse* er mange gange almindeligere i juridisk fagsprog end i genteknologernes fagsprog.



Figur 3.

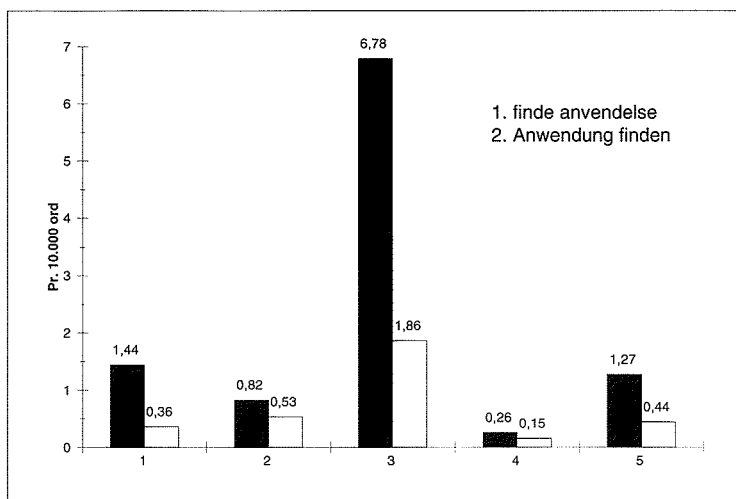
I fig. 3 ses fordelingen af de 44 eksempler på det tyske *Anwendung finden*. Vi ser her en fordeling, som minder meget om den danske, specielt ses, at det juridiske rager meget højere op end de andre. Alligevel er der forskelle, og dem ser vi tydeligst, hvis vi sammenligner fordelingen på dansk og tysk, som det er vist i fig. 4.

Vi konstaterer, at *Anwendung finden* ikke er nær så hyppigt som *finde anvendelse*, at forskellen mellem niveauerne ikke er nær så stor som på dansk, og endelig, at der er færre forekomster på niv. 1 end på niv. 2A.

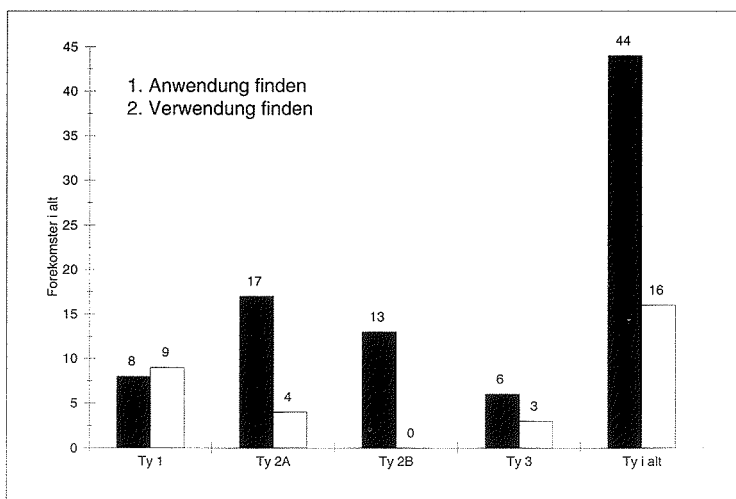
Nu ved vi, at *finde anvendelse* også svarer til *Verwendung finden*, og det kunne måske være forklaringen på forskellen til dansk. Derfor har jeg sammenlignet forekomsterne af *Anwendung finden* og *Verwendung finden*, og resultatet ses i fig. 5 (i absolutte tal). Det ændrer ikke noget særligt i det samlede billede. Det mest bemærkelsesværdige er, at man i de juridiske tekster ikke finder et eneste eksempel på *Verwendung finden*, mens der hos genteknologerne er en lille overvægt af eksempler på *Verwendung finden*.

I begyndelsen af artiklen blev det nævnt, at udtrykket *finde anvendelse* er mere eller mindre synonymt med passiven *anvendes*. Derfor er det nærliggende at sammenligne de to udtryk.

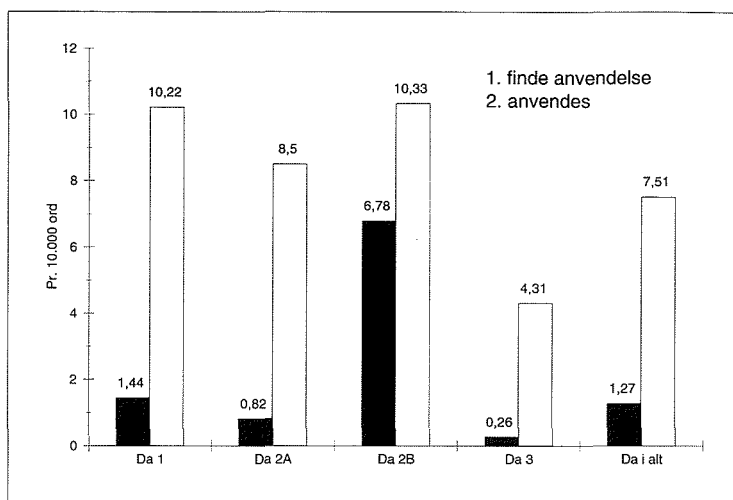
I fig. 6 ses, at ved passivkonstruktionen er forskellen på niveauerne ikke nær så stor som ved FVK'en, specielt er der her ikke nogen forskel på de to fagsprog, og vi ser, at passiven er meget almindeligere end FVK'en. Selv hos juristerne finder man endnu flere passiver end FVK'er.



Figur 4.



Figur 5.



Figur 6.

Hvad kan man nu som leksikograf bruge disse statistiske oplysninger til?

For det første kan de understrege vigtigheden af, at man i ordbøgerne sørger for at medtage en type konstruktioner, som er almindelige i videnskabelige tekster. For det andet viser de også, at forskellen mellem forskellige fagsprog ikke kun er et spørgsmål om fagtermer i snæver forstand.

Nu kunne man ud fra det, vi her har set, være fristet til at slutte, at FVK'er åbenbart er særlig karakteristiske for det juridiske fagsprog. Jeg har tidligere vist, at det fx gælder for konstruktioner med det danske FV *foretage* (Boje 1993:116f.). Men desværre er det mere kompliceret end som så. Det fremgår af det sidste statistik-diagram, fig. 7.

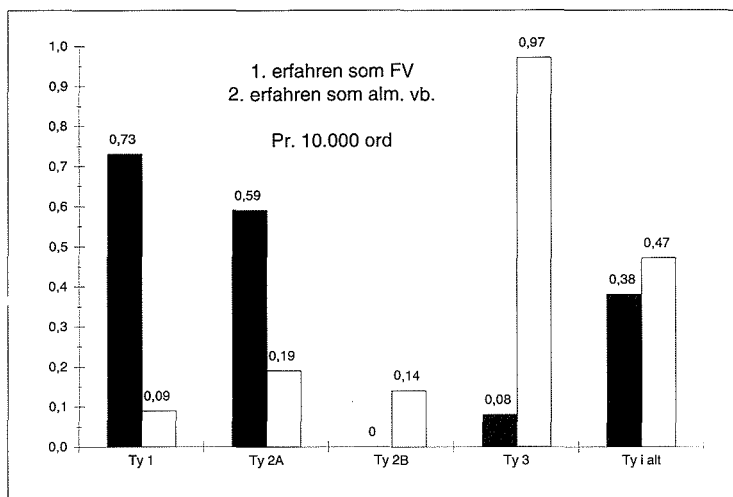
Det tyske verbum *erfahren* kan — i modsætning til det danske *erfare* — også bruges som et passiv-synonymt FV, og hvis vi ligesom ved *finde anvendelse* først nøjes med at se på niv. 1, 2A og 3, finder vi en fordeling af FV og alm. verbum, der nogenlunde svarer til, hvad man kunne forvente.

Igen er det juristerne, der står for overraskelsen. I de juridiske tekster er der ikke ét eneste eksempel på *erfahren* som FV. Jeg vil afstå fra at forsøge en forklaring, men blot sige, at det understreger nødvendigheden af at se differentieret på fagsprog.

4.2.2.3 Kontrastive undersøgelser

4.2.2.3.1 Baggrund

Der er en høj grad af parallelitet mellem de FVK'er, der bruges på tysk, og dem, der bruges på dansk. Da samtidig FVK'er, som beskrevet i indledningen, er kompositionelt forståelige, vil det i en passiv ordbog ikke være nødvendigt at gøre noget særligt ud af dem. Men da paralleliteten er langt fra de 100%, er sagen ikke nær så enkel i en aktiv ordbog.



Figur 7.

Ingemar Persson (1992) gør sig til talsmand for den opfattelse, at man skal give en grundig beskrivelse af konstruktionstyperne under de pågældende FV'er, men at det for de fleste substantiver er overflødig at angive mulighederne for dannelse af FVK. Nu beskæftiger Persson sig ganske vist kun med konstruktionerne med *bringen* og *kommen*, men selv her kan jeg ikke give ham ret. Jo, nok i, at der skal gives en generel og eksemplificeret beskrivelse ved verberne, men ikke i, at det er overflødig at give oplysningerne ved substantiverne.

For ordbøger til maskinoversættelse er det under alle omstændigheder forkert, for de skal have alle oplysninger eksplicit, men selv hvis vi kun tænker på den menneskelige oversætter og i første omgang begrænser os til FVK type 2, vil vi på den måde ikke kunne undgå fejlversættelser. Sagen er jo, at reglerne ikke er faste regler, men beskrivelser af dannelsesmønstre, der bruges tit, men netop ikke 100%. Der er tale om idiosynkratiske begrænsninger, som man i reglen ikke skænker en tanke, før man oplever, at en udlænding overtræder dem.

Hvis man fx til at udtrykke kausativitet altid kunne bruge *bringen*, eller i det mindste altid, når man på dansk i den tilsvarende konstruktion bruger *bringe*, og så kunne bruge *setzen*, når vi på dansk bruger *sætte*, så var problemet til at overse. Men det er ikke tilfældet. Se fx på eks. 6, som er taget fra Persson. Medmindre man vælger friere omskrivninger, bliver det på dansk til:

(6a) *sætte under mistanke — komme under mistanke — være under mistanke*

En dansker ville altså umiddelbart forvente, at den kausative FVK på tysk skulle hedde **unter Verdacht setzen*.

Persson udtaler sig slet ikke om type 1, men dér gælder det i endnu højere grad, at valget af FV på L2 er uforudsigeligt ud fra verbet på L1.

4.2.2.3.2 Egne undersøgelser

Dette er baggrunden for, at jeg især har interesseret mig for de konstruktioner, hvor jeg vidste eller formodede, at der ville være problemer med at vælge rigtigt ved „Hinübersetzung“.

Tag fx det danske verbum *foretage*. Det er langt det almindeligste FV i forbindelse med type 1. På tysk har vi tilsvarende verbet *vornehmen*, som i de fleste tilfælde er en rimelig oversættelse af *foretage* — men netop ikke altid. I en hel del tilfælde skal eller kan man bruge *unternehmen*, og selv om man har regler, som giver et fingerpeg om, hvad man skal vælge, er det ikke sikre regler, og man kan næppe blot opstille en liste over substantiver, der forbindes med *unternehmen*. En sammenligning af eksemplerne i korpus med *vornehmen* og *unternehmen* viser fx, at flere substantiver optræder med begge verber, i nogle tilfælde med en betydningsnuance, i andre tilfælde tilsyneladende uden betydningsforskel.

Noget tilsvarende gælder, når man skal oversætte det danske *komme*. I de fleste tilfælde er *kommen* en udmærket løsning, men korpus viser også eksempler, hvor det samme substantiv forbindes med både *kommen* og *geraten*, *kommen* og *gelangen* eller *geraten* og *gelangen*. Derimod er der intet substantiv, som i korpus forbindes med alle tre verber. Også her kan der gives tendens-regler, men hverken sikre regler eller udtømmende lister.

Som sidste eksempel kan jeg nævne oversættelsen af dansk *sætte* til tysk *setzen* eller *versetzen*, som heller ikke kan klares med enkle regler eller lister. Og så har jeg endda — bortset fra eksempel (6a) med *in Verdacht bringen* — ikke omtalt de tilfælde, hvor man på tysk skal vælge et FV som ikke minder om det danske, eller hvor man ved oversættelsen slet ikke kan bruge et FV. Jeg har heller ikke her kunnet komme ind på forskelle i artikelbrug eller i muligheder for modifikation.

Desværre kan jeg endnu ikke fremlægge færdige resultater, og når undersøgelserne er afsluttet, kan resultaterne højst kaldes endelige set i relation til Gentek-korpus, ikke set i relation til de pågældende konstruktioner generelt. Men selv med et væsentligt større korpus som baggrund vil det givetvis være nødvendigt at supplere med informanttests, før man med rimelig sikkerhed kan beskrive de acceptable variationsmuligheder.

5 Konsekvenser for ordbogsarbejdet

Det er forhåbentlig fremgået af det foregående, at hvis man vil sikre sig, at en oversætter ved hjælp af den bedste grammatik, der kan opdrives, plus en ordbog skal kunne producere alle korrekte og kun korrekte FVK'er, skal ordbogen indeholde flere oplysninger end nogen i øjeblikket eksisterende ordbog. Man er i så fald efter min mening nødt til for hvert substantiv, der kan indgå i en FVK, at have oplysninger om i hvert fald:

- (1) hvilket / hvilke FV kan bruges
- (2) hvilken betydningsforskel er forbundet hermed, fx aktionsart, stillag
- (3) hvilken / hvilke præpositioner kan bruges (fx *in* / *zu*) — og med hvilken forskel.
- (4) hvordan er artikelbrugen
- (5) hvilke valensbundne led kan tilknyttes
- (6) hvilke modificerende led er mulige (ingen / fra lukket liste / fra semantisk begrænset liste / frit)

Disse punkter er ikke lige relevante for alle tre konstruktionstyper. Fx er (4)–(6) især vigtige for type 1, mens (2)–(3) især har betydning for type 2.

De praktisk arbejdende ordbogsforfattere vil sikkert nu sige, at det forslag er aldeles urealistisk: en ordbog, som skulle opfylde alle disse krav, er umulig at lave, umulig at sælge og umulig at bruge, og det har de ganske ret i — i hvert fald hvis man tænker på traditionelle papirbaserede ordbøger.

Mit forslag er tænkt som den ideale fordring, men samtidig mener jeg alvorligt, at alle disse oplysninger burde stå til rådighed i hvert fald for ordbogsforfatterne, som så kunne foretage en afvejning af behov og muligheder, men helst også for „krævende“ brugere.

For mig er det vigtigste i første omgang, at så mange som muligt af disse oplysninger bliver indsamlet og registreret i et fastlagt maskinlæsbart format.

Jeg tror ikke, at vi skal ret langt ind i det næste årtusinde, før de elektroniske ordbøger til mennesker har nået et stade, så man kan lagre alle relevante oplysninger i pladsbesparende formalismer og hente dem frem i en brugervenlig form — på skærm eller papir — på en måde, så brugeren selv bestemmer, hvilke oplysningstyper han vil have præsenteret, og hvordan, altså bl.a. med hvilken detaljeringsgrad.

Jeg tror med andre ord ikke, at det vil vare længe, før alle disse oplysninger vil kunne *finde anvendelse!*

Litteratur

Ordbøger

Bork = Egon Bork 1990: *Dansk-tysk Ordbog*. (Gyldendals Røde Ordbøger) 9. udg., 4. opl. København: Gyldendal.

DUW = *Duden Deutsches Universalwörterbuch* 1989, hrsg. von . . . Günther Drosdowski. 2. Aufl. Mannheim: Dudenverlag.

Hjorth = Grethe Hjorth 1993: *Dansk-Tysk Ordbog. Undervisning*. (Gyldendals Røde Ordbøger) 1. udg. København: Gyldendal.

Munksgaard = Henrik Bergstrøm-Nielsen/Henrik Lange/Henry Verner Larsen 1991: *Dansk-tysk ordbog*. (Munksgaards store ordbøger) 1. udg. København: Munksgaard.

Norsk-tysk = Tom Hustad 1979/1984: *Stor Norsk Tysk Ordbok*, Studentutgave. Oslo: Universitetsforlaget.

Wahrig = Gerhard Wahrig 1986/91: *Deutsches Wörterbuch*. Neuausgabe 1986 = Jubiläumsausgabe. Gütersloh: Bertelsmann Lexikon Verlag.

Anden litteratur

Boje, Frede 1993: Funktionsverboms-konstruktioner i dansk og tysk. I: Gert Engel (red.): *Proceedings af seminar om korpuslingvistik i fagsprogsforskningen*. Hindsgavl Slot nov. 1992, 104–121.

Boje, Frede/Anna Braasch 1992: Hvad får man skudt i skoene? Flerordsenheder i aktive ordbøger for mennesker og maskiner. I: R. V. Fjeld (udg.): *Nordiske studier i leksikografi*. Rapport fra Konferanse om leksikografi i Norden 28.–31. maj 1991. Oslo: Nordisk Forening for Leksikografi, 483–500.

Boje, Frede/Susanne Nøhr Pedersen 1989: *NPs with predicative nouns as heads*. Intern EUROTRA-rapport. København.

- Dyhr, Mogens 1980: Zur Beschreibung von Funktionsverbgefügen. I: *Festschrift für Gunnar Bech*. Kopenhagener Beiträge zur Germanistischen Linguistik, Sonderband 1. København: Institut for germansk filologi, Københavns universitet, 105–122.
- Feil, Ruth 1994: Undersøgelse af danske funktionsverber. I: *Udforskning af dansk ordforråd og grammatik*. UDOG. Rapport nr. 2. København: Center for Sprogteknologi, 127–136.
- Hansen, Gyde 1986: *Kontrastive Analyse des Artikelgebrauchs im Dänischen und Deutschen*. København: Nyt Nordisk Forlag Arnold Busck.
- Helbig, Gerhard/Joachim Buscha 1984: *Deutsche Grammatik*. 8. Aufl. Leipzig: VEB Verlag Enzyklopädie Leipzig.
- Lauridsen, Ole/Ingrid Skytte Andersen 1993: *Rapport over oprettelsen af et dansk og et tysk maskinlæsbart tekstkorpus inden for området genteknik 1989*. ARK nr. 66. København: Det erhvervsproglige Fakultet, Handelshøjskolen i København.
- Lauridsen, Ole/Theis Riiber/Henning Søndergaard 1991: Erstellung eines dänischen und eines deutschen Textkorpus — Fachsprache Gentechnik. I: *HERMES, Tidsskrift for Sprogforskning nr. 6*, 125–137.
- Nøhr Pedersen, Susanne 1990: The Treatment of Support Verbs and Predicative Nouns in Danish. I: Jörgen Pind/Eiríkur Rögnvaldsson (ed.): *Papers from the Seventh Scandinavian Conference of Computational Linguistics, Reykjavík 1989*. Reykjavík: Institute of Lexicography/Institute of Linguistics, 208–217.
- Nøhr Pedersen, Susanne 1992: Frozen Expressions in Danish and French. I: Steen Jansen & al. (eds.): *Computational Approaches to Text Understanding*. København: Museum Tusulanum Press, 75–88.
- Nøhr Pedersen, Susanne/Frede Boje/Henrik Selsøe Sørensen 1989: *On Support Verb Constructions in Danish*. Intern EUROTRA-rapport. København.
- Nøhr Pedersen, Susanne/Sabine Kirchmeier-Andersen 1991: Contrastive Linguistics in EUROTRA. I: Karen M. Lauridsen/Ole Lauridsen (eds.): *Papers from the CL-Symposium at The Aarhus School of Business 28–30 August 1989*. Århus: Det Erhvervsproglige Fakultet, Handelshøjskolen i Århus, 155–186.
- Persson, Ingemar 1992: Das kausative Funktionsverbgefüge (FVG) und dessen Darstellung in der Grammatik und im Wörterbuch. I: *Deutsche Sprache 2/1992*, 153–171.
- v. Polenz, Peter 1963: Funktionsverben im heutigen Deutsch. *Sprache in der rationalisierten Welt*. Wirkendes Wort Beiheft 5. Düsseldorf.
- v. Polenz, Peter 1987: Funktionsverben, Funktionsverbgefüge und Verwandtes. Vorschläge zur satz-semantischen Lexikographie. I: *Zeitschrift für Germanistische Lexikographie* 15, 169–189.
- Seelbach, Dieter 1990: Zur Entwicklung von bilingualen Mehrwortlexika Frz-Dtsch: Stützverbkonstruktionen und adverbiale Ausdrücke. I: Burkhard Schaefer/Burghard Rieger (udg.): *Lexikon und Lexikographie*. Sprache und Computer Bd. 11. Hildesheim: Georg Olms Verlag, 179–207.
- Weinrich, Harald 1993: *Textgrammatik der deutschen Sprache*. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

	Niv. 1	Niv. 2A	Niv 2B	Niv 3	I alt
Excerpter, danske	65	189	15	254	523
Excerpter, tyske	87	114	7	208	416
Tekstord i procent (dansk)	18	38	9	35	100
Tekstord i procent (tysk)	22	32	7	39	100
anvendelse, finde (abs.)	26	31	61	9	127
anvendelse, finde (rel.)	1,44	0,82	6,78	0,26	1,27
Anwendung finden (abs.)	8	17	13	6	44
Anwendung finden (rel.)	0,36	0,53	1,86	0,15	0,44
Verwendung finden (abs.)	9	4	0	3	16
Verwendung finden (rel.)	0,41	0,13	0	0,08	0,16
Anw. + Verw. finden (abs.)	17	21	13	9	60
Anw. + Verw. finden (rel.)	0,77	0,66	1,86	0,23	0,6
anvendes (abs.)	184	323	93	151	751
anvendes (rel.)	10,22	8,5	10,33	4,31	7,51
erfahren FV (abs.)	16	19	0	3	38
erfahren FV (rel.)	0,73	0,59	0	0,08	0,38
erfahren (alm.) (abs.)	2	6	1	38	47
erfahren (alm.) (rel.)	0,09	0,19	0,14	0,97	0,47

Bilag.

Arbejdet med „Forslag om dansk standard for lagring og udveksling af leksikalske data“

This paper gives a brief report on the work done so far within the ongoing STANLEX project which is aiming at a proposal for standardization of storage and exchange of lexical data. The STANLEX group is affiliated to the Danish Standard Association. The current goal is to develop a clearly defined format (comprising both structure and content aspects) to support efficient sharing of machine readable dictionary data. Also the reusability and multifunctionality aspects of these resources will be strengthened by means of a general taxonomy covering both printed dictionaries and current requirements of lexicon modules integrated with natural language processing systems.

1 Indledning

Sprog i almindelighed — forståelse og håndtering af information spiller en stadig større rolle i det internationaliserede samfund, vi lever i. Der er stadig større behov for hurtig og præcis tekstbehandling, herunder produktion af tekster, oversættelse osv. „I den situation kan sprogteknologiske hjælpemidler få stor betydning“, fastslås det i en rapport om „Sprogteknologi“ fra Teknologinævnet. Computeren har for længst vundet indpas i al slags arbejde med tekster.

Sprogteknologiske produkter er bl.a. elektroniske ordbøger, termbaser, systemer for stave-, grammatik- og stilkontrol, samt systemer for maskinstøttet oversættelse og maskinel oversættelse. Disse produkter bruges, når man arbejder med sprog på elektronisk medium: fx skriver sine tekster på computer, anvender orddeling, stavekontrol eller slår op i en elektronisk ordbog, når man skriver på fremmedsprog.

Dansk er et lille sprogområde, og det betyder i mange henseender, at udvikling af sprogteknologiske produkter vil halte bagefter de store sprogs udvikling, om ikke andet så af økonomiske grunde. Det er dyrt at udvikle sådanne hjælpemidler, derfor er det vigtigt at økonomisere med de ressourcer, landet har.

2 Teknologinævnet

Den før nævnte pjece er udarbejdet af Teknologinævnet, som er en offentlig dansk organisation under Forskningsministeriet, og som iværksætter helhedsvurderinger af den teknologiske udviklings muligheder og konsekvenser.

Teknologinævnet har i 1993 nedsat et tværfagligt udvalg bestående af eksperter og repræsentanter for erhvervslivet (dvs. producenter og nuværende samt potentielle brugere af sprogteknologiske produkter) til at undersøge sprogteknologiens rolle i erhvervsmæssig tekstproduktion og oversættelsesarbejde.

Undersøgelsens resultater er beskrevet i en situationsrapport om anvendelsen af eksisterende sprogteknologiske værktøjer for dansk og vurderer, hvilke strategier der gør det muligt at komme på forkant med udviklingen med henblik på at opfylde fremtidige behov.

3 Teknologinævnets rapport

I rapporten *Dansk sprogteknologi — status, perspektiver og handlemuligheder* (Bech m.fl. 1994) fremsættes anbefalinger vedr. en koordineret dansk handlingsplan for fremtidige aktioner i.r.den for det sprogteknologiske område, og der peges på fire særlige satsningsområder:

- Sprogteknologiske demonstrationsprojekter (afprøvning og systematisk evaluering)
- Etablering af en fælles standard for lagring og udveksling af leksikalske data (ved arbejdsgruppen STANLEX¹)
- Oprettelse af en stor dansk maskinbrugbar ordbog
- Oprettelse af netværk af danske termbanker

Det er åbenlyst, at især punkterne 2 og 3 kan være af interesse for leksikografer. Dette indlæg fokuserer på punkt 2.

Arbejdet med etableringen af en fælles standard for lagring og udveksling af leksikalske data er startet i september 1994 og forventes afsluttet ved omkring årsskiftet 1995/96. Vi er i skrivende stund (sommeren 1995) endnu ikke nået frem til et konkret, afrundet forslag om en sådan standard, derfor kan der her kun gives en redegørelse for selve målsætningen og den proces, arbejdsgruppen STANLEX er i gang med.

Alle leksikografer ved, hvor arbejdskrævende det er at lave ordbøger af høj kvalitet, og at materialet fra gode ordbøger bliver brugt igen og igen som grundlag eller kilde for nye slags udgaver, reviderede udgaver, som model for andre ordbøger osv. Dette er en form for genbrug af det arbejde, den viden og erfaring, der er opsamlet i en enkelt, stor ordbog, som måske er blevet til gennem indsats fra flere generationer af leksikografer.

I den efterfølgende fremstilling nævnes forskellige slags ordbøger. Nedenfor beskrives ganske kort, hvad de enkelte betegnelser står for.

Humanordbøger:

Fokus på *bruger*: ordbøger beregnet for mennesker

Medium: traditionelt tryk; i dag også elektronisk udgave

Maskinlæsbare ordbøger:

Fokus på *medium*: Ordbogsdata fremstillet og/eller lagret på elektronisk medium

¹STANLEX gruppens medlemmer (i alfabetisk orden): Anna Braasch (CST; Terminologigruppen), Ib Elfving (Info:Partner A/S), Gert Engel (HHS, Institut for Erhvervsforskning), Hanne Jensen (IBM; efter d. 1.4.95 Munksgaards Forlag), Claus Bo Jørgensen (Textware), Bente Maegaard (CST), Bodil Nistrup Madsen (Kontaktperson; HHK, Institut for Datalingvistik; Terminologigruppen), Ole Norling-Christensen (Den Danske Ordbog), Hanne Ruus (KUA, Institut for Nordisk Filologi), Ragnar Heldt Nielsen (Teknologinævnet), Klaus Søndergaard (Dansk Standard), eksterne medlemmer.

Maskinbrugbare ordbøger:

Fokus på *bruger*: leksikalske datasamlinger beregnet til anvendelse i sprogteknologiske produkter

Termsamlinger:

Fokus på *indhold*: leksikografisk og faglig beskrivelse af termer

Medium: i denne sammenhæng fortrinsvis terminologi i databaseform

Bruger: udformet principielt til både mennesker og maskiner

Denne opstilling kan på ingen måde betragtes som typologi, den er snarere en ad hoc liste over de eksempler på leksikografiske produkter, som indgår i STANLEX-gruppens arbejdsgrundlag.

Sprogteknologiske produkter (fx et oversættelsessystem eller en såkaldt forfatterhjælp) inkluderer altid en eller anden form for maskinbrugbar ordbog, som et af systemets centrale moduler.

Sprogets hastige udvikling — måske især på det terminologiske område — kræver, at ordbogsarbejdet følger med. I fundamentet for maskinel sprogbehandling indgår også store basisordbøger for moderne almensprog, og da det er meget tidskrævende at udarbejde sådanne ordbøger, er det ønskeligt at kunne drage nytte af de eksisterende traditionelle humanordbøger (Braasch 1994a).

Den krævende systematik, der kendetegner maskinbrugbare ordbøger kan på den anden side virke gavnligt — eller bevidstgørende — ved udarbejdelse af ordbøger for mennesker (Braasch 1994b). Der åbnes nye specielle perspektiver i leksikografien, som også støttes af den moderne teknologi.

Der stilles store krav til den moderne leksikografi, både hvad oplysningsmængde og arbejdets kvalitet og hurtighed angår. Det er vigtigt, at kravene til det færdige produkt beskrives på et så tidligt tidspunkt i et leksikografisk projekt som overhovedet muligt. En dansk standard for lagring og udveksling af leksikalske data vil også kunne bidrage til hensigtsmæssig planlægning af nye typer ordbogsopgaver.

4 STANLEX arbejdsgruppen

På Teknologinævnets initiativ er der blevet nedsat en arbejdsgruppe, der fik navnet **STANLEX**, dels fordi gruppen skal beskæftige sig med **standardisering af leksikalske oplysninger**, dels også fordi gruppen opfatter sig som en slags fortsættelse af DANLEX-gruppen, som udførte pionerarbejde inden for analyse og klassificering af ordbogsoplysninger til elektronisk behandling af ordbogsdata.

Arbejdsgruppen omfatter repræsentanter for institutioner, herunder også Center for Sprogteknologi (CST), virksomheder og organisationer, der har viden om og interesse i sprogteknologi, datamatisk leksikografi og terminografi samt standardisering. Arbejdsgruppen er således meget bredt sammensat. Projektet er placeret hos Dansk Standard, under standardiseringsudvalget for informationsteknologi, under udvalget for teksthåndtering.

5 Målsætningen

Hovedopgaven er at udarbejde „et forslag til indhold i en dansk standard for indholds- og strukturbeskrivelse af leksikalske oplysninger“. Formålet med en fælles standard er „at lette genbrug og udveksling af leksikalske oplysninger“ (citerer fra intern målsætningserklæring). Da ordbogs- og terminologiarbejde er meget ressourcekrævende, kunne der opnås store besparelser ved at benytte ensartede metoder ved indholds- og strukturbeskrivelse af leksikalske data, der fx kan tænkes at indgå i sprogteknologiske værktøjer.

I denne sammenhæng er det måske på sin plads at give en kort definition af, hvad der her betegnes som **leksikalske data**. Vi har tidligere nævnt de ordbogstyper, som STANLEX-gruppen primært arbejder med. Leksikalske data er ganske enkelt elementer eller helheder i forskellige slags beskrivelser af „ord“ — eller opslagsenheder — indeholdt i disse ordbogstyper eller leksikalske datasamlinger.

I det efterfølgende redegøres for de udførte arbejdsopgaver og delresultater der er opnået inden den 1. maj 1995, og som kan anses for at være væsentlige led i udarbejdelsen af et forslag om standardisering af leksikalske data.

Vi har i projektet indtil nu koncentreret os om at indsamle, sammenligne og beskrive eksempler på tilgængelige maskinlæsbare ordbogsdata ud fra indholds- og strukturmæssige kriterier, primært på grundlag af **DANLEX-taksonomi** (DANLEX 1987). Taksonomien er præsenteret i tabelform i tabel 1.

6 Arbejdsproces og status

6.1 Arbejdsprocessen

En grundlæggende erfaring er, at datalagring i forskellige systemer besværliggør dataudvekslingen; med andre ord for at kunne sikre problemfri dataudveksling er det nødvendigt at nå til enighed om formater, data er beskrevet i, og måder, de bliver lagret på. Dette gælder naturligvis også leksikalske datasamlinger. På dette grundlag arbejder vi på et forslag til harmonisering af beskrivelses- og lagringsmetoder.

Arbejdsprocessen er planlagt til at bestå af to hovedfaser:

- I den første fase behandles projektdeltagernes leksikalske datasamlinger enkeltvis og i sammenhæng: der foretages indholdsanalyse og -beskrivelse med henblik på at nå frem til en fælles klassifikation, der kan dække alle oplysningstyper der forekommer i datasamlingerne.
- I den anden fase arbejdes med en SGML-baseret (Standard Generalized Markup Language) strukturbeskrivelse, der skal føre frem til forslag til standardformat(er) til lagring og udveksling af data.

Målet er at producere en teknisk rapport med vejledning om muligheder og fordele ved udveksling og genbrug af data, samt et egentligt forslag til standardiserede lagrings- og udvekslingsformater.

Hovedkategorier	Kategorier	Subkategorier
etymologiske oplysninger	<ul style="list-style-type: none"> • parallel • oprindelse • datering 	
fonetiske oplysninger	<ul style="list-style-type: none"> • prosodiske træk • segmentale træk 	
grafiske oplysninger	<ul style="list-style-type: none"> • ortografiske opl. 	<ul style="list-style-type: none"> • stavning • orddeling
	<ul style="list-style-type: none"> • grafisk symbol 	
grammatiske oplysninger	<ul style="list-style-type: none"> • ordklasse 	
	<ul style="list-style-type: none"> • bøjningsopl. 	<ul style="list-style-type: none"> • paradigmeopl. • bøjningsform
	<ul style="list-style-type: none"> • orddannelsesopl. 	
	<ul style="list-style-type: none"> • syntaktiske opl. 	<ul style="list-style-type: none"> • valens • syntaktisk funktion
pragmatiske oplysninger	<ul style="list-style-type: none"> • tekstlige opl. 	<ul style="list-style-type: none"> • citat • mulig kontekst
	<ul style="list-style-type: none"> • brugsopl. 	<ul style="list-style-type: none"> • tidlig dimension • rumlig dimension • social dimension • frekvens
	<ul style="list-style-type: none"> • eksternt henvisning 	<ul style="list-style-type: none"> • litteraturhenvisning • kildehenvisning
	<ul style="list-style-type: none"> • evalueringsopl. 	
	<ul style="list-style-type: none"> • administrative opl. 	<ul style="list-style-type: none"> • opl. om indsamling og bearbejdning af data • intern henvisning • homografnummer • tekniske oplysninger
semantiske oplysninger	<ul style="list-style-type: none"> • emneklassificerende opl. 	
	<ul style="list-style-type: none"> • oplysninger om semantiske relationer 	<ul style="list-style-type: none"> • generisk over/underordningsrelation • partitiv relation • successiv relation • kausalrelation • antonymi
	<ul style="list-style-type: none"> • indholdsspecificerende oplysninger 	<ul style="list-style-type: none"> • leksikalsk parafrase • analytisk/syntetisk definition • denotativ definition • ostensiv definition • faglig forklaring • semantiske træk
	<ul style="list-style-type: none"> • ækvivalensopl. 	<ul style="list-style-type: none"> • ækvivalensrelation inden for ét sprog • ækvivalensrelation mellem to eller flere sprog

Tabel 1: DANLEX-taksonomien, generelle oplysningstyper.

6.2 Udgangspunkt

Vi tager udgangspunkt i den tidligere nævnte DANLEX-taksonomi (tabel 1), der er opstillet af en gruppe forskere og leksikografer. Den grundlæggende undersøgelse omfattede en bred vifte af traditionelle leksikografiske opslagsværker, men maskinbrugbare ordbøger blev — naturligt nok — kun perifert nævnt. Siden dette arbejde er afsluttet (1986), har udviklingen af sprogteknologiske værktøjer medført et stadig stigende behov for maskinbrugbare ordbøger.

Derfor indgår nu også denne type ordbøger i STANLEX-gruppens arbejdsgrundlag, og det medfører, at den oprindelige taksonomi skal opdateres på forskellige punkter.

Den grundlæggende beskrivelse generaliserer oplysningskategorierne og indordner dem i en hierarkisk klassifikation, denne er DANLEX-taksonomien. Betegnelserne for opstillingens hovedkategorier er baserede på lingvistikens deldiscipliner, såsom etymologi, grammatik, fonetik, semantik, pragmatik osv.

Denne taksonomi er derefter — stadig inden for rammerne af DANLEX-projektet — blevet afprøvet som beskrivelsesværktøj på forskellige ordbogstyper, der inddrager det danske sprog, dvs. tosprogs- og etsprogsordbøger; ordbøger ordnet efter onomasiologiske hhv. semasiologiske principper; almensproglige og fagsproglige, synkrone og diakrone, præskriptive og deskriptive typer osv. På denne måde har man undersøgt, hvor bredt et spektrum af ordbøger den opstillede taksonomi kan siges at dække.

Undersøgelsen førte til den konklusion, at det var muligt og hensigtsmæssigt at opstille en sådan generel taksonomi, der viste sig også at kunne dække de andre germanske sprogs og i stor udstrækning de romanske sprogs leksikografiske behov. Desuden antog man, at taksonomien også var anvendelig til beskrivelse af leksikalske datasamlinger „til brug i informationsteknologiske systemer“ (DANLEX 1987). Da DANLEX-gruppen imidlertid afsluttede sin undersøgelse af ordbøger i 1985/86, har man følgelig ikke haft adgang til større sådanne datasamlinger for dansk. STANLEX-gruppen derimod, som i 1994, næsten 10 år senere er gået i gang med en videreførende undersøgelse, har adgang til et bredt udvalg af leksikalsk materiale udarbejdet til sprogteknologiske formål.

6.3 STANLEX materialet

Den største del af STANLEX-gruppens materiale stammer fra leksikografiske datasamlinger som deltagerne arbejder med til daglig. Hvert sæt data er forsynet med to slags beskrivelser.

For det første en *overordnet beskrivelse* af den pågældende datasamling, omfattende

- kort redegørelse for dens karakter (indhold, struktur, størrelse, status, medium, evt. grafisk fremtrædelsesform. . .)
- oversigt over datamatiske, lingvistiske og leksikografiske principper, der danner grundlaget for datasamlingen

Denne overordnede beskrivelse er udarbejdet for at sikre fælles teoretisk basis til den videre behandling af materialet.

For det andet en mere individuel, *detaljeret beskrivelse* af det pågældende leksikografiske materiale, som de enkelte medlemmer råder over, der omfatter

- fortegnelser over oplysningstyperne til sammenligning med DANLEX-taksonomiens kategorier hhv. subkategorier
- træk/værdiliste-erklæringer (dog ikke helt ned til de mindste detaljer)
- færdige ordbogsartikler (også kodet som databaseposter eller leksikalske regler) til illustration af brugen af træk i de forskellige oplysningstyper

Denne del af materialet er det centrale i arbejdsprocessen, idet DANLEX-taksonomien ønskes videre afprøvet og udbygget ved at inddrage sådanne ordbogstyper eller projekter i undersøgelsen, der endnu ikke var tilgængelige i den første halvdel af firserne — herunder især de sprogteknologiske.

Den nyeste viden stammer i høj grad fra arbejdet med datamatstøttet leksikografi og korpusundersøgelser samt maskinel sprogbehandling og andre grene af sprogteknologi. Det er naturligt, at der blandt STANLEX-gruppens medlemmer er mange repræsentanter for anvendt sprogteknologi, især maskinel og maskinstøttet oversættelse.

- Oversættelsessystemerne der indgår i undersøgelsen er:

PATRANS (Center for Sprogteknologi (CST) og Lingtech): maskinoversættelsessystem til patenttekster

METAL (Siemens, Handelshøjskole Syd): maskinoversættelsessystem til tekniske tekster

WINGER/Info:Partner: støtteværktøjer til maskinoversættelse

LMT (Logic-based Machine Translation) og TM2 (Translation Manager): maskinoversættelsesprogrammel fra IBM.

- Andre sprogteknologiske produkter, såsom elektroniske ordbøger, tekstsamlinger, leksikografisk software osv. er også repræsenterede (TextWare A/S)

- Desuden har gruppen adgang til terminologi i databaseform, nemlig til

DANTERM-basen (Handelshøjskolen i København), som har til hensigt at tilgodese både menneske og maskine som bruger.

På den anden side deltager nogle relevante projekter inden for den sproglige dimension med betydelig vægt på det datamatiske aspekt i det leksikografiske arbejde:

- Moderne dansk sprog

Den Danske Ordbog (under udarbejdelse), der bygger på et artikelformat, som beskrives som en SGML-struktur. Det benyttede datamatiske redigeringsværktøj (GestorLex) er et avanceret sprogteknologisk produkt.

- Historisk sprog

er repræsenteret ved *Folkeviseprojektet*; her arbejdes med elektronisk tekstbase og ordbase. Projektet rejser specielle lingvistiske og datamatiske problemer (fx flere ortografiske sideformer og dertil hørende komplekse søgerutiner), som også indgår i STANLEX-gruppens overvejelser.

Fra forlagssiden deltager Munskgaards Forlag.

Hovedkategorier	Kategorier	Subkategorier
etymologiske oplysninger	<ul style="list-style-type: none"> • parallel • oprindelse • datering 	+ (under udarbejdelse)
fonetiske oplysninger	<ul style="list-style-type: none"> • prosodiske træk • segmentale træk 	
grafiske oplysninger	<ul style="list-style-type: none"> • ortografiske opl. 	<ul style="list-style-type: none"> • stavning • orddeling
grammatiske oplysninger	<ul style="list-style-type: none"> • grafisk symbol 	
	<ul style="list-style-type: none"> • ordklasse + køn 	
	<ul style="list-style-type: none"> • bøjningsopl. 	<ul style="list-style-type: none"> • bøjningsparadigme • bøjningsform
	<ul style="list-style-type: none"> • orddannelsesopl. 	
	<ul style="list-style-type: none"> • syntaktiske opl. 	<ul style="list-style-type: none"> • syntaktisk ramme • syntaktisk funktion
pragmatiske oplysninger	+ grammatisk specifikation	
	<ul style="list-style-type: none"> • tekstlige opl. 	<ul style="list-style-type: none"> • citat • mulig kontekst + kontekstbegrænsning
	<ul style="list-style-type: none"> • brugsopl. 	<ul style="list-style-type: none"> • tidslig dimension • rumlig dimension • kommunikativ dimension • frekvens + teksttypologisk opl.
	<ul style="list-style-type: none"> • eksternt henvisning 	<ul style="list-style-type: none"> • litteraturhenvisning • kildehenvisning
	<ul style="list-style-type: none"> • evalueringsopl. 	+ (under udarbejdelse)
semantiske oplysninger	<ul style="list-style-type: none"> • administrative opl. 	<ul style="list-style-type: none"> • opl. om indsamling og bearbejdning af data • intern henvisning • homografnummer • tekniske oplysninger
	<ul style="list-style-type: none"> • emneklassificerende opl. 	+ (under udarbejdelse)
	<ul style="list-style-type: none"> • oplysninger om semantiske relationer 	<ul style="list-style-type: none"> • generisk over/underordningsrelation • partitiv relation • successiv relation • kausalrelation + associativ relation • antonymi + metonymi
	<ul style="list-style-type: none"> • indholdsspecificerende oplysninger 	<ul style="list-style-type: none"> • leksikalsk parafrase • definition • denotativ definition • ostensiv definition + udvidelse af definition • faglig forklaring • indholdsspecif. træk
	<ul style="list-style-type: none"> • ækvivalensopl. 	<ul style="list-style-type: none"> • ækvivalensrelation inden for ét sprog • ækvivalensrelation mellem to eller flere sprog + ækvivalensbegrænsning

Tabel 2: STANLEX-taksonomien, generelle oplysningstyper, status: sommeren 1995. Tilføjelser i forhold til DANLEX-taksonomien er markeret med + og fremhævet med fed. Andre ændringer er fremhævet med fed.

6.4 Sammenligning af STANLEX-materialet med DANLEX-taksonomien

Arbejdsfase 1, dvs. sammenligning af STANLEX-materialet med DANLEX-taksonomien er nu gennemført, og resultaterne er nedfældet i tabelform. Derefter arbejder vi på at opstille den nye STANLEX-taksonomi. En foreløbig version af denne taksonomi er præsenteret i tabel 2. (Ændringer i forhold til DANLEX-taksonomien er fremhævet med fed.)

Især analysen af maskinbrugbare ordbøger bragte nye aspekter ind i opstillingen af den taksonomiske hierarki, derfor fokuserer fremstillingen nedenfor på denne ordbogstype. Et af de centrale spørgsmål var følgende: På hvilke punkter skal DANLEX-taksonomien udbygges for også at dække maskinbrugbare ordbøger?

De grundlæggende krav til formuleringen af oplysninger i maskinbrugbare ordbøger er, at den leksikografiske beskrivelse skal være

- eksplicit
- entydig
- udtømmende
- konsistent
- formaliseret

To vigtige aspekter bør fremhæves i forhold til traditionelle (humane) ordbøgers måde at bringe oplysninger på, nemlig ekspliciterings- og formaliseringsgraden. Disse to aspekter er specifikke for oplysninger i leksikografiske datasamlinger, der skal kunne bruges af maskiner, dvs. udnyttes i sprogteknologiske værktøjer. Maskinen skal have direkte adgang til hver bid af relevant information, som ellers kan gives implicit eller vha. eksempel, omskrivning eller forklaring i ordbøger for mennesker (jf. Braasch 1994) samtidig med at informationen skal gives i en fuldstændig fastlagt form i overensstemmelse med systemets formaliseringskrav.

Arbejdsforløbet omfattede blandt andet en detaljeret sammenligning mellem oplysningstyperne i de undersøgte datasamlinger og den oprindelige taksonomis kategorier. Det første spørgsmål var:

- Hvorvidt og hvordan kan oplysningstyper, der forekommer generelt i maskinbrugbare ordbøger, tilordnes DANLEX-taksonomiens kategorier hhv. subkategorier?

I dette skridt blev overensstemmelser og afvigelser mellem den oprindelige taksonomi og det nye sprogteknologiske materiale registreret.

Vi opstillede derefter sammenlignende tabellariske oversigter ud fra to synsvinkler:

1. Oversigter over DANLEX-taksonomiens kategorier/subkategorier og deres repræsentation i hver individuel datasamling, der indgår i STANLEX-materialet
2. Oversigter over den enkelte datasamlings oplysningstyper grupperet efter DANLEX-taksonomiens kategorier/subkategorier

Vi analyserede derefter nogle typiske ordbogsartikler fra hver datasamling og forsøgte at forsyne deres oplysningstyper med kategori/subkategori-betegnelser fra taksonomien. Det viste sig i en række tilfælde, at det ikke var nogen triviell opgave at tilordne en oplysning til en bestemt type og dermed en bestemt taksonomisk subkategori. Andre gange manglede der i taksonomien en dækkende subkategori. Især oplysninger, der hører til under hovedkategorien *Grammatiske oplysninger* viste sig at være vanskelige at håndtere.

Maskinbrugbare ordbøger fx til oversættelsessystemer er særlig detaljerede og eksplicite inden for hovedkategorien *Grammatiske oplysninger*. De indeholder en række oplysninger, fx oplysning om ledfunktionen af valensbundne argumenter ved verber, der håndteres ud fra forskellige lingvistiske principper i de enkelte maskinbrugbare ordbøger. De kan derfor enten tilordnes den taksonomiske kategori *Syntaktiske oplysninger* (derunder subkategorien *Syntaktisk funktion*), eller kategorien *Grammatisk specifikation*.

Vi arbejder stadig med en udvidet liste af (sub)kategorier, der foreløbig omfatter oplysningstyper under kategorierne

- Syntaktisk ramme
- Syntaktisk funktion
- Grammatisk specifikation

under hovedkategorien *Grammatiske oplysninger*.

Disse oplysningstyper er velegnede til formalisering, fordi de forholdsvis enkelt kan systematiseres, fx ud fra nogle lingvistiske eller datamatiske principper, og derefter kan de repræsenteres i en enkel formel kodeform i maskinbrugbare ordbøger.

De andre ordbogstyper gav anledning til at gå i dybden med andre taksonomiske kategorier:

- Humanordbøger bidrager primært med oplysningstyper inden for hovedkategorierne *Etymologiske oplysninger*, *Semantiske oplysninger* og *Pragmatiske oplysninger*. Under den sidst nævnte er en ny subkategori indført i taksonomien, nemlig *Teksttypologisk oplysning*. Arbejdet på *Den Danske Ordbog* er banebrydende på dette punkt, idet hvert tekststykke i korpuset er forsynet med 'Headers' — en slags overskrifter — som fastholder tekstkildens relevante egenskaber. Denne oplysning er systematisk opført i den leksikografiske beskrivelse (af eksempelmaterialet). Desuden er fx normative oplysninger også kun repræsenteret i humanordbøger, hvor de til gengæld spiller en vigtig rolle. Dette har vi endnu ikke arbejdet med i detaljer.
- Terminologibaser stiller særlige krav til detaljeringsgraden i kategorien *Indholds-specificerende oplysninger* (især til subkategorierne for definitionstyper). De hører hjemme under hovedkategorien *Semantiske oplysninger*. Fagsystematisk beskrivelse hører til under den anden subkategori, nemlig *Oplysninger om semantiske relationer*. Vi har endnu ikke beskæftiget os mere dybtgående med disse oplysningstyper, men som det kan ses ved en sammenligning af den oprindelige (DANLEX) taksonomi og den opdaterede version (STANLEX), er der også her kommet nogle nye subkategorier til.

Det turde fremgå af fremstillingen ovenfor, at de forskellige ordbogstyper bidrager til opdateringen af taksonomien under hver sin(e) hovedkategori(er); man kan tale om en synergi-effekt.

På grundlag af de indtil nu udførte analyser og sammenligninger kan vi konkludere, at DANLEX-taksonomiens *hovedkategorier* og (med en enkelt undtagelse også) *kategorier* har vist sig at være dækkende for STANLEX-materialet. I kolonnen for *subkategorier* (dvs. på tredje niveau i hierarkiet) har vi fået brug for en del tilføjelser. For især de grammatiske oplysninger gælder det, at selvom kriterierne for maskinbrugbarhed af leksikalske datasamlinger i høj grad bygger på kravet om eksplicit og udtømmende formulering, er det ikke hensigtsmæssigt at udbygge taksonomien med et fjerde hierarkisk niveau, svarende til subkategorier af anden grad.

Grunden til denne beslutning er for det første, at taksonomien bliver tungere at arbejde med (især i de sammenhænge, hvor man ikke har brug for denne yderligere detaljeringsgrad); for det andet viste det sig, at oplysningstyperne på dette niveau i høj grad er system- og teoriafhængige, og derfor vil man ikke kunne opretholde princippet om generaliserbarhed.

7 Hvad skal/kan en sådan standard bruges til og hvordan?

Den her følgende redegørelse er baseret på en foreløbig skitse over STANLEX-gruppens opgaver nedfældet i notatet *Forslag til indhold i en dansk standard for indholds- og strukturbeskrivelse af leksikalske oplysninger*. Standarden vil omfatte

- klassifikation af oplysninger (i hovedkategorier, kategorier og subkategorier), dvs. en taksonomi
- opstilling af beskrivelsesmodeller for ordbogsstrukturer (makro- og mikrostruktur)

STANLEX-gruppen har foreløbig defineret to primære anvendelsesområder for en sådan standard. Det første er:

- *Standarden skal bruges ved udveksling af leksikalske data. . .*

En sådan situation foreligger fx, når man ønsker at udvide ordbogen i et maskinoversættelsessystem (det såkaldte *lexicon*) med ordbogsindgange fra et andet system eller fra en offentligt tilgængelig termbase. Der kan også være tale om at flette to maskinlæsbare ordbøgers materiale sammen til én enkelt, ny og større ordbog med det formål at udgive den i trykt eller elektronisk form. Hurtig udveksling af store mængder data er ikke længere et teknisk problem, men det forbliver en omfattende opgave at analysere og harmonisere datasamlinger som er meget forskellige, hvad indhold og struktur, men også hvad beskrivelsesmåde, dvs. præsentation af oplysninger, angår.

Inden sammenlægning af to eller flere leksikalske datasamlinger er det altså nødvendigt at analysere materialet og dertil har man brug for taksonomien. Analysen viser

- på hvilke punkter datasamlingernes oplysninger er kompatible
- på hvilke punkter der er indholdsmæssige afvigelser (fx i detaljeringsgrad af samme oplysning eller grundet forskelle i lingvistisk tilgang osv.)

- hvilke oplysninger der mangler i det samlede materiale i forhold til den planlagte beskrivelsesmodel

Datasamlinger bør i øvrigt forsynes med udførlig (og præcis) dokumentation og kan med fordel blive struktureret og beskrevet iht. den givne taksonomi.

Det andet anvendelsesområde skitseres på følgende vis:

- *Standarden kan . . . også bruges ved planlægning, udvikling og lagring af leksikalske data.*

Det er klart, at når man indser fordelene ved at have adgang til flere (indholds- eller størrelsesmæssigt — eller på anden måde) forskellige men kompatible leksikalske datasamlinger, vil man bestræbe sig på at udforme fremtidige ordbøger sådan, at de kan kombineres indbyrdes og udnyttes i flere sammenhænge. Resultatet vil blive at materialet kan bruges og genbruges i forskelligartede leksikografiske produkter.

Inden for standarden er det hensigten, at den opdaterede taksonomiske opstilling over leksikalske oplysningstyper skal fungere som generelt beskrivelsesværktøj og skal være udbygget med

- definitioner for alle kategorier og subkategorier
- eksempler på klassificering af faktiske oplysninger
- kommentarer, når der synes at være brug for uddybende forklaring

For at øge standardens anvendelighed i praktisk planlægning og udarbejdelse af leksikalske datasamlinger gives der supplerende eksempel materiale i et anneks. Dette materiale vil efter de nuværende planer omfatte

- typeeksempler på beskrivelsesstrukturer i leksikografiske datasamlinger
- sammensætning af standardelementer til forskellige leksikografiske formål
- eksempler på hensigtsmæssige lagrings- og udvekslingsformater bygger på standard-elementer

STANLEX-gruppen arbejder i indeværende, anden arbejdsfase på at løse ovennævnte opgaver.

8 Afsluttende bemærkninger

Da STANLEX-projektet befinder sig midt i en arbejdsproces er der endnu ikke så meget andet at konkludere på nuværende tidspunkt end det jeg har kunnet sige sammenfattende om de delresultater, der er nået indtil nu.

Det danske initiativ står naturligvis ikke alene i den faglige omverden. Der er på internationalt plan forskellige typer projekter i gang inden for leksikografi og standardisering; nedenstående nævnes et par eksempler på denne aktivitet.

TEI: (Text Encoding Initiative, under organisationerne ACH, ACL, ALLC) har efterhånden løbet gennem en del år. Der udgives med mellemrum en ny (opdateret) version af publikationen „Guidelines For the Encoding and Interchange of Machine-Readable Texts“ (TEI 1994). Foruden andre emner, der er relevante for forskellige typer af dokumenter har rapporten også et større afsnit om „Print Dictionaries“. Afsnittet giver gode retningslinjer til at arbejde med makro- og mikrostruktur, hierarkisk strukturering af artikler og oplysningstyper og meget andet.

En anden type projekt er **EAGLES** (Expert Advisory Group on Language Engineering Standards), et komplekst EU-støttet projekt. En af dets undergrupper beskæftiger sig med at sammenligne hvordan human- og maskinbrugbare ordbøger koder morfosyntaktiske oplysninger. Alle EU-sprog er blevet inddraget. Formålet med dette delprojekt er at udarbejde et forslag til international standard til repræsentation af morfosyntaks i leksikalske datasamlinger.

Litteratur

- Bech, Annelise/I. Elfving/G. Engel/J. Lund/B. Mægaard/B. Nistrup-Madsen/A. Melchior/A. Møller/F. Svanholm/K. Aakjær/R. Heldt Nielsen 1994: Dansk sprogteknologi — status, perspektiver, handlemuligheder. *TeknologiNævnets rapporter 1994/1*. København.
- Braasch, Anna 1994a: How Far Do Printed Dictionaries and MT-Lexicons Share Information? I: Paolo Alberto/Paul Bennett (ed.): *Lexical Issues in Machine Translation. Studies in Machine Translation and Natural Language Processing*. Vol. 8. Luxembourg: CEC, 117–133.
- Braasch, Anna 1994b: Når maskinen tager én på ordet — ordbogsarbejde for maskinoversættelse. I: Anna Garde/Pia Jarvad (udg.): *Nordiske studier i leksikografi II. Skrifter undgivet af Nordisk Forening for Leksikografi*. Skrift nr. 2. København: LEDA, 29–38.
- The DANLEX-GROUP (Ebba Hjorth/B. Nistrup-Madsen/O. Norling-Christensen/Rosenkilde Jacobsen/H. Ruus) 1987: *Descriptive Tools for Electronic Processing of Dictionary Data*. Studies in Computational Lexicography. Lexicographica Series Maior 20. Tübingen: Max Niemeyer Verlag.
- TEI 1994 = C. M. Sperberg-McQueen/Lou Burnard (ed.): *Guidelines For the Encoding and Interchange of Machine-Readable Texts*. Chicago/Oxford.

Hovedværk og supplement

A number of greater dictionaries have found themselves compelled to publish a supplement in one or more volumes in order to improve the original work. This paper deals with three Danish dictionaries and, concerning an integration of the main dictionary and the supplement, *The Oxford English Dictionary*. The paper argues against a similar project regarding to *Ordbog over det danske Sprog*.

Tillæg eller supplementer til ordbøger bliver ofte glemt af ordbogsbrugerne. Det synes derfor at være en dårlig idé at trykke supplementer. Når det alligevel sker har det naturligvis forskellige gode grunde. Under udarbejdelsen af en ordbog dukker der nemt nogle ting frem som det var ønskeligt at få med, eller at få korrigeret, men som ikke mere kan placeres på rette sted. Disse ting kommer ofte i et tillæg, og de vil blive optaget i en senere, revideret udgave. Mindre ordbøger kan naturligvis lettere end de større værker udkomme i nye udgaver, der forbedrer og udvider en ældre. I det følgende vil jeg kun beskæftige mig med de større ordbogsværker.

1 Tre større danske ordbogsværker

Otto Kalkars *Ordbog til det ældre danske Sprog* udkom i fire bind fra 1881 til 1907. Et femte bind, som er det tykkeste, med supplerende stof til hele alfabetet, er trykt 1908–18. Dette bind kalder sig blot „Femte Bind“. Allerede i slutningen af første bind er der supplerende stof, nemlig 65 sider med titlen „Tillæg og Rettelser“; de er almindelig kendt som I.-stjerne. Dette tillæg er optaget i bind V på den måde at der gives sidehenvisninger til I.*

H. F. Feilbergs *Bidrag til en Ordbog over jyske almuesmål* udkom i tre bind fra 1886 til 1911. Et fjerde bind, der af omfang nærmest er et halvt bind, kalder sig „Tillæg“ og er trykt 1912–14. Bagest i dette bind er der syv og en halv siders yderligere tillæg med overskriften: „Ord, ordformer, oplysninger og henvisninger som er kommet forf. i hænde eller fundet under og efter den sidste korrekturlæsning af Tillægget.“

Ordbog over det danske Sprog, ODS, grundlagt af Verner Dahlerup omkring århundredskiftet, består af 27 egentlige ordbogsbind, som er udgivet fra 1918 til 1954. Et supplement på fem bind er under udgivelse, og jeg er redaktør ved det; de to første bind er udkommet i 1992 og 1994.

2 Supplementernes indhold i forhold til hovedværkerne

De oplysninger som et supplement bringer kan i princippet henregnes til følgende arter:

1. Afhjælpning af tilfældige mangler i hovedværket.
2. Rettelse af fejl.
3. Færdiggørelse af henvisningsapparatet inden for værket.

4. Tilføjelse af opdaterende oplysninger.
5. Ændring af hovedværkets sigte.

Om disse punkter vil jeg foreløbig sige i al almindelighed, at de tre første nærmest følger af sig selv. Ved ordbøger med oplysninger om samtidssproget vil desuden en opdatering hurtigt blive ønskelig. En ændring af hovedværkets sigte er en betænkelig sag og noget der kan true det gode forhold mellem hovedværk og supplement.

Redaktionsarbejdet ved et supplement er præget af både en tvang og en frihed. Tvangen ligger i bundetheden til hovedværkets struktur. Materialet skal fremlægges inden for de rammer der er fastlagt i hovedværket. I makrostrukturen gælder det at opslagsordenes form og rækkefølge vanskeligt kan ændres, fx har ODS-Supplementet bevaret et dobbeltskrevet *a* i stedet for *å* i opslagsordene, og det alfabetiseres som *a+a*. I mikrostrukturen bør fx hovedværkets betydningsinddeling så vidt muligt respekteres. Friheden ligger i retten til at lade en artikel stå som den gamle redaktør udformede den. Det forholder sig helt anderledes ved redigering af en ny udgave, hvor hver linje bliver den nye redaktørs tekst, også det der overtages uændret fra forgængerens. Et supplement kan vælge sin medtagelsespolitik og lade visse sider af den gamle ordbog stå ukommenteret.

Materialet til et supplement kan dels være tilkommet i løbet af hovedværkets redaktionstid, dels være fremskaffet med henblik på et supplement.

3 Kalkar og Feilberg

I Kalkars ordbog (se figur 1) er en gradvis udvidelse af rammerne tydelig. I værkets efterskrift beklager Kalkar at indsamlingen af materiale er sket efter tre forskellige planer. Om dette forhold vil jeg citere Marie Bjerrums afhandling (1976:99):

Orprindelig var det Kalkars mening at lave en ordsamling af gamle ord fra tiden 1300–1700, dvs. ord, som ikke mere kendtes eller brugtes, og desuden ord, som i de gamle tekster havde en fra den nuværende afvigende betydning. Denne plan afløstes af den, som han fik pålagt af Videnskabernes Selskab i 1873 som betingelse for støtte til ordbogen, nemlig at han skulle udvide samlingerne ved at gennemgå flere kildekrifter og medtage ikke blot alle ord, som ikke fandtes mere, eller som havde fået en afvigende betydning, men også ord, som afveg fra nydansk med hensyn til genus og bøjningsformer, eller hvis ortografi var udtryk for en fra nyere dansk afvigende lydlig form. Den tredje plan er den, som ordbogen med tillægsbindet er udtryk for, nemlig at give en oversigt over hele ordforrådet 1300–1700.

Feilbergs tillæg (se figur 2) går derimod ikke ud over de principielle rammer for hovedværket. I indledningen til første bind (s. IV) gør Feilberg rede for hovedpunkterne i sin plan; det hedder her blandt andet:

Hele det jyske ordforråd indtages og opsamles såvidt muligt; altså ikke blot de ejendommelige eller sjældne ord og talemåder, som glossariet indeholder, men lige så fuldt de mest dagligdags. [...] De kulturminde, sproget gæmmer, fremdrages og forklares, såvidt jeg evner det.

— **Bygning**, no. hunk. (∞ sv. bygning.); om the samme bygningh købe wille, tha schulle the werre hinde negesth at købe (1541). N. D. Mag. I. 188; forstyrrede den bygning før, end hun bleff fuldkommen. Hvitf. II. 61. — ik.; vaar dette gantske bygning, deelt met tuende skeluerck. Vedel, Saxo. 434; det biugning kallis Hierusalem. Ranch. 30. — ft.; for de konstlige bygning skyld. Ranch. 11; bygninger. sst. — biugning. Røsenv., Gl. L. IV. 105; V. 507; (1645). Orion (Kvartalskr.) II. 196; Wal-lensbek, Ptochotrophæi Hol-mensis topogr. 5; Geh. Ark. Årsb. II. 288.

— **Bygning**, no.; beprydet med skøne bygninge. Bunting, Itinera-rium. II. 109.

Bygning, no.; — hank.; HM 177 (ovf. III. 814a7). — ik. (1542). RdU IV. 14 (ovf. IV. 160a20); (1585). DS V. 173 (ovf. II. 391b40). — ft. ∞; de bygning her baade gamle oc ny. VdP 12v; — se I.* 45b. Jf for- (I.* 61b), himmel-, hus-, ind-, mølle-, skib-, sten-, træ-, udb.

Figur 1. Otto Kalkar: *Ordbog til det ældre danske Sprog*. I, I* og V.

blind, to. *blīn -n* (D.); *bleōn* (Åbenrå, Ang.); *bleñ -ñ* (Løgst., Sams); *blej bleñc* *blēj* (Vens.) = rgsn., *de kan æn bleōn fæl å æ veke* (Vlb.), om hvad der er overmåde let at se el. forstå (også vestj.); *æn bliñ karm viñaror*, ruder malede udenpå væggen; *æ ruðor æ bleñ* (vestj.) uigennem-sigtige, anløbne; *æn bliñ Batøl*, en blinde Bertel (s. d.); *o gyjor nåvot i bliñ*; *go i blēj* (Vens.); *bl—mañ* i kortspil; *bl—basøl*, se barsel; „blind“ kaldes konen, når hun er kirkegangskone og har mistet sit barn (N. Sams); *bl—is*, se blind-skred.

blind brændenælde, J. T. 312 = blind nælde smst., en plante, døvnælde tvetand, lamium album L. (Haderslev, Østerlinnet).

blind, to. vittighed: *de ka dær go manø bliñ fābi-j udøn o sije* (Askov); ordspil: *æn ku gāt tol o vær bliñ, wår ot et får æn sājens skyl* (vestj.); — man bliver bl-, om man sover i måneskin (vist alm.), jfr. Trevelyan Wales 40; ved sporerne af bovister (vist alm.), se ulvefis; af svaletæg, Tobias II. 11; om man til-stopper kilde, Kr. Sagn III. 208.47; spot-ter med hellig kildes vand, Thiele II. 25; vil prøve vandet jule- el. nyårsnat, FbJ. I. 270. 356; ariansk biskop, s. opbryder kirke, Meyer Abergf. 160; hær blindes, da spyd kastes over den, Rafn, Oldn. S. V. 222; — msker blindes som straf: Rafn, Oldn. S. IV. 138, Olaf Hell. S. kap. 73, Rørik, jfr. V. 240, se Gr. RA. 707.5; Helge Thorersen var blind, fordi Ingeborg Gudmundsdatter rev hans øjne ud, Rafn, Oldn. S. III. 126; den dygtige mester blindes, se mester II. 584.40 a,

Figur 2. H. F. Feilberg: *Bidrag til en Ordbog over jyske Almuesmål*. I og IV.

Dette punkt, folkemindestoffet, kom til at spille en stadig større rolle for Feilberg. Når det i tillæggets slutord (s. II) hedder:

Så er jo da bogen udkommet. Tillægget er, såvidt jeg har ævnet og magtet, ført op til dato

så tænkes der (ifølge sammenhængen) på den folkloristiske litteratur. Videre slås det fast at

Det var fra begyndelsen af min tanke ej blot at søge og muligst nøjagtig gengive ord og ordformer, men bruge dem i de enkelte artikler som et midtpunkt, hvorom folkets tænke måde gennem eksempler fandt et nogenlunde fyldigt udtryk.

4 ODS

ODS-Supplementets forhold til sit hovedværk er ikke helt så enkelt. I indledningen til ODS (I s. XXI) hedder det at ordbogens opgave er

at give en fremstilling af det danske ordforraad, saaledes som dette er, og som det har udviklet sig fra Holbergs tid til vor, saa vidt muligt oplyst ved hjemmelsteder fra den almindelige og faglige litteratur i dette tidsrum.

I indledningen til ODS-S (I s. IX) siges at formålet er „at supplere ODS’ oplysninger om det danske sprog i perioden 1700–1955“. Men der siges ikke noget om sproget på supplementets redaktionstid. ODS-artiklerne til de første bogstaver er redigeret færdig omkring 1920 og beskriver sproget dengang, samtidig med at de foregående 200 år inddrages. Supplementets artikler til de første bogstaver er færdigredigeret omkring 1990, men formålet er nu at give supplerende oplysninger om sproget i et afsluttet tidsrum der jo efterhånden ligger ret langt tilbage i tiden. ODS-S ser således bort fra et af ODS’ formål, og vel hovedformålet, nemlig at beskrive sproget som det er nu. Man kan sige at ODS omdefineres til at være en periodeordbog over tiden 1700–1955, og det er denne ordbog der suppleres.

Men hvorfor er det blevet sådan? Hvorfor opererer ODS-S ikke med et nyt ajourført nutidssprog?

Det Danske Sprog- og Litteraturselskab, som udgiver ordbogen, ønskede ved værkets afslutning en afrunding af ODS, sådan at hele perioden op til tiden for det sidste ordbogsbind blev behandlet m.h.t. hele alfabetet, og sådan at visse mangler i ODS’ citatgrundlag blev afhjulpet. Det blev præciseret fra bevillingsgivernes side at formålet for det nye projekt kun skulle være at give værket en afrunding, ikke at følge sprogets udvikling ud over tiden for de sidste ODS-bind.

Ingen gjorde sig dengang helt klart at den ordbogstype som ODS tilhører er karakteriseret ved en beskrivelse under ét af sproget nu og sproget før. Hans Jonsson (1992) har kaldt det en „sambeskrivning“. Hvordan så når det aktuelle sprog ikke skulle beskrives? Redaktør Jørgen Glahder, den daværende ordbogsleder, handlede leksikografisk korrekt ved i nogle prøveartikler at negligere kravet om en tidsgrænse. Han fortsatte altså med at beskrive sin samtids sprog. Han medtog de nyeste ord, og hvis en tidspræget ODS-oplysning, fx en udtale- eller brugsbetegnelse, ikke passede med sproget o. 1960, så blev den rettet. I det supplement som nu er under udgivelse vil en sådan tidsbundet ODS-oplysning derimod blive stående som den gamle redaktørs bedømmelse af sin tids sprog. Noget andet er at

der eventuelt kan komme en tilføjelse, hvis et ords brug er blevet markant ændret op imod slutningen af ODS-perioden; fx har ordet *billedkunstner*, der i 1920 er anset for sjældent, fået tilføjelsen „senere alm.“. Nu retter vi kun hvad der var forkert dengang det blev trykt. Det var naturligvis tanken at supplementet skulle færdiggøres ret hurtigt, om ikke i Glahders resterende funktionstid på cirka 10 år (indtil 1964) så i hvert fald 5–10 år senere. Men selv om dette havde været realistisk, så var der dog mere end en menneskealder imellem redaktionstiden for de første bogstavers artikler og supplementsartiklerne til dem. Det nye nu-sprog ville på mange punkter være forskelligt fra det gamle.

Det er et spørgsmål om man overhovedet kan ajourføre en aktuel håndbog eller ordbog ved hjælp af et tillæg. Den rigtige løsning er naturligvis en ny udgave eller en helt ny bog. Hvis der bare skal tilføjes noget, går det an, men hvis der også skal ændres eller stryges noget, bliver det besværligt. Hvis alle ændringerne skulle meddeles i form af et supplement, så blev det et supplement der ret beset var surrogat for en ny udgave. Nu er det som nævnt ikke meningen at ODS skal ajourføres i den forstand, for vi beskriver ikke sproget i 1990'erne. Ligesom Jørgen Glahder vil vi gerne redigere leksikografisk forsvarligt, men det bliver på et andet grundlag. I Supplementets indledning (s. XXV) hedder det:

ODS er ikke mere en gyldig nutidsordbog, og den var det heller ikke da de sidste bind udkom. [...] Men ODS har blivende værdi som en samling af oplysninger vedrørende dansk sprog og kultur gennem de 250 år der kan kaldes for ODS-perioden. Det er denne side af ODS, det sprog- og kulturhistoriske opslagsværk, der kan suppleres; og den gamle samtidsordbog kan ses som et tidsdokument der gør første halvdel af det 20. århundrede til en særlig veldokumenteret periode i det danske sprogs historie.

Et supplement kan som nævnt vælge at interessere sig mere eller mindre for de forskellige arter af oplysninger i hovedværket. ODS-S vil udbygge fremstillingen af det danske ordforråd gennem de 250 år, så godt materialet nu tillader det. Man kunne også sige: fremlægge det vigtigste af den materialesamling som på en del punkter er noget tilfældig og utilstrækkelig. Hovedvægten kommer til at ligge på følgende forhold:

- nyordsstoffet i slutningen af perioden, altså ord og ordanvendelser der endnu ikke var i sproget da vedkommende afsnit af ODS blev trykt
- almensprogets fremmedord gennem hele perioden
- henvisninger inden for værket

Men også andre forhold får en bedre belysning, bl.a. ordenes datering (se Duekilde 1994). Og citatstoffet udvides, især til de første bogstaver. Derimod kan vi ikke være forpligtet til at følge hele det gamle ordstofs skæbne op til periodens afslutning, 1955. Det gælder måske især det faglige ordstof, som ODS lagde stor vægt på at registrere fra sin samtid. I mange tilfælde er oplysningerne til disse ord blevet forældet et par årtier senere. ODS-S-redaktionen ved det ikke altid og kan tillade sig ikke at tage stilling til det.

ODS-Supplementet kan siges at rumme en ændring af hovedværkets sigte. Skuden drejes bort fra samtidsretningen og får retning bagud i tiden. Den kolliderer ikke med et af Det

Danske Sprog- og Litteraturselskabs nye flagskibe: *Den Danske Ordbog*, en 6-bindes ordbog over moderne dansk, som er under udarbejdelse her i 1990'erne.

5 Hovedværk + supplement = ?

Det er ubekvemt med supplementer. Det kan synes ønskeligt at få oplysningerne fra et supplement flettet ind i hovedværket, så brugerne slipper for at slå op flere steder. Resultatet af en sådan sammenfletning forventes at blive en ordbog af samme karakter som hovedværket, men udvidet og forbedret sådan som supplementet nu har kunnet gøre det. Hvis et supplement kun indeholder oplysninger af en art som kunne have været medtaget i første omgang, så kan det være en god idé at flette dem ind i hovedværket. Dette forudsætter for øvrigt også at værket er af en sådan leksikografisk kvalitet at det kan bære udvidelsen.

Både Kalkars ordbog og Feilbergs ordbog er genudgivet i fotografisk optryk, og det vil jo sige at brugerne stadig må trækkes med supplementerne. Til gengæld er der ved at blive udarbejdet helt nye ordbøger til gammeldansk, som er den ældre del af Kalkars periode, og til jysk (se litteraturfortegnelsen).

En sammenkobling af ODS og ODS-S er fra flere sider blevet nævnt som et ønskeligt projekt, selv om det vel af markeds-mæssige grunde må siges at være urealisabelt. ODS er ved at udkomme i fotografisk optryk for fjerde gang, og her medtages naturligvis supplementsbindene. Jeg vil alligevel diskutere sammenkoblingsprojektet, fordi jeg er blevet lidt forskrækket over at møde dette ønske fra flere sider og fra personer som jeg troede nemt kunne se de principielle uoverensstemmelser mellem ODS og ODS-S. Men først vil jeg vise hvordan en sådan sammenføring af oplysningerne er gennemført i England.

6 Oxford-ordbogen

The Oxford English Dictionary udkom i tolv store bind 1888–1928, og de fik et supplementsbind i 1933. Et større supplement, i fire bind, udkom 1972–86, og det overflødig-gør opslag i 1933-bindet. Derefter blev det hele gjort maskinlæsbart. Næste trin blev en integreret udgave af hovedværk og supplement (med tilføjelse af ca. 500 nye ord) i 1989; den kalder sig 2. udgave af OED og foreligger også som CD-Rom. Der er her foretaget en sammenfletning med den mindst mulige ændring af de to tekster.

I eksempelartiklen (figur 3) står ordet „also“ i stedet for: „Common in recent use and frequently having the sense of“. Et gammelt „nu“ er altså fjernet. Men den nye udgave har langt fra i alle tilfælde ajourført de gamle oplysninger i definitioner og brugsangivelser. En oplysning som „now rare“ kan enten betyde sjældent omkring 1900 eller sjældent i 1989. Noel Osseltons anmeldelse (Osselton 1993) påviser dette problem og konkluderer at den omhyggelige læser ofte må søge tilbage til de oprindelige bind for at tolke indholdet i den nye udgave. Oxford-redaktionen kender naturligvis godt disse svagheder. Der er gjort rede for de manglende moderniseringer og forbedringer i slutningen af indledningen til 2. udgave, og den kommende, reviderede, 3. udgave nævnes. Der er altså sparet på redaktionsarbejdet ved den engelske sammenfletning. Formålet har ikke været at give en dækkende beskrivelse af moderne engelsk sprogbrug, kun at være et skridt på vejen mod en egentlig ny udgave.

soulful ('səʊlful), *a.* Also soul-full. [f. SOUL *sb.* + **Fig. 1.**]

1. Full of soul or feeling; of a highly emotional, spiritual, or ~~esthetic~~ nature or character, or marked by this, ~~(also)~~ affectedly or unduly ~~esthetic or emotional~~.

1860 PRINCESS Victoria *Lit.* 6 Feb. in R. Fulford *Dearest Child* (1964) 23: The dear Princess said to Fritz Karl while they were singing something of Mozart 'You are looking soulful.' 1867 GILCHRIST *Blake* xxii, Very striking and soulful is the general effect (of an engraving). 1882 L. C. LILLIE *Prudence* 48 Who can be soulful and an athlete? 1882 B. HART *Flip* iii, The poetic, soulful side of his mission was delicately indicated by a pale blue necktie. 1897 *Naturalist* 84 For all who live by it will be manful, soulful, honest, and without fear. 1931 A. HUXLEY *Music at Night* 18 When the great obvious truth is affirmed... in a series of soulful close-ups... the sensitive can only wince. 1951 *Sunday Pictorial* 21 Jan. 6/4 His eyes become deceptively soulful.

trans. 1869 W. CORY *Litt. & Journl.* (1877) 263 This bit of the earth's expression, this soul-full bit of the earth, the quiet bay.

2. Expressive or indicative of deep feeling or emotion.

1868 TUCKERMAN *Collector* 312 The soulful glow of expression in the inspired countenance of the Apollo. 1869 Mrs. WHITNEY *W. Girls* xii, Mother and Madam Pennington looked at each other with soulful eyes. 1878 *The Choir* 16 Nov. 730 One of those soulful tunes which cannot fail to engage attention.

3. Expressive of Black feeling; characteristic of Black music. Also as quasi-*adv.*

1964 *Amer. Folk Music Occasional* i. 46, I sing my song more soulful. 1973 *Black Panther* 15 Sept. 15/1 The audience is encouraged to join hands with their neighbors and all standing away with the soulful music. 1973 *Black World* Apr. 9/1 'Soul' is a highly valued concept among Afro-Americans. Soulful behavior may be called something else by some Afro-Americans, but its value remains.

Hence 'soulfully *adv.*', 'soulfulness.

1880 W. S. GILBERT *Patience* i. 11, I am soulfully intense. 1882 *Advance* (Chicago) 30 Apr. 249 She... carried the audience by her dignity, earnestness and soulfulness. 1893 *Nation* 9 Feb. L.VI. 110/2 Languages vary, as do individuals, in genius or soulfulness. 1923 JOYER *Olympe* 379 Bronze, listening by the beer-pull, gazed far away Soulfully. 1979 *Daily Tel.* 23 Apr. 15/2 Her two daughters, ... one bitter and promiscuous, one soulfully virginal, appeared as remote from everyday living as folklore princesses.

Figur 3. OED + Supplement. (Efter Osselson 1993.)

7 ODS + -S?

De ODS-brugere der har nævnt det ønskelige i en sammenfletning af hovedværk og supplement, har naturligvis især tænkt på at slippe for altid at have flere bind fremme ad gangen. ODS' supplementsartikler er endda udformet sådan at en sammenfletning med de gamle artikler ser ud som en nærliggende mulighed (se figur 4).

Artiklerne indeholder et eller flere plusser. Læseren møder et plus når en tilføjelse sætter ind. Når et plus står på højt niveau i artiklen rækker det langt, står det på lavere niveau rækker det kortere. Et plus foran opslagsordet gælder naturligvis en hel artikel. Plus foran et betydningsnummer markerer at denne betydning, inklusive eventuelle underordnede citatgrupper, ikke findes i ODS. Hvis der derimod skal tilføjes citatstof til en betydning der findes i ODS, så føres læseren hen til betydningsnummeret, før plusset kommer. En ODS-spalte skulle tilsyneladende nemt kunne klippes i stykker og få tilføjet de oplysninger som ODS-S har at byde på.

Processen forudsætter naturligvis at alle artiklerne er gjort maskinlæsbare (se Norling-Christensen 1988), og at det nødvendige værktøj er udviklet.

-dygtig,
adj. især [II.å.1]: (*dværgfrugttræer*) er bæredygtige i en meget ung Alder. *Hjemmet*. 1907.231.

-kraftig, adj.
[II.4.1] (*især* Ω , l. br.) som er i stand til at taale vægten af noget ell. holde noget oppe. de bærekraftige Skuldre. *JV Jens*. SS.65. || overf.: Denne Grundvold for den „vigtige Theori“ er kun i ringe Grad bærekraftig. *FrNiels.R.III.44*. mere bærekraftige Ideer, der kan tage den unge . . . Slægt i sin Tjeneste. *Tidsk.* 1914. II.323.

-dygtig, adj. + (*if* -kraftig). + || (*til* II. bære 4.1). Naar Betonen er blevet tilstrækkelig bæredygtig, fjernes Forskallingen. *TeknVidunderland*. (1928). 570. der dannes en meget elastisk og bæredygtig Flade. *Textilhog-n.* (1944-46). 730. overf., som er holdbar, pålidelig, tilstrækkelig stærk: Hvor .. kunstneriske Evner mangler eller ikke er tilstrækkelig bæredygtige. *Freud. Det Ubevidste*. (overs. 1920). 58. en sund og bæredygtig Balance mellem Prisniveau og Lønniveau. *Socialdem.* 1/3 1933. 3.sp. 2. et bæredygtigt Kompromis (om grundloven). *RaunkjærsKonv.* XII.1311.

-kraftig, adj. + (*rednings*) Baadene var sikre og bærekraftige. *Myl Erich*. S.122. || 236,33 overf. + (*if* -dygtig S). et Læsedrama .. med store rige Roller, der dog ved Opførelse har svært ved at virke bærekraftige. *EFraderiksen*. UL.239. nogen bred, bærekraftig Uvilje mod Tilstandene sporedes ikke. *AndNx*. FG.70.

Figur 4. ODS og ODS-S.

En sammenfletning kan tænkes foretaget efter forskellige principper. Ved Oxfordordbogens 2. udgave kan man ikke se hvad der kommer fra hovedværket og hvad der kommer fra supplementet, og dette viser sig som nævnt at medføre kritisable forhold. Hvis man nu tydeligt markerer hvad der kommer fra supplementet, fx ved at bevare alle plusserne og trykke med en helt anden sats, så er det vel, bortset fra småproblemer undervejs, kun økonomiske forhold der hindrer projektet? For mig at se ville en sådan ODS-udgave nu i alle tilfælde blive et misfoster.

Når en ca. 1920-ordbog suppleres med noget der først kommer ind i sproget senere, så bliver en sådan integreret udgave naturligvis ikke en 1920-ordbog. Men det bliver heller ikke en ordbog over sproget på et senere tidspunkt, for 1920-præget er jo ikke fjernet, idet hovedværkets tekst stadig står der og skal benyttes. Fortalerne for en sammenfletning kan så sige: Vi er naturligvis ikke interesseret i en ny 1920-ordbog, eller en 1955-ordbog, men vi ville gerne have en ordbog til de 250 år som er mere bekvem at slå op i end ODS med ODS-S. Ja, men en sådan historisk periodeordbog er jo endnu ikke redigeret. Det vi kunne få, blev den gamle 1920-ordbog, 1930-ordbog osv. suppleret med noget stof fra hele perioden 1700–1955.

Set fra ODS-S-redaktionens ståsted falder ODS-perioden i tre afsnit:

- tiden før ODS-nutiden, fra 1700 til ca. 1900, såkaldt „ældre sprog“
- ODS-nutiden, fra ca. 1900 til artiklernes trykår
- tiden efter ODS-nutiden, fra artiklernes trykår til 1955(ff.)

ODS-nutiden er omdrejningsaksen, fordi den trods de supplerende oplysninger stort set bliver stående som ordbogsværkets nutid. Ordenes senere skæbne er der kun af og til givet oplysninger om (se indledningen til ODS-S I s. XLII og XLIII). Og der er jo ikke en ny nutid

som alt skal relateres til i stedet for den gamle nutid. Supplementets redaktionssprog gør ikke brug af det lille ord „nu“, der i så høj grad præger ODS, for vi står uden for den periode vi behandler. Tidsangivelsen „det 20. århundrede“ er derimod anvendelig; hvis en nærmere angivelse er påkrævet, kan det fx hedde „beg. af 20. årh.“, „i 1930'erne“, „efter 1945“ osv. Sprogbrug der kun tilhører tiden før ODS-nutiden kaldes „ældre spr.“ eller „ældre kem.“, „ældre mil.“ og lignende.

Med hensyn til sammenfletning ville der næppe være de store problemer med det stof der hører til den ældste del af ODS-perioden. Der er her tale om tilfældige mangler i hovedværket. Det samme gælder til dels for ODS-nutiden. Men når vi kommer til tiden efter ODS-nutiden bliver der tilføjet noget som ligger uden for hovedværkets univers, og et sådant misforhold mellem de to dele gør helheden problematisk. Det der står i vejen for en sammenfletning af ODS og ODS-S er altså samtidsordbogen ODS, beskrivelsen af hvordan sproget er „nu“. Og denne gamle nutid gennemtrænger ODS i en sådan grad at det næppe er muligt ved en overkommelig redigering i forbindelse med fletteprocessen at løfte artiklerne ud af den oprindelige tidsbundethed og gøre værket til en periodeordbog.

8 Samtidsordbog og periodeordbog

Som nævnt har den nye Oxford-ordbog problemer med at holde styr på de forskellige nutider der er involveret. Når det ikke går helt galt, skyldes det, så vidt jeg kan se, at den engelske ordbog aldrig har lagt så stor vægt på at sætte alt i forhold til samtidssproget, som ODS har. Den redaktionelle ramme omkring citatstoffet er enklere i Oxford-ordbogen. Der er en vis angivelse af ordenes status i samtidssproget; de kan bl.a. betegnes som døde eller arkaiske eller dialektale eller som kometord, og de kan henføres til et fagområde, men i sammenligning med ODS eller *Svenska Akademiens Ordbok* er bedømmelsen af ordene i deres forskellige anvendelser meget lidt nuanceret (jf. Hast 1993 og se for eksempel supplementsartiklen *automobile*, hvor læseren ikke gøres opmærksom på at ordet nu står svagt i sprogborgen, især i britisk engelsk).

Det historiske element i de sambeskrivende ordbøger har ikke samme funktion i dem alle. I Oxford-ordbogen har de historiske principper bevirket at redaktøren ligesom løfter sig op over sin egen tid, han ser den i næsten samme perspektiv som de foregående århundreder. I ODS er samtidsperspektivet derimod det dominerende; ODS kan virke som en samtidsordbog hvortil der er føjet noget historisk stof.

En ordbog vil naturligvis altid være præget af sin redaktionstid. Dette præg kan et supplement med god samvittighed lade stå, for brugeren vil jo vide at hovedværket er gammelt og kan indeholde oplysninger der ikke mere er helt i orden. Men i en sammenkobling der ser ud som en ny udgave forventer man at den nye redaktion står inde for hver linje. Tilfældigt så jeg at *Berlin* i Oxford-ordbogens 2. udg. 1989 stod præsenteret som hovedstaden i Preussen! Det er ganske vist en encyklopædisk detalje, men karakteristisk for en nyudgave som ikke er en sand ny udgave.

En ny udgave af ODS, forstået som en ny nutidsordbog med inddragelse af de nærmest foregående århundreder, har aldrig seriøst været på tale. Men der vil forhåbentlig komme nye projekter vedrørende periodens ordforråd. ODS bør under alle omstændigheder gøres maskinlæsbar og tilgængelig for hurtige søgninger i både hovedværk og supplement.

Supplementer er altid nødløsninger, eventuelt forarbejder til senere nye udgaver. Et supplement kan være den eneste mulighed for at få noget værdifuldt stof fremlagt. Brugere kan jamre over besværet, men de må trods alt være glade for at de får nogle yderligere oplysninger. Feilberg beder så smukt i sit slutord brugere om at tage imod hvad der er givet „og tage tiltakke dermed, og ikke klage for meget over hvad der mangler.“

Litteratur

- Bjerrum, Marie 1976: Otto Kalkar og hans ordbog. I: Otto Kalkar: *Ordbog til det ældre danske Sprog*. Fotografisk optryk. VI, 5–167.
- Den Danske Ordbog*. I–VI. Under udarbejdelse ved Det Danske Sprog- og Litteraturselskab. Kbh.
- Duekilde, Anne 1994: Om datering af ordene i dansk. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II*. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993. Skrifter udg. af Nordisk Forening for Leksikografi. Skrift nr. 2. Kbh., 53–60.
- Feilberg, H. F. 1886–1914: *Bidrag til en Ordbog over jyske Almuesmål*. I–IV. Fotografisk optryk 1977. *Gammeldansk Ordbog*. Under udarbejdelse ved Det Danske Sprog- og Litteraturselskab. Kbh.
- Hast, Sture 1993: SAOB och dess likar. I: *Ord och lexicon*. Festskrift till Hans Jonsson. Stockholm: Norstedts, 55–84.
- Jonsson, Hans 1992: De stora historiska ordböckernas särart. I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi*. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991. Skrifter udg. af Nordisk Forening for Leksikografi. Skrift nr. 1. Oslo, 231–34.
- Jysk Ordbog*. Under udarbejdelse ved Institut for Jysk Sprog- og Kulturforskning, Aarhus Universitet.
- Kalkar, Otto 1881–1918: *Ordbog til det ældre danske Sprog*. I–V. Fotografisk optryk med ændringer. I–VI. 1976.
- Norling-Christensen, Ole 1988: Optical Character Reading of Complex Dictionary Texts. I: Karl Hyldgaard-Jensen/Arne Zettersten (eds.): *Proceedings of the Fourth International Symposium on Lexicography* April 20–22, 1988 at the University of Copenhagen. *Lexicographica*. Series Maior 26. Tübingen: Niemeyer, 219–32.
- ODS = *Ordbog over det danske Sprog*. I–XXVIII. 1918–56. Fotografiske optryk 1966–70, 1975–77, 1981–83, 1993–96.
- ODS-S = *Ordbog over det danske Sprog. Supplement*. I–V. 1992ff.
- OED = *The Oxford English Dictionary*. I–XII. 1888–1928. Supplement 1933. Supplement I–IV. 1972–86. Second Edition I–XX. 1989. CD-Rom 1989.
- Osselton, Noel 1993: The Oxford English Dictionary, Second Edition. I: *International Journal of Lexicography*. Vol. 6. nr. 2. 1993, 124–31.

Tekstbaseret lemmaselektion

Based on a research project at The Aarhus School of Business, which includes a text corpus of about 1 800 000 words established in preparation for lexicographical processing, problems regarding the use of a text corpus for lemma selection are discussed. Among other things, the priority of a text corpus compared to that of other sources in this phase is discussed, and an efficient search method used as an alternative to traditional selection methods is suggested.

1 Oprettelse af tekstkorpus

1.1 Baggrund

I 1991 og 1992 blev der på Handelshøjskolen i Århus gennemført undersøgelser, der fokuserede på brugen af ordbøger. Formålet med disse undersøgelser var bl.a. at kortlægge de problemer, som opstår under oversættelse af fagtekster samt at få viden om, hvordan oversætterne løser problemerne, og om hvilke hjælpemidler de benytter sig af.

Resultaterne fra analysen af brugerundersøgelserne, som bl.a. er beskrevet i Duvå/Laursen/Maidahl (1991), blev bl.a. brugt som udgangspunkt for udarbejdelse af forslag til en forbedring af den eksisterende tosproglige fagleksikografi. Forslagene tog hensyn til de behov, som undersøgelserne viste, at oversætterne havde.

For at imødekomme disse behov skulle en ideel fagordbog bl.a. indeholde definitioner, kollokationer og eksempler. Derudover skulle den kunne formidle fagindsigt til brugerne, da undersøgelserne viste, at oversætternes største problem var manglende faglig viden, og vi hævder, at uden denne viden og indsigt i det pågældende fagområde, kan der ikke produceres oversættelser af tilfredsstillende kvalitet. I de nævnte undersøgelser kunne oversætterne ikke hente hjælp til erhvervelse af denne viden i de eksisterende ordbøger, idet ordbøgerne ikke indeholdt informationer om sagen eller de såkaldte encyklopædiske oplysninger.

Vi besluttede på denne baggrund at undersøge, om et tekstkorpus kunne være et nyttigt redskab for den tosprogede fagleksikografi, dvs. om et sådant korpus via bearbejdning kunne danne basis for generering af en specialordbog inden for regnskabsvæsen, som kunne formidle brugerne den fagindsigt, som de ifølge vore brugerundersøgelser manglede. Vi vil i det følgende diskutere korpus egnethed i forhold til andre typer materiale i forbindelse med især lemmaselektion.

Vi mener, at det er af stor betydning for brugeren, at fagordbøger på det mikrostrukturelle plan indeholder ovennævnte oplysninger, men vi mener derudover, at opbygningen af ordbogens makrostruktur spiller en væsentlig rolle i denne forbindelse. Således argumenterer vi i det følgende for, at en fagordbog skal være tematisk eller onomasiologisk opbygget, da vi mener, at en systematisk makrostruktur umiddelbart vil kunne bidrage til at formidle en del af den faglige indsigt, som en fagoversætter har behov for (cf. Christiansen/Duvå/Laursen 1994). Vi går således et skridt videre end Pedersen (1995), der argumenterer for, at fagets

systematik ikke skal afspejles i makrostrukturen, men derimod præsenteres i ordbogens omtekst i form af faglig indledning eller lignende.

Det traditionelle ordbogskoncept med en semasiologisk opbygning bidrager ikke umiddelbart til at formidle faglig viden, idet en alfabetisk ordliste ikke giver brugeren information om de begrebsmæssige relationer mellem termene. Manglen ved en sådan alfabetisk liste skulle være evident. Her følger lemmata i fuldstændig vilkårlig rækkefølge, dvs. de forskellige faglige deluniverser blandes totalt sammen, og i bedste fald forsøger leksikografen at afhjælpe denne mangel ved en eller anden form for henvisningssystem til faglige informationer.

Eksempel på alfabetisk liste (Fryd 1990):

[...]
 entreprenørvirksomhed
 entreprise
equity metode = indre værdis metode
 erfare
 erfaringens vej, ad
 erholde

Et eksempel skal illustrere fordelene i denne forbindelse ved en systematisk opbygning frem for en alfabetisk (eksemplet er konstrueret):

- 1 konsolideringsformer
- 1.1 linie for linie konsolidering
- 1.1.1 (fuldstændig) konsolidering
- 1.1.2 pro rata konsolidering
- 1.2 konsolidering på én linie
- 1.2.1 **equity-metode**
- 1.2.1.1 eliminering af interne fortjenester og tab

Fordelen ved den systematiske opbygning skulle være klar: en bruger, der skal oversætte fx *equity-metoden*, får umiddelbart begrebet præsenteret i sin begrebsmæssige sammenhæng, dvs. vedkommende får alene ved at se på lemmaudsnittet indsigt i, hvilket begrebsmæssigt univers equitymetoden hører hjemme i. Er vedkommende regnskabsmæssig ukyndig får han/hun her straks en viden om, at det drejer sig om en konsolideringsmetode og kan ved at bevæge sig op eller ned i den systematiske lemmaliste supplere sin faglige viden alt efter behov. Den systematiske ordning forudsætter naturligvis en begrebsmæssigt udtømmende lemmabestand.

Vi vil i det følgende se på, hvilke typer materiale der kan komme i betragtning til generering af en sådan **systematisk lemmaliste** og senere se på, hvilke kriterier og metoder der kan anvendes i den forbindelse.

1.2 Ordbøgers materiale

Den empiriske basis kan bestå af følgende elementer cf. Bergenholtz/Tarp (1995), herefter betegnet Manualen, hvis opdeling vi — med visse ændringer — i store træk følger.

1. Egen fagsproglige og faglige kompetence eller viden om det pågældende område (i manualen kaldet introspektion — et begreb, som vi mener snarere refererer til metoden, cf. nedenfor)
2. Belægsamlinger
3. Foreliggende beskrivelser, der kan deles i a) ordbøger b) håndbøger, leksika, encyklopædier med flere c) lærebøger og monografier og d) videnskabelige artikler
4. Tekstkorpora

1.2.1 Egen fagsproglige og faglige kompetence

Om det er muligt at selekttere lemmata til en fagsproglig ordbog uden anden basis end leksikografens eget aktive og passive ordforråd inden for et givet fagsprogligt område, må i princippet afhænge af, i hvor høj grad den faglige kompetence er til stede. Hvis det handler om en ekspert på fagområdet, må man gå ud fra, at en så godt som udtømmende bestand af det til faget hørende register er til stede hos den pågældende, men som det anføres i manualen (p. 92), kan modersmålskompetencen være inkomplet. Dette er bl.a. tilfældet, hvis fx undervisning og publikation af artikler foregår på et andet sprog end modersmålet. En dansk ekspert i afsætningsøkonomi kan fx i sit aktive ordforråd rumme termen *skimming price*, uden at kunne gøre rede for, hvad begrebet hedder på dansk.

1.2.2 Belægsamlinger

Belægsamlinger er den klassiske basis for fagsprogsordbøger. Det drejer sig om eksemplarsamlinger, som fx oversættere — typisk translatører — via deres daglige arbejde har nedfældet på kartotekskort eller i en database. De udgør en slags erfaringsbank, og da de afspejler helt konkrete erfaringer fra de daglige arbejdsopgaver, er de særdeles relevante. På den anden side vil de som følge af en fagoversætters mangeartede oversættelsesopgaver som regel være tilfældigt udvalgte og vil derfor sjældent rumme et fuldstændigt register af termer inden for et konkret område. Og hvis en leksikograf, der samtidig er oversætter, oven i købet vælger ukritisk at publicere en komplet belægsamling af ovennævnte blandede art, bliver resultatet tilfældigt, metodisk kritisabelt, og undertiden ganske morsomt.

Vi har hentet nogle eksempler på en sådan uheldig lemmaselektion fra en dansk-spansk ordbog (Galindo 1989) der angiveligt skulle dække området *erhverv og økonomi*:

alfonseri	postbedrageri
bordel	svangerskabsforebyggelse
diamantbryllup	rufferi
posebæreevne	vielse
positionslys	

Sådanne belægsamlinger vil ofte afspejle leksikografens eget univers snarere end det univers, ordbogen angiver at dække.

1.2.3 Foreliggende beskrivelser

Under denne overskrift inkluderer manualen et vidt spektrum af fagtekster, herunder ordbøger, håndbøger, leksika samt lærebøger og relevante artikler.

At anvende ordbøger som eneste empiriske basis ville være kontradiktorisk, idet motivationen for at gå i gang med en ordbog ofte er manglende eller mangelfulde ordbøger på det pågældende område. Evt. kan eksisterende ordbøger fungere som inspirationskilde eller som *reminders*, det vil fx sige som kontrolmateriale for en lemmaselektion.

Anvendelse af indeks eller stikordsregistre til håndbøger og lærebøger som basis for en lemmaselektion og strukturering heraf kan virke relevant. Vi har imidlertid foretaget adskillige tjek inden for vores fagområde, der viser, at sådanne indeks er ufuldstændige og i alt fald kun dækker fagområdets øverste begrebsmæssige niveau. Hvis sådanne indeks anvendes, er det derfor nødvendigt at gå videre fra opslagsordet til selve teksten for at supplere bredden og dybden af lemmata. Den egentlige basis for selektionen vil derfor være **teksten** som sådan med indekset som det strukturerende udgangspunkt.

1.2.4 Tekstkorpus

Ved tekstkorpus forstår vi en samling tekster, der er produceret i en kommunikativ sammenhæng (cf. Bergenholtz/Pedersen 1994), og som desuden er specielt sammensat. Som sådan adskiller det sig fra belægsamlinger samt fra ordlister og ordbøger. Et sådant korpus bestående af en velovervejede samling af fagtekster mener vi, er det bedste udgangspunkt for arbejdet med fagsprogsordbøger.

Hvis man vælger tekster af forskellige forfattere og med forskellige funktioner (lovtekster, lærebøger, håndbøger mv.), der begrebsmæssigt dækker det pågældende fagområde, vil man være nogenlunde sikker på, at de fleste centrale termer i fagsystematikken bliver repræsenteret. Det skal tilføjes, at det ikke, som det er praksis for korpora, der er beregnet på syntaktiske analyser, er hensigtsmæssigt at tage stikprøver ud fra en større samling tekster. Eftersom det handler om et velafgrænset fagområde, der begrebsmæssigt skal dækkes helt, vil det rigtige være at vælge en række hele tekster, der repræsenterer termsystematikken i det pågældende fagområde.

I det tidligere nævnte ordbogsprojekt vedrørende regnskabsterminologi vil der således ikke være nogen ide i at tage et afsnit om balancen fra én håndbog og så vælge fx afsnit om noter, værdiansættelse og afskrivninger fra tre andre håndbøger. Det drejer sig derimod om at vælge et par anerkendte håndbøger, der formodes at have terminologisk gennemslagskraft, dvs. som fx bliver anvendt i uddannelsen af regnskabseksperter.

Ved sammensætning af sådanne korpora, dvs. ved bevidst at vælge nogle enkelte centrale og anerkendte værker inden for tekster med forskellig funktion, kan man ikke tale om, at korpus er repræsentativt i traditionel forstand, idet dette forudsætter stikprøveudtagning fra en kendt grundmængde (Manualen s. 95). Valg af typiske tekster, der „efter bedste skøn“ dækker det pågældende fagområde kaldes i Manualen for *eksemplarisk*, og vi kan tilslutte os denne skelnen.

Med hensyn til omfang af et sådant tekstkorpus nævner manualen, at hvor det til almensproglige ordbøger vil være nødvendigt med et korpus på 50 millioner tekstord, er det vanskeligere at afgrænse dette omfang for fagsprogskorpora. Der nævnes et konkret eksempel i Manualen på, at et korpus på 500 000 ord har været dækkende til en fagordbog.

1.3 Beskrivelse af *Regnskabskorpus*

1.3.1 Emneområde

Emneområdet for vores tekstkorpus er som nævnt regnskabsvæsen, i første omgang med hovedvægt på eksternt regnskabsvæsen.

Da der er tale om en kontrastiv bearbejdning (da-fr, da-sp) af tekstmaterialet, er der etableret tre delkorpora.

1.3.2 Sammensætning af korpus

Korpus er sammensat af følgende teksttyper, som er de centrale for området:

- love
- kommenterede lovtekster
- regnskabsstandarder
- regnskabsvejledninger
- autentiske regnskaber
- lærebøger
- artikler

Udarbejdelse af bibliografi og sammensætning af korpus er sket i samarbejde med fageksperter.

1.3.3 Omfang

Det danske delkorpus er på ca. 800 000 løbende ord, hvilket stort set svarer til lignende korpora. Om korpus i denne størrelsesorden er tilstrækkeligt til at afspejle områdets lemmata må bero på et skøn foretaget af leksikografen og fageksperter i fællesskab under hensyntagen til det givne formål.

I øvrigt mener vi ikke, at det er muligt at give præcise kriterier for, hvor stort et omfang et korpus bør have, idet der er mange faktorer, der spiller ind såsom formål, omfang af det faglige univers, man vil undersøge, ressourcer, lagerkapacitet etc. Men efter vort skøn er et korpus på omkring 1 mio. løbende ord et rimeligt udgangspunkt for et eksemplarisk fagkorpus.

Der er af komparative hensyn tilstræbt symmetri mellem delkorporaerne L1 og L2 på det tematiske og teksttypologiske niveau.

2 Fra korpus til lemmasektion

2.1 Kriterier for lemmasektion

Der kan opstilles forskellige kriterier for opstilling af en lemmaliste til en fagordbog. Vi mener, at følgende to hovedkriterier er relevante i forbindelse med generering af en systematisk specialordbog:

1. alene fagord, men derimod *alle* fagord uanset frekvens, der indgår i fagsystematikken, medtages, eller
2. fagord som beskrevet i 1) samt andre ord udvalgt efter andre kriterier, herunder verber, adjektiver, adverbier og orddele (affikser, ordstammer mv.). Andre kriterier kan fx være frekvens eller *pædagogiske* erfaringer for, hvad der almindeligvis volder problemer.

Vi mener ikke, at kun fagord/termer skal medtages, men at det af hensyn til brugers behov er væsentligt, at også andre ord, der er specielt relevante for det pågældende fagområde, medtages som lemmata. Men vi understreger, at de skal være specielt relevante. Vi mener således ikke, at almindelige økonomiske fællesord (som fx *pris, beregning, beregne, kontanter, købe, sælge, låne, lån*) skal medtages som lemmata, men kun de ord, der har en specifik betydning eller en specifik funktion inden for området regnskab, fx verber som *udgiftsføre, afskrive, konsolidere, værdiansætte, ansætte, optage til* [. . .].

I vores tematisk/onomasiologiske koncept vil vi i første fase selekttere termer som lemmata, og det betyder, at flerordstermer selekteres og lemmatiseres på lige fod med enkeltordstermer. Andre relevante ord vil vi senere lade indgå i ordbogen enten som fx kollokationer eller i særlige systematiske registre (fx verber systematiseret efter proces: *aktivere, værdiansætte, værdiregulere (nedskrive, opskrive, afskrive* [. . .]).

2.2 Metoder til lemmasektion

Ved selektion af lemmata til en fagsproglig ordbog, forekommer der *a priori* tre metoder:

1. Introspektion (her forstået som brug af den kompetence, leksikografen selv er i besiddelse af)
2. Multispektion (her forstået som et samarbejde mellem en eller flere fagsprogskyndige og en eller flere fagkyndige)
3. Automatisk sortering

Vi forudsætter, at den valgte empiriske basis er et til formålet sammensat tekstkorpus, idet vi som nævnt mener, at et sådant er det bedste grundlag for generering af en fagordbog. Den leksikografiske bearbejdning af tekstkorpus frembyder to åbenbare muligheder.

Man kan excerperere ved en manuel gennemgang af materialet, idet man ved en form for introspektion el. evt. ved multispektion afgør, hvilke lemmata der skal medtages. Denne proces er umådeligt tidskrævende.

Den anden mulighed er, at man indscanner teksterne og udskriver en alfabetisk ordliste, som leksikografen kan sortere ud fra. Denne fremgangsmåde har den ulempe, at listen kun indeholder enkeltord, således at man ved potentielle flerordstermer skal ind i teksten igen for at kontrollere ordets omgivelser. Dette kræver gode associerende evner hos leksikografen, hvilket følgende eksempel fra vort regnskabskorpus kan illustrere:

afskrive	afskrivninger
afskrivelige	afskrivningsberettigede
afskriveligt	afskrivningsfond
afskriver	afskrivningsforløb
afskrives	[...]
afskrivning	afskrivningsmetode
[...]	[...]
	afskrivningsperiode
	[...]
	afskrivningsplan

Blandt de anførte ord i listen er der flere, som vi på forhånd ved indgår i flerordstermer. *Afskrivning* fx indgår i *lineær afskrivning* og i *skattemæssig afskrivning*, *afskrivelig* indgår fx i *afskrivelig omkostning*. Men en sådan introspektion vil næppe nogensinde afsløre samtlige flerordstermini omkring en hovedterm og ej heller afdække alle de listede ord, der er kandidater til en flerordsterm. I alle tilfælde vil behovet for multispektion til bedømmelse af termsystematik og synonymi være til stede.

En fuldstændig automatisk sortering kan finde sted, hvis man sammenholder den alfabetiske lemmaliste med et korpus fra det almensproglige område/et andet fagsprogligt område. Vi foretog et sådant eksperiment i starten af vores projekt, hvor vi sammenlignede de danske regnskabsstandarder med et eksisterende korpus af almensproglige tekster. Man kan på denne måde få en koncentreret liste af fagtermer, men maskinen sorterer på den anden side mange termer fra, der indgår i både det almensproglige og det specifikke fagsproglige univers. Der kan undertiden være tale om homonymer, og i andre tilfælde fagord, der er blevet optaget i almensproget. Eksempler på sådanne termer, der blev sorteret fra som alment ordforråd var *aktiv*, *post* og *skat*.

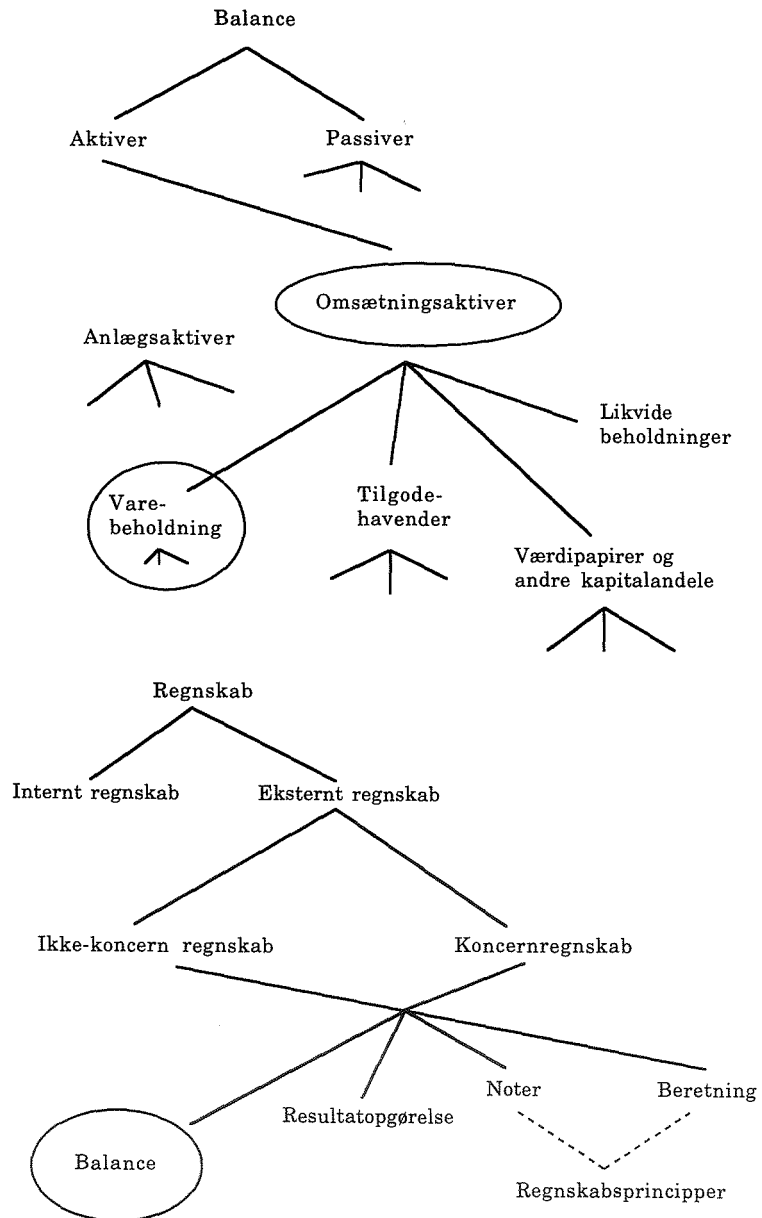
Det siger sig selv, at flerordsproblematikken og behovet for multispektion til opstilling af begrebssystematik og vurdering af synonymymer også er til stede ved en automatisk sortering.

2.3 Udnyttelsen af tekstkorpus i lemmaselektionen

Vi mener på trods af ovennævnte skepsis omkring metoderne til lemmaselektion via et tekstkorpus fortsat, at et specielt sammensat tekstkorpus er det mest systematiske empiriske materiale og dermed bedste grundlag for udarbejdelsen af en systematisk fagordbog, og vi vil derfor give et bud på, hvorledes man også i selektionsfasen kan få en optimal udnyttelse af tekstkorpus. Vi vil give et par eksempler herpå fra vort regnskabskorpus, hvis danske delkorpus er på ca. 800 000 løbende ord.

I stedet for at anvende de to ekstreme muligheder, altså gennemlæsning af hele korpus og automatisk udskrivning af ordlister, har vi forsøgt at lade det tematiske/systematiske princip, som vi vil basere vores ordbog på, være styrende allerede i lemmaselektionsfasen. Vi har a

priori opdelt det univers, vi vil behandle, *det eksterne regnskab*, i en række deluniverser, se figur 1.



Figur 1.

Vi vil dernæst lade det indscannede tekstkorpus vise vej til de begreber, der indgår i de enkelte deluniverser. Vi vil illustrere denne fremgangsmåde ved at tage nogle eksempler fra deluniverset *omsætningsaktiver*. Vi søger ved hjælp af programmet *free-text-browser* på overbegrebet *omsætningsaktiver*. Begrebet indgår i årsregnskabet som en del af balancen, og da den danske årsregnskabslov regulerer opstillingen heraf, dvs. gruppering i romertalsposter (= de overordnede poster) og arabertalsposter (= underposterne) får vi herved foræret en struktur på dette delunivers. En række eksempler fra vort korpus viser, at terminologien her er meget ensartet, idet lovgivningens terminologi går igen i såvel lærebøger som autentiske regnskaber.

OMSÆTNINGSAKTIVER

- I Varebeholdninger
 - 1 Råvarer og hjælpematerialer
 - 2 Varer under fremstilling
 - 3 Fremstillede færdigvarer og handelsvarer
 - 4 Igangværende arbejder for fremmed regning
 - 5 Forudbetaling for varer

OMSÆTNINGSAKTIVER

VAREBEHOLDNINGER

Råvarer og hjælpematerialer

Varer under fremstilling

Færdig- og handelsvarer

Igangværende arbejder for fremmed regning

Acontofakturering af igangværende arbejder

Forudbetalinger for varer

I ALT

Vi sætter os for at undersøge, om terminologien uden for opstillingerne er lige så konsekvent. Vi søger på *varelager*, som vi formoder er synonymt med de i opstillingerne anvendte *varebeholdninger* og finder et eksempel, der understøtter vor formodning:

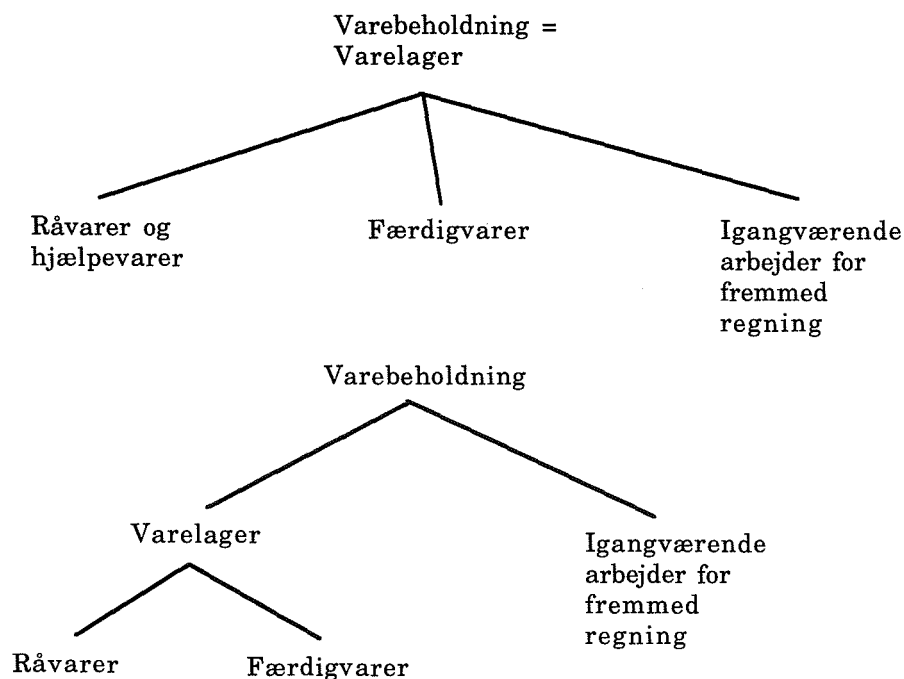
Varebeholdninger

Varelagre er optaget til laveste værdi af anskaffelses- eller genanskaffelsespris med nedskrivning for ukurans og langsom omsætningshastighed. (Noter — Autentisk regnskab)

Andre eksempler peger derimod i retning af, at dette ikke er tilfældet.

Når resultatopgørelsen således er påvirket af f.eks. afskrivningspolitik med hensyn til anlægsaktiver og værdiansættelsesmetoder for *varelager og igangværende arbejder*, vil valget af regnskabsprincipper på disse områder også påvirke balancen. (Lærebog)

hvor *varelager* og *igangværende arbejder* er sideordnede og dermed underbegreber til varebeholdninger. Vi har således to minibegrebssystemer som det vises på figur 2:



Figur 2.

Vi søger dernæst på nogle typiske kollokater til *varebeholdninger/varelager* (nedskrivning/værdiansættelse) og finder, at begge termer forekommer i denne kontekst, og ydermere dukker en tredje term op, som vi formoder er et synonym til varelager, nemlig *lagerbeholdninger*.

Det drejer sig om *nedskrivning på varelager*, henlæggelse til investeringsfond samt forskellen mellem skattemæssige og regnskabsmæssige afskrivninger på driftsmidler mv. (Lærebog)

Eventualskatten er beregnet som 40% af skattemæssige merafskrivninger på materielle anlægsaktiver og skattemæssige nedskrivninger på *varebeholdninger*. (Autentisk regnskab)

Når dækningsbidragstankegangen derfor anvendes ved *værdiansættelse af varebeholdninger* — dvs. uden tillæg af IPO — er der en inkonsekvens ved udarbejden af resultatopgørelsen. (Lærebog)

LIFO er ikke en alment accepteret metode til *værdiansættelse af lagerbeholdninger*, da den hverken viser den sædvanlig praksis eller giver de præcise resultatomkostninger. (Lærebog)

Figur 18.: Diagram over *værdiansættelse af varelager*. (Lærebog)

Vi har således en mistanke om, at der er forvirring omkring begrebsindholdet af disse termer og beslutter at gemme dem til eksperten (multispektion), idet de eksisterende

definitionsordbøger ikke kan klare spørgsmålet for os. Vi mener under alle omstændigheder, at alle tre termer skal medtages i lemmalisten, idet de, uanset frekvens, hører til begrebssystemet.

Vi vil vise endnu et eksempel på, hvordan vi via søgning på en overordnet term har fundet frem til en række termer, der kan bruges til tentativt at opstille en begrebssystematik i relation til det pågældende delunivers. Vi har valgt termen *afskrivning* og har ved søgning herpå fundet følgende flerordstermer:

Hvis der i stedet anvendes mere komplicerede afskrivningsregler, f.eks. *saldoafskrivning* i de første år og herefter *lineær afskrivning*, må redegørelsen være mere detaljeret. (Lærebog)

Eventualskat beregnes med 34 pct. af forskellen mellem den regnskabsmæssige og skattemæssige værdi af varelageret, omsætningsaktiver og hensættelser mv. samt af forskellen mellem de *regnskabsmæssige og skattemæssige afskrivninger* på materielle anlægsaktiver. (Autentisk regnskab)

Den skat, der kan henføres til årets resultat, er endvidere påvirket af varierende skattesatser i de enkelte lande samt af forskellen mellem *skattemæssige og driftsmæssige afskrivninger*. (Autentisk regnskab)

Akkumulerede afskrivninger 1. juni 1990

Kursregulering primo

Årets afskrivninger

Tilbageførte afskrivninger på solgte aktiver

Akk. afskrivninger 31. maj 1991

Bogført værdi 31. maj 1991 (Autentisk regnskab)

I praksis er denne regel kun aktuel for såkaldt *genvundne afskrivninger på bygninger mv.* (Lærebog)

Afskrivninger primo

Afskrivninger

Afgang

Afskrivninger ultimo (Autentisk regnskab)

Ekstraordinære afskrivninger (nedskrivninger)

De forudgående bemærkninger har taget sigte på de *ordinære afskrivninger*, dvs. afskrivninger, som foretages regelmæssigt (systematisk) år efter år. (Lærebog)

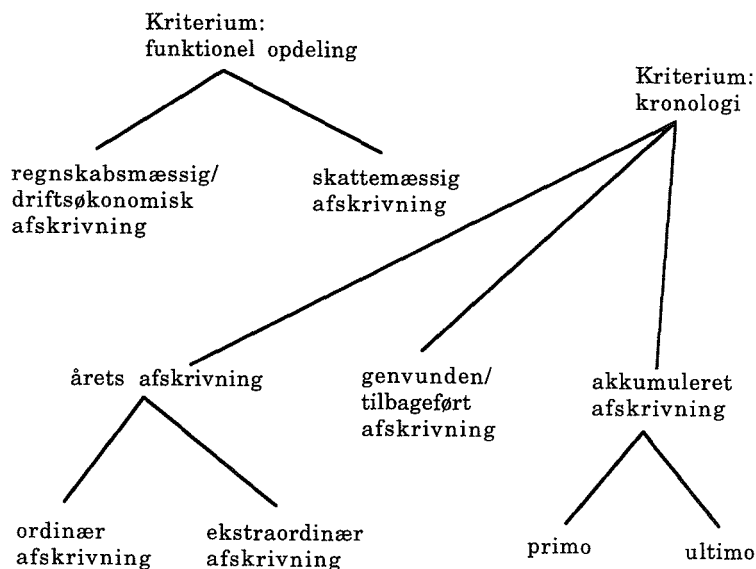
På basis heraf vil vi opstille et tentativt begrebssystem, som vises på figur 3.

Vi vil herefter lade en ekspert se på, om vores systematisering holder, herunder også bekræfte/afkræfte vores formodning om synonimi med hensyn til termerne *driftsmæssig/regnskabsmæssig afskrivning* og *genvundne/tilbageførte afskrivninger*.

Ved søgningen i tekstkorpus støder man i konteksterne til de pågældende termer ofte på de processer, der knytter sig til deluniverset *omsætningsaktiver*. For det første er der spørgsmålet, om selve **aktiveringen** af et beløb, dvs. om et beløb skal opføres som et aktiv (= **aktiveres**) eller ej. I tilfælde af aktivering skal genstanden/beløbet, etc. **opføres** i balancen, **værdiansættes** og beløbet kan senere **værdireguleres**. Sådanne termer tilknyttet

fordi disse er termer med en specifik betydning i forhold til andre universer (aktivering, optagelse, ansættelse), og dels fordi de refererer til generelle processer i forbindelse med de termer, vi fokuserer på. Ydermere har vi erfaring for, at især verberne (*aktivere, optage, ansætte*) volder problemer ved oversættelse, og vi vil derfor optage disse verbalformer i lemmalisten, ikke mindst på grund af deres høje frekvens i forhold til substantivformerne.

AFSKRIVNING



Figur 3. Tentativt begrebssystem

Andre værdipapirer og kapitalandele

Under denne post *opføres* f.eks. børsnoterede obligationer og aktier

Verbalformer: 268

Substantivformer: 0

Råvarer, hjælpematerialer og handelsvarer *optages til* anskaffelsespris eller dagspris, hvor denne er lavere.

For GN Store Nord er den væsentligste virkning af ændringer i årsregnskabsloven *optagelsen af* associerede virksomheder efter indre værdis metode

Verbalformer: 144 forekomster

Substantivformer: 1 forekomst

Råvarer og hjælpematerialer samt handelsvarer er *værdiansat til* indkøbspriser eller dagspriser, hvor disse er lavere.

LIFO er ikke en alment accepteret metode til *værdiansættelse* af lagerbeholdninger

Verbalformer: 130

Substantivformer: 185

3 Konklusion

Konkluderende kan vi sige, at undersøgelsen har vist, at et maskinlæsbart tekstkorpus *kan* være et nyttigt og hensigtsmæssigt redskab i forbindelse med selektion af lemmata til en specialordbog inden for fagsprog.

Korpus kan således *støtte*, dvs. af- og bekræfte leksikografens viden, og det kan *supplere* dennes viden (fx med hensyn til at finde — ukendte — underbegreber og synonymer), og desuden er det et hurtigt og effektivt redskab. Men en hensigtsmæssig anvendelse af korpus forudsætter bl.a. (1) at korpus er eksemplarisk, dvs. begrebsmæssigt dækkende sammensat og (2) at den leksikograf, der anvender korpus, har stor fagsproglig og stor faglig viden samt stor indsigt i terminologiske principper.

Litteratur

Ordbøger

Fryd 1990 = Ejnar Fryd: *Anglo/Amerikansk-Dansk, Dansk-Anglo/Amerikansk specialordbog inden for revision, regnskabsvæsen mv.* Tønder: FSRs Forlag.

Galindo 1989 = Orlando Galindo: *Erhvervs- og økonomiordbog Dansk-Spansk.* København: GTO.

Anden Litteratur

Bergenholtz, Henning/Jette Pedersen 1994: Zusammensetzung von Textkorpora für die Fachlexikographie. I: Burkhard Schaefer/Henning Bergenholtz (Hrsg.): *Fachlexikographie Fachwissen und seine Repräsentation in Wörterbüchern.* Tübingen: Gunter Narr Verlag, 161–176.

Bergenholtz, Henning/Sven Tarp m.fl. 1994: *Manual i fagleksikografi, Udarbejdelse af fagordbøger — problemer og løsningsforslag.* Herning: Systime.

Christiansen, Lisbet Maidahl/Grete Duvå/Anna-Lise Laursen 1994: Das Translationswörterbuch für Fachsprachen: Ein integriertes Konzept I: Burkhard Schaefer/Henning Bergenholtz (Hrsg.): *Fachlexikographie Fachwissen und seine Repräsentation in Wörterbüchern.* Tübingen: Gunter Narr Verlag, 269–282.

Duvå, Grete/Anna-Lise Laursen/Lisbet Maidahl 1992: Brugerundersøgelse vedrørende oversættelse af fagtekst. I: Ruth Vatvedt Fjeld (udg.): *Nordiske studier i leksikografi, rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991.* Oslo: Nordisk forening for leksikografi, 105–133.

Manual = Bergenholtz/Tarp 1994.

Pedersen, Jette 1995: *Koncernspecifik brancheordbog med særligt henblik på teknisk engelsk. Prækonceptionelle overvejelser og koncept.* Århus: Handelshøjskolen i Århus.

Den fagsproglige ordbogsartikels fysiognomi

The purpose of this paper is to discuss various problems arising when working with bilingual lexicography. The discussion is based on a current project: a bilingual French-Danish dictionary in the field of law intended for translators. Before embarking on our main subject (the structure of the dictionary article), we briefly examine the lexicographic structures currently used in the metalexigraphic literature and present the lexicographic structures which we have chosen to operate with. We then describe the macrostructure of our dictionary and the principles used when selecting the lemmata of the dictionary (the mediostucture). We base our examination on concrete examples from the French-Danish dictionary and concentrate on the microstructure, that is the article notes following the lemma and the structure of this information. The essential element of the microstructure being the statement of equivalents, we focus on the information to be provided in the equivalent field. Finally we examine the collocation field which includes both the collocations as such and such material as we consider illustrative of the use of the different lemmata.

1 Indledning

Formålet med dette indlæg er at se nærmere på artikelstrukturen i en fagsproglig ordbog. Artikelstrukturen er et meget centralt problemfelt, fordi det jo er afgørende for hele ordbogsarbejdet, at man sørger for — med udgangspunkt i ordbogens målgruppe og formål — at medtage oplysninger i et sådant omfang og på en sådan måde, at brugeren får mest muligt ud af at konsultere værket.

Indledningsvis skal vi dog se på, hvilken type ordbog der danner grundlag for vore overvejelser. Vi skal placere den i den gængse ordbogstypologi, se på dens målgruppe samt kort behandle dens opbygning i sammenligning med de muligheder, andre leksikografer arbejder med.

Dernæst vil vi kort berøre de problemstillinger, man står overfor i forbindelse med udvælgelsen af lemmabestanden. Og endelig vil vi så som sagt koncentrere os om artikelstrukturen og fremdrage nogle af de mange konkrete problemer, man støder på under arbejdet med struktureringen af fagsproglige ordbogsartikler.

2 Ordbogstype og målgruppe

Den ordbog, der er udgangspunkt for vore overvejelser, og som vi er ved at udarbejde, er en bilingval fransk-dansk oversættelsesordbog. Det drejer sig om en udpræget specialordbog, en juridisk ordbog, med en primær målgruppe bestående af dansksprogede ældre fransk-studerende samt færdiguddannede oversættere af juridiske tekster. Der er således tale om en målgruppe, der ganske vist har generel modersmålskompetence, men de er ikke specialister

inden for jura — hverken på kildesproget eller på målsproget. Jurister kan naturligvis også anvende ordbogen, men deres baggrund og viden har ikke haft indflydelse på valg eller fravalg af oplysningerne i ordbogen. De vil derfor nok ikke have brug for alle de forklaringer, som findes i værket af hensyn til den primære målgruppe.

3 Ordbøgers opbygning

Der er almindelig konsensus om, at ordbøger, herunder i særdeleshed fagordbøger, ikke blot består af selve de leksikografiske enheder, sådan som det ellers fremgår af de traditionelle definitioner af ordbøger, fx *Petit Roberts* definition:

recueil d'unités significantes de la langue (mots, termes, éléments . . .) rangées dans un ordre convenu, qui donne des définitions, des informations sur les signes.

Ud over de leksikografiske enheder, selve ordbogens hjerte, af mange kaldet **ordlisten**, er der eller burde der være udarbejdet et forord, en brugervejledning og eventuelt andre afsnit med vigtige oplysninger vedrørende ordbogens indhold og strukturering.

3.1 Grundbegreber vedrørende ordbøgers strukturering

Inden vi ser på opbygningen af den omhandlede fransk-danske juridiske ordbog, skal vi minde om, at det i den metaleksikografiske litteratur diskuteres, hvordan grundbegreber vedrørende ordbøgers strukturering defineres og anvendes. Det drejer sig blandt andet om begreberne makrostruktur og mikrostruktur.

Termen **makrostruktur** anvendes i mindst tre forskellige betydninger. Således defineres makrostrukturen af fx Hartmann (1983:225) som det samlede antal indgange i selve ordlisten. Andre fx Hausmann (1977:3), Wiegand (1983:432), Svensén (1987:215), Hausmann/Wiegand (1989:328) og Bergenholtz/Tarp (1994:11 og 205) anvender termen om den ordnede mængde af lemmaer, dvs. om de enkelte opslagsords placering i forhold til hinanden — fx alfabetisk opstilling contra systematisk opstilling — og om strukturering i en eller flere ordlister.

Og der er endnu en mulighed. Således foreslår Nielsen (1994:76), at termen makrostruktur benyttes om en ordbogs overordnede opdeling i delkomponenter, dvs. opdeling i ordliste og de andre komponenter med vigtige oplysninger til værket (forord, brugervejledning osv.).

Hvad angår udtrykket **mikrostruktur**, var det franskmanden Josette Rey-Debove (1971:151ff.), der som den første anvendte begrebet i leksikografien. Ifølge Rey-Deboves definition omfatter mikrostrukturen den enkelte leksikografiske enheds informationsstruktur efter indgangsordet. Det er en definition, der stadig benyttes af mange leksikografer, fx Hausmann/Wiegand (1989:328 og 340), Bergenholtz/Tarp (1994:11 og 216) og Nielsen (1994:223–224). Andre leksikografer medregner imidlertid lemma til mikrostrukturen, fx Hausmann (1977:4), Svensén (1987:194 og 202ff.) og Zgusta (1988:IX).

Vi skal ikke komme nærmere ind på denne diskussion her. Det vigtigste er at vide, hvad man taler om. Derfor skal vi kort redegøre for, hvordan vi benytter disse metaleksikografiske termer og baggrunden herfor.

3.2 Det valgte struktureringshierarki

I vores arbejde med fagsproglig oversættelsesleksikografi opererer vi med tre struktureringsniveauer. På det overordnede plan er det som anført relevant at tale om struktureringen af selve ordbogen i forord, brugervejledning, samlingen af leksikografiske enheder, bibliografi mv. Dernæst er det på sin plads at se på, hvilke lemmaer der skal indgå i samlingen af leksikografiske enheder, og hvordan disse lemmaer skal struktureres. Og sidst men ikke mindst skal der fastlægges en hensigtsmæssig udvælgelse og strukturering af de oplysninger, der skal anføres til de enkelte lemmaer.

I lyset af disse tre leksikografiske struktureringsniveaers indbyrdes relationer har vi fundet det logisk at anvende følgende udtryk for de tre struktureringsniveauer:

Makrostruktur	Mediostruktur	Mikrostruktur
Forord	afsnit 1 afsnit 2 afsnit 3 osv.	
Brugervejledning	do.	
Forkortelsesliste	do.	
Faglig indledning	do.	
Samling af leksikografiske enheder	lemma 1 lemma 2 osv.	gram.felt ækvi.felt kollo.felt osv. gram.felt osv.
Bibliografi	område 1 område 2 område 3 osv.	

Figur 1. Ordbogsstrukturer

Vi har i lighed med Nielsen (1994:114) i vores fremstilling valgt udtrykket **makrostruktur** om den overordnede strukturering af de enkelte delkomponenter, da det semantiske indhold af denne term indikerer, at der er tale om det overordnede niveau. Bergenholtz/Tarp (1994:12) anvender udtrykket **rammestruktur** om dette struktureringsniveau, og Hausmann/Wiegand (1989:330) taler blot — ud over ordlisten — om forskellige **functional component parts**, men angiver ingen term for selve den overordnede strukturering.

Om det laveste niveau i hierarkiet, dvs. om oplysningerne til lemma og disse oplysningers strukturering, har vi valgt udtrykket **mikrostruktur**, altså en anvendelse der er i overensstemmelse med den klassiske definition af udtrykket inden for leksikografien.

Om det mellemniveau, som det bliver relevant at arbejde med i vores struktureringshierarki, såfremt det accepteres, at lemma ikke hører til mikrostrukturen og ej heller til

makrostrukturen, benytter vi **mediostruktur**, som foreslået af Nielsen (1994:116). Vi finder denne terminologi semantisk mere logisk end den, der er anvendt i Bergenholtz/Tarp (1994:11–12), hvor struktureringen af lemmaerne som anført benævnes makrostruktur, og hvor ordbogens strukturering i delkomponenter kaldes rammestruktur.

Endelig skal det bemærkes, at udtrykket **artikelstruktur** i lyset af den valgte strukturering benyttes omfattende lemma plus de mikrostrukturelle informationskomponenter.

3.3 Den fransk-danske juridiske ordbogs makrostruktur

Vi skal nu se konkret på opbygningen af vores fransk-danske juridiske ordbog. I de senere år anbefales det helt naturligt, at der gøres mere ud af de andre komponenter i ordbøgers makrostruktur, dvs. forord, brugervejledning, faglig indledning osv. (Svensén 1987:222, Hausmann/Wiegand 1989:330, Bergenholtz/Tarp 1994:173ff., Nielsen 1994:85ff.).

Det er vi enige i. Og vi lægger da også stor vægt på, at alle de beslutninger, der er taget med hensyn til ordbogens strukturering, artiklernes opbygning, henvisningsstruktur osv. klart fremgår og forklares i en grundig brugervejledning, samt at der i forordet gives vigtige informationer om ordbogen. Men vi vil gerne understrege, at det nu engang er selve de leksikografiske enheder og deres opbygning, der for os er det centrale i en ordbog.

Vi har valgt at undlade en delkomponent, som det af nogle (Bergenholtz/Tarp 1994:160f., Nielsen 1994:98ff.) anbefales at medtage som introduktion til fagsproglige ordbøger. Det drejer sig om en egentlig faglig indledning til ordbogen indeholdende encyklopædiske oplysninger om værkets områder, altså i vores tilfælde komparative betragtninger inden for de forskellige retsområder.

Dette fravalg skyldes flere forhold. Først og fremmest behandler ordbogen terminologi inden for temmelig mange retsområder (se nedenfor). Vi mener, at det vil føre for vidt at give en brugbar introduktion til så mange områder. En meget kort introduktion til de forskellige områder har nemlig efter vores opfattelse ikke nogen stor nytteværdi. Vores målgruppe forudsættes at have en vis basal viden om juridiske sammenhænge, ellers kan de slet ikke oversætte juridiske tekster. En introduktion af et rimeligt omfang omfattende komparative beskrivelser af områderne vil optage for megen plads i ordbogen, og der er fare for, at den forældes hurtigt på grund af lovændringer, EU-harmoniseringer osv. En sådan behandling hører således hjemme i en lærebog.

Vores udgangspunkt er endvidere, at en multiområdeordbog som vores bør tilstræbe, at den viden, der er nødvendig for, at brugerne kan benytte de angivne ækvivalenter korrekt, indlægges i de enkelte artikler. Og i de forhåbentlig få tilfælde, hvor der er behov for yderligere oplysninger, må brugeren konsultere de lærebøger o.l., der er angivet i den emneinddelte bibliografi.

3.4 Arbejdsmetode og dokumentationsmateriale

Selve arbejdet med ordbogen, dvs. såvel lemmaselektionen som arbejdet med de enkelte leksikografiske ordbogsartikler, er fortrinsvis foregået ved hjælp af edb.

Lemmaselektionen er foregået ud fra et tekstkorpus, som er afgrænset ud fra de retsområder, vi har valgt at lade indgå i ordbogen (se nedenfor). Korpus består af såvel skrevne tekster som elektronisk læsbart materiale. Det skrevne tekstkorpus er sammensat af de

vigtigste danske og franske love og andre regler, lærebøger, franske og danske etsprogede juridiske opslagsværker samt domme og artikler i danske og franske juridiske tidsskrifter inden for de behandlede områder.

Det elektroniske materiale udgøres af tre maskinlæsbare korpora inden for området, som er indlagt i søgeprogrammet WordCruncher. Det drejer sig dels om et dansk og et fransk maskinlæsbart tekstkorpus inden for aftaleret (Dyrberg/Faber/Hansen/Tournay 1991), dels om en basal fransk lærebog i civilret der anvendes på cand.ling.merc.-studiet i København, og dels om et fransk korpus der går under betegnelsen Lecticiel.

Selve ordbogsbasen findes i en til formålet udarbejdet database i databaseprogrammet DataPerfect.

4 Ordbogens mediostruktur

Inden vi går over til at behandle selve artikelstrukturen, skal vi kort se på lemmaselektionen og struktureringen af lemmaerne, dvs. på mediostrukturen.

4.1 Ordbogens emneområder

Det juridiske univers er meget omfattende, det består af en lang række retsområder, der hver især kunne gøres til genstand for en særskilt ordbog. Vi har derfor været nødt til dels at begrænse antallet af retsområder, der medtages i ordbogen, dels at begrænse ordbogens empiriske basis.

Med udgangspunkt i nedenstående grovmaskede oversigt over det franske retssystems områder har vi fastlagt, hvilke hovedområder der ud fra deres relevans for målgruppen først og fremmest skal behandles i ordbogen.

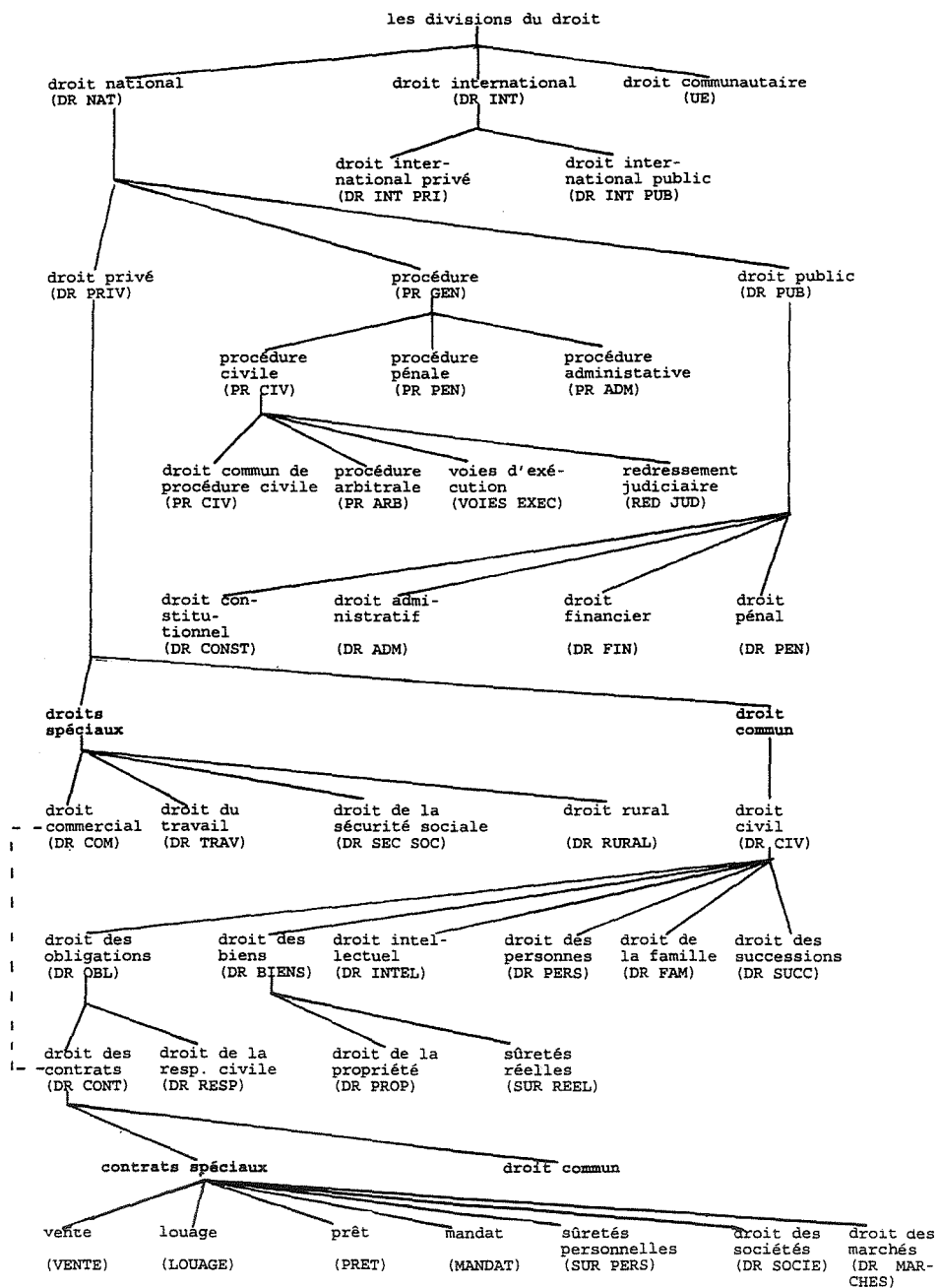
Udvælgelsen af emneområder er foretaget med afsæt i de områder, der undervises i inden for disciplinen juridisk sprog på cand.ling.merc.-studiet på Handelshøjskolerne i Danmark.

Vi skal ikke her give en detaljeret gennemgang af, hvilke områder der er medtaget i ordbogen, men blot understrege, at hovedvægten er lagt på områder, der er basale for erhvervslivet, dvs. privatret (*droit privé*), herunder især obligationsret (*droit des obligations*), og civilproces (*procédure civile*).

4.2 Afgrænsning af termer contra kollokationer

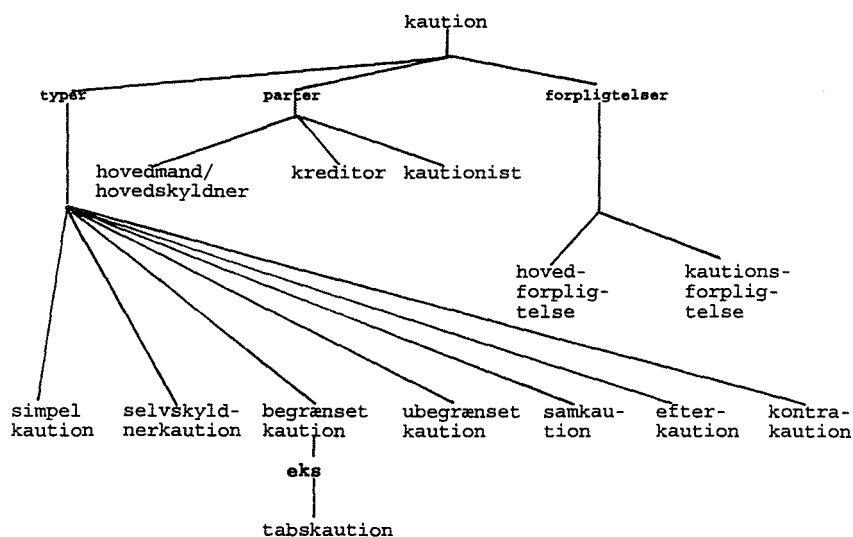
Et problem ved selve lemmaselektionen består i at afgøre, hvilke ord og ordforbindelser der anses for at være termer og derfor skal have en selvstændig ordbogsartikel, og hvilke forbindelser der må anses for at udgøre kollokationer til en term. Vi baserer vores afgrænsning af lemmaer og kollokationer på vores termbegreb.

Vores udgangspunkt for udvælgelsen af ordbogens lemmaer er således, at ord og ordforbindelser, der af eksperterne, dvs. juristerne, anses for at betegne et afgrænset definerbart begreb inden for et givet retsområde, skal medtages som term, uanset om det drejer sig om etleddede eller flerleddede udtryk. Det lyder ligetil, men som så ofte er grænserne flydende, og der er tendens til, at jo mere man ved om et emne, jo flere udtryk anses for at betegne et fast definerbart begreb og ikke blot være en ordforbindelse inden for det omhandlede område.



Figur 2. Oversigt over det franske retssystem.

Som støtte ved udvælgelsen af termbestanden i ordbogen har vi ud over vores erfaring som undervisere inden for områderne i vid udstrækning anvendt begrebsplaner, som vi har opstillet over de enkelte delområder. Et eksempel på et sådant begrebsplan er følgende trædiagrammer over kaution på henholdsvis dansk (figur 3) og fransk (figur 4):



Figur 3. Begrebsplan over kaution.

Ud over termene har vi valgt som opslagsord at medtage elliptiske former af termene, akronymer og udtryk for fænomenet metonymi. Det vil vi komme mere ind på i afdelingen om ordbogsartiklen.

4.3 Strukturering af lemmer

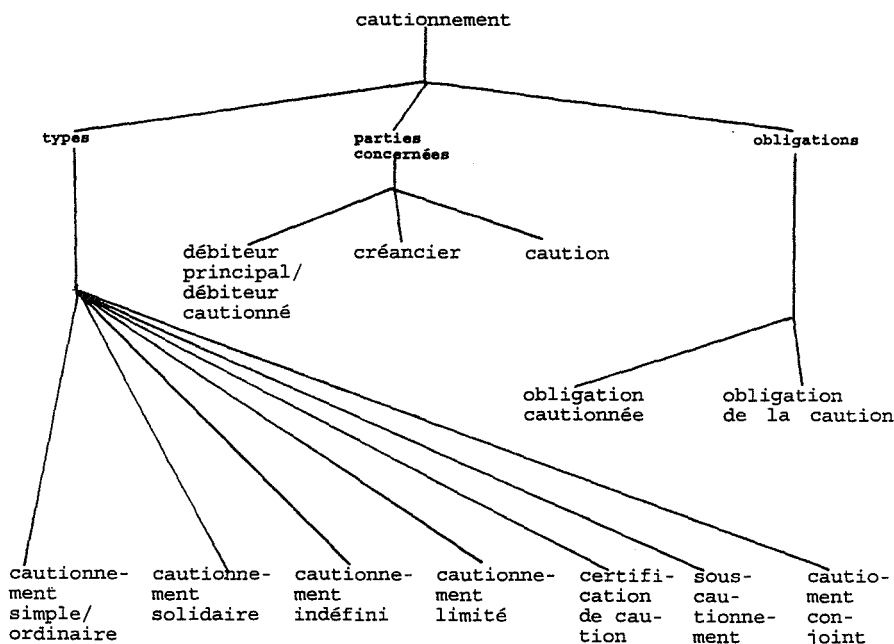
Vi har valgt at gennemføre en alfabetisk opstilling af lemmerne, en alfabetisering der i øvrigt foretages maskinelt ved hjælp af databaseprogrammet DataPerfects alfabetiseringssystem. Princippet for den alfabetiske opstilling er ord-for-ord-metoden (Svensén 1987:215), ifølge hvilken blanktegn ordnes før skrifttegn.

De flerleddede størrelser opføres således i alfabetisk rækkefølge efter første led uanset ordklasse, bortset fra sammensatte præpositioner. Fx opføres udtrykket *en vertu de* på følgende måde: **vertu, en ~ de**.

Dette alfabetiseringsprincip kombineret med princippet om, at flerleddede størrelser opføres som opslagsord, er pladsøkonomisk. Det er således ikke nødvendigt at opføre en flerleddet størrelse bestående af fx substantiv 1 + præposition + substantiv 2 både under substantiv 1 og under substantiv 2, som visse leksikografer gør det (Henriksen 1991). Eftersom et meget stort antal opslagsord i en specialordbog rent faktisk er flerleddede størrelser, er pladsbesparelsen betragtelig.

Samtidig lettes brugerens søgning i ordbogen betydeligt. Når brugeren ved, at ordbogen som opslagsord indeholder flerleddede størrelser, og at disse er ordnet alfabetisk efter det

første led, behøver han ikke at gætte på, under hvilket led leksikografen har valgt at anbringe den flerleddede størrelse.



Figur 4. Begrebsplan over *cautionnement*.

5 Specialordbøgers artikelstruktur

Målet med ordbogen er som nævnt at sætte brugeren i stand til at oversætte en fransk juridisk tekst til en ækvivalerende dansk juridisk tekst uden i særlig vidt omfang at skulle konsultere andre værker. Det betyder, at den enkelte ordbogsartikel — ud over naturligvis at have en præcis, overskuelig og dækkende ækvivalentside — skal indeholde formidling af et fagligt indhold, som giver brugeren den forståelse for begrebet, som er nødvendig for, at oversættelsen kan blive vellykket.

Hvilken indflydelse har så denne præcisering af ordbogens type, formål og primære målgruppe?

5.1 Feltopdeling af ordbogsartiklen

Der er flere forhold, der skal fremhæves. Vi skal indledningsvis kort beskrive ordbogsartiklens interne struktur. Hver artikel består af to hoveddele, nemlig et lemmafelt hvor opslagsordet figurerer, og en ækvivalentside der er underinddelt i et egentligt ækvivalentfelt og i et felt, som vi har valgt at kalde kollokationsfelt.

lemmafelt	gram. felt	ækvivalentfelt	kollokationsfelt
lemma	ordklasse, køn	1. a. retsområde- angivelse	kollokationer og eksempler
		b. ækvivalent c. forklaring til både lemma og ækvivalent	
		2. do.	kollokationer og eksempler
		osv.	

Figur 5. Artikelskema

5.1.1 Lemma

Vi har allerede været lidt inde på de kriterier, vi anvender ved udvælgelsen af lemmer. Vi skal her ikke give en udtømmende beskrivelse af de konstituer, der udfylder lemmafeltet, men vi vil dog ikke undlade at berøre et par fænomener, som er meget fremherskende i fransk fagsprog.

5.1.1.1 Elliptiske former

Det drejer sig for det første om anvendelsen af elliptiske former af flerleddede størrelser. Dette er et særdeles almindeligt fænomen i fransk juridisk sprog. Da vores målgruppes faglige viden ofte ikke er stor nok til, at de kan afgøre, hvordan den fulde term er konstrueret, medtages de elliptiske former som opslagsord med henvisning til det fulde udtryk (se (1) og (2)):

(1)

personnes f.pl.

DR PERS = **droit des personnes**.

droit m. des personnes

DR PERS personret (retsområde).

(2)

acceptation f. bénéficiaire

DR SUCC = **acceptation de succession sous bénéfice d'inventaire**.

acceptation f. de succession sous bénéfice d'inventaire

DR SUCC „overtagelse af [døds]bo uden vedgåelse af gældsansvaret“, „tiltrædelse af arv med beneficium inventarii“.

Ved henvisning fra den afkortede form til det fulde udtryk er henvisningstegnet et lighedstegn (=). Det skal bemærkes, at afkortningsproblematikken bliver yderligere interessant og vanskelig ved, at en elliptisk form i mange tilfælde kan siges at være polysem og altså anvendes om flere forskellige fulde termer, som illustreret i nedenstående eksempler:

(3)

prescription f.

1. DR GEN forskrift, bestemmelse (fx i lov, aftale);
2. DR PROP = **prescription acquisitive**;
3. DR PRIV/PR GEN = **prescription extinctive**.

prescription f. acquisitive

DR PROP hævd (...)

prescription f. extinctive

DR PRIV/PR GEN forældelse (...)

(4)

capacité f.

1. DR GEN kompetence, evne, ~ **de tester** testationskompetence;
2. DR PERS = **capacité de jouissance**;
3. DR PERS = **capacité d'exercice**.

capacité f. de jouissance

DR PERS retsevne (evne til at være berettiget el. forpligtet i et retsforhold).

capacité f. d'exercice

DR PERS handleevne, habilitet (evne til at råde over sin formue og forpligte sig ved retshandler).

Ved disse polyseme elliptiske udtryk skulle forklaringerne ved de forskellige fulde termer kombineret med den konkrete kontekst gerne kunne sætte brugeren i stand til at vælge den rigtige løsning.

5.1.1.2 Akronymer

For at gøre ordbogen så brugervenlig som muligt medtages som lemmaer også akronymer med henvisning til den uforkortede form:

(5)

SNC f.DR SOCIE = **société en nom collectif**.

Ved henvisning fra forkortelsen til den uforkortede form er henvisningstegnet ligeledes et lighedstegn (=).

5.1.1.3 Metonymi

Et tredje særligt område, som vi har forsøgt at tilgodese i så høj grad som muligt, er fænomenet metonymi, fx brug af gadenavne og navne på bygninger til angivelse af ministerier og offentlige institutioner (Dyrberg/Tournay 1990b). Metonymi anvendes meget hyppigt på fransk, men medtages kun i ringe omfang i eksisterende et- og flersprogede ordbøger.

Det kan således være vanskeligt for en udlænding at finde frem til, hvad fx de franske udtryk *Bercy* og *Quai d'Orsay* står for. I disse tilfælde angiver vi — som vist i (6) — hvad udtrykket betegner samt baggrunden herfor.

(6)

Bercy m.

DR ADM det franske økonomi- og finansministerium (opkaldt efter det kvarter i Paris hvor det ligger).

Quai m. d'Orsay

DR ADM det franske udenrigsministerium (opkaldt efter den gade i Paris hvor det ligger).

5.1.2 Grammatiske oplysninger

I det grammatiske felt har vi, eftersom vores målgruppe består af brugere med gode almensproglige franskkundskaber, gennemført en temmelig enkel grammatisk glossering af opslagsordene. Vi skal ikke komme nærmere ind på det her.

5.1.3 Ækvivalentsiden

Det er ækvivalentsiden og dens struktur, som på flere planer er den, der volder os de største problemer. Som tidligere anført, og som det fremgår af skemaet over artikelstrukturen, kan ækvivalentsiden opdeles i to delfelter, nemlig dels selve ækvivalentfeltet, og dels et felt til angivelse af kollokationsmateriale. Først kort om ækvivalentfeltet og de problemer, man skal have løst.

5.1.3.1 Ækvivalentfeltet

De typer oplysninger, den enkelte ordbogsforfatter vælger at medtage i en ordbog, kan naturligvis struktureres på en række forskellige måder. Vi skal ikke her give en samlet fremstilling af alle mulighederne, men illustrere en mulighed med udgangspunkt i vores ordbog.

En konkret konsekvens af vores specialordbogs type, formål og målgruppe er efter vores opfattelse, at der er behov for glosseringer af i hvert fald to typer. Derfor underopdeles ækvivalentfeltet i tre delfelter, dvs. et felt til ækvivalenten og to felter til glosseringer.

Den første glossering, der indføres i felt 1a (se figur 5), udgøres af en angivelse af retsområde i forkortet form, som gør brugeren opmærksom på, hvor i retssystemet vi befinder os, hvilket bør være en meget nyttig faglig oplysning for brugeren. I andet felt (1b) anføres selve ækvivalenten, og når vi skønner det nødvendigt, udfyldes det tredje felt (1c) med en forklaring af de ofte meget specialiserede juridiske udtryk, der indgår i ordbogen. Da der er tale om endog meget specialiserede begreber, anser vi det for nødvendigt for opfyldelsen af vores hovedmål med ordbogen at medtage faglige forklaringer, selv hvor der må anses at eksistere fuld ækvivalens mellem lemma og målsprogsterm.

(7)

compétence f. d'attribution

PR GEN saglig kompetence (udtryk for at en bestemt type organ, fx byretten, er kompetent til at behandle en sag), **la matière entre dans la ~ du tribunal de grande instance** sagen hører under byretten for store sagers saglige kompetence.

5.1.3.1.1 Rækkefølgen af synonyme ækvivalenter

Først lidt om ækvivalenterne. Lemma kan være enten monosemt eller polysemt. I begge tilfælde skal der tages stilling til, i hvilken rækkefølge synonyme ækvivalenter skal præ-senteres, og ved polyseme lemmaer skal rækkefølgen af delækvivalenter også afgøres. Der findes flere muligheder ved begge spørgsmål.

For de synonyme ækvivalenteres vedkommende kan man vælge at anføre dem i alfabetisk orden eller efter frekvens. Det alfabetiske princip finder vi ikke hensigtsmæssigt, da brugeren i så tilfælde ingen vejledning får overhovedet. Opretholdelse af alfabetisk rækkefølge med angivelse ved hver ækvivalent af fx stilistiske forskelle mener vi ikke er mulig, da nuancerne kan være meget små og intuitivt baseret og således vanskelige at beskrive.

En rækkefølge, der er baseret på egentlige frekvensanalyser, ville være det ideelle, men er ikke praktisk mulig, når det drejer sig om en ordbog af det omfang, vores ordbog har.

Vi har derfor valgt en tredje løsning, nemlig at anføre den ækvivalent, som vi på baggrund af vores arbejde med de pågældende områder vil anbefale, på førstepladsen i ækvivalentfeltet. De øvrige ækvivalenter angives efter samme princip, dvs. med faldende „anbefalingsstyrke“. Det skal dog bemærkes, at forskellen mellem de opgivne ækvivalenteres anbefalingsgrad ikke er særlig stor, da mindre anbefalelsesværdige termer normalt ikke medtages.

5.1.3.1.2 Rækkefølgen af delækvivalenter

Vi vil nu se på rækkefølgen af delækvivalenter ved polyseme lemmaer. Bergenholtz/Tarp (1994:222–223) angiver to mulige rækkefølger i det såkaldte polysemiindeks, nemlig enten en alfabetisk rækkefølge, eller en rækkefølge baseret på delækvivalenternes frekvens. Ingen af disse løsninger finder vi acceptable eller håndterbare i vores tilfælde. Vi har derfor forsøgt at opstille en række kriterier med henblik på at gøre rækkefølgen præcis og ensartet. Vi har gennemført forsøg med en kombination af flere kriterier, fx at betydninger der ligger tæt ved hinanden følger efter hinanden, den almindelige betydning før den specielle betydning, konkreter før abstrakter. Men vi er kommet frem til, at det er meget vanskeligt at bruge for mange kriterier, og er endt med følgende kriterium, som forsøges gennemført konsekvent: det generelle før det specielle, dvs. det mindre faglige og teoretiske før det mere faglige og teoretiske.

(8)

défaillance f.

1. DR GEN vanskeligheder, krise, ~ **financière** økonomiske vanskeligheder;
2. DR OBL misligholdelse, manglende opfyldelse af forpligtelse, ~ **contractuelle** misligholdelse af aftale;
3. PR GEN ubeblivelse (fra fx et retsmøde).

Et rækkefølgeproblem af lignende art står vi overfor, når vi skal indplacere henvisningen fra elliptiske former til deres fulde termer i polysemiindekset. Skal vi samle alle udtryk — og dermed også elliptiske udtryk med deres henvisning — der hører til samme semantiske felt (dvs. hos os typisk inden for samme retsområde), eller skal man placere de elliptiske udtryk samlet sidst i polysemiindekset, og i så fald i hvilken rækkefølge?

Den første løsningsmodel vil betyde, at en artikel kan blive splittet op, så der skiftes mellem angivelse af ækvivalenter og henvisninger til fulde termer. Det er et princip, der

nogle gange virker meget hensigtsmæssigt, men andre gange bliver grafisk rodet og upraktisk for brugeren. Vi har derfor valgt at samle alle henvisninger til sidst i artiklen. Også her har vi valgt mellem en slags semantisk gruppering og en alfabetisk angivelse.

Det kan forekomme hensigtsmæssigt at anføre eksempelvis en henvisning til et overbegreb før henvisning til underbegreber. Det vil imidlertid ikke altid være sådan, at der henvises til et overbegreb. Undertiden kan der ved flere henvisninger være tale om sideordnede udtryk, blandet med overordnede og underordnede udtryk.

Derfor har vi valgt at angive henvisningerne alfabetisk. Det er også det hurtigste for brugeren, når han skal slå de fulde termer op.

(9)

adjudication f.

1. VENTE tilslag (ved auktion);
2. DR MARCHES „udbud [hvor det billigste tilbud skal antages]“ (opfordring til at afgive tilbud, hvor udbyder er forpligtet til at antage det billigste tilbud);
3. VOIES EXEC = **adjudication judiciaire**;
4. VENTE = **adjudication volontaire**.

5.1.3.1.3 Faglige forklaringer

Nu er vi kommet til det sidste delfelt i ækvivalentfeltet, nemlig forklaringsfeltet. Mange vælger, at kommentarer til lemma placeres foran ækvivalenterne, mens kommentarer til ækvivalenterne følger efter disse. Undertiden kan en faglig forklaring imidlertid angå såvel lemma som ækvivalent.

Vi mener derfor, at det er mest brugervenligt og overskueligt at anføre **en** samlet evt. kontrastiv forklaring efter ækvivalenten. Det betyder, at såfremt der intet er angivet, angår forklaringen både franske og danske forhold. Det vil fx være tilfældet ved fuld ækvivalens mellem lemma og målsprogsterm (se (7)).

I tilfælde af, at der på målsproget ikke findes nogen ækvivalenter til lemma, kan man anvende forskellige strategier til gengivelse af lemma, fx låneord, låneoversættelse og analogoversættelse. Kan disse muligheder imidlertid ikke benyttes, må der udarbejdes et oversættelsesforslag, som markeres grafisk fx ved anførselstegn. Faglige forklaringer kan her være til stor hjælp for brugeren, når han skal vurdere, om det anførte oversættelsesforslag kan anvendes i den kontekst, han står overfor. Forklaringer ved oversættelsesforslag vedrører naturligvis franske forhold, og intet angives:

(10)

référé m.

- PR CIV „hastesagsprocedure“ (særlig hurtig procedure der kan anvendes i preserende sager til at træffe en foreløbig afgørelse, som senere kan indbringes for domstolen), **ordonner** (el. **statuer**) **en** ~ afsige kendelse i hastesag.

Ved partiel ækvivalens fremgår det af forklaringen, om den går på danske eller franske forhold, idet vi angiver DK og F (se (11) og (12)).

(11)

huissier m. **de justice**

1. PR GEN „person med offentlig bemyndigelse til at varetage tvangsfuldbyrdelser, forkynde stævninger o.l.“;
2. PR GEN ≈ stævningsmand;
3. VOIES EXEC „retsmedhjælper“, ≈ kontorfoged (varetager i F de fleste af de opgaver, som fogeden gør i DK, men er ikke dommer som den danske foged, se også **juge de l'exécution**);
4. PR GEN = **huissier de justice audiencier**.

5.1.3.2 Kollokationsfeltet

Der er to typer materiale, som vi ønsker at medtage for at illustrere brugen af termerne. Det drejer sig dels om kollokationer, dels om eksempler der kan bestå af hele sætninger eller brudstykker af sætninger.

Vi skal her kort se på, hvordan vi anfører dette illustrationsmateriale. Nogle ordbøger vælger at angive alle kollokationer og eksempler til sidst i den enkelte artikel efter de forskellige delbetydninger (vel at mærke også ved polyseme lemmaer). Vi mener imidlertid, at brugeren har størst glæde af at få alle relevante oplysninger samlet ved den enkelte delbetydning.

Et andet spørgsmål, der skal tages stilling til, er, i hvilken rækkefølge illustrationsmaterialet skal angives. Man kan igen vælge at arbejde med to felter, et til kollokationer og et til eksempler. Vi har af flere grunde valgt at anføre kollokationer og eksempler samlet i alfabetisk rækkefølge efter det ord, som vi anser for at være nøgleudtrykket i henholdsvis kollokationen og eksemplet.

(12)

appel m.

1. PR GEN appel (i DK overbegreb for anke og kære);
2. PR GEN anke (i DK appel af dom), **faire** (el. **former** el. **interjeter** el. **relever**) ~ **contre un** (el. **d'un**) **jugement** anke (el. appellere) en dom, **statuer à charge d'** ~ afsige dom som kan ankes (el. appelleres);
3. PR GEN kære (i DK appel af kendelse el. beslutning).
4. (...)

Man kunne vælge at samle synonyme kollokationer, fx flere synonyme verber til et substantiv (conclure, former, passer un contrat), og således bryde den strengt alfabetiske rækkefølge. Ved lange artikler med mange kollokationer og eksempler kan denne løsning imidlertid gøre artiklen uoverskuelig. Derfor anser vi en strengt alfabetisk rækkefølge for at være den mest hensigtsmæssige løsning i vores tilfælde.

Ud over kollokationer har vi som sagt også medtaget eksempler bestående af hele sætninger eller dele af sætninger. Eksemplerne skal vise, hvordan man i en oversættelses-situation fx skal skifte ordklasse eller omformulere i forhold til kildeteksten.

(13)

mandataire m.

MANDAT fuldmægtig, fuldmagtshaver, den befuldmægtigede, mandatar, **par les présentes le constituant institue pour ~ X et lui donne les pouvoirs ci-après à l'effet de ...** undertegnede befuldmægtiger herved X til ...

6 Afsluttende bemærkninger

Vi har her fremlagt de hovedprincipper, som vi bestræber os på at følge under udarbejdelsen af den fransk-danske juridiske ordbog. Vi skal afslutningsvis bemærke, at det, der har voldt os de største problemer og fortsat volder problemer under arbejdet med ordbogen, er en konsekvent implementering af vore principper. Vi mener, at det er vigtigt at forsøge at følge de fastlagte principper, men at det ikke kan undgås, at man må ty til pragmatiske løsninger.

Bibliografi

- Bergenholtz, Henning/Sven Tarp 1994: *Manual i fagleksikografi. Udarbejdelse af fagordbøger — problemer og løsningsforslag*. Herning: Systime.
- Dyrberg, Gunhild/Joan Tournay 1990a: Définition des équivalents de traduction de termes économiques et juridiques sur la base de textes parallèles. I: *Cahiers de lexicologie* 56–57, 261–274.
- Dyrberg, Gunhild/Joan Tournay 1990b: Bercy en retard sur Gorby. I: *SPRINT* 2, 17–25.
- Dyrberg, Gunhild/Dorrit Faber/Steffen Leo Hansen/Joan Tournay 1991: *Oprettelse af fagsproglige tekstkorpora: Engelsk – fransk – dansk juridisk sprog: Aftaleret*. ARK 60. København: Handelshøjskolen i København.
- Hartmann, R. R. K. (udg.) 1983: *Lexicography: Principles and Practice*. London: Academic Press.
- Hausmann, Franz Josef 1977: *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Max Niemeyer Verlag.
- Hausmann, Franz Josef/Herbert Ernst Wiegand 1989: Component Parts and Structures of General Monolingual Dictionaries: A Survey. I: Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta (udg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie* I. Berlin/New York: Walter de Gruyter, 328–360.
- Henriksen, Torben 1991: *Juridisk ordbog spansk – dansk*. København: Gads Forlag.
- Nielsen, Sandro 1994: *The Bilingual LSP Dictionary: Principles and Practice for Legal Language*. Tübingen: Gunter Narr Verlag.
- Petit Robert 1993 = Josette Rey-Debove/Alain Rey/Paul Robert: *Le Nouveau Petit Robert*. Paris: Dictionnaires Le Robert.
- Rey-Debove, Josette 1971: *Étude linguistique et sémiotique des dictionnaires français contemporains*. The Hague/Paris:
- Svensén, Bo 1987: *Handboki i leksikografi. Principer och metoder i ordboksarbetet*. Stockholm: Esselte Studium/Tekniska nomenklaturcentralen.

- Wiegand, Herbert Ernst 1983: Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. I: Herbert Ernst Wiegand (udg.): *Studien zur neuhochdeutschen Lexikographie* III. Hildesheim/New York: Olms, 401–474.
- Wiegand, Herbert Ernst 1989a: Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. I: Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta (udg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie* I. Berlin/New York: Walter de Gruyter, 371–409.
- Wiegand, Herbert Ernst 1989b: Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. I: Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta (udg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie* I. Berlin/New York: Walter de Gruyter, 409–462.
- Zgusta, Ladislav 1988: *Lexicography Today. An annotated bibliography of the theory of lexicography*. Tübingen: Max Niemeyer Verlag.

A Concordance to Old Icelandic Texts and its Lexicographic Value

Den første del af dette foredrag er en beskrivelse af projektet *Konkordanse til de Islandske sagaer*, som udkommer på CD-ROM i slutningen af 1995. I den anden del af foredraget spekulerer forfatteren på den nytte som ordbogsredaktører kan have af en konkordanse som denne, og giver nogle konkrete eksempler som skal vise at den foreliggende konkordanse vil muliggøre en bedre og nøjagtigere ordbogsbeskrivelse af gammelislandsk, både i syntaktisk og semantisk henseende.

The main subject of my paper will be a new concordance to the *Íslendinga sögur* (Icelandic Family Sagas), which will be published on CD-ROM later this year. In the first part of the paper, I will describe the concordance, but in the second part, I will consider its potential use in dictionary making.

1 The concordance

The concordance to the Family Sagas (Eiríkur Rögnvaldsson et al. 1995) is one of the first concordances to be published in Iceland. The very first computerized concordance to an Icelandic text is the one which professor Baldur Jónsson and his collaborators made to the novel *Hreiðrið* by Ólafur Jóhann Sigurðsson. This concordance, which was published in a limited number of copies in 1978 (Baldur Jónsson 1978), differs however from the present one in various respects, the most important difference being that it is not lemmatized.

The first lemmatized concordance to a text in Icelandic appeared just before last Christmas. This is the concordance to the latest edition of the Bible, which was made by a group of specialists from different institutions (*Bibliulykill* 1994). This work is in many ways comparable to ours, but there are, however, several important differences. First, many of the most frequent words are omitted; for instance, all prepositions, conjunctions, and several adverbs, and also a few frequent verbs and nouns. All such words are included in our concordance. Second, the ordering of the occurrences of each word is different. In the concordance to the Bible, the examples are ordered according to the order in which they appear in the Bible. In our concordance, on the other hand, the examples are alphabetically ordered according to the following word.

A group of scholars started working on the concordance to the Icelandic Sagas in 1989. This group consists of Bergljót Kristjánsdóttir, Guðrún Ingólfsdóttir, Örnólfur Thorsson, and myself, but several others have also worked more or less on the project, which has been generously supported by the Icelandic Science Fund. It is based on a new edition of the Sagas, which appeared in 1985 and 1986 (*Íslendinga sögur* 1985–86). Some of the editors of that edition are also among the leaders of the present project, which can thus be seen as a continuation of the edition.

The main work on the concordance was done in 1989, and in November that year, the lemmatization was almost finished, so that preliminary results of some frequency studies on the vocabulary of the Sagas could be presented at a conference in Reykjavík; these results have been published in the journal *Skáldskaparmál* (Eiríkur Rögnvaldsson 1990). At that time, the grants that had been given to the project had been used up, but the project itself, however, was far from finished. The lemmatization had to be carefully checked and proof-read, the computer files had to be corrected, etc. But due to lack of money, the editors of the concordance have only been able to work on it in their spare time the last five years.

This does not mean, however, that the concordance has been inaccessible up to now. Since 1992, when the work was practically finished, it has been preserved at the Institute of Linguistics at the University of Iceland, both on a computer and in a laser print-out. Everybody has had unlimited access to both versions. In addition, the editors have answered numerous questions from all over the world, concerning words and phrases in the Sagas. However, with respect to the usefulness of the concordance, it has of course been a major drawback that it is not publicly available.

We have recently made a contract with the publishing house Mál og menning, which holds the copyright to the editions on which the concordance is based. Later this year, they are planning to publish a CD, which will include the concordance and also a text version of the Sagas. Both will be easily searchable by means of special Windows-based programs. There will be links between the text and the concordance, so that it will be possible to click on a certain word in the text and get all the examples of that word on the screen; or to click on a word in the concordance and get the surrounding text on the screen. The CD will also include several lists, such as a frequency list, a list of compounds, etc.

1.1 The making of the concordance

The concordance comprises all the texts in the edition on which it is based, except the *þættir*; the poetry is also omitted. The Sagas are around 40, but some of them exist in two widely different versions, so that 50 different texts are printed in the edition. This is around 5 megabytes of text, or nearly 900 000 running words; 2 079 pages.

We started by inserting special markers for each Saga, chapter numbers and page breaks. Then we could use WordCruncher, from Johnston & Co. in the United States, to generate a list of all the occurrences of each individual word-form. In this list, we have the word-form in question in the middle, with approximately 40 characters context in each direction, and references to Saga, chapter and page in the beginning of each line. At this stage, the file looks like the picture in (1a).

The next step is to prepare this file for lemmatization. We used WordPerfect macros to boldface the word in the middle, and to sort the examples of each word-form alphabetically, according to the following context. After that, the file looks as in (1b).

Up to this point, the process has been reasonably mechanic, but now comes the difficult part; the lemmatization itself, where we group together all the forms belonging to each individual lexeme, and make a distinction between all homonyms belonging to different lexemes. We considered using computer programs to make this easier, and we actually tried one such program, but we soon found out that its benefits did not compensate for the errors it made. So, the lemmatization had to be done manually, which was quite a task, considering

the size of the corpus. When it was finished, the concordance files were printed on a laser printer, giving the result shown in (1c).

(1)a

			I*þak (1)	
Laxd	64:1634	stóðu út af ásendamir og var einart	I*þak	á húsinu og ekki gróíð. Þá mætti
			I*þakið (6)	
Njála	77:213	eigi fyrr en þeir höfðu undið allt	I*þakið	af skálanum. Gunnar skýtur þá af
Egla	22:392	þurt og bræddur víðurinn en nærum	I*þakið	um mærið, Þórólfur bað menn sína
Gísl	13:864	húsin að drjúpa sem líklegt var er	I*þakið	tók að rofna. Gísl spratt upp
Gísl	18:917	þvo mikill að þegar tekur af skálanum	I*þakið	öðrum megin og hver að öðrum þar til
Gísl	18:917	og hver að öðrum þar til er allt er	I*þakið	af húsinu. Vatnfall fylgdi hér og svo
Vígl	2:1957	gulli í skurðina. Þetta herbergi var	I*þakið	blíyi og steint allt innan. Skíðgarður
			I*þakka (14)	
Njála	44:177	þá til nokkurs." "Allvel skal eg	I*þakka	þeim af þeir segja mér heim víg
Egla	16:384	öllu best varló er eg hef gert til	I*þakka	yðvarra." Konungur segir að ekki
Finnb	31:654	jarl hefa gefið sér, mág sinn. Þeir	I*þakka	honum harðla vel og ríða heim. Hafði
Fóstb	5:786	farló hafði með þeim Vemundi. Þeir	I*þakka	henni sitt tillag er hún hafði þeim
Fóstb	5:786	og er þeir voru búnir til ferðar þá	I*þakka	þeir henni þarvist sína og allan
Fóstb	20:818	búnir þá fara þeir á konungs fund og	I*þakka	alla vingan þá er hann hafði þeim
Gísl	9:860	það var vant að þorkell var vanur að	I*þakka	bróður sínum verkið en nú er hann
Grætt	91:1093	ekki víst um aturkomu sína "Vil eg nú	I*þakka	yður öllum," segir Þorsteinn, "Þversu
GunkV	11:1160	honum því að eg á honum lífgjöf	I*þakka."	Karl segir: "Eg vil fá að sjá
Halm	6:1205	sem konungur er og vil eg	I*þakka	hérvist mína." Þá kom það fram
Halló	7:1235	þar nær sem konungurinn er. Vil eg	I*þakka	honum hirðvist." Þá kom það fram
Laxd	40:1686	hann vildi. Þá mætti Kjartan:	I*þakka	viljum vér yður konungur er þér gefið
Svarf	8:1789	og ríðu í braut. Nú	I*þakka	menn Þorsteini fyrir er þér gefið
Vígl	4:1911	þetta vil eg eigi veita." Þeir	I*þakka	honum vel og þótti þeim þó mikill í
			I*þakkað (1)	
Egla	65:469	voru. Var þetta verk honum allvel	I*þakkað.	Þá kvað Egill:

(1)b

þak (1)				
Laxd	64:1634	stóðu út af ásendamir og var einart	þak	á húsinu og ekki gróíð. Þá mætti
þakið (6)				
Gísl	18:917	og hver að öðrum þar til er allt er	þakið	af húsinu. Vatnfall fylgdi hér og svo
Njála	77:213	eigi fyrr en þeir höfðu undið allt	þakið	af skálanum. Gunnar skýtur þá af
Vígl	2:1957	gulli í skurðina. Þetta herbergi var	þakið	blíyi og steint allt innan. Skíðgarður
Gísl	13:864	húsin að drjúpa sem líklegt var er	þakið	tók að rofna. Gísl spratt upp
Egla	22:392	þurt og bræddur víðurinn en nærum	þakið	um mærið, Þórólfur bað menn sína
Gísl	18:917	þvo mikill að þegar tekur af skálanum	þakið	öðrum megin og hver að öðrum þar til
þakka (14)				
Fóstb	20:818	búnir þá fara þeir á konungs fund og	þakka	alla vingan þá er hann hafði þeim
Gísl	9:860	það var vant að þorkell var vanur að	þakka	bróður sínum verkið en nú er hann
Fóstb	5:786	farló hafði með þeim Vemundi. Þeir	þakka	henni sitt tillag er hún hafði þeim
Halm	6:1205	sem konungur er og vil eg	þakka	hérvist mína." Þá kom það fram
Finnb	31:654	jarl hefa gefið sér, mág sinn. Þeir	þakka	honum harðla vel og ríða heim. Hafði
Halló	7:1235	þar nær sem konungurinn er. Vil eg	þakka	honum hirðvist." Þá kom það fram
Vígl	4:1911	þetta vil eg eigi veita." Þeir	þakka	honum vel og þótti þeim þó mikill í
GunkV	11:1160	honum því að eg á honum lífgjöf	þakka."	Karl segir: "Eg vil fá að sjá
Svarf	8:1789	og ríðu í braut. Nú	þakka	menn Þorsteini fyrir er þér gefið
Laxd	40:1686	hann vildi. Þá mætti Kjartan:	þakka	viljum vér yður konungur er þér gefið
Grætt	91:1093	ekki víst um aturkomu sína "Vil eg nú	þakka	yður öllum," segir Þorsteinn, "Þversu
Egla	16:384	öllu best varló er eg hef gert til	þakka	yðvarra." Konungur segir að ekki
Njála	44:177	þá til nokkurs." "Allvel skal eg	þakka	þeim af þeir segja mér heim víg
Fóstb	5:786	og er þeir voru búnir til ferðar þá	þakka	þeir henni þarvist sína og allan
þakkað (1)				
Egla	65:469	voru. Var þetta verk honum allvel	þakkað.	Þá kvað Egill:

After that, the concordance was proof-read and the lemmatization rechecked. During that process, all available dictionaries were consulted, especially Fritzner's *Ordbog over det gamle norske sprog* (Fritzner 1954), of course, but also Ásgeir Blöndal Magnússon's etymological dictionary *Íslensk orðsifjabók* (Ásgeir Blöndal Magnússon 1989), and several other works. This was a very time-consuming process, as one can imagine given the fact that the paper version of the concordance is more than 7 000 pages with 100 lines on each page and ca. 100 characters per line. Since Old Icelandic is a highly inflected language, homonyms of different lexemes are very frequent, and therefore, it was necessary to read most of these lines carefully, because it is very often possible that a rare inflectional form

of a verb, for instance, is homonymous with a form that one would a priori think that could only be a noun.

(1)c

þak hk; þak (1); þakið (4)			
Laxd 64;1634	stóðu út af ásendamir og var einart	þak	á húsinu og ekki gróið. Þá mælti
GíslL 18;917	og hver að öðrum þar til er allt er	þakið	af húsinu. Vatnfell fylgdi hér og svo
Njála 77;213	eigi fyrir en þeir höfðu undið allt	þakið	af skálanum. Gunnar skýtur þá af
GíslS 13;894	húsin að djúpa sem líklegt var er	þakið	tók að rofna. Gísli spratt upp
GíslL 18;917	svo mikill að þegar tekur af skálanum	þakið	öðrum megin og hver að öðrum þar til
þakka so; þakka (13); þakkað (1)			
Fóstaþ 20;818	búinir þá fara þeir á konungs fund og	þakka	alla vingan þá er hann hafði þeim
GíslS 9;880	það var vant að borkell var vanur að	þakka	bróður sínum verkið en nú er hann
Fóstaþ 5;786	ferið hafði með þeim Vemundi. Þeir	þakka	henni allt tillag er hún hafði þeim
HallM 6;1205	sem konungur er og vil eg	þakka	hérvið mína." Þá kom það fram
Finnb 31;654	jarl hafa gefið sér, mér sínu. Þeir	þakka	honum harða vel og ríða heim. Hafði
HallÓ 7;1235	þar nær sem konungurinn er. Vil eg	þakka	honum hriðvíst." Þá kom það fram
YggL 4;1911	þetta vil eg eigi veita." Þeir	þakka	honum vel og þótti þeim þó mikilið í
GunnKV 11;1160	honum því að eg á honum lífgjafi að	þakka."	Kari segir: "Eg vil fá að sjá
Svarf 8;1789	og ríðu í braut. Nú	þakka	mann Þorsteini fyrir er hann tókst
Laxd 40;1696	hann vildi. Þá mælti Kjartan:	þakka	villjum vér yður konungur er þér gefið
Grett 91;1093	ekki víst um ásturkomu sína "vil eg nú	þakka	yður öllum," segir Þorsteinn, "hveru
Njála 44;177	þá til nokkurs." "Allveg skal eg	þakka	þeim ef þeir segja mér heim víg
Fóstaþ 6;786	Og er þeir voru búinir til ferðar þá	þakka	þeir henni þarvíst sína og allan
Egla 65;469	voru. Var þetta verk honum allivel	þakkað.	þá kvað Egill:
þakja so; þakið (2)			
Vígl 2;1957	gulli í skurðina. Þetta herbergi var	þakið	blíði og steint allt innan. Skíðgarður
Egla 22;392	þunt og bræddur víðurinn en nærum	þakið	um næfri. Þórdífur bað menn sína
þökk kv; þakka (1)			
Egla 16;384	öllu best varið er eg hefði gert til	þakka	yðvarra." Konungur segir að ekki

The final step was to make the necessary corrections to the computer files. As I said before, this was practically finished in 1992, even though individual corrections are still being made. Users of the concordance have sometimes noticed errors and inconsistencies which they have told us about. I can particularly mention Þórdís Úlfarsdóttir, who went carefully through the concordance in connection with Jón Hilmar Jónsson's work on his book *Orðastaður*, which was published last year (Jón Hilmar Jónsson 1994). Þórdís gave us a list of errors that she had found, and we are very grateful to her and others who have assisted us in eliminating errors as far as possible.

1.2 Vocabulary and word frequency

The concordance has already proved to be very useful in itself. Let me first mention its use as a frequency dictionary. For the first time, we now have an overview of the vocabulary of a whole literary genre; the Icelandic Family Sagas. Of course, there exist dictionaries of Old Icelandic, especially Fritznér's (1954) *Ordbog over det gamle norske sprog*; and as is well known, the Arnarnagnæan Commission in Copenhagen has been working on a dictionary of Old Norse for several decades. These works, however, comprise not only narrative texts like the Sagas; they also cover other genres such as the law, lives of saints, etc. The vocabulary of these genres is remarkably different from that of the Sagas.

Now we know that the vocabulary of the Sagas is somewhere between 12 000 and 12 500 words — the exact figure depends on our definition of lexeme, and besides, differences between manuscripts can of course affect the figure. We can also find out the vocabulary of each individual Saga. *Njáls saga*, for instance, uses around 3 200 different words. It appears that the Sagas use unusually few words. Unfortunately, however, we cannot show this statistically, since there exist no comparable studies of Modern Icelandic texts — except for the Bible, which is hardly representative of Modern Icelandic.

The Institute of Lexicography has recently published a frequency dictionary of Modern Icelandic, *Íslensk orðtíðnibók* (Jörgen Pind et al. 1991). It is possible to compare several figures from this work to the results of our study of Old Icelandic. This is done in (2) below.

(2)

	<i>Íslendingasögur</i>			<i>Íslensk orðtíðnibók</i>
	<i>Lexemes</i>	<i>Running words</i>	<i>%</i>	<i>%</i>
<i>Nouns</i>	7 292	117 252	15.63	20.58
<i>Verbs</i>	1 447	203 148	27.09	20.65
<i>Adjectives</i>	2 851	31 947	4.26	7.14
<i>Adverbs</i>	706	173 484	23.13	23.25
<i>Pronouns</i>	52	94 800	12.64	14.88
<i>Conjunctions</i>	20	113 823	15.18	12.01
<i>Numerals</i>	33	6 292	0.84	1.18

The first column shows how many lexemes in the Sagas belong to each part of speech. As you see, the nouns make up almost 60% of the vocabulary. I must point out that adverbs and prepositions are grouped together. This is done to facilitate the comparison with the results from *Íslensk orðtíðnibók*, and besides, it is often very difficult or even impossible to draw a line between these two parts of speech.

In all the other columns, the figures refer to running words but not to lexemes. In the second column, we see that the relative frequency of running words belonging to each part of speech is widely different from the relative frequency of lemmas. The last two columns show percentages; the first of them shows the percentage of running words in each part of speech in the Sagas, whereas the second shows comparable figures from *Íslensk orðtíðnibók*.

As you see, the figures are rather similar. There is, admittedly, a considerable difference in the relative frequency of nouns. The reason is that we have omitted all proper names from our figures for nouns in the Sagas. It must be noted that proper names are no doubt much more common in the Sagas than they are in the texts on which *Íslensk orðtíðnibók* is based. If we had chosen to include proper names in our figures, the relative frequency of nouns would have been higher in the Sagas than in *Íslensk orðtíðnibók*.

It must also be noted that we have chosen to count all instances of participles, both past and present, as verb forms; the only exception being present participles used as nouns, such as *eigandi*. The obvious alternative would have been to classify the participles as either verbs or adjectives according to their syntactic status in each case, as is done in *Íslensk orðtíðnibók*. We actually tried this in the beginning, but we soon came to the conclusion that it was impossible to make a principled decision in all cases, and the only consistent solution would be to count all participles as verbs. This decision, of course, results in relatively more occurrences of verbs and fewer occurrences of adjectives than it would have done if we had followed the same principles as the authors of *Íslensk orðtíðnibók*; but if we take this difference into account, I think we can say that the figures in the last two columns are very similar.

1.3 Other uses

We are pleased to say that the concordance has already been used and quoted in numerous publications in different disciplines, such as medieval literature, historical syntax, history, folklore, ethnography, law, zoology, physics, etc. In our view, one of the most important features of the project is its interdisciplinary character. It brings together scholars from various fields of study, who are working on some aspects of Medieval Iceland. They can use the concordance to locate places of interest in the Sagas, and thus, they can get a unique overview of their subject. Thus, the concordance has already inspired several studies, and the insights these scholars get by using the concordance will in turn be of tremendous use in the semantic description of numerous words in the Sagas.

2 How the material affects the structure of the dictionary

My second main subject in this talk is the use of this kind of material, i.e. a concordance, in dictionary making. In what way does it affect the final form of the lemmas in a traditional dictionary if the material is a concordance, but not accumulated by a traditional excerption? The effects are numerous and of various kinds, but the most important are those listed under (3):

- (3) a. Frequency information facilitates the selection of citation forms
- b. The semantic description of very common words will be more accurate; different senses of a word can be more easily ordered by importance, and various subtle semantic differences can be more easily detected
- c. Formal categorization will be more prominent, and syntactic features (such as case government) are listed more systematically
- d. The selection of text examples (citations) will be more accurate, and the examples will be more typical
- e. All kinds of collocations and word patterns are more obvious, and therefore more likely to be mentioned

In the following, I will discuss each of these effects in turn.

2.1 Frequency

As is well known, it is not always necessary nor feasible to list every word that occurs in a given corpus as a separate dictionary entry with its own description. On the contrary, there are numerous cases where two or more words which differ somewhat in form should rather be considered as belonging to the same lexeme, and listed under one citation form. Such examples can be of various types, and some of them are shown in (4) below.

- (4) a. hofgoði – hofsgoði
atgervimaður – atgervismaður
höfuðbani – höfuðsbani
hugboð – hugarboð
- b. aðdráttamaður – aðdráttarmaður
affaradagur – affarardagur
- c. drukknan – drukkun, geipan – geipun
auðigur – auðugur, ástúðigur – ástúðugur
hraustleikur – hraustleiki, hvatleikur – hvatleiki
- d. atgervi – atgjörvi
gagnvert – gagnvart
- e. heyrinkunnigur – heyrunkunnigur
hlægur – hlæglegur – hlægilegur

In Icelandic, such formal differences are often due to different ways of compounding. In the Icelandic Sagas, both *hofgoði* and *hofsgoði* are found, as shown in (4a). We can explain this difference by saying that in the former, the first constituent of the compound is the stem, whereas in the latter, the first constituent is the gen.sg. form. However, there is little doubt that these two should be considered as belonging to the same lexeme.

We also find a number of compounds where the first constituent sometimes has the gen.sg. form, but in other cases the gen.pl. form. This is most frequent in words where the first part has a gen.sg. ending in *-ar*; then the only difference between the gen.sg. and the gen.pl., which always ends in *-a*, is the *-r*. Since the number (sg. or pl.) of the first constituent in such compounds is (usually) not semantically distinctive, and since the *-r* is often not clearly pronounced, such vacillation in number is common in Modern Icelandic; and many similar examples can also be found in the Sagas, such as *aðdráttarmaður* and *aðdráttamaður*, which are shown under (4b) above.

There are also various examples of suffixes having more than one form in the Sagas; for instance *-anl-un*, *-igurl-ugur*, *leikurl-leiki* and others, in words like *geipanlgeipun*, *auðigur/auðugur*, *hraustleikur/hraustleiki*, as shown in (4c) above. We also find words with and without breaking, such as *atgjörvi* and *atgervi*, as shown under (4d); and various other types, cf. (4e).

In cases like these, the lexicographer is often faced with several problems. It is often difficult to decide whether to group two or more different forms under one headword. Even if it can be shown that two different forms stem from the same lexeme historically, it is by no means evident that they should be given a single lexical entry in the dictionary. It is perfectly possible that each form has developed a special meaning which makes it natural to list both forms separately.

If we decide to group the different forms together in a single dictionary entry, it is often difficult to select the headword. The most straightforward solution would perhaps be to select the most frequent form as the citation form, but it is often not easy to find out which of the forms is most frequent. It would for instance not be wise to base the choice on the number of examples that have been excerpted from texts.

A concordance can be of a great help in solving these problems. Since the concordance contains all the occurrences of every single word in a given corpus, it is easy to find the frequency of any particular form. This gives the lexicographer a more solid ground, on which to build the selection of a citation form. However, it is clear that frequency is not the only factor to consider in this respect; the selection must also fit into the system, so to speak.

A concordance also makes it easier to decide whether two different, but related forms actually mean the same, and hence should be listed under the same dictionary entry. By the careful examination of all the examples that the concordance makes possible, one can sometimes detect subtle semantic differences that would otherwise not be noticed.

2.2 Meaning

It is a well-known tendency for traditional excerption to give a somewhat skewed picture of the meaning or use of individual words. Lexicographers tend to pick up unusual examples, and hence, such examples often get a more prominent status in the dictionary than they deserve. When a dictionary is based on a concordance, this problem can be avoided, because in principle, at least, the description is based on all the examples found in the corpus. Therefore, the most frequent meaning and use should get prominent status in the description.

It is very important in this respect that the lexicographer who writes the final description for the published dictionary has access to all stages of the material. When a lexicographer is writing a dictionary entry on the basis of examples that have been collected in a traditional excerption, he is completely dependent on his examples. Of course, he can, in principle, look up the citations in the excerpted texts, but in practice, it is impossible to do so, except in a limited number of cases. Therefore, the lexicographer does not know how typical his data are.

I'll just show you one example. Fritzner (1954) gives the following semantic definition of the word *heimsókn*:

- (5)
- heimsókn*: 1. Besøg
2. Besøg som man aflægger i retslig Øiemed, for at fremme en Retssag o. desl.
3. voldeligt Overfald paa en i hans hjem, hjemsogelse hvorunder man bruger Magten mod dem som ere i Huset

In our corpus, we find the following examples of this word:

(6)

<i>heimsókn</i> kv;	<i>heimsókn</i> (5);	<i>heimsóknar</i> (1);	<i>heimsóknina</i> (1);	<i>heimsóknir</i> (1)	
Laxd	63;1632	og eigi mundi eg veita honum slíka	<i>heimsókn</i> .	Eða hvað er enn þá?" Hann svarar	
Gullþ	17;1138	að víg þeirra Helga skyldu á standast	<i>heimsókn</i>	og fjórráð við Þóri. Síðan var saman	
Vopnf	17;2003	eg brátt fara til Hofa og veita Bjarna	<i>heimsókn</i>	og sækja hann með eldi ef vér getum	
HallM	10;1218	komi fyrir víg Galta og þar með	<i>heimsókn</i>	við Kolfinnu fyrir mannamun en fyrir	
HallÓ	12;1249	komi víg Einarssonar og þar með	<i>heimsókn</i>	við Kolfinnu fyrir mannamun. En fyrir	
Svarf	15;1800	og snýr hann þá málinu öllu til	<i>heimsóknar</i> .	þetta voru ráð Hrólfs nefglitu og	
Reykd	14;1755	víggunum en víg Narfa skal koma fyrir	<i>heimsóknina</i>	og brotttöku konunnar með en hinir	
Vatn	29;1878	Már kvað þá sýna af sér óvingjarlegar	<i>heimsóknir</i>	og lést eigi mundu láta sinn hlut	

From these examples, it looks as if the most frequent meaning is the one under 3. in (5) above. Admittedly, Fritzner bases his description on many more texts, but however, there are

reasons to believe that a close consideration of all the examples would change the structure of the lexical entry.

When we started preparing the concordance, we were planning to exclude most function words; conjunctions and prepositions, and also many or most adverbs and pronouns. We did not think that examples of these words would be of any interest, since there are many examples of some of them on each page of the text. But when the real work started, we soon found out that a concordance could tell us many things about these words.

It is evident that only a limited number of examples of these words make the basis of their description in a traditional dictionary. Here we can, in principle, base our description on all the examples, and hence, we should be able to present a much more coherent description, both formally and semantically.

2.3 Syntactic characteristics and formal classification

In linguistic definitions of the lexicon, we usually read that this is the place where information on all unpredictable features of individual words is stored. This includes phonetic and phonological features (pronunciation), inflection, syntactic features, and meaning. This definition of course applies to the mental lexicon, but not to dictionaries, but by and large, I think we can say that these features are also the ones we can expect to find in a good dictionary.

Traditional dictionaries usually do justice to three of the above-mentioned fields. The phonological features can often be deduced from the spelling, and of course, many dictionaries show phonetic transcription. Inflection is usually shown by mentioning inflectional class, showing the principal parts of verbs, etc. The main part of the entry is, then, the semantic description.

Syntactic features, however, are usually not systematically represented. It is of course shown to which part of speech each lexical entry belongs; but features such as the case government and argument structure of verbs, for instance, are usually not mentioned. True, we can often see from the citations whether some verb takes one or two objects, or whether it governs accusative, dative, or genitive; but the point is that information on this is not systematically present, and it is sometimes lacking. One of the reasons for this is probably that the excerption of texts is not done with syntactic characteristics in mind, and therefore, there is simply no basis for including such things in the dictionary.

Here we have, once again, one of the problems with traditional excerption. Lexicographers have the tendency to pick up unusual or exceptional examples. This is fine, of course; but the danger is that such examples will be overrepresented in the material, at the expense of the normal use of words. If we find, for instance, one example where a certain verb governs a different case than it usually does, we are likely to pick up this example; and later, it might end up in a published dictionary, perhaps as the only text example which shows the case government of this verb.

By using a concordance, such dangers can be avoided. Since we have direct access to all the examples of each individual word in the corpus, we can simply count how often each verb takes each case, and make that information a part of the lexical entry, either directly or indirectly.

When the structure of a dictionary is based on a concordance of the kind we have made, it is bound to affect the final form of the lemmas in various ways. The main effect is probably that syntactic characteristics will be more prominent than they would otherwise, but semantic characteristics will tend to be less prominent. However, it must be emphasized that syntactic and semantic characteristics often go together, of course.

It is likely that in a traditional excerption, the meaning will be the dominant factor. The lexicographer tends to pick up those examples that exemplify the meaning of the word in question; but he will be less likely to select his examples according to their syntactic status.

In many ways, it is more straightforward to let formal characteristics govern the structure of the lemma than to let the semantics do the job. One reason is that the formal classification is usually rather clear-cut; the syntactic status of the word in question is normally reasonably clear, so that the formal classification is not problematic. Semantic classification often presents much more difficult problems, and the lexicographer will have to rely on his intuitions to a much greater extent.

I can mention here that in the *Sýnihefti sagnorðabókar* (Ásta Svavarsdóttir et al. 1993), which Orðabók Háskólans published two years ago, formal classification is dominant, but semantic classification subordinate. I think this booklet shows well the merits of that structure. However, it must be kept in mind that this work is not based on a concordance, but rather on material from a traditional excerption of texts; and as I said above, this might mean that certain syntactic constructions are not justly represented.

2.4 Selection of text examples

It is very important that the text examples in a dictionary are carefully chosen. The appropriate examples can shed a new light on the meaning of a word, and be more illuminating than a long and tedious definition or explanation. In a dictionary which is based on material from a traditional excerption, we can always expect the selection of examples to be more or less arbitrary. The examples in the material can have been collected for various reasons; they may be of interest semantically, syntactically, or morphologically, for instance, but that does not mean that they are typical of the use of the word in question.

2.5 Collocations

It is by no means obvious in what order the examples of each word form should appear in a concordance. We decided to order the inflectional form of each lexeme alphabetically, as shown in (7) below. There we have first all the examples of the form *heimil*, then all the examples of the form *heimila*, and so on. If you look at the examples of each form, you see at once that they are alphabetically ordered according to the following word or words.

It must be admitted that the decision to choose this particular order was not built on much considerations, but nevertheless, we think that this decision has proved to be correct. The reason is that this ordering reveals how common it is that the same string of words occurs many times in the corpus. The reasons for such recurrent patterns can of course vary. In some cases, it is fairly clear that one author is imitating another, and even though that can be of a great interests to philologists, such information should hardly enter the dictionary.

(7)

heimill lo; heimil (3); heimila (5); heimilan (1); heimilast (1); heimill (4); heimilt (10); heimul (2); heimull (1); heimult (6)	
Ljósc 23:1697	Þorkell hitti þá brátt og mælti:
Ljósc 13:1673	hugsað hafa hver niður skal koma en
Vatn 16:1862	mælti: "Það er vel gert. Er þér og
Valjó 3:1831	Ljótur svarar: "Þú skalt eiga
Laxd 29:1575	beiða yður herra að þér létuð oss
Ljósc 8:1667	fylkismenn sögðu að þeir bræður sætu
Fósb 23:832	yfir þér að fjandinn á ekki þig svo
Vatn 17:1863	hitti stýrimann að máli og lét honum
GíslS 12:863	þar til húas og lét Þorvaldur honum
Hænsþ 5:1421	til væri." "Það muntu eiga allra
Valjó 8:1839	ná nokkrum þeirra." Narfi kvaðst
Fjót 19:712	"Óngvan sá eg nema sjálfan mig en
Vatn 31:1881	þar var öllum mönnum matur
VígGl 2:1907	konunga eða annarra höfðingja? Er þar
Eirík 7:529	vorum malt og mjöl og korn og er yður
Fósb 11:800	og skemmta þér hér? Nú er það
Njála 13:142	fyrir Glúmi ef hann er lengur, en
Hænsþ 10:1427	reka hingað fé sitt. Skal þeim það
GíslS 32:890	þú hefir sagt. Og mun þér nú þykja eg
HalMV 1:1221	Stýrimaður mælti: "Það skal ykkur nú
Eyrb 49:601að	maður skyldi jafnörgum mönnum eiga
Fígam 30:759	þar að vera "og allt mitt góss er þér
Valjó 6:1835	á skipinu fara. Ásmundur kvað þeim
Vatn 12:1858	rammra hluta en fési. En það er
Fósb 3:779	Húskarlinn mælti: "Gakk inn þú.
Laxd 62:1630	segir hann, "að fylgja þér heim því að
Njála 136:296	mælti til Fíosa: "Því eru borð sett að
Laxd 70:1642	gangi fram." Þorkell svarar:
Egla 61:458	innngöngu í höllina. Honum var það og
Egla 71:478	til þessar ferðar þá mun yður það
Egla 73:482	sæja hestum sínum. Þorfinnur bóndi lét
Egla 33:408	og sagði svo að það var skylt og
Egla 41:417	umsjá." Þórir sagði að það var
Heimil	mun vist með mér sem fyrir þótt vant sé
Heimil	munu þar til vor orð." Hann
Heimil	vor mörk sem þú vilt höggva láta en eg
Heimila	hegning ef oftar verður en legg nú
Heimila	mörk ybra að höggva húsavið.
Heimila	sveitarvist þeim sem þeir viðu.
Heimila	til iltra hluta sem þú viðir gert hafa
Heimila	vist með sér ef hann vildi. Hrafn kvað
Heimila	hest sinn. Ríður hann nú við hrynjandi
Heimilast	að veita öðrum þitt en eigi mitt."
Heimill:	"Til þess er eg nú búinn. Er mér og
Heimill:	er greiði við ykkur sá sem þér viljið
Heimill	og hestaskipti og allur annar
Heimill	vor tilbeini" - þá var Hákon
Heimill	að hafa af sífti sem þér viljið og
Heimill	að þú sért hér ef þú vilt það því að
Heimill	á Glúmur að lofa það, en ekki er það
Heimill:	Eg hefi hey ærin. Eru hér og nógar
Heimill	eiga að gera af sífti er mér sýnist."
Heimill	og kann vera að þið séuð hér eigi verr
Heimill	róm í himinríki sem standa mættu í
Heimill	til þess að þér megi þá betur líka við
Heimill	þar að vera. Þóðvar kvaðst það mundu
Heimill	þeim er fara vilja með mér. Hinum er
Heimul	mun þér gisting." Þorgeir segir:
Heimul	mun þér gisting hér vera nálangt. En
Heimull	er matur þeim er hafa þurfa."
Heimult	er þér mágur að eg gangi með máli þessu
Heimult	gert. Gangu þeir Egill inn með helming
Heimult	og allan farargreiða þann er þér viljið
Heimult	skyldi það. Gangu þeir Egill þá inn í
Heimult	um systur Þóris fóstbróður síns að hann
Heimult	þó að Þórólfr vildi fleiri menn hafa

Often, however, it is evident that some word pattern or collocation is at stake, and that kind of information should be a part of the dictionary. A few examples of such patterns are shown in (8) and (9).

(8)

alldjarflega ao; alldjarflega (8)

Þórb 13:2043	Síðan ganga þeir saman og berjast
Eyrb 62:616	menn til varnar og barðist sjálfur
ÞorSH 2:2062	koma milli klæða sér og barðist þá
Gullþ 15:1136	manna. Þá kom Gunnar að og barðist
Grett 2:955	fram þá hélt ekki við. Þórir barðist
Þórb 11:2039	öðrum stór sár. Steingrímur barðist
Grett 4:957	af skipunum. Víkingar lögðu að
Gullþ 12:1131	við túngarð í Rauðsdal og börðust

alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega

. Fann Þórbur það brátt að Stóri var
. Hann gekk mjög út á virkið er hann
. Lítu síðar heyrðu þeir mælt í
. Þórir það sína menn hlífa sér og gæta
og féll á skipi sínu með mikilli
og varð fjögurra manna bani. Skiptist
og þóttu hinir komnir í stilli.
því að Raubur var frækn maður. En svo

(9)

alldrengilega ao; alldrengilega (7); alldrengilega (1)

HallM 4:1199	að og sækja að þeim en þeir verjast
Kjaln 16:1457	Kolfinnur hjó hart og tíðum og sótti
Eyrb 58:613	sköfustu en þeir Óspakur vörðust
JökBú 2:1463	þeir báðir að Eitli en hann varðist
Gullþ 20:1142	hvortveggi mjög sár. Þorgils varðist
Gullþ 8:1127	Hyrningur Hallsson kom heim og segir
Gullþ 13:1132	hjalla einum. Varðist Þorbjörn þaðan
Grett 82:1079	sótu að fast en Illugi varði þá báða

alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega

. En þó kom þar sem mælt er að ekki má
. Hjóst þá allmjög skjöldur Búa. En er
. Varð þeim þó handfátt og urðu þeir
. Þar kom um síðir að þeir gátu drepið
en féll þó fyrir þeim Gunnari og Grími.
frá rtr þeirra Þóris. Hallur
með stokkinum því að vopn hans höfðu
. En Grettir var með öllu óvígur bæði

In (8) we see that the adverb *alldjarflega* and the verb *berjast* almost always go together; and (9) shows that the same goes for the adverb *alldrengilega* and the verb *verjast*. This is not mentioned in any existing dictionary of Old Icelandic or Old Norse, as far as I know, and

it would probably not be fair to claim that it should. On the contrary; I think that we must have access to a concordance to see this. However, there can be no doubt that this is not a coincidence, and information on this should be found in a dictionary. It may be noted in this connection that the concordance to the Sagas has already been used in a published dictionary; this is Jón Hilmar Jónsson's (1994) *Orðastaður*, which is a dictionary of collocations.

Examples of this kind are numerous in the concordance. It is true, of course, that one can sometimes infer something of this kind in the citations in the published dictionaries. The trouble is, however, that it is difficult to know what these examples really show; how typical they are. It is not clear on which principles the excerption has been based, and which of the excerpted examples actually appear in the dictionary.

In (10) we see another similar example. The word under consideration is the conjunction *uns*.

(10)

<i>uns</i> st; <i>uns</i> (36)			
Fljót	18:709	Arneiðarstaði í kvöld og þóð mún þar	<i>uns</i>
Njála	142:306	Vil eg eiga rétting allra orða minna	<i>uns</i>
Grett	52:1033	enginn varðhöld á sér. Hann fór nú	<i>uns</i>
Þórð	9:2032	og ríður undan sem mest getur hann	<i>uns</i>
Eirík	3:522	ríða skyldu. Einar fór suður aftur	<i>uns</i>
Eirík	4:524	sagt. Þýr Þorbjörn skip sitt og fer	<i>uns</i>
Vopnf	2:1988	bolði mikla á hávu skafti. Hann fer	<i>uns</i>
BandK	4:32	þá fer Oddur með hinn tuttuganda mann	<i>uns</i>
GunKe	6:1153	þau skildu. Gekk Gunnar þá leið sína	<i>uns</i>
Þorhv	6:2056	yfir Lagarfljót og upp með fljótinu	<i>uns</i>
Fljót	21:720	Mývatns og fara svo vestur til sveita	<i>uns</i>
Vatn	40:1891	taka við honum og fór hann um sveitir	<i>uns</i>
Fljót	18:711	vermans. Nú dregur þó saman með þeim	<i>uns</i>
Svarf	24:1817	Karl Karisson vex upp með móður sinni	<i>uns</i>
GíslS	24:879	fer hann í Geirþjófsfjörð og er þar	<i>uns</i>
Grett	30:1001	annan mann. Þeir voru fimm saman, ríðu	<i>uns</i>
GíslS	3:854	er til komu. En þau Gíslí fara	<i>uns</i>
Njála	6:132	Hún bjó sig skjótt og síðan ríða þau	<i>uns</i>
Grett	90:1092	með góðu foruneyti og fóru alla leið	<i>uns</i>
Fljót	23:723	þeir ríða nú út eftir héraði	<i>uns</i>
Hrafn	14:1414	undan. Ríða þeir Sámur þá allt að einu	<i>uns</i>
Þórð	9:2033	ríða eftir þeim. Ríða þeir leið sína	<i>uns</i>
Fljót	13:697	hvorutveggju með hinu vestra landinu	<i>uns</i>
GíslS	7:857	þeir fara dag þann og um nóttina	<i>uns</i>
GíslS	29:887	þessu máli. Sveinarnir fara nú	<i>uns</i>
BjHft	30:114	þær er upp liggja í dalinn frá Völlum	<i>uns</i>
Korm	7:1476	Þetta ríð taka þeir, fara	<i>uns</i>
Vopnf	18:2004	Fara þeir nú sem þeir mega mest	<i>uns</i>
Fljót	18:708	til Óss og upp í heiði til Gínguskarðs	<i>uns</i>
Harð	36:1290	sté á skip reiður mjög og fara nú	<i>uns</i>
GíslS	11:862	Hann ljær þeim hestana og ríða þeir	<i>uns</i>
Fljót	26:727	nú ofan að ánni og upp eftir nesinu	<i>uns</i>
Njála	131:287	í braut og fyrir norðan jökul og svo	<i>uns</i>
Heið	22:1373	og styðjið hana á baki og ríðið svo	<i>uns</i>
Grett	19:985	hans alla hluti vel. Leið nú svo fram	<i>uns</i>
Harð	38:1292	ofan Indriðastíg hjá Þyrli og beið þar	<i>uns</i>
		eg kem á morgun, nema eg finni eigi	<i>uns</i>
		eg kem máli mínu til rétra laga.	<i>uns</i>
		er hann kom í Vatnsfjarðardal og fór	<i>uns</i>
		hann kemur heim til Þverár og unnr	<i>uns</i>
		hann kemur heim. Nokkuru síðar	<i>uns</i>
		hann kemur í Brattahlíð. Tekur Eiríkur	<i>uns</i>
		hann kemur í sauðabás og rekur þaðan	<i>uns</i>
		hann kemur skammt frá bæ þeim er	<i>uns</i>
		hann kom heim. Urðu hans menn honum	<i>uns</i>
		hann kom í Atlavík snemma morguns.	<i>uns</i>
		hann kom til Hjelgafells, því að hann	<i>uns</i>
		hann kom til Ottars og það hann	<i>uns</i>
		hann kom upp í ásinu. Þá hallar burt af	<i>uns</i>
		hann var tólf vetra gamall og töluðu	<i>uns</i>
		haustar. Þá fer hann enn til Þorkels	<i>uns</i>
		þar til er þeir komu á	<i>uns</i>
		þau koma í Friðarey til Styrkás og	<i>uns</i>
		þau koma til þings. Unnur gekk	<i>uns</i>
		þau komu í Noreg. Tóku	<i>uns</i>
		þeir koma á Arneiðarstaði. Er gengið	<i>uns</i>
		þeir koma á heiðarbrúna. Sá hann þá	<i>uns</i>
		þeir koma á Miklabæ í Óslandshlíð.	<i>uns</i>
		þeir koma fyrir nes það er gengur	<i>uns</i>
		þeir koma í dal þann er gengur upp af	<i>uns</i>
		þeir koma í Geirþjófsfjörð og liggja	<i>uns</i>
		þeir koma í Hólmsland og töluðu þá	<i>uns</i>
		þeir koma í Saurbæ. Tók Bersi vel við	<i>uns</i>
		þeir koma mjög svo í öndverðan dalinn.	<i>uns</i>
		þeir koma svo upp að ein brekka var	<i>uns</i>
		þeir koma þar að er Geir flaut dauður	<i>uns</i>
		þeir komu á Mosvöllu og þaðan inn	<i>uns</i>
		þeir komu á Völvöllu. Þá var af	<i>uns</i>
		þeir komu til Svínafells. Flosi sendi	<i>uns</i>
		þér komið fram að Faxalæk. Hann fellur	<i>uns</i>
		Þorfinns var heim von.	<i>uns</i>
		Þorsteinn fór til blóthúss síns sem	<i>uns</i>

When we look at the examples we see that in a great majority of them, or 31 out of 36, the verb *koma* follows *uns*. Note that *uns* is a temporal conjunction, and it is impossible to deduce from its meaning that it has closer ties to *koma* than to any other verb. Another temporal conjunction in Old Icelandic, *þar til*, for instance, does not have any comparable ties to any particular verb.

I could add hundreds of examples similar to those that I have mentioned. In some of the cases, information on word combinations or ties between words clearly should be found in a dictionary; in other cases, this may not be so clear. The point is, however, that the concordance gives us a unique overview of such patterns, and makes it possible to see things that simply could not be seen without such a tool.

3 Conclusion

In the first part of this paper, I described the making and the structure of a forthcoming concordance to the *Íslendinga sögur*, whereas in the second part, I talked about its potential use in dictionary making, as I see it. During the last few years, Guðrún Ingólfssdóttir, Bergljót Kristjánsdóttir and others have actually been using the concordance as a basis for a new dictionary of the Sagas. In another paper in this volume, Guðrún gives a short description of their work.

References

- Ásgeir Blöndal Magnússon 1989: *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Ásta Svavarsdóttir/Guðrún Kvaran/Jón Hilmar Jónsson/Kristín Bjarnadóttir 1993: *Sýnihefti sagnorðabókar*. Rannsóknar- og fræðslurit 3. Reykjavík: Orðabók Háskólans.
- Baldur Jónsson 1978: *Orðstöðulykill að Hreiðrinu*. [Fjölrít.]
- Biblúlykill*. 1994. Orðalyklar að Biblíunni 1981. Reykjavík: Biblíulykilsnefnd/Hið íslenska Biblíufélag.
- Eiríkur Rögnvaldsson 1990: Orðstöðulykill Íslendinga sagna. *Skáldskaparmál* 1, 54–61.
- Eiríkur Rögnvaldsson/Bergljót Kristjánsdóttir/Guðrún Ingólfssdóttir/Örnólfur Thorsson 1995: *Orðstöðulykill Íslendinga sagna*. Forthcoming on CD-ROM.
- Fritzner, Johan 1954: *Ordbog over Det gamle norske Sprog*. Nytt uforandret opptrykk av. 2. utgave (1883–1896) med et bind tillegg og rettelser redigert av Didrik Arup Seip og Trygve Knudsen. Oslo: Trygve Juul Møller.
- Íslendinga sögur*. 1985–86. Editors: Bragi Halldórsson, Jón Torfason, Sverrir Tómasson, Örnólfur Thorsson. Reykjavík: Svart á hvítu.
- Jón Hilmar Jónsson 1994: *Orðastaður*. Orðabók um íslenska málnotkun. Reykjavík: Mál og menning.
- Jörgen Pind/Friðrik Magnússon/Stefán Briem 1991: *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.

Funktionsverber i det danske sprog

This paper gives an outline of a research project concerning Danish support verb constructions (German: Funktionsverbgefüge), a phenomenon previously confined to the periphery of Danish linguistics. The following issues are addressed: problems in defining support verb constructions, the inventory of a list of Danish support verbs, their treatment in monolingual and bilingual dictionaries, and an ongoing study of three Danish support verbs (*foretage* 'undertake/make/do', *lave* 'make/do', *gøre* 'do/make'). The paper suggests a grading of support verb constructions according to scope for modification (unrestrictedness/idiomaticity) and it advocates a semantic approach. A semantic classification of verbs can constitute the basis for predicting which verbal nouns, i.e. the substantival counterparts of verbs, may combine with the different support verbs.

1 Introduktion til begrebet funktionsverber

En kort introduktion til **funktionsverber** og **funktionsverbumbonstruktioner** (FVK) kan gives ved hjælp af følgende eksempler: *føre forhandlinger* — *stå til ansvar* — *sætte i bevægelse* — *bringe i system* — *begå en letsindighed* — *foretage gensplejsning*. Verberne i disse konstruktioner har som fuldverber en meget konkret betydning, f.eks. *føre* 'bringe af sted, lede, stå i spidsen for'; *stå* 'have en opret stilling'; *sætte* 'anbringe noget'; *bringe* 'levere, overdrage noget'. Med verberne *begå* og *foretage* forholder det sig anderledes, idet de har en mere uspecificeret betydning: 'udøve/udføre en handling'.

I de ovennævnte eksempler kaldes de anvendte verber for funktionsverber, og de adskiller sig fra de tilsvarende fuldverber ved at være ret indholdstomme, et kendetegn, der også gælder for hjælpeverber. Forskellen på et fuldverbumbes og et funktionsverbumbes betydning kan illustreres med de følgende eksempler, hvor a-eksemplerne viser fuldverbernes betydning, mens b-eksemplerne illustrerer samme verber anvendt som funktionsverber. Her spiller fuldverbets betydning ikke nogen rolle, idet hverken *komme* eller *give* bidrager til sætningens betydning med den konkrete betydning, som fuldverbet har i a-eksemplet.

- (1) a Han *kom til et møde* på skolen
b Det *kom til et møde* mellem dem
- (2) a Hun *gav sit bidrag* til formanden for foreningen
b Hun *gav sit bidrag* til, at festen lykkedes

Et andet lighedspunkt til hjælpeverberne findes i forbindelse med oversættelse af funktionsverber mellem forskellige sprog, hvor der, ligesom det hyppigt er tilfældet med hjælpeverberne, ikke findes én enkelt fremgangsmåde. De følgende eksempler kan illustrere to problemer med oversættelse af funktionsverber: (1) det, der er en FVK på det ene sprog, skal af og til gengives med et leksem på et andet sprog (jf. (3a)), (2) en FVK på det ene sprog skal muligvis også på et andet sprog gengives med en FVK, men det er ikke altid det verbum, der anses som den direkte ækvivalent for funktionsverbet, der er det idiomatisk korrekte (jf. (3b)):

- | | | | |
|-------|--|---------------|--------------------------------|
| (3) a | foretage udgravninger
lave fusk | eng.:
ty.: | excavate
pfuschen |
| b | foretage afstemning
gøre indvendinger | eng.:
ty.: | take a vote
Einwurf erheben |

Specielt i den tyske sprogvidenskab, men også i den engelske og franske, har funktionsverber været et emne, man har beskæftiget sig med gennem længere tid. Den tyske tradition går tilbage til von Polenz (1963), der indførte begrebet 'Funktionsverb', mens Engelen (1968) skabte begrebet 'Funktionsverbgefüge'.

2 Typer af funktionsverber/FVK'er

Der er enighed om, at et funktionsverbum er et verbum, der i forhold til dets brug som fuldverbum har et blegnet semantisk indhold, men derudover er der ikke enighed om definitionen på en FVK. Ser man på de indledningsvis nævnte eksempler og deres mulige omskrivninger,

føre forhandlinger	forhandle
stå til ansvar	være ansvarlig
sætte i bevægelse	bevæge
bringe i system	systematisere
begå en letsindighed	være letsindig
foretage gensplejsning	gensplejse

så kan dette illustrere et hyppigt anvendt definitions-kriterium, nemlig at det er muligt at parafasere konstruktionen med et simpleks-/fuldverbum eller et kopulaverbum + adjektiv, hvor simpleksverbet eller adjektivet har fælles rod med det verbalsubstantiv, som optræder i FVK'en.

Et andet almindeligt anvendt kriterium er, at der skal være samme reference for den prædikative gruppe og FVK'en, med andre ord samme subjekt for det finitte verbum og for verbalsubstantivet, hvilket kan ses ved følgende tests:

han foretager en rejse	= hans rejse	han rejser
han begår en letsindighed	= hans letsindighed	han er letsindig
han står til ansvar	= hans ansvar	han er ansvarlig
han er til hjælp	= hans hjælp	han hjælper

MEN:

han sætter båden i bevægelse	≠ hans bevægelse	?han bevæger båden
han bringer hende til vanvid	≠ hans vanvid	≠ han er vanvittig

De sidste FVK'er kan ikke parafaseres med simpleksverbet, idet de indeholder et kausativt element, som de tilsvarende simpleksverber ikke har. Der findes således ikke en fælles reference mellem subjektet for det finitte verbum og verbalsubstantivet (*han bringer hende til vanvid* = *hendes vanvid* — *hun bliver vanvittig/han gør hende vanvittig*).

En FVK denoterer altså den betydning, der ligger i den rod, som verbalsubstantivet/adjektivet er afledt af (*han fører forhandlinger* = *han forhandler*) og kan derudover indeholde et kausativt element (*han bringer hende til vanvid* = *han gør hende vanvittig*),

et durativt aspekt (*markerne står i blomst = markerne blomstrer*) eller et inkoativt aspekt (*hesten sætter i trav = hesten begynder at trave*).

Hverken den definition, der bygger på, at en FVK kan gengives vha. en parafrase, der indeholder simpleksverbet, eller den definition, der kræver et og samme subjekt for det finitte verbum og for verbalsubstantivet, er således dækkende for de konstruktioner, som man intuitivt regner til FVK'erne. Konstruktionerne er altså, til trods for at de har fælles træk, for forskellige til at blive omfattet af en og samme definition.

Dette er måske grunden til, at FVK'erne ofte bliver defineret vha. lister over funktionsverber og derpå inddelt efter syntaktisk type. På dansk er der primært tale om to typer: funktionsverbum + NP og funktionsverbum + præposition + NP (af og til nævnes tre typer, hvor den tredje er NP + funktionsverbum af typen *der består/opstår enighed*). Som vist ovenfor, så kan man ikke konsekvent sige, at funktionsverbum + NP kan parafraseres med simpleksverbet og funktionsverbum + præposition + NP med et andet verbum + simpleksverbet — selvom begge dele ofte er tilfældet.

Efter min mening bør man i højere grad inddele FVK'erne på grundlag af både deres betydning og deres idiomacitet og sproglige fasthed. Ser man på følgende FVK'er:

foretage bedømmelse: hun foretager bedømmelse af eksaminanden
stille til rådighed: hun stiller sin bil til rådighed
bringe til vanvid: hun bringer ham til vanvid

kan man konstatere, at verbalsubstantivet i FVK'erne i forskellig grad lader sig modificere mht. nøgen, bestemt og ubestemt form og mht. tilføjelse af attributter:

hun foretager (en) bedømmelse/bedømmelsen/bedømmelser af eksaminanden
 hun foretager en overfladisk bedømmelse/grundige bedømmelser af eksaminanden
 hun stiller sin bil til *en rådighed/*rådheden
 *hun stiller sin bil til hyppig rådighed
 hun stiller sin bil til hans rådighed
 hun bringer ham til *et vanvid/*til vanvidet
 *hun bringer ham til hyppig vanvid

På grundlag af disse eksempler kunne man skelne mellem modificerbare, i mindre grad modificerbare og ikke modificerbare FVK'er. Derudover kan man se på, hvordan FVK'erne kan parafraseres, om det kan ske udelukkende med simpleksverbet, eller om andre konstruktioner er nødvendige for at gengive en eventuel kausativitet, durativitet eller inkoativitet.

3 Danske funktionsverber

Denne undersøgelse drejer sig om at finde frem til en opdeling eller graduering af potentielle FVK'er. I sidste ende vil nok ikke alle potentielle FVK'er blive betegnet som FVK'er eller typiske FVK'er, idet nogle sandsynligvis skal regnes til idiomere.

Til dette formål er der opstillet en liste over potentielle danske funktionsverber, dels ud fra lister over funktionsverber fra andre sprog og dels efter gennemsyn af verberne i NDO (*Nudansk Ordbog*)¹⁰. Der er i listen ikke angivet, hvorvidt de medtagne funktionsverber optræder i konstruktioner bestående af funktionsverbum + NP og/eller i konstruktioner bestående af funktionsverbum + præposition + NP:

affatte (-)	frembringe	lægge	stå
afgive	fremføre	løbe	sætte
afgå (-)	frem sætte	nedlægge	søge
afholde	fælde (-)	nyde	tage
aflægge	fæste (-)	nære	tegne
afsigte (-)	føle	nå	træde (-)
anbringe (-)	føre	opføre (-)	træffe
anstille	få	opnå	uddele (-)
bedrive (-)	geråde (-)	overbringe (-)	uddrage
begå	give	overføre (-)	udfolde
blive	gøre	pleje (-)	udføre
bringe	gå	præstere (-)	udstede
bære	have	rejse	udstråle
danne	holde	rende (-)	udsætte
drage	indføre	samle (-)	udtrykke
drive	indgive (-)	skabe	udvise
erlægge (-)	indlede	skride (-)	underkaste
falde	indrykke (-)	slutte	vente
fange (-)	kaste	slå	vinde
finde	knytte (-)	sprede (-)	være
foranstalte (-)	komme	spænde (-)	yde
foretage	lave	starte (-)	øve
forrette	lide	stikke	åbne
forøve (-)	ligge	stille	

Med fed er de funktionsverber markeret, der er de mest frekvente i det undersøgte materiale, nemlig to korporaer (det danske genteknologiske korpus BIODA, og et mindre skønlitterært korpus (Holmboe 1988)), NDO og Dansk sprogbrug, for så vidt som funktionsverberne her optræder i FVK'er. Hyppighed er arbitrært sat til 20 eller flere forekomster i et af korporaene eller NDO/Dansk sprogbrug. Med (-) angives, at verberne ikke optræder som funktionsverber hverken i BIODA, det skønlitterære korpus eller NDO/Dansk sprogbrug.

Disse verber kan inddeles i grupper efter forskellige kriterier, f.eks. syntaktisk efter konstruktionstype i funktionsverbum + NP og funktionsverbum + præposition + NP. De kan også inddeles efter andre kriterier, f.eks. om de kan parafraseres med fuldverbet eller med et andet verbum + fuldverbet. Grundlæggende er der den forskel på de to grupper af FVK'er, der tegner sig efter parafraseringskriteriet, at FVK'en i den første type i forhold til simpleksverbet kan indgå i eller åbne for nogle syntaktiske konstruktioner, som simpleksverbet ikke kan, mens FVK'en i den anden type muliggør at medtage nogle semantiske træk, nemlig kausativitet, inkoativitet og durativitet, som simpleksverbet ikke indeholder. I nærværende undersøgelse anses parafraseringsmåden sammen med FVK'ernes modifikationsmuligheder som de relevante og hensigtsmæssige kriterier til en inddeling.

Den første gruppe, der er blevet undersøgt, består af funktionsverber, der optræder i FVK'er, der uden videre kan parafraseres vha. simpleksverbet. Disse funktionsverber kan igen inddeles efter deres produktivitet og efter forholdet mellem funktionsverbets betydning som simpleks-/fuldverbum og som funktionsverbum. Det er tidligere blevet nævnt, at et

verbum som funktionsverbum har en blegnet betydning i forhold til samme verbum som fuldverbum. Også mht. dette forhold kan der gradueres, idet der ved nogle funktionsverber er tale om, at deres fuldverbumbetydning i høj grad er til stede, hvilket f.eks. gælder for *føle* og *lide*; også *finde* og *søge* indeholder i mange FVK'er en del af deres fuldverbumbetydning, mens andre funktionsverber ser ud til ikke at have en fuldverbumbetydning.

Indtil videre er fortrinsvis en lille gruppe blevet undersøgt, der skiller sig ud på følgende måde: på den ene side er der tale om verber, der ikke i sig selv kan siges at forekomme som fuldverbum med en betydning, der adskiller sig fra funktionsverbets, og på den anden side gælder for næsten alle verber i denne kategori, at de hører til de produktive sproglige fænomener, som tilsyneladende ikke har nogen idiomatisk karakter. Det drejer sig om *foretage*, *lave* og *gøre*. For disse verber gælder begge de tidligere nævnte kriterier om parafraseringsmuligheden vha. simpleksverbet alene og også den fælles reference af subjekterne til det finitte verbum og verbalsubstantivet.

Da FVK'erne med disse verber kan parafraseres vha. simpleksverbet, uden at der sker en betydningsændring, kan disse FVK'er betegnes som typiske strækformer: simpleksverbets semantiske indhold og morfologiske funktion strækkes ud og fordeles på to sproglige størrelser, hvilket grafisk kan illustreres på følgende måde:

<p>Simpleksverbet:</p> <p>bærer af de morfologiske kategorier person, numerus, tempus, modus, diatese</p> <p>+</p> <p>bærer af det semantiske indhold</p>	<p>Funktionsverbet:</p> <p>bærer af de morfologiske kategorier person, numerus, tempus, modus diatese</p> <hr/> <p>Verbalsubstantivet:</p> <p>bærer af det semantiske indhold</p>
---	---

Anvendelsen af de FVK'er, som disse verber indgår i, har til formål at forhindre tunge konstruktioner og åbne for nogle syntaktiske muligheder, som simpleksverbet ikke har. F.eks. kan konstruktioner „forkortes“ bl.a. gennem udeladelse af led:

Hun undersøger noget på en indgående måde.

→ Hun foretager indgående undersøgelser.

De stemte, og resultatet af afstemningen afgjorde deres fremtid.

→ De foretog den afgørende afstemning.

4 Ordbogsdefinitioner

I leksikografien behandles funktionsverber ikke selvstændigt, hvilket kan være anbragt, da det langt fra gælder for alle FVK'er, at de er rent idiomatiske og som sådanne bør medtages som idiomatisk sprogbrug. Da der i en del tilfælde er tale om regulære og produktive fænomener, bør det overvejes at give dem en anden behandling i ordbøgerne.

I det følgende gennemgås kort behandlingen af de tre aktuelle verber i nogle ordbøger, der skønnes at burde kunne give oplysninger om de FVK'er, disse verber optræder i. Først ses der på tre ensprogede ordbøger, nemlig ODS (*Ordbog over det danske sprog*) som den

største danske ordbog, Dansk sprogbrug, en stil- og konstruktionsordbog, der netop burde give oplysninger om verbernes konstruktionsmønstre, og NDO, derpå på to tosprogede ordbøger, en dansk-tysk (Munksgaard) og en dansk-engelsk (V & B).

4.1 ODS

I ODS fylder artiklen med *foretage* to spalter og indeholder fire forskellige betydninger, hvoraf de tre må anses som forældede (jf. NDO), mens den fjerde er den væsentlige i denne sammenhæng: „(nu kun m. abstr. obj. og hyppigst m.h.t. hvad man udfører uden (større personlig) anstrengelse, haardt arbejde osv.) være beskæftiget med; være i færd med; udføre; gøre“. Ud over citater fra litterære værker angives som eksempler: *foretage et eksperiment, forsøg, en sammenligning, udgravning, undersøgelse, øvelse* o.a.

lave behandles med angivelse af syv betydninger på godt otte spalter, hvoraf den første („bringe i en vis orden, i rette stand olgn. (jf. ogs. bet. 2); indrette; arrangere (1); ordne; (til)berede; forberede“) beslaglægger fire spalter. Flere af de andre betydninger synes at dække over funktionsverbumbetydningen, men specielt ved den fjerde er der tale om *lave* som funktionsverbum: „i al alm.: beskæftige sig med ell. udføre noget“. Følgende konstruktioner angives: *lave et overslag, udkast, bedrageri, underslæb, fejl*.

gøre tildeles i alt godt 36 spalter med angivelse af 13 betydninger, der deles i to overordnede betydninger: „bevirke, at noget bliver til“ og „m.h.t. handling“. I den sidste gruppe findes den væsentlige i forbindelse med funktionsverber: „udføre (en handling); forrette; foretage (sig); øve“. Specielt undergruppen 9.7 er interessant („m. andet obj., der betegner en handling, ofte blot som omskrivning af det til vedkommende subst. svarende verbum ...“). Der nævnes mange konstruktioner, f.eks. *gøre adskillelse, angreb, anmærkninger, anskrig, anstrengelser, bebrejdelse(r)* o. m. a.

4.2 Dansk sprogbrug

Under *foretage* listes nogle eksempler (bl.a. *foretage en afstemning, et angreb, en sammenligning, en udgravning*). Ved *lave* opføres fem betydninger, hvoraf den tredje dækker funktionsverbumbetydningen: ‘være beskæftiget med, bestille’, dog uden angivelse af en eneste FVK (under betydningen ‘frembringe’ angives *lave vrøvl*, der evt. kan anses som en FVK). Også her tildeles *gøre* mest spalteplass, men det eneste eksempel på en FVK er *gøre et arbejde*.

4.3 NDO

De tre verber får alle en ret kort beskrivelse: ved *foretage* angives ud over frasen *hvad foretager du dig?* et eksempel (*foretage et eksperiment*); *lave* behandles lidt udførligere med tre angivne betydninger, hvoraf den tredje dækker funktionsverbumbetydningen (‘bestille, arbejde med’), men der gives intet eksempel på en FVK; og *gøre*, der også her får den længste beskrivelse, beskrives bl.a. med ‘udføre, udrette’, og også her er der ingen eksempler på en FVK.

I de ensprogede ordbøger er der altså ikke megen hjælp at hente, når man slår op under verberne, til gengæld gives der fyldigere oplysninger under de fleste af de verbalsubstantiver, der kan findes som selvstændige opslagsord.

4.4 Munksgaard

For *foretage* angives der to grupper af ækvivalenter: „1. unternehmen, machen (fx rejse, spadseretur, udflugt; en ekskursion), 2. (gennemføre, iværksætte, lave) vornehmen, durchführen, anstellen, (især mil) unternehmen (fx et angreb, fremstød)“. Desuden indeholder artiklen „udtryk“, der består af *foretage beregninger* (= Berechnungen anstellen), *foretage et eksperiment* (= ein Experiment durchführen), *foretage de nødvendige skridt* (= die nötigen Schritte unternehmen/tun/einleiten), *foretage åbningen af udstillingen* (= die Eröffnung der Ausstellung vornehmen, die Ausstellung eröffnen). Hvorledes „udtrykkene“ adskiller sig fra de to først nævnte grupper er uklart, idet der ved næsten alle „udtryk“ anvendes de samme verber som under 1. og 2. Det er dog bemærkelsesværdigt, at artiklen afsluttes med: „se substantiverne“.

Under *lave* sker der en betydningsopdeling på følgende måde: „(skabe, fremstille, udføre) machen (fx ure; guld, undersøgelse; penge), (i forb. med adv) [. . .], (mere alment, foretage sig, især) tun, treiben, schaffen, (reparere) [. . .], (opdigte) [. . .], (producere) [. . .]“. Derefter følger nogle „udtryk“, hvoraf *lave sit arbejde* og *lave brok* evt. kan opfattes som FVK'er.

Ved *gøre* deles betydningen op i tre punkter, hvoraf to har underpunkter indeholdende *gøre* som funktionsverbum: „1. machen a) med subst som obj (Ex *gøre en bemærkning* = eine Bemerkung machen; *gøre store øjne* = große Augen machen) [. . .] 2. tun a) med verbalsubstantiv som obj (Ex *gøre et spring, et træk* = einen Sprung, einen Zug tun) se under substantiverne [. . .] 3. forvolde skade [. . .]“. Derefter nævnes mange „udtryk“, men ingen med *gøre* som funktionsverbum.

4.5 V & B

Ved *foretage* angives følgende ækvivalenter: *undertake, carry out, perform, make, do*. Derefter følger mange „udtryk“, hvoraf næsten alle er typiske FVK'er, som f.eks. *foretage af-skrivninger, afstemning, en omgående bevægelse, forandringer, forsøg*.

lave optræder med følgende betydninger: „1. (forfærdige, fremstille) make [. . .] 2. (tilberede) prepare, make [. . .] 3. (reparere) mend, repair [. . .] 4. (opdigte) make up [. . .] 5. (begå) do, commit [. . .] 6. (forrette nødtørft) [. . .]“. Derefter følger mange „udtryk“, men ingen FVK'er.

Ved *gøre* angives syv betydninger: „1. (udføre, besørge) do [. . .] 2. (volde, øve) do [. . .] 3. (besøge som turist) do [. . .] 4. (frembringe, fremkalde) make [. . .] 5. (bringe i en vis tilstand) make, render [. . .] 6. (foretage, fx en bevægelse, en rejse) make (fx a gesture, a journey), 7. (tilbagelægge) do [. . .] 8. (som gentagelse af et vb) do [. . .] 9. (mar.: nå) make [. . .]“. Derefter følger først „udtryk“, hvori *gøre* forbindes med substantiver, og her optræder en del FVK'er: *gøre en aftale* (= agree on, make an agreement), *gøre én bebrejdelser* (= reproach sby), *gøre indvendinger* („se indvende, indvending“), *gøre en opfindelse* (= make an invention), *gøre ham et spørgsmål* (= ask him a question) osv.

Generelt må man sige, at disse verber som funktionsverber ikke behandles på en systematisk måde, og det er ikke angivet, at de optræder i produktive konstruktioner. Også ved de funktionsverber, der kan angive kausativitet, durativitet eller inkoativitet, findes der ifølge en kursorisk undersøgelse af ordbøgerne åbenbart ingen systematisk fremstilling.

5 Undersøgelse af de tre funktionsverber:

foretage, lave og gøre

Ved undersøgelsen af de tre funktionsverber *foretage*, *lave* og *gøre* kan det konstateres, at der er forskel på, hvilke verbalsubstantiver, der kan konstrueres med dem. Nogle verbalsubstantiver kan konstrueres med alle tre og nogle kun med to eller et af dem:

foretage/lave/gøre	{	en beregning forberedelser et forsøg en vandring						
foretage/lave	{	en aftale en beskrivelse en blanding udsætning	lave/gøre	{	bemærkninger opdagelser vold undtagelse			
foretage	{	dåb vedligeholdelse åbning	lave	{	en debat et forbud en forstørrelse	gøre	{	en antagelse krav nytte

Hypotesen er, at verbalsubstantivets betydning afgør, hvilke funktionsverber det kan indgå en FVK med. En typologisering af verberne efter deres betydning (som tillige er en typologisering af deres afledninger) må kunne anvendes i forbindelse med udsagn om verbalsubstantivernes kombinationsmuligheder med de enkelte funktionsverber, så man evt. kan nå frem til udsagn, som f.eks. at afledninger af verber, der betegner en tilstand, ikke kan konstrueres i en FVK med *foretage*, *gøre* eller *lave*. Almindeligvis inddeles verber i handlings-, tilstands- og procesverber, men denne inddeling er ikke tilstrækkelig til dette formål, da en tredeling er for unuanceret og indebærer nogle uløste inddelingsproblemer (som f.eks. behandlingen af verber, der betegner egenskaber eller relationer). Man kan f.eks. ikke generelt sige, at alle handlingsverber kan kombineres med de tre verber, som vist ovenfor. Hertil kræves en finere inddeling.

Med inspiration fra Levin (1993) og især Ballmer og Brennenstuhls inddeling af de tyske verber (1986) har jeg forsøgt at kategorisere i første omgang ca. 1000 danske verber og har foreløbig kunnet konstatere følgende forskelle mellem de tre verber:

De handlinger/situationer, som ikke kan udtrykkes vha. konstruktioner med de tre verber, er næsten alle handlinger, som subjektet ikke kan kontrollere. Der er nogle undtagelser, hvor handlinger, der normalt ikke kan kontrolleres af subjektet, tilsyneladende kan frembringes kunstigt. I disse tilfælde er en kombination af det pågældende verbs afledning mulig med *lave* eller *gøre*, men ikke med *foretage*: *lave et bøvs*, *lave et hik*, *gøre et fund*, *gøre sig håb*, *gøre en opdagelse*, *gøre en opfindelse* (*lave en opfindelse* indebærer fremstillingen/produktionen af selve opfindelsen). Dette kan også belyses med eksempler på verbalsubstantiver afledt af verberne *finde ngt*, *opdage ngt*, *opfinde ngt* og *halshugge ngn*.

Følgende acceptable/uacceptable FVK'er indeholdende *gøre*, *lave* og *foretage* kan dannes:

gøre et fund	*lave et fund	*foretage et fund
gøre en opdagelse	?lave en opdagelse	*foretage en opdagelse
?gøre en opfindelse	lave en opfindelse	*foretage opfindelse
*gøre en halshugning	lave en halshugning	foretage en halshugning

De fire verbale handlinger/situationer befinder sig på en skala, der går fra [-con] (*finde noget*) til den „mest“ kontrollable handling (*halshugge ngn*). Eksemplerne tyder på, at en FVK med *gøre* gengiver en mindre bevidst og en mindre kontrolleret eller kontrollerbar handling end en FVK med *foretage*.

gøre kan altså kombineres med nogle verbalsubstantiver, der er afledt af verber, der må karakteriseres som [-con], hvilket til en vis grad også gælder for *lave*, mens *foretage* slet ikke kan optræde sammen med disse verbalsubstantiver. Kun *gøre* kan f.eks. kombineres med *opdagelse/opdagelser*, *fund*, *håb*, *bekymringer*, og man kan f.eks. ikke **foretage fejl*, men både *gøre fejl* og *lave fejl*. Også de følgende eksempler tyder på, at den handling, der denoteres af konstruktionen med *lave*, er mere kontrollerbar end den handling, der denoteres af konstruktionen med *gøre*:

Han har med vilje lavet mange fejl	↔	?Han har med vilje gjort mange fejl
Han har prøvet at undgå at lave fejl	↔	*Han har prøvet at undgå at gøre fejl

Andre eksempler viser det samme, f.eks. følgende, der denoterer en situation, som subjektet ikke har kontrol over, og hvor en parafrase kun er mulig vha. en FVK med *gøre*:

- Hun sukkede/skreg/brølede af smerte. →
 Hun gjorde et suk/et skrig/et brøl af smerte.
 *Hun lavede et suk/et skrig/et brøl af smerte.
 *Hun foretog et suk/et skrig/et brøl af smerte.

Er den samme situation bevidst fremkaldt, er mulighederne for omskrivning med en FVK lidt anderledes:

- Hun sukkede/skreg/brølede for at fange hans opmærksomhed. →
 Hun gjorde et suk/et skrig/et brøl for at fange hans opmærksomhed.
 Hun lavede et suk/et skrig/et brøl for at fange hans opmærksomhed.
 ?Hun foretog et suk/et skrig/et brøl for at fange hans opmærksomhed.

foretage må derfor siges snarere at kunne anvendes i situationer, der betegner en meget bevidst handling, mens *lave* har en mellemposition og *gøre* hyppigt anvendes i ikke-agenskontrollerede situationer.

Med hensyn til artikelbrug forholder de tre verber sig også „skalært“: *gøre* anvendes med verbalsubstantiver i ubestemt form eller pluralis, *lave* ligeså, dog også med verbal-substantivet i bestemt form og i enkelte tilfælde også med verbalsubstantivet i nøgen form, mens verbalsubstantivet i konstruktionerne med *foretage* kan optræde både i nøgen, bestemt og ubestemt form.

I de tilfælde, hvor det er muligt at anvende alle tre verber, er der en forskel mht. tilknytning af præpositionsled til verbalsubstantivet. Ved *foretage* og til dels ved *lave* er det muligt at tilføje et præpositionsled, der er knyttet til verbalsubstantivet med præpositionen *af*. Dette er den præposition, der ved nominalisering almindeligvis anvendes til tilknytning

af et verbums direkte objekt. I konstruktionerne med *gøre* bliver en anden præposition foretrukket.

Han bevæger armen hurtigt. →
 Han foretager en hurtig bevægelse af /med armen.
 Han laver en hurtig bevægelse ?af/med armen.
 *Han gør en bevægelse af armen.
 MEN: Han gør en bevægelse med armen.

Hun vender bilen hurtigt. →
 Hun foretager en hurtig vending af/med bilen.
 Hun laver en hurtig vending af/med bilen.
 *Hun gør en hurtig vending af bilen.
 MEN: Hun gør en hurtig vending med bilen.

I de tilfælde, hvor basisverbet kan optræde både transitivt og intransitivt, ser det ud til, at det er det intransitivt brugte verbum, der er basis for verbalsubstantivet i konstruktionen med *gøre*. Som det ses i eksemplet ovenfor med *vende*: *gøre en vending*, bliver det direkte objekt ikke tilsluttet med præpositionen *af*, som det ellers er tilfældet med de direkte objekter, men med en anden præposition, hvilket peger på, at det er det intransitive *vende*, der er basis for *vending* i denne konstruktion. Denne „intransivering“ af verbalsubstantivet ses også i de følgende eksempler:

Hun tilstår mordet.
 Hun foretager en tilståelse af mordet.
 Hun laver en tilståelse (af mordet).
 *Hun gør en tilståelse af mordet. MEN: Hun gør en tilståelse.
 Han opfinder en ny motor.
 *Han gør en opfindelse af en ny motor. MEN: Han gør en opfindelse.

Det kan tyde på, at *gøre*-FVK'erne i mindre grad end *foretage* og *lave* er synonyme med simpleksverbet, og at der sker en intransivering af FVK'en, når den konstrueres med *gøre*.

6 Afsluttende bemærkninger

Min undersøgelse er langt fra afsluttet, men jeg mener, at en systematisering af verber + deres afledninger i bestemte semantiske typer kan anvendes til at udsige noget om mulige og umulige FVK'er. Dette gælder formentlig også for de funktionsverber, som jeg ikke har behandlet endnu, f.eks. kan man formode, at det kausative *bringe i* eller *bringe til*, hvor *bringe i* betegner begyndelsen af en proces eller tilstand (*bringe i anvendelse*), og *bringe til* betegner det opnåede mål eller den opnåede tilstand eller proces (*bringe til anvendelse*), kun kan konstrueres med afledninger af verber, der denoterer en tilstand eller proces, men ikke en handling.

Oplysninger som disse må være vigtige at medtage i ordbøger, i ensprogede og måske især i de to- eller flersprogede. En ting, der ofte er væsentlig for ordbøgers mere eller mindre

uddybende behandling af forskellige sproglige fænomener, er deres fysiske begrænsning, sidetallet. Ved anvendelse af elektroniske ordbøger slipper man i høj grad for dette traditionelle pladsproblem, da det ikke er noget problem at registrere og finde oplysninger flere steder (for FVK'ers vedkommende f.eks. både under verbet og verbalsubstantivet). Men et problem i forbindelse med funktionsverber vil stadig være uløst, nemlig at forudse og at angive endnu ikke realiserede kombinationer, specielt i tosprogede ordbøger — derfor mener jeg, at en systematisk undersøgelse af verber/verbalsubstantiver og en „typologisering“ af disse er nødvendig.

Bibliografi

Korpora/opslagsværker

- Bergenholtz, H. 1992: *DK87-90*. Århus: Handelshøjskolen.
- BIODA = O. Lauridsen/T. Riiber/H. Søndergaard 1991: *BIODA — det danske genteknologiske korpus*. Handelshøjskolen, Århus.
- Dansk sprogbrug = Bruun, E. 1977: *Dansk Sprogbrug. En stil- og konstruktionsordbog*. Gyldendal: København.
- Holmboe, H. 1988 (= påbegyndt 1988): *Skønlitterært korpus* (bestående af romaner og noveller). Handelshøjskolen, Århus.
- Munksgaard = *Dansk-tysk ordbog* 1993. 1. elektroniske udgave. Munksgaard: København.
- NDO = Chr. Becker-Christensen/P. Widell (red.) 1991: *Nudansk Ordbog*. 2. elektroniske udgave. København: Politikens forlag.
- ODS = V. Dahlerup 1919-1956: *Ordbog over det danske sprog*. København: Gyldendal (28 bd.)
- V & B = *Dansk-engelsk ordbog* (Vinterberg/Bodelsen) 1992. 1. elektroniske udgave. Gyldendal: København.

Sekundærlitteratur

- Ballmer/Brennenstuhl 1986: *Deutsche Verben. Eine sprachanalytische Untersuchung des Deutschen Verbwortschatzes*. Tübingen: Gunter Narr.
- Boje, F. 1992: Funktionsverboms-konstruktioner i dansk og tysk. I: G. Engel (red.): *Korpuslingvistik i fagsprogsforskningen*. Proceedings af seminar afholdt på Hindsgavl Slot 26. og 27. nov. 1992, 104-121.
- Dyhr, M. 1980: Zur Beschreibung von Funktionsverbgefügen. I: M. Dyhr/K. Hyldgaard-Jensen/J. Olsen (udg.): *Festschrift für Gunnar Bech*. Kopenhagener Beiträge zur Germanistischen Linguistik, Sonderband 1. København, 105-121.
- Engelen, B. 1968: Zum System der Funktionsverbgefüge. I: *Wirkendes Wort*, Vol. 18, 289-303.
- Esau, H. 1976: Funktionsverbgefüge revisited. I: *Folia Linguistica*, IX-1-4, 135-160.
- Giry-Schneider, J. 1987: *Les prédicats nominaux en français*. Genève: Librairie Droz.
- Günther, H./S. Pape 1976: Funktionsverbgefüge als Problem der Beschreibung komplexer Verben in der Valenztheorie. I: H. Schuhmacher (udg.): *Untersuchungen zur Verbvalenz*. Tübingen: Gunter Narr, 92-128.
- Helbig, G. 1979: Probleme der Beschreibung von Funktionsverbgefügen im Deutschen. I: *Deutsch als Fremdsprache*, Vol. 16, Nr. 5, 273-286.

- Herrlitz, W. 1973: *Funktionsverbgefüge vom Typ ?in Erfahrung bringen?* Tübingen: Niemeyer.
- Honnef-Becker, I. 1993: Die Behandlung von Funktionsverbgefügen im Unterricht Deutsch als Fremdsprache. I: *Zielsprache Deutsch*, 24, 2, 103–107.
- Klein, W. 1968: Zur Kategorisierung der Funktionsverben. I: *Beiträge zur Linguistik und Informationsverarbeitung*, Vol. 13, 7–37.
- Levin, B./M. Rappaport 1988: Nonevent-er nominals: a Probe into Argument Structure. I: *Linguistics*, Vol. 26, 1067–1083.
- Levin, B. 1993: *English Verb Classes and Alternations, a Preliminary Examination*. Chicago: University of Chicago Press.
- Nøhr Pedersen, S. 1991: The Treatment of Support Verbs and Predicative Nouns in Danish. I: K. M. Lauridsen/O. Lauridsen (udg.): *Contrastive Linguistics*. Århus, 155–186.
- Nøhr Pedersen, S. 1993: On Valency and the Danish Support Verb 'være'. I: M. Herslund/F. Sørensen (udg.): *Papers from the CL-symposium at the Aarhus Business School, 28–30 August 1993*. *Lambda*, nr. 18, 74–98.
- Nøhr Pedersen et al. 1989: *On Support Verb Constructions in Danish*. Intern EUROTRA-rapport. København.
- Persson, I. 1975: *Das System der kausativen Funktionsverbgefüge*. Kristianssand: CWK Gleerup.
- Persson, I. 1992: Das kausative Funktionsverbgefüge (FVG) und dessen Darstellung in der Grammatik und im Wörterbuch. I: *Deutsche Sprache*, Vol. 2, 153–171.
- v. Polenz, P. 1963: Funktionsverben im heutigen Deutsch. Beiheft zu *Wirkendes Wort*, Heft 5.
- v. Polenz, P. 1987: Funktionsverben, Funktionsverbgefüge und Verwandtes. I: *ZGL (Zeitschrift für germanistische Linguistik)*, Vol. 15, 169–189.
- Richter, G. 1988: Funktionsverbgefüge in der gegenwärtigen deutschen Alltagssprache und frei gesprochenen Wissenschaftssprache. I: *Deutsch als Fremdsprache*, Vol. 25/26, 337–341.

Noen vanskelige adjektiver i leksikografisk perspektiv

The main purpose of this paper is to discuss how some special adjectives are described in dictionaries. A field-semantic analysis of non-restrictive adjectives shows inconsistencies and incompleteness in lemmaselection and semantic description in two authorized Scandinavian dictionaries. According to a cognitive analysis of adjectives it shows that non-restrictiveness is a matter of use more than of adjective type. The paper concludes that this should also be taken into account when choosing examples and polysemes in dictionary definitions.

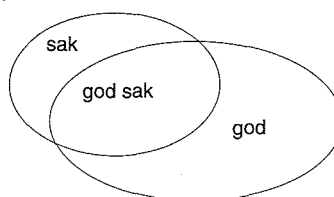
1 Innledning

På Euralex-konferansen i Amsterdam i 1994 satte Henk Verkuyl (1994:2f) fram noen tanker om kunnskapsrepresentasjon i ordbøker. Han refererte til "de fem C-er": "completeness", "consistency", "correctness", "currency" og "citation". Dette illustrerte han ved definisjoner av brikkene i sjakkspillet på den måten at **fullstendighet** krever at alle sjakkbrikkene blir tatt med i ordboka, dvs en skal stille krav til lemmaseleksjonen. Videre var **konsistens** illustrert med at hver brikke bør få samme type beskrivelse, slik at ikke en brikke blir definert etter måten den kan flyttes på, en annen etter utseendet eller bare som "sjakkbrikke". **Korrekthet** ble illustrert med at sjakkbrikkene skal beskrives på samme abstraksjonsnivå, noe som vel så godt illustrerer konsistensprinsippet. Manglende korrekthet hadde blitt tydeligere i tilfelle brikkene blir definert med feil flyttingsregler. Det fjerde prinsippet, "currency", kan kanskje oversettes med **gangbarhet**, noe som innebærer at termene skal defineres under det mest gangbare uttrykket (f eks at en bestemt brikke kalles *springer* og ikke *hest* på norsk). Hans femte krav dreide seg om **sitatteknikk**, og i den forbindelse kildeutvelgelse. Jeg vil ikke gå nærmere inn på kriteriene til Verkuyl, men konsentrere meg om to av dem, nemlig konsistens og fullstendighet, og overføre dem på en vanskelig gruppe adjektiver som av Manfred Pinkal (1985:53) er kalt ikke-restriktive adjektiv.

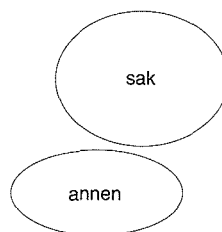
2 Definisjon av ikke-restriktive adjektiv

Et vanlig restriktivt adjektiv har som hovedoppgave å avgrense betydningen til det substantiviske leddet det står til. I tillegg fins det en liten, men lite homogen gruppe ikke-restriktive adjektiv som kan stå attributivt til et substantiv uten å fylle adjektivets hovedoppgave: å beskrive eller avgrense det substantiviske overleddet nærmere. Disse adjektivene gir ofte problemer i leksikografisk arbeid. De ikke-restriktive adjektivene blir dels klassifisert som adjektiv, dels som pronomen i ordbøkene. Jeg velger å kalle dem alle for adjektiv, idet de

kan stå attributivt til et substantiv. Restriktive adjektiv kan rent grafisk framstilles slik etter en mengdeteoretisk tankegang:



mens de ikke-restriktive adjektivene kan framstilles slik:



Figurene illustrerer to hypotagmer som formelt sett er like: *en god sak*, *en annen sak*, men mengdeteoretisk og semantisk er de svært forskjellige. De ikke-restriktive adjektivene kan normalt ikke stå predikativt.

I forbindelse med en analyse av adjektiv i tre norske lovtekster (lov om folketrygd 1987¹, lov om offentlighet i forvaltningen 1985, lov om behandlingsmåten i forvaltningssaker 1986) har jeg kartlagt en del slike ikke-restriktive adjektiv. Det er meg bekjent ikke gjort noen større eller sammenhengende analyse av denne typen adjektiv basert på tekstundersøkelser, selv om både Lundbladh (1988:108) og Malmgren (1990:62) har antydnet kategorien i redegjørelsen for Noréens "placerande assumptiva glosor". Forekomstene av ikke-restriktive adjektiv i lovtekstene var behandlet på ganske forskjellige måter i ordbøkene uten at det var åpenbart for meg hva som var grunnlaget for det.

3 Feltteoretisk analyse

En måte å komme disse uregjerlige adjektivene litt nærmere på, er å forsøke å beskrive dem etter en feltteoretisk modell. Hovedpoenget med analyse i semantiske felt er nettopp at ord og ordbetydning ikke kan beskrives som isolerte fenomener, men må betraktes som deler av et relasjonelt system. Det totale relasjonelle systemet er delt inn i mindre semantiske felt. Dessverre er semantisk feltteori forholdsvis vagt beskrevet, og det har ikke lyktes meg å finne en konsistent og klar definisjon av et semantisk felt. Mel'čuk (1988:170) er den som innen leksikografisk arbeid har gitt den mest operasjonelle definisjonen på et semantisk felt:

A semantic field is the set of all lexical units that share an explicitly distinguished non-trivial semantic component.

¹ Årstallene refererer seg til den utgaven av lovteksten som jeg har undersøkt, ikke til det året loven ble gitt.

Definisjonen stemmer forholdsvis bra med det Lyons (1977:250) kaller Trier-Weisgerber-teorien, men den er lettere å anvende i og med kravet om en felles, betydningsskillende komponent. Problemet er å finne denne komponenten. Når den er funnet, kan en utnytte relasjoner som antonymi, hyponymi, synonymi og polysemi ved utvelgelsen av de enhetene som hører til et felt, og for å beskrive forholdet mellom enhetene.

Det viste seg at mange av de ikke-restriktive adjektivene i mine tekster kan henføres til noen få semantiske områder eller felt, eller at de tilhører en gruppe der det råder familielikheter, og jeg har her forsøkt å sette opp de ikke-restriktive adjektivene jeg fant i lovene for to semantiske felt som jeg har kalt EKSISTENS, HER OG NÅ². Denne oppstillingen sammenlikner jeg så med hensyn til dekning og beskrivelse i den største autoriserte ordboka for norsk bokmål, *Bokmålsordboka* (heretter forkortet BOB). I noen tilfeller trekkes sammenlikning med *Svensk Ordbok* (heretter forkortet SO) inn, da redaksjonen der delvis har anvendt semantisk feltteori i sin beskrivelse og har definert en del av adjektivene feltteoretisk ved at det oppgis antonymer, noe som er svært opplysende i betydningsbeskrivelsen.

Feltet for EKSISTENS kan skisseres som et kontinuum fra det fraværende til det eksisterende, eller for den saks skyld fra 0 til 1, dvs som et skalært forhold inndelt blant annet ved følgende adjektiver (de adjektivene som er tilnærmet synonyme er satt tilnærmet under hverandre):

manglende	hypotetisk	eventuell	mulig	tenkelig	sannsynlig	virkelig
	potensiell		tenkbar	rimelig	faktisk	
fraværende						
	imaginær	innbilt				

Det andre feltet jeg vil prøve å sette opp, er mer komplisert enn det første, nemlig feltet for utpeking eller deiksis, som jeg har kalt HER OG NÅ. Jeg har valgt å dele det i fire underområder, da de nok har det deiktiske som en klar felleskomponent, men dette deiktiske refererer seg til fire forskjellige forhold:

Situasjonssted:

viss	[denne/dette/disse]	annen	
bestemt	vedkommende	nærmere	øvrig
spesiell	nærværende		
særlig	tilstedeværende		
felles	egen		

Subfeltet dekker utpeking av det som er innen rekkevidde i en gitt situasjon til det som foreligger eller gjelder i en annen situasjon (her-der-forholdet). Jeg har valgt å ta med de påpekende pronomenene *denne*, *dette*, *disse*, da de semantisk fyller mye av den samme funksjonen som de andre leksemene i feltet har, selv om de formelt sett neppe kan klassifiseres som adjektiv.

²Disse feltene er definert etter min intuisjon og språkfølelse. Det kan sikkert diskuteres om inndelingen skulle vært annerledes. Hvor store områder et felt skal utgjøre, er fortsatt uklart i og med at betydningsmessig opposisjon forekommer på mange nivåer. Det er på dette punktet jeg mener feltteorien er metodisk svak.

Situasjonstid:

tidligere	foregående	nåværende		neste	seinere
	forrige	aktuell	følgende		
gammel			nærmeste		

Subfeltet er ordnet etter en kronologisk felleskomponent fra noe fortidig til noe framtidig, med et nåpunkt i fokus og en ubestemt fortid og en ubestemt framtid som ytterpunktene (nå-da-forholdet).

Del av mengde:

alle	noen	ingen
samtlige	diverse	enkelt
hver		

Dette subfeltet dekker det som både uttrykker noe deiktisk og samtidig markerer det som en del av en helhet.

Sammenlikning:

forskjellig	liknende	maken
		samme
		tilsvarende

Dette subfeltet dekker først og fremst en pekende funksjon, men elementene uttrykker også en sammenlikningskomponent.

Man kan sikkert tenke seg en oppdeling av disse subfeltene, slik at f eks motsetningen mellom *egen-felles* danner et eget felt, men da blir forbindelsen mellom *egen* og *særskilt* borte. Det råder så vidt jeg kan se altså en kryssklassifisering her som ikke er så lett å fange med en feltanalytisk modell, da den forutsetter et hierarkisk oppbygd betydningssystem, noe som ikke alltid stemmer for naturlig språk.

4 Klassifisering

Uansett hva slags pronomendefinisjon man bygger på (semantisk eller etymologisk), vil det bli problematisk å skille mellom hvilke som klassifiseres som adjektiv og hvilke som klassifiseres som pronomen, noe en oversikt over klassifiseringen i BOB viser:

aktuell	adj	felles	adj	maken	adj
alle	ub.pron	foregående	adj	neste	adj
annen	ub.pron	forrige	adj	noen	ub.pron
bestemt	adj	forskjellig	adj	ny	adj
denne/dette/		følgende	adj	nærmeste	adj
disse	pron	gammel	adj	nårværende	adj
diverse	adj	hver	pron	nåværende	adj
egen	adj	ingen	pron	nærmere	adj
enkelt	adj	liknende	adj	samme	pron

samtlige	adj	særlig	adj	viss	adj
seinere	adj	tidligere	adj	øvrig	adj
slik	pron	tilsvarende	adj		
spesiell	adj	vedkommende	adj		

Det kan være vanskelig å forstå hvorfor noen er blitt klassifisert som adjektiv og noen som pronomen av disse lemmaene, da de må sies å ha tilnærmet samme funksjon, først og fremst det å peke ut en referent. Klassifiseringsproblemene her svarer vel også til en diskusjon om hva som er beskrivende ledd, og hva som er bestemmende ledd. Semantisk kan de ha noenlunde samme funksjon. Vi ser i hvert fall at klassifiseringen i pronomen og adjektiv kan diskuteres.

Denne inkonsistensen kan tyde på at det er kryssende hensyn som skal ivaretas. For det første har tradisjonell klassifisering i ordbøker ofte vært retningsgivende, videre blir det tatt hensyn til etymologiske forhold, og noe kan kanskje tilskrives tilfeldighetene, ikke minst tilfeldigheter i det eksempelmateriale som ordbøkene bygger på. Spørsmålet er hva som skal være det styrende prinsippet i en leksikografisk beskrivelse av et språks ordinventar. Ved en semantisk-systematisk synkron beskrivelse, f eks ut fra feltteori, vil beskrivelsen lettere bli mer konsistent.

5 Lemmatisering

Når det gjelder lemmatiseringen, er det heller ikke lett å oppdage konsistensen. Brukes morfologiske kriterier, kan vi tenke oss at bare grunnformer skal gi opphav til lemma. Men presens partisipp-formene *vedkommende* og *liknende* er ført opp som egne lemmaer, mens *tilsvarende* er sublemmatisert under verbet å *tilsvare* med opplysning om at presens partisipp-formen forekommer i adjektivisk bruk. Det er for så vidt en generell grammatisk opplysning som gjelder for alle tre leksemene. Flere presens partisipp-former er ført opp som egne lemmaer: *foregående*, *nåværende*, *følgende*. Konsistenskravet er dermed ikke tilfredsstillt. Man kan tenke seg at inkonsistensen kommer av kompleksiteten i orddanningen. Men verbet *likne* er et ettordslemma, og det er vel så god grunn til å tenke seg *liknende* som en avledet form av å *likne*, som at *manglende* er presens partisipp av å *mangle*. *Komme ved* og *svare til* er komposisjonsmessig sett like komplekse. I BOB er *manglende* og *tilsvarende* ikke oppført som egne lemmaer, mens de andre presens partisippene her er det. Det kan forklares med at de andre presens partisippene ikke kan henføres direkte tilbake til et enkelt verb, vi har ikke former som *å *fravære*, *å *tilstedevære*, men semantisk er *fraværende* og *manglende* synonymer (med *tilstedeværende* som antonym). Men om direkte avledning av et ettordslemma skulle være et prinsipp for lemmatiseringen, er det heller ikke selvfølgelig at *potensiell* skal føres opp som eget lemma. Her kommer kanskje vurdering av frekvens og leksikalisering av bøyde former inn, men helt uproblematisk er forholdet ikke. Bortsett fra morfologiske kriterier burde det ført til at også *manglende* ble opphøyet til et lemma, siden det er mer frekvent enn *fraværende* (både i mitt lovtekstmateriale og i *Norsk frekvensordbok*). Det er heller ikke noen klar grunn til å vurdere *potensiell* som avledning og *manglende* som bøyning, siden begge er leksikalisert som adjektiver. For meg ser det ut til at konsistenskravet ikke er oppfylt på dette punktet, noe som ville vært lettere å avdekke dersom feltanalyse

hadde vært brukt. SO gir ikke noe sammenlikningsgrunnlag her, da *manglende* ikke fins i svensk.

Av andre bøyde og leksikaliserte former er *nærmere* en komparativ form av *nær*, og det er vel ingen klarere grunn til å føre det opp som lemma enn f eks *nyere*, jf *nærmere regler*, *nyere regler*, men her er altså bare *nærmere* eget lemma. Følgende tre former mener jeg er leksikalisert i gradbøyd form: *tidligere*, *seinere*, *nærmeste*. De to som er leksikalisert i komparativ, er ikke oppført som egne lemmaer, mens *nærmeste*, som er i superlativ, er eget lemma. Det er vanskelig å se noen morfologisk begrunnelse for denne lemmatiseringen.

Det jeg hittil har pekt på, antyder ønskeligheten av å finne et konsistent system for lemmatiseringen av de problematiske ikke-restriktive adjektivene som fører til en konsistent beskrivelse av dem. Jeg mener det forutsetter at man tar utgangspunkt i det semantiske mer enn i det formelle. Da vil det kanskje også vise seg at det er et slags samsvar mellom formelle (nemlig syntaktiske) og innholdsmessige egenskaper.

6 Kognitiv analyse

De aller fleste adjektivene jeg har nevnt, er ført opp med to eller flere betydningsnummer. Det vil si at de anses som polyseme. Men det er vanskelig å se noe system i denne polysemien, noe som stemmer med det redaktørene selv sier om å skille en definisjon opp i betydningsnummer. I SO er det innført et mer komplisert system enn i BOM, da polysemene der både deles inn i betydningsnummer, og også har en egen kategori for betydningsnyanse markert med en liten firkant.

De ikke-restriktive adjektivene kjennetegnes syntaktisk ved at de ikke kan anvendes predikativt. Når slik bruk likevel forekommer, er det med en annen betydning, det vil si at et annet betydningsnummer realiseres. Taylor (1992) beskriver dette forholdet innenfor rammen av kognitiv grammatikk. Han opererer med adjektiv som absolutte eller syntetiske uttrykk, et skille som langt på vei svarer til restriktive og ikke-restriktive adjektiv. Men i tillegg legger han vekt på at mange adjektiv kan være både det ene og det andre, avhengig av hvilken tolkning det har. Dermed er han inne på adjektivenes polysemi, noe Pinkal (1985) ser helt bort fra i sin behandling av de ikke-restriktive adjektivene. Taylor viser i sin analyse av det engelske adjektivet *old* at det har forskjellig betydning når det brukes predikativt og når det brukes attributivt. Det vises likeledes tydelig når det norske leksetet *gammel* settes opp med sine to motsetningspoler, nemlig

gammel — ung
gammel — ny

Det kan være betydningsmessig forskjell på *gammel* i *en gammel venn* og *en gammel lærer*. I attributiv stilling har vi nemlig en tvetydighet, den gamle vennen kan ha høy alder (syntetisk tolkning), eller være en tidligere venn (absolutt tolkning), den gamle læreren kan ha høy alder (syntetisk tolkning), eller ha vært lærer tidligere (absolutt tolkning), men i predikativ stilling kan det bare være snakk om at personene har høy alder:

læreren er gammel
vennen er gammel

Det samme gjelder selvsagt ved *ny*, som kan brukes om noe som nylig er anskaffet, eller det kan være ubrukt. Verken semantisk eller formelt er det grunnlag for å snakke om homonymi i slike tilfeller, så det må være korrekt å konkludere med at det foreligger polysemi. BOB har her satt opp syv betydningsnummer for *gammel*:

- 1 som har levd el. vært til lenge
- 2 tradisjonell, vanlig, velkjent
- 3 som har vart lenge, som har sitt opphav i fortiden
- 4 erfaren
- 5 utslitt, ubrukkelig, avleggs
- 6 tidligere, forrige
- 7 av en viss alder

SO har imidlertid bare to betydningsnummer:

- 1 som har eksisterat under lång tid <mots **ung**, ny 1>
- 2 som i förflyuten tid fyllde funktionen som framgår av sammanhanget <mots **ny2**>

Derimot har betydningsnummer 1 mange underordnede betydningsnyanser markert med "äv. neutralt, äv. med tonvikt på funktionen [...]".

De to ordbøkene har altså en ganske forskjellig praksis for beskrivelsen av det semantiske innholdet i adjektiv. En kritisk gjennomgang av den norske definisjonen viser at flere av betydningsnummerne kan føres sammen, f. eks 1 og 3, mens 6 skiller seg tydelig ut betydningsmessig, en betydning som er nær 2 i den svenske (med forbehold om min manglende evne til å oppfatte nyanser i svensk). Det er tydelig at de to betydningsnummerne i den svenske ordboka nettopp svarer til den syntetiske og den absolutte bruken av adjektivet, og at dette skillet egentlig gir tilstrekkelig definisjon. Den er dessuten symmetrisk med definisjonen av antonymet *ny* i BOB:

- 1 som er laget eller kommet til for kort tid siden
- 2 som kommer etter noe av samme art

Betydning 1 er synonym med *ung*, som er den syntetiske betydningen, mens 2 gir den absolutte betydningen. SO har her fire betydningsnummer og gir en mindre ryddig oversikt over betydningen:

- 1 som för helt kort tid sedan kommit att existera
- 2 som ej tidigare är känd eller använd
- 3 som tillkommer utöver det tidigare
- 4 mest i sms som återigen finns

7 Sammenliknende oversikt

En oppstilling over definisjonene i BOB og SO for de semantiske feltene, og med Taylors (1992) betydningsdifferensiering i tankene, skulle gi grunnlag for å vurdere i hvilken grad hans systematikk kunne vært fulgt ved beskrivelsen av de adjektivene som forekommer i feltet. Av plasshensyn avgrensers jeg meg til ett felt, nemlig EKSISTENS:

fraværende	BOB	1	ikke til stede	A
		2	uttrykksløs	S
frånvarande	SO		som ej är närvarande <mots. närvarande> på en viss plats, trots att dette förväntas	A
hypotetisk	BOB		som knytter seg til hypoteser, mulig, tenkt	S
	SO		som blott grundar sig på antaganden som gäller eller är av vikt endast om vissa förutsättningar oppfylles	S A
potensiell potentiell	BOB		som eksisterer som en mulighet, utviklingsdyktig	A
	SO		som kan utvecklas till, är teoretisk mulig men ej fullt förverkligad	A
rimelig	BOB	1	sannsynlig	A
		2	billig, moderat	S
rimlig	SO	1	som det finns (fornuftig) anledning att tro på	A
		2	som har sådan omfatning som det finns fornuftig anledning att kræva	?
mulig	BOB	1	som lar seg gjennomføre	S
		2	som kan hende el. tenkes	A
möjlig	SO		som kan tänkas förekomma i visst sammanhang	A
tenkelig	BOB		(som kanskje er) mulig	A
tenkbar tänkbar	BOB		tenkelig	A
	SO		som man kan tänka sig och oppfatta som rimlig eller möjlig	A
eventuell	BOB		som kan forekomme	A
	SO		som kanskje kommer att inträffa, om händelse, resultat e.d. <SYN. möjlig> äv. som kanskje existerar	A
sannsynlig sannolik	BOB		trolig, rimelig, som en kan vente seg	A
	SO		tämligen trolig med hänsyn till föreliggande fakta	A
tilstedeværende	BOB		nærværende, som er til stede	A
nærværende	BOB	1	som er til stede, som er framme i bevisstheten	S
		2	som er her nå, som en har for seg i øyeblikket	A
närvarande	SO	1	som är på platsen	S
		2	ofta substantiverat: som nu förlöper, i nuläget	A
virkelig	BOB	1	som eksisterer, fins faktisk	A
		2	ekte, sann	S

verkelig	SO	1	som finns (i världen) och kan uppfattas eller påvisas med sinnena <jfr. faktisk>	A
		2	som i själva verket utgör <jfr. sann 2, riktig 2>	S
faktisk	BOB		virkelig, reell	A
	SO		som verkligen existerar <SYN. existerande>	A
reell	BOB	1	virkelig, faktisk, mots. imaginær	A
		2	redelig, ærlig	S
	SO	1	som finns (i världen) och kan uppfattas eller påvisas med sinnena <SYN. verkelig> <fackspr. jfr. faktisk>	A

Hver av de forskjellige polysemene er til høyre i oppstillingen merket med A eller S for absolutt eller syntetisk betydning av adjektivet. Oversikten viser flere eksempler på at bare den ene betydningen er beskrevet i ett av språkene, mens det andre har med begge to (*fraværende*, *hypotetisk*). Det er rimelig å anta at norsk og svensk er så like språk at begge betydningene forekommer i begge språkene, i hvert fall i de eksemplene som er nevnt her. Betydnings skillet er så vesentlig både semantisk og syntaktisk at det bør være viktig å få det med i definisjonen, eventuelt til foretrengsel for en del andre og finere betydningdifferanser det er spandert plass på. En systematisk gjennomgang av de andre feltene viser også der at det ofte mangler oppgave over en av betydningene. Jeg mener at man ved å bruke en systematisk metode i beskrivelsen vil unngå inkonsistens og ufullstendighet i ordbøkene av denne typen.

Litteratur

Ordbøker

BOB = M. I. Landrø/B. Wangenstein 1993: *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. Oslo: Universitetsforlaget.

SO = *Svensk Ordbok*. 1986. Utarbeidet ved Språkdata, Göteborgs universitet. Solna: Esselte.

Annen litteratur

Heggstad, Kolbjørn 1982: *Norsk frekvensordbok*. Oslo: Universitetsforlaget.

Lundbladh, Carl-Erik 1988: *Adjektivets komparation i svenskan. En semantisk beskrivning*. Lund: Lund University Press.

Lyons, John 1977: *Semantics* 1. Cambridge: Cambridge University Press.

Malmgren, Sven-Göran 1990: *Adjektiviska funktioner i svenskan*. Göteborg: Acta Universitatis Gothoburgensis.

Mel'čuk, Igor 1988: Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria. I: *International Journal of Lexicography* Vol. 1 No. 3, 165–188.

Pinkal, Manfred 1985: *Logik und Lexikon*. Berlin/New York: Walter de Gruyter.

Taylor, John R. 1992: Old problems: Adjectives in Cognitive Grammar. *Cognitive Linguistics* 3–1, 1–36.

Verkuy, Henk 1994. *Knowledge Representation in Dictionaries. Keynote lecture*. Delivered at the 6th Euralex International Congress Amsterdam, 30.8–3.9.1994.

Brolexikon

The Collins COBUILD English Language Dictionary (1987) introduced a new concept in lexicography in the shape of a new style of definition. A new step has now been taken in the direction of a bilingual dictionary based on this new type of definition, the so-called "bridge dictionary". These dictionaries — one of them has been published already — are meant to "bridge" the gap between monolingual and bilingual dictionaries by translating definitions. This will make it easier for the reader to use the dictionary, since definitions have been a stumbling-block for foreign readers. The bridge dictionary style of definition is discussed, especially the syntactic problems caused by mixing two languages. A new borderline between the languages is suggested by the author.

1 Från Collins COBUILD till brolexikon

När Collins COBUILD English Language Dictionary kom år 1987 innebar detta på många sätt en revolution inom den lexikografiska världen. Hörnpelarna i det nya konceptet var (a) autentiska textexempel direkt hämtade ur strategiskt valda moderna engelska texter, samlade i en mycket omfattande textdatabas, (b) ett helt nytt definitionssystem som kombinerar ordförklaringar på enkel prosa med en exakthet som går utöver traditionella definitioner, och (c) en formaliserad grammatisk beskrivning av varje betydelse som ger läsaren de exakta gränserna för användningen av ett ord.

COBUILD-idén har fått en mycket stor betydelse inom den lexikografiska branschen. Arkitekten bakom lexikonet — professor John Sinclair — har nu tagit ytterligare ett steg för att möjliggöra att lexikonet används på ett mer rationellt sätt också för andraspråksinläring. Problemet med enspråkiga lexikon för andraspråksinläring är kort uttryckt att de är för svåra — läsaren stupar framför allt på ordförklaringarna som ju ges på det främmande språket. Om läsaren kunde ledas vid handen in i ett enspråkigt lexikon på andraspråket, så skulle detta kunna få stora konsekvenser också för tvåspråkig lexikografi. Enspråkiga, omsorgsfullt utarbetade lexikon skulle på detta sättet kunna utnyttjas på en mycket större marknad än vad som är möjligt idag.

Detta är John Sinclairs tanke: att skapa vad han kallar "brolexikon" (bridge dictionaries) som för läsaren innebär en bro över till de helt enspråkiga lexikonerna. Fördelarna för utgivarna av det enspråkiga lexikonerna ligger framför allt i en större marknad (och ett mindre behov av riktiga tvåspråkiga lexikon som är utomordentligt kostsamma). Läsaren erbjuds ett lexikon utarbetat helt enligt andraspråkets förutsättningar och med ett autentiskt språkligt exempelmaterial. Svårigheten för läsaren att få tillgång till nyttigheterna i lexikonet undanröjs genom vissa, systematiska kopplingar till läsarens modersmål: översättning av ordförklaringar eller ekvivalenter på modersmålet till de olika ordbetydelserna. Till detta kommer att lexikonet på sätt och vis går i bägge riktningarna: det blir inte bara ett "passivt" lexikon som kan användas för att slå upp okända ord i texter, utan det blir också ett "aktivt" lexikon genom att de svenska motsvarigheterna till de engelska orden ges i ett index. Vi får

ett engelsk-svenskt och ett svensk-engelskt lexikon och en definitionsordbok med exempel i en och samma volym.

För närvarande pågår arbete att omsätta denna idé i praktiken i bl.a. Brasilien, Tyskland och Danmark. Det första lexikon som publicerades blev ett portugisiskt-engelskt COBUILD som kom ut i Brasilien under våren 1995. Utgångspunkten är i dessa fall en förkortad version benämnd Collins COBUILD Student's Dictionary.

2 Strukturen i brolexikon

Det viktigaste inslaget i brolexikon är att definitionerna ges i användarens eget språk, allt enligt principen att det som är okänt för läsaren skall förklaras med hjälp av sådant som är känt. Och användaren är normalt bättre förtrogen med sitt eget modersmål än med det främmande språket.

COBUILD-definitioner består typiskt av två delar: den första delen ger uppslagsordet i sitt typiska syntaktiska och lexikaliska sammanhang. Till exempel:

- (1) If a level or amount drops, it becomes less
- (2) If you drop a hint or piece of advice, you give it to someone in a casual way

I det första exemplet ges huvudsakligen en syntaktisk upplysning om transitiv/intransitiv användning, i det andra exemplet ges uppslagsordet i en typisk lexikalisk kollokation som *drop a hint*. Tanken är att läsaren på ett enkelt sätt skall få en uppfattning om ordets naturliga kontext utan att egentligen behöva känna till termer som **transitivitet** eller **kollokation**. Denna del av definitionen är den kanske viktigaste nyheten i COBUILD-definitioner.

Den andra delen i definitionen är betydelsebeskrivningen, dvs. det som normalt kallas definition. Hela COBUILD-definitionen skiljer sig dessutom från gängse lexikografisk praxis däri att den är utformad i vanlig prosaform och ofta i en slags interaktiv form:

- (3) If you ask what someone is **driving at** ...
- (4) When you **breath in** ...

Grundidén i definitionerna behålls i brolexikonet, med den skillnaden alltså att det engelska uppslagsordet fogas in i en icke-engelsk språklig kontext. Så här kan det se ut i det portugisiska lexikonet under huvudordet *piecemeal*:

- (5) Um processo é descrito como **piecemeal** quando se desenvolve de um modo gradual, geralmente a intervalos irregulares
(En process beskrivs som **piecemeal** när den utvecklas stegvis, vanligtvis med oregelbundna mellanrum)

I exemplet fogas det engelska ordet lätt in i sin portugisiska kontext, även om läsaren kanske till en början känner sig ovan vid språkblandningen. Men en sådan blandning är inte okänd i språkpedagogiska sammanhang. En lärare kan mycket väl säga något i stil med "to pull someone's leg är att lura någon (på ett skämtsamt sätt)".

Men övergången mellan språken är inte alltid så omärklig som i exemplet. När det gäller förklaringen av verb inträffar ofta att den syntaktiska ramen skiljer sig åt i de bägge språken.

Här följer ett exempel på lite hårdare syntaktisk kollision, hämtat ur det danska COBUILD-projektet (som ännu befinner sig i början på sitt arbete och inte nödvändigtvis kommer att publiceras i det här skicket).

- (6) Hvis noget, fx mad eller fornøjelse, **satiates** (mætter) én, betyder det, at man har fået så meget af det, at man er blevet træt af det

Observera att ordet dels förklaras med en definition, dels en dansk synonym. Skillnaden mot tidigare exempel är att det engelska ordet här är ett transitivt verb som utrustas med danskt objekt. Också detta kan man emellertid vänja sig vid. Värre är det när objektet är t.ex. en *att*-sats (här och i fortsättningen markeras ett fingerat svenskt exempel med en stjärna):

- (7) *Om någon **thinks** att något är fallet, så tycker man att det är så

Än värre blir det om det är skillnad på transitivitet i engelska och svenska:

- (8) *Om någon **flies** en drake, så låter han/hon den stiga i luften [. . .]

Lägg också märke till att uppslagsordet presenteras i varierande form beroende på hur subjektet ser ut: om någon **flies**/om du **fly**(!).

3 Utkast till ett engelsk-svenskt brolexikon

Även om det finns många stötestenar — som vi snart skall återkomma till — för den som vill konstruera ett brolexikon så bör man inte glömma fördelarna. En uppenbar sådan fördel har med ekonomi och arbetsinsats att göra. Ett tvåspråkigt lexikon av hygglig kvalitet är en stor investering för ett förlag. Om det därför finns en möjlighet att utnyttja enspråkiga lexikon också för tvåspråkiga ändamål är detta en lockande utväg. Vi har under senare tid sett en rad enspråkiga engelska lexikon av typ ”inlärningslexikon” som vänder sig till den stora icke-engelska publik som vill lära sig engelska. Därvid har det visat sig att dessa lexikon inte är fullt så enkla att använda som förlagen hade tänkt sig. Tanken på brolexikon är en logisk konsekvens av denna insikt: man slipper därigenom att investera i ett fullt utbyggt tvåspråkigt lexikon.

För användaren av lexikonet ligger troligen det främsta värdet i att man kan få tillgång till ett lexikon utvecklat inom målspråket med allt vad detta innebär av kompetens och aktuella språkexempel. Trots allt är de bästa beskrivarna av engelska språket engelsmännen själva. Man kommer också närmare betydelsen i de engelska orden om man utnyttjar riktiga definitioner snarare än mer eller mindre exakta ekvivalenter.

Brolexikonet är därför värt en genomtänkt granskning. Det bästa sättet att få grepp om utformningen av sådana lexikon är att själv göra ett försök. Här följer några exempel (se figur 1) ur ett tilltänkt engelsk-svenskt brolexikon med en del anpassningar som har befunnits nödvändiga vid utformningen av ordförklaringarna.

nag Här har definitionen ändrats från ”if you nag someone” till ”to nag”. Objektet finns bara i betydelsedelen (vilket kanske är en nackdel). Å andra sidan slipper man ”Om du **nag** någon”.

BILAGA

Collins COBUILD Student's Dictionary
Bridge Bilingual: Swedish (utkast juni 95)

N. N är förkortning för norr(a)

nadir /nə'diːd/ SING N **Nadir** är lägsta punkten på något =NADIR *The government was at the nadir of its unpopularity.*

nag /næg/, *nags, nagging, nagged.* 1 VB with or without OBJ **To nag** är att på ett tröttande sätt söka övertala någon =TJATA *He used to nag me endlessly about money. ...having nagged for an invitation.* 2 VB with at Om tvivel eller misstanke **nags at you**, så känner du en gnagande oro eller misstanke =GNAGA *Something that she had said had been nagging at him.* o **nagging** ATTRIB ADJ =GNAGANDE *She had a nagging sense of inadequacy.*

nail /neɪl/, *nails, nailing, nailed.* SUBST =SPIK See picture headed TOOLS. *...the mirror that hung from a nail on the wall.* 2 VB with OBJ and ADJUNCT =SPIKA *They nail plastic sheets over their windows. There were signs nailed to the trees.* 3 COUNT N =NAGEL *He keeps biting his nails.* 4 PHRASE **hit the nail on the head** slå huvudet på spiken (=säga precis det rätta) 5 a **nail in something's coffin**: see coffin.

nail down. PHR VB =SPIKA NER, SPIKA FAST

nail up. PHR VB =SPIKA UPP *...the warning notice that he had nailed up on the pole.*

naive /naɪv/, also spelled **naïve**. ADJ En person är **naive** som tror att allting är mycket enklare än de i verkligheten är =NAIV *You're surely not so naive as to think that this will change anything.*

o **naively** ADV *They naively assume that things can only get better.*

naivety /naɪvɪti/, also spelled **naïvety** or **naïveté** UNCOUNT N En person visar **naivety** när hon tror att allting är mycket enklare än det i verkligheten är =NAIVITET *In this he showed political naivety.*

naked /neɪkɪd/. ADJ Den som är **naked** är utan kläder =NAKEN *He was naked except for a pair of underpants. ...the men's naked bodies.* o **nakedness** UNCOUNT N *They seized towels to hide their nakedness.* 2 ATTRIB ADJ Man beskriver föremål som **naked** när de inte är täckta eller skyddade...*naked light bulbs. Never look for a gas leak with a naked flame.* 3 ATTRIB ADJ Man använder **naked** för att

beskriva uppförande eller starka känslor som inte alls döljs *The home employment offered to housewives is naked exploitation. His face broke into an expression of naked anxiety.* 4 ATTRIB ADJ Om man kan se något med **the naked eye**, så kan man se det med blotta ögat (=utan hjälp av glasögon, mikroskop etc.)

name /neɪm/, *names, naming, named.* 1 COUNT N =NAMN *His name is Richard Arnason.* 2 VB with OBJ and COMPLEMENT **To name** någon el. något betyder att ge dem namn *She wanted to name the baby Colleen.* 3 VB with OBJ **To name** någon eller något betyder också att identifiera dem genom att ge dem namn =NÄMNA...*a Minister, whom he did not name. ...various flowers: roses, tulips and snapdragons, to name only a few.* See also **named.** 4 VB with OBJ and after **To name** någon el. något efter en person el. sak, betyder att ge dem samma namn som den personen el. saken. =UPPKALLA *The College in Holborn is named after her.* 5 VB with OBJ **To name** något, t.ex. tidpunkt för ett möte eller priset på något, betyder att man säger vad man vill att det skall vara. =ANGE *He named a price he thought would scare me off.* 6 COUNT N =RYKTE *Grey spoke out in public to clear Haldane's name. They were giving the country a bad name.* 7 COUNT N Man kan säga att den som har blivit känd är ett **name**. 8 See also **brand name, Christian name, maiden name.** o Om man nämner någon **by name**, så säger man namnet snarare än att hänvisa till det indirekt. o **To call someone names** betyder att förolämpa någon genom att beskriva honom med nedsättande ord. o Om något är registrerat i ditt **name**, så tillhör det dig officiellt och är reserverat för dig. *The room was reserved in the name of Peters.* o Om du gör något **in the name of** ett ideal eller en grupp människor, så gör du det för att du tror på detta ideal eller representerar folket. *The group claims to speak in the name of 'the simple people of the country'.* o **To make a name for oneself** betyder att bli känd och beundrad för något man gjort *George Eliot had already made a name for herself as a writer.*

named /neɪmd/. PRED ADJ När man säger vad en person, en sak eller en plats är **named** så ger man deras namn =(SOM) HETER *...a lecturer named Harold Levy.*

Figur 1. Collins COBUILD Student's Dictionary Bridge Bilingual: Swedish (utkast juni 95).

nail Här finns ingen definition alls vilket beror på huvudprincipen att mer eller mindre kulturoberoende konkreter enbart ges en ekvivalent.

name To name någon el. något, jfr under **nag**.

name ”Om något är registrerat i ditt **name**”; här är problemet att man måste ge ett engelskt ord ett svenskt genus vilket är nödvändigt kopplat till pronomenet. Brytningen mellan engelska och svenska känns inte riktigt bra.

Kanske finns det ett naturligare snitt mellan engelskan och svenskan. Vi skall se närmare på ett sådant förslag.

4 Förslag till naturligare snitt mellan språken

Dessa problem är naturligtvis inte okända för konstruktörerna av brolexikon. Inför det portugisiska lexikonet (Guidelines for the Bilingualization of COBUILD Dictionaries, 1992, unpublished) påpekas exempelvis följande av Mona Baker, en av de inblandade från COBUILD:s sida:

It is not always easy to fulfil the requirements of addressing the user directly in the target language at the same time as fulfilling the requirements of retaining the headword, together with its syntax and collocates, in English.

Denna underdrift exemplifieras med sådana språkliga skiljaktigheter som att vissa språk böjer ord i person, numerus och genus. I ett språk som portugisiska får man t.ex. avhjälpa genusproblem genom att utveckla betydelsedelen av definitionen:

(9) **A cat** é un gato ou um gata
(Ungefär: **A cat** är en hankatt eller honkatt)

Vidare pekar man på lösningar som innebär att man istället för en finit form använder en infinitiv inledd med *att* (jfr ändringen ovan av ”if you nag someone” till ”to nag someone”).

Samtidigt är det uppenbart att man inte utan vidare vill släppa den del av definitionen där ordet presenteras i en typisk kontext. Detta är ju faktiskt COBUILD:s speciella signum. Hur skall då problemet lösas?

Jag tror att lösningen ligger i förlängningen av de medel man redan tar till för att klara av omöjliga syntaktiska klyvningar mellan språken: man utvidgar helt enkelt räjongen för det engelska huvudordet. Det finns flera goda skäl till detta:

- (a) De engelska orden skall presenteras i sin engelska kontext, inte i en svensk (även om det är en hygglig översättning).
- (b) Det ordförråd som finns i den första delen av definitionen är ofta begränsat (*If you, When you, someone* etc.) och lätt att lära. Rör det sig om olika subjekt så kan man ge dessa på svenska inom parentes. T.ex.:

(10) Someone or something (en person, anordning eller maskin) that is **accurate**, kan utföra en uppgift utan att göra fel

- (c) Det är naturligare att växla mellan språken just i det läge där betydelsebeskrivningen kommer. Det är ju denna del som är den egentliga förklaringen och det är också här som läsaren har störst behov av hjälp.

Låt oss se vad detta skulle innebära för de exempel på hårdhänt syntaktisk klyvning som tidigare redovisats.

- (6a) If something (t.ex. mat eller nöje) **satiates** you, så betyder det att man har fått så mycket av det att man har blivit trött på det

När det gäller utformningen av vänsterledet kan man tänka sig att inom parentes översätta exempel på vad subjekt eller objekt (eller andra medspelare) närmare bestämt kan vara. Vidare kan den vanliga uppsättningen medspelare möjligen förkortas på lämpligt sätt, även om detta strider mot COBUILD:s tanke att undvika "ordboksspråk".

- (7a) If someone **thinks** that something is the case, så tycker man att det är så
 (8a) If someone **flies** a dragon (=drake), så låter han den stiga i luften

I det här fallet är det inte ett exempel på vad objektet kan vara för ord, utan det rör sig om ett enda ord. I sådana fall kan objektet lämpligen direkt översättas inom parentes.

5 Sammanfattande synpunkter

Det finns många goda argument för den typ av lexikon som här kallas "brolexikon", dvs. en slags hybrid mellan det enspråkiga och det tvåspråkiga lexikonet. Det hittills publicerade lexikonet (portugisisk-engelska eller vad man nu skall kalla den här typen) och pågående arbete på andra lexikon visar samtidigt på svårigheten att förena två nästan oförenliga uppgifter: att ge det engelska uppslagsordet i en naturlig *engelsk* kontext samtidigt som man översätter denna kontext till *svenska* (för att göra den enklare att förstå för läsaren). Varje par av språk kommer härigenom att upptäcka nya besvärliga kollisioner mellan språken som gör kontextdelen av ordförklaringen ohanterlig. Ett sätt är då att backa från de mest hårdhänta infogningarna av uppslagsordet och exempelvis välja en infinitiv (t.ex. *to nag*) framför ett finit verb (om någon *nags*). Detta är inte en bra lösning eftersom man därigenom börjar tumma på inte mindre än två av huvudreglerna i COBUILD-definitionen, nämligen att ge ordet i sin naturliga kontext och att utforma kontexten som naturligt tilltal till läsaren.

Det förslag som här framförs innebär att språkgränsen går mellan kontext (på engelska) och förklaring (på svenska). Dock kan man även i kontextdelen vid behov infoga svensk översättning, nämligen då förlagan antingen exemplifierar möjligt subjekt eller objekt etc., eller anger just det ord som måste vara subjekt eller objekt etc.

En sådan markering av språkgränsen bör inte vara mer komplicerad för läsaren eftersom kontextdelen ofta består av formellt angivna medspelare till uppslagsordet. Dessutom kan man behålla de två grundpelarna i COBUILD-definitionen: den naturliga kontexten och det direkta tilltalet.

Litteratur

Collins COBUILD English Language Dictionary. 1987. London/Glasgow: Collins.

Collins COBUILD Student's Dictionary. 1989. London/Glasgow: Collins.

Collins COBUILD Student's Dictionary. Bridge Bilingual: Portuguese. 1995. London/Glasgow: Collins.

A Dictionary of the Icelandic Family Sagas and Semantic Classification

Artikkelen handler om ett prosjekt som består i en leksikografisk beskrivelse av vokabularet i islendingesagaene på grunnlag av et korpus som inneholder en standardisert tekstutgave av samtlige sagaer. Beskrivelsen av substantivene utgjør den første fasen i prosjektet. For å komme fram til mest mulig presise definisjoner, blir substantivene først klassifisert ut fra forskjellige semantiske kriterier, noe som gir en oppdeling i omtrent 50 ulike semantiske kategorier. Denne framgangsmåten, som ikke har vært praktisert tidligere i en islandsk ordbok, bidrar til en konsistent behandling av de ord som knytter seg til de enkelte kategorier. Samtidig blir det lettere å avdekke relevante semantiske skilnader. Den semantiske klassifiseringen har vist seg å være til stor nytte for forskere på forskjellige områder.

1 Introduction

In this paper I will describe a part of a work in progress, bearing the working title *A Dictionary of the Icelandic Family Sagas*.¹ The project, at least in its first phase, only covers the Icelandic Family Sagas. The corpus is therefore finite, and furthermore, the Sagas only describe a rather limited subset of human affairs and actions. Because of this, a detailed semantic description of the vocabulary gives a clear and detailed picture of the world of the Icelandic Family Sagas. The dictionary is developed as a database.

The concordance itself is described by Eiríkur Rögnvaldsson (1995; this volume), and this is the first time that a concordance has been used as a basis for a dictionary in Iceland. The concordance facilitates the work on the dictionary considerably, in terms of explaining meanings, classifying into grammatical categories and, last but not least, of grouping formulaic (or quasi-formulaic) expressions or unusual syntactic structures. We have also noticed that semantic explanations are more precise when working on every example of a word at one time.

The following examples (cf. fig. 1–6) shows how the database is structured. As each word is entered into the database it is divided into its components. Affixes and inflectional endings are indicated with special symbols, '+' is placed between a suffix and a root, and '-' between a stem and an inflectional ending. Prefixes and suffixes are also entered into separate fields. Each word is marked for gender, and the frequency of each case-form is recorded. This work can proceed fairly quickly, as all words are already classified in the concordance, which also gives the frequency of every form of the word. Grammatical categories are given

¹The main participants in this project are Bergljót Soffía Kristjánsdóttir and myself, but Margrét Guðmundsdóttir, Svanhildur Óskarsdóttir, Eiríkur Rögnvaldsson and Örnólfur Thorsson have also taken part. None of the participants is either a lexicographer or an engineer, and our specialization is mainly in medieval literature, apart from Eiríkur Rögnvaldsson and Margrét Guðmundsdóttir who are linguists.

categories are given as in other dictionaries, the difference being that in this dictionary the number of categories is greater, and the analysis is based on syntactic context. This will be useful for linguists or scholars studying the literary characteristics of the sagas. The definitions are arranged according to frequency, and phrases and constructions are explained under separate headings following the basic definition. There are also fields for linguistic, semantic, literary and other remarks.

2 The Corpus

Before coming to my main subject, i.e. the semantic classification system, a few words on the nature of the corpus are due. Many people think that the true meaning of every word in a certain text can be found by looking the word up in a dictionary. While it is of course true that the basic meanings of words, and even some constructions, are defined in dictionaries, these definitions usually only take ordinary, every-day language into account. The problem is that words are not isolated objects, and in fact it can be maintained that words only have a meaning in some context. A word in isolation is therefore only a skeleton, without flesh and blood. In traditional dictionaries, however, the symbolic meaning or various connotations of words are rarely given, because these are more closely connected with fiction and often refer to mythical reality. The lexicographer, on the other hand, is of course always interpreting ordinary human language, and trying to find the basic meaning of words, and in a certain sense also the true meaning.

In our project, the fields for semantic and literary remarks can be used for explanations that fall outside the scope of an ordinary dictionary. There we can play with the fiction which is a part of every word. To find out the basic meaning of words, the lexicographer must often tear the words from the earth that in fact gives them their life and colour. The greater flexibility of our form gives full scope to the context of the words, and therefore we may be seeing some new possibilities in dictionary making. We must admit that this form is in many ways both complicated and time-consuming, but as pointed out by Eiríkur Rögnvaldsson (1995; this volume), the soil, i.e. the syntactic structure, is more significant here.

Orð	brjóst% Kyn	hk	Fsk	Vsk	
Nf.et.	brjóst	2		Pf.et.	brjóst 32 18
Bgf.et.	brjósti	17	3	Ef.et.	
Nf.ft.				Pf.ft.	
Bgf.ft.				Ef.ft.	
e Með fs.			á/fyrir/í brjóst; á/í brjósti, !e-ð gengur úr brjósti		
l Orðasb.			e-m býr e-ð í brjósti		
m Klasi			venja af brjósti, bera e-ð í brjósti		
Merking 1			bringa, barmur		
Merking 2			kirtill sem framleiðir mjólk		
Merking 3			hugur, innræti		
Merking 4			fremsti hluti e-s %hlíf, vörn		
Merkingarl. aths.			í Flóamanna sögu (24;754,772) er Þorfinnur vaninn af brjósti föður síns svo það eru ekki bara konubrjóst sem framleiða mjólk: 'Latur hann [Þorgils] nú saxa á geirvörtuna á sér og kemur þar blóð út. Síðan latur hann teygja það og kom þar út blanda og eigi lét hann af fyrr en það var mjólk og þar fæddist sveinninn við.' (Flóam 23;750)		

Figure 1.

Some disadvantages of working with a closed corpus should be mentioned, as shown in the example of the word *brjóst* 'breast' (cf. fig. 1). By looking the word up in the concordance, two examples of the construction *að venja af brjósti* 'to wean (a child)' can be found. In the first one, from *Fjótsdæla saga*, a woman weans her daughter. In the second example, from *Flóamanna saga*, a young boy is being weaned, but the breast in question is that of his father. These examples can indeed give a defect picture of medieval society, but on the other hand they also show that there is nothing new beneath the sun.

The advantages of working with a closed corpus are in fact far greater than the disadvantages. To name an example, the word *rass* 'ass' only appears four times in the Icelandic Family Sagas (cf. fig. 2). In all cases it refers to the ass of *Guðmundur ríki Eyjólfsson*, a great chieftain in the northern part of the country in the beginning of the 11th century. The word is used by the enemies of *Guðmundur* when trying to make fun of him, humiliate him, and thus question his power and authority. Their point of attack is *Guðmundur*'s ass, the true symbol of his supremacy and position, doubly apt as he was a very fat man.

Therefore the word *rass* does not really appear in its basic meaning in the texts, but only as a symbol of something else, i.e. as a personification. In our dictionary or database we show the basic meaning, but in addition the contextual meaning of these examples is shown in the field **Bókmenntl. aths.** 'literary remarks'. Such connotations are not usually found in ordinary dictionaries, but our database gives us the scope for such contextual meanings alongside the basic meaning.²

Orð	rass	Kyn	kk	Fsk	Vsk
Nf.et.	rass		2	Df.et.	rass 2
Dgf.et.				Ef.et.	
Nf.ft.				Df.ft.	
Dgf.ft.				Ef.ft.	
a Frl.so.	rass sveitist				
q Acccinf	kveða rass hafa freistað/leitað flestra lækjanna				
Merking 1	afturendi, both				
Bókmenntal. aths.	2LjósA, 2LjósC. Orðið er einungis notað um afturendann á <i>Guðmundi ríka</i> á <i>Möðruvöllum</i> , en hann er áþreifanlegt tákni um veldi hans (sbr. <i>Guðrún Ingvólfsdóttir</i> . 1993. <i>Að eiga mikið undir sér</i> . Orðaforði heyrjaður <i>Guðrúnu</i> kvaran. Reykjavík:37-38.)				

Figure 2.

Another example, showing the importance of the context, is the word *brottaka* 'removal of something from somebody' (cf. fig. 3). There are three examples of the word in the corpus, from three different sources, i.e. *Kjalnesinga saga*, *Kormáks saga*, and *Reykdale saga*. Although the basic meaning of the word is 'the removal of something from somebody', all the examples are of men stealing other men's wives. This specific negative usage is dominant in the Icelandic Family Sagas, and the usage shows that a wife, in medieval society, was often thought of as her husband's property. These examples also show that one result of working with a rather closed world is that a very clear picture emerges of at least some aspects of the words, making clear the interactive influence of language and society.

²Similar observations are made in the description of the Cobuild Project (cf. Moon 1987:99).

Orð	brott tak-að	Kyn	kv	Fsk	Vsk
Nf.et.				Þf.et.	brotttöku 2
Þgf.et.	brotttöku	1		Ef.et.	
Nf.ft.				Þf.ft.	
Þgf.ft.				Ef.ft.	
c Andlag spyrja brottöku e-s					
e Með fs. fyrir brottöku e-s; í brottöku e-s					
Merking 1	brotnám; rán				
Bókmenntal. aths.	1Kjaln, 1Korm, 1Reykd. Vert er að gefa gaum að því að orðið er einungis haft um rán á konum				
Aðrar athugas.	sjá brotttekja				

Figure 3.

It is not only the syntactic structure or the reference to symbolic reality that has effect on the meaning of words. Things like human behaviour also characterize the words and form their meaning. This shows better than anything else how complicated the meaning of every word can be, how delicate the interplay of the corpus and the words of that corpus can be, and how difficult the task of the lexicographer is. Despite all the traditions in dictionary making, it is indeed good to question them from time to time.

In our project, the text leads the way, and a broad overview of the whole text is necessary, as the true meaning of a word can be hidden at some distance to the actual occurrence of the word itself. To name an example, the word *blíðmæli* is explained in Fritzner's *Ordbog over Det gamle norske Sprog* (1883–96) as 'venlig Tale' (i.e. 'gentle talk'), quoting one example from *Flateyjarbók* and two from *The Lives of Saints*, but none from the Icelandic Family Sagas. At first sight, this meaning is also the right one for the occurrence in the Icelandic Family Sagas (cf. fig. 4). Closer scrutiny of the story itself indicates that the word may have the supplementary meaning 'flattery', i.e. 'gentle talk and flattery.'

Orð	blíð+mæli-ið	Kyn	hk	Fsk	Vsk
Nf.et.				Þf.et.	
Þgf.et.				Ef.et.	
Nf.ft.	blíðmæli	2		Þf.ft.	
Þgf.ft.				Ef.ft.	
a Erl.so.	blíðmæli tjóa				
o Samsetn	#ógnir né blíðmæli				
Merking 1	vingjarnleg orð				
Málfræðil. aths.	ef marka má Fritzner er orðið ótvírætt ft í fornu máli				
Merkingarl. aths.	vert er að kanna hvort orðið hefur merkingarpáttinn smjaður, fagurgali!				
Bókmenntal. aths.	1LjósA, 1LjósC				

Figure 4.

Another such example is the word *útibúr* 'out-house, store-house' (cf. fig. 5). Dictionaries, including Fritzner's, explain it as 'a store-house, detached from other houses on the farm'. In the Icelandic Family Sagas, however, there is one example where the store-house is adjacent to the kitchen.

Orð	úti+búr%	Kyn	hk	Fsk	Vsk
Nf.et.	útibúr	4		Pf.et.	útibúr 11 4
Þgf.et.	útibúri	12	4	Ef.et.	útibúrs 3 2
Nf.ft.				Þf.ft.	útibúr 2
Þgf.ft.				Ef.ft.	
a Frl.so.	útibúr stendur í túni				
b Frl.sf.	útibúr er opið/óbrunnið				
c Andlag	brjóta upp útibúr, eiga/sjá útibúr				
e Með fs.	+í útibúr; að/af/á/í/úr útibúri, vaka yfir útibúri; til útibúrs				
f Með ek.	mikill/(ákaflega) stór				
g Sem ek.	dyr útibúrs				
j Ópersl.	útibúr skortir eigi				
Merking 1	afhýsi eða lítið hús (oftast) laust frá íbúðarhúsum þar sem matur var geymdur & matarbúr, matargeymsla				
Merking 2	skemma, geymsla				
Málfræðil. aths.	vera áfast útibúri er þgf með lo				
Aðrar athugas.	í einu tilvikum er útibúrið áfast eldhúsi. Sjá búr				

Figure 5.

Working so closely with the text, examples of meaning not found in other dictionaries are often found. Such an example is the word *tiltekja* (literally, 'what a man takes to doing'), where extant dictionaries have the explanation 'an expedient, contrivance', but not the additional meaning 'intention, plan' (cf. fig. 6).

Orð	til+tek+j-a%	Kyn	kv	Fsk	til	Vsk	j
Nf.et.	tiltekja	5	1	Pf.et.			
Þgf.et.	tiltekju	1		Ef.et.			
Nf.ft.				Þf.ft.	tiltekjur	8	
Þgf.ft.	tiltekjum	1		Ef.ft.			
a Frl.so.	e-m mislíkar tiltekja						
b Frl.sf.	tiltekja er meyleg, tiltekja þykir ill/óvarleg/undarleg						
c Andlag	hafa/spyrja tiltekju, vita tiltekju e-s; ráða tiltekju sinni						
e eð fs.	fyrir tiltekju sína, (ræða) um tiltekju; yfir tiltekju e-s						
f Með ek.	annar/hann/hún/sá/sjá/sinn/þessi/pinn						
o Samsetn	tiltekja og drottinssvik/hneisa						
p Andh.							
Merking 1	tiltæki, uppátæki; verk, athöfn						
Merking 2	áform						
Merkingarl. aths.	merk. 2 er ekki að finna í öðrum orðabókum						

Figure 6.

3 The List of Semantic Classification

I will now turn to the list of semantic classification. At the outset of the project we decided to define each part of speech separately, starting with the nouns. We thought this procedure would make the task of defining words easier and ensure greater consistency. Soon we also saw that we could improve this method considerably by classifying all the nouns semantically. Such a semantic classification of nouns has never before been used in an Icelandic dictionary.

It has often been maintained that the subject matter of the Icelandic Family Sagas is very narrow, and this is in some respects true. The list below, containing the table of contents of the semantic classification, shows, however, that the semantic classes touch upon nearly every aspect of life.

As shown below in (1), there are 21 primary classes in the list as it now stands, and the total number of classes comes to 54, when the secondary classes are included. The classes are ordered by size, and the relative size of the semantic classes is interesting in itself. The largest class (about 8 pages) is the class *Human nature*. For readers of the Icelandic Family Sagas, this is hardly surprising, as their main subject is man and nature, and *Nature* is the second largest of the classes (about 4 pages). Next in line are the classes *Culture and Learning* and *Conflict* (about 3 pages each). Then comes *Economy* (2½ pages), and *Language of Legal Affairs and Administration* (more than 2 pages). *Human Condition* and *Buildings* are 2 pages each, but other semantic classes are 1 page or less.

The size of the classes emphasizes the well-known fact that the Icelandic Family Sagas describe a feudal-rural community, where man, nature and feelings are in place of honour. It is, however, rather difficult to guess what the result of applying the same classification to all the words in a dictionary such as Fritzner's *Ordbog* would be, as his corpus is very much larger. The outcome would certainly be bound to be interesting!

(1)

<i>Human nature</i>	People Characteristics Thinking, Perception, Feelings
<i>Nature</i>	Animals Heaven and Earth Weather Vegetation Source of Energy, Metals, Minerals
<i>Conflict</i>	Warfare Armament
<i>Culture and Learning</i>	Story-telling and Studies, Poetry and other Arts Language Written Language Quantity, Directions, Numbers Time Games and Sports
<i>Economy</i>	Agriculture Fishing and Hunting Trade and Merchandise Handwork (Needlework, Sewing, Knitting) Construction
<i>Language of Legal Affairs and Administration</i>	
<i>Buildings</i>	Houses and Dwellings Other Buildings

<i>Human condition</i>	Death and Life Healing Wounds and Injury Diseases Human body Cleaning and Leftovers
<i>Transport</i>	Travels (in general) Travels on Land Ships and Sailings
<i>Tools, Instruments and Vessels</i>	
<i>Respect and Power</i>	Wealth Gifts Festivals, Celebrations and Hospitality
<i>Clothes and Ornament</i>	
<i>Religion</i>	Places/Houses Objects Creatures Events
<i>Family</i>	Children and Upbringing Family Mariage
<i>Food and Drink</i>	
<i>Women</i>	
<i>Fortune and Misfortune</i>	
<i>Inhabited Area</i>	Countries Urban Area and Rural Area
<i>Dreams</i>	
<i>Form</i>	
<i>Deficiency/Lack, Difference</i>	

When we started to classify all the nouns semantically, we tried to give an overview of most of the human and cultural aspects of everyday life, but we also had to take notice of the kind of corpus we were working on. We then proceeded to classify every noun according to its basic meaning. In the pilot list we did not take any notice of secondary meaning, but each word could be placed in more than one semantic class. Three versions of this list have now been made, and in the second version the semantic classes are arranged like in a thesaurus.

The semantic classes are not all classified by this system (e.g. *Language of Legal Affairs and Administration, Clothes and Ornament, Food and Drink, Tools, Instruments and Vessels, Deficiency/Lack, Difference*, etc.). Some of them are too small, and others have not been fully defined yet. The question of how exact the classification should be also remains open.

In the beginning, the classification was not very thorough, mainly because the size of most of the semantic classes did not really call for too much detail, i.e. the lexicographer could easily gain a fairly good grasp of the whole of each class without it.

We have, however, made some experiments with a more detailed classification, such as shown in the class *Religion*. After finishing the work on this semantic class, the person concerned revised the classification, giving the end result shown in (1) above. This semantic class is rather simple, and I think that the result is convincing. I am not totally satisfied, however, because this kind of classification is often bound to make things very complicated. The problem is that the system allows words to be a part of more than one primary class, and furthermore, they can also be a part of more than one secondary class. This, I fear, will make the use of this list rather confusing. I am aware of the fact that not everyone agrees with me.

The list of semantic classification has been revised a number of times, in order to make it a better tool for scholars who are working on the Icelandic Family Sagas. Many scholars from different fields of study have obtained material from the list, using the information in connection with the concordance. This considerably speeds up the search for all examples pertaining to the things they are researching.

There are mainly two advantages of using a list of semantic classification, such as ours, in a dictionary project. Firstly, it gives the lexicographer a unique perspective of the semantic field being worked on each time, and prevents concepts with the same meaning from being defined in different ways. At the same time, it can reveal subtle semantic differences, which would otherwise be unnoticed. Secondly, it can sometimes reveal more clearly how related concepts differ, e.g. in usage, etc. — Those are the same features as found in the Cobuild Project.

When we decided to classify the nouns semantically, we mainly had two things in mind. In the first place, we wanted to improve the processing method, and secondly, we felt the need for some progressive steps to be seen, by making some parts of the project useful before the completion of the project, especially as it is such a huge one. The list of semantic classification is such a progressive step. It is both a useful tool in the dictionary project itself, and in various research on the Icelandic Family Sagas. — Later this year, when the concordance will be published, the list of semantic classification will also be published with it, and thus get a still greater distribution.

References

- Eiríkur Rögnvaldsson 1995: A Concordance to Old Icelandic Texts and its Lexicographic Value. [This volume, 123–135.]
- Fritzner, Johan 1883–96: *Ordbog over Det gamle norske Sprog. Omarbejdet, forøget og forbedret Udgave*. I–III. Kristiania: Den norske Forlagsforening.
- Fritzner, Johan 1972: *Ordbog over Det gamle norske Sprog IV. Rettelser og tillegg ved Finn Hødnebo*. Oslo/Bergen/Tromsø: Universitetsforlaget.
- Moon, Rosamund 1987: The analysis of meaning. In: J. M. Sinclair (ed.): *Looking up*. An account of the COBUILD Project in lexical computing. London: Collins.

En historisk ordbog og dens aner

The subject of this paper is a historical dictionary being worked on at present at the Institute of Lexicography in Reykjavík, and the value of older dictionaries as sources for that work. The problems of excerpting from older material, especially from early periods where sources can be insufficient or incomplete, inevitably makes older dictionaries a prominent source for historical dictionaries, no matter how deficient or incomplete they are. An account is given of the major Icelandic dictionaries, both in manuscript and in print, from the 17th to the early 19th century.

1 Indledning

Interessen for historiske ordbøger kan spores tilbage til midten af forrige århundrede da brødrene Grimm tog fat på arbejdet med deres store ordbog, *Deutsches Wörterbuch*. Det første hæfte udkom 1852 og det første bind blev afsluttet 1854. I romantikkens ånd havde brødrene den hensigt at indsamle materiale til og redigere en ordbog som ikke kun var beregnet for de lærde men for almindelige mennesker uden sociale skel (Dückert 1987:10). En ordbog skulle ikke være et opslagsværk, men den skulle læses i sammenhæng som en almindelig bog, igen og igen, og på den måde skulle den tjene til at oplyse almenheden om folkearven, sproget. I ordbogens fortale skrev Jacob Grimm bl. a.:

warum sollte sich nicht der vater ein paar wörter ausheben und sie abends mit den knaben durchgehend zugleich ihre sprachgabe prüfen und die eigne anfrischen? die mutter würde gern zuhören. frauen, mit ihrem gesunden mutterwitz und im gedächtnis gute sprüche bewahrend, tragen oft wahre begierde ihr unverdorbenes sprachgefühl zu üben, vor die kisten und kasten zu treten, aus denen wie gefaltete leinwand lautere wörter ihnen entgegen quellen (Grimm 1854:XIII).

I fortalen redegør han også for formålet med ordbogen:

Es soll ein heiligthum der sprache gründen, ihren ganzen schatz bewahren, allen zu ihm den eingang offen halten. das niedergelegte gut wächst wie die wabe und wird ein hehres denkmal des volks, dessen vergangenheit und gegenwart in ihm sich verknüpfen (Grimm 1854:XII).

Selv om Jacob Grimms høje tanker, som kom til udtryk i fortalen, ikke alle nåede at blive realiseret, og selv om han i begyndelsen ikke havde en anelse om hvor lang tid arbejdet ville tage, havde det med det samme indflydelse på ordbogsarbejder i andre lande.

Richard Trench var en af dem som blev begejstret for dette tyske ordbogsværk. I to kendte artikler 1857 kritiserede han engelsk ordbogsarbejde, og han havde en væsentlig andel i grundlæggelsen af den ordbog som senere blev udgivet under titelen *Oxford English*

Dictionary. Trench's artikler blev publiceret af The Philological Society under titelen *On Some Deficiencies in Our English Dictionaries* (Landau 1989:67).

Interessen spredte sig også til Norden. Omkring 1900 startede Verner Dahlerup arbejdet med *Ordbog over det danske sprog* under indflydelse fra de førnævnte ordbøger, og *Svenska Akademiens ordbok* var man gået i gang med lidt før. Man lagde vægt på at se nutidssproget i historisk perspektiv og basere alle oplysninger på autentiske belæg.

2 En historisk ordbog på Island

På Island gjorde interessen sig også gældende. Her begyndte man at arbejde på en historisk ordbog i 1917 og fortsatte indtil 1928, da der blev lukket for de bevillinger som der blev givet til projektet. Her gik man igen igang omkring 1944, og *Orðabók Háskólans* fejrede i fjor sit 50 års jubilæum. Man kan sige at de ordbøger jeg allerede har nævnt er vores ordbogs aner, men det er først og fremmest dens islandske aner som jeg vil beskæftige mig med i det følgende.

Som allerede er fremkommet er hensigten med en historisk ordbog at give så udførlig fremstilling som muligt af et sprog i den periode vedkommende ordbog skal dække — for Universitetets ordbogs vedkommende perioden fra 1540 til nutiden. For at opnå dette må der indsamles materiale fra alle periodens vigtigste kilder, således at samlingen indeholder et materiale som er tilstrækkeligt til at man kan følge ordforrådets udvikling og de enkelte ords eventuelle betydningsændringer. Ved ordforrådets udvikling forstås, at nye ord bliver til, både låneord og nydannelser, medens andre ord bliver forældet og forsvinder ud af sproget. Betydningsudviklingen af enkelte ord skulle fremgå af en historisk ordbogs materialesamling, hvor man kan finde det samme ord i forskellig kontekst fra forskellige tider. I visse tilfælde vil det fremgå af disse belæg at ordets betydning ikke altid er den samme, den kan skifte således, at en gammel betydning går af brug og en ny afløser den. Fremstillingen af materialet skal være deskriptiv, den skal vise ordforrådets historiske udvikling på en objektiv måde, ikke være normativ ved at udtrykke en positiv eller negativ vurdering. Jo større og bedre det indsamlede materiale er desto holdbarere skulle de oplysninger være som man kan uddrage af det.

Orðabók Háskólans begynder sin excerpering ved året 1540, det år den første bevarede trykte islandske bog udkom, Oddur Gottskálkssons oversættelse af det Ny testamente, som blev trykt i Roskilde. Island fik sit første trykkeri allerede i 1530-årene, men indtil 1770-årene var der kun ét trykkeri i landet, som ejedes det meste af tiden af individuelle biskopper, og som følge heraf blev der nærmest kun trykt kirkelig litteratur i denne periode. Denne litteratur var i mange tilfælde oversat fra dansk eller tysk og dens ordforråd og stil var ofte i høj grad påvirket af forlæggene. Såvel denne omstændighed som disse teksters forholdsvis begrænsede emneomfang gør at teksterne kun kan give et begrænset billede af det levende sprog.

Hele ordforrådet kan derfor aldrig excerperes fra trykte tekster, hverken enkeltord eller ordforbindelser. Man må supplere med materiale fra andre kilder. En meget omfattende litteratur fra 1500-, 1600- og 1700-tallet er overleveret i håndskrifter. En del af disse tekster er kommet på tryk senere — især i dette århundrede — og er stort set blevet excerperet for *Ordbogen*. En overvejende del er imidlertid uudgivet og må derfor excerperes fra håndskrifterne (eller i senere tid normalt håndskriftfotografier). En vis excerpering af dette

materiale har fundet sted. Det drejer sig f.eks. om digtsamlinger, brevbøger og visse oversatte verdslige værker. Denne excerpering kunne udvides, også når det gælder utrykt materiale fra de sidste århundreder, f.eks. privatbreve og dagbøger. Sådanne kilder kan udvide vores kendskab til det ordforråd der var knyttet til folks daglige liv og arbejde.

Andre kilder, som vi kan søge til, er ordbøger og ordsamlinger som kan supplere vores viden om det ældre ordforråd. Disse samlinger bygger ikke kun på trykte og utrykte skrevne kilder. De meddeler også ordforråd som hørte til forfatternes talte sprog men ikke viser sig i deres samtids trykte bøger. I visse fald er sådanne ord heller ikke fundet i senere trykte bøger, men er blevet noteret i de sidste årtier gennem Ordbogens indsamling fra talesproget. Jeg har tidligere på en tilsvarende konference fortalt om, hvordan Ordbogen har draget nytte af den almene sproglige interesse og indsamlet værdifuldt materiale fra talesproget. Mit foredrag denne gang vil handle om ældre ordbøger og ordsamlinger, såvel trykte som håndskrevne, og deres betydning ved indsamling af materiale til en historisk ordbog. For en leksikograf, som arbejder på en historisk ordbog, er det meget vigtigt at kende kontinuiteten i leksikografien og forstå hvorledes det ene ordbogsværk bygger på det andet fra de ældste ordbøger og op til nutiden.

3 Det 17. århundrede

3.1 Specimen lexicis runici

Fra 1600-tallet er der bevaret tre islandske ordsamlinger, to trykte og én i manuskript. Den ene af de udgivne samlinger var udarbejdet af præsten Magnús Ólafsson på Laufás. Efter hans død lod Ole Worm samlingen trykke i København 1650 under titelen *Specimen lexicis runici*, og forinden havde han ansat en ung islænding, Guðmundur Andrésón, til at gennemgå håndskriftet. Det vides at Guðmundur foretog en vis supplerings af materialet og der er også tilføjelser fra andre samtidige, både islandske og danske.

Der er foretaget en grundig undersøgelse af hvilke kilder samlingen er baseret på. Det drejer sig om nogle islændingesagaer, kongesagaer, eddaerne og gamle lovtekster. Imidlertid forekommer også ord, talemåder og ordsprog som det ikke lykkedes at finde i disse gamle kilder, og her er der sandsynligvis tale om levende sprog på 1600-tallet som er en gevinst for en historisk ordbog (Faulkes 1964). *Specimen* giver ikke noget større tilskud til vores ordbogs eksempelsamlinger, idet det først og fremmest skulle dække det gamle sprog og deres kilder derfor hovedsagelig fra før 1540. Excerpteringen for ordbogen gav dog godt 200 ord som ikke forekom i de middelalderlige tekster som *Specimen* for det meste byggede på. I 130 tilfælde drejede det sig om Ordbogens ældste belæg. Blandt dem er der 11 verber og af dem er der 8 som har været almindelige i sproget til den dag i dag. 34 er adjektiver, langt de fleste sammensatte af substantiv + adjektiv eller to adjektiver; kun 5 er usammensatte, hvoraf 4 stadig er almindelige i sproget. Af substantiverne er der 30 usammensatte; nogle af dem er kendt fra andre middelaldertekster end dem der *Specimen* byggede på, men i de fleste tilfælde er belæggene i *Specimen* Ordbogens ældste, og adskillige af dem er hapax-legomena i Ordbogens samling.

3.2 *Lexicon Islandicum*

Guðmundur Andrússon arbejdede selv på en islandsk ordsamling, men han døde (1654) inden den var trykkesfærdig. Den udkom i København 1683 på professor P. H. Resens initiativ under titelen *Lexicon Islandicum*. Resen har ikke haft tilstrækkelig kendskab til det islandske sprog og tilsyneladende har han heller ikke haft nogen kvalificeret medhjælper. Bogen vrimler med trykfejl, og den har længe fået dårligt omdømme. F.eks. Rasmus Kristian Rask erklærede den for ubrugelig (1811:XXXVIII), og sådanne vurderinger har bevirket at den gennem tiderne nærmest forblev upåagtet af forskere. Leksikografer har dog i de senere årtier gjort sig klart at *Lexicon Islandicum* indeholder mange ord fra 1600-tallets talesprog, hvoraf en del ikke dukker op i samtidens trykte skrifter. Herfra har Ordbogen i mange tilfælde sine ældste belæg på ord og ordforbindelser som får støtte af yngre skriftlige kilder, i nogle tilfælde fra de sidste årtiers talesprog og i andre tilfælde viser de sig at have paralleler i nabosprogene. Alt dette kan man tydeligt se eksempler på i Ásgeir Blöndal Magnússons etymologiske ordbog, *Íslensk orðsifjabók*, som blev udgivet af Orðabók Háskólans 1989.

Guðmundur Andrússon angiver kilderne til sine talesprogseksempler kun generelt ved at sige at sådan og sådan siger folk (aiunt, vocant, ait plebs) (Gunnlaugur Ingólfsson 1993:45). Iøvrigt er *Lexicon Islandicum*s ordforråd for størstedelens vedkommende hentet fra middelalderlitteraturen.

Lexicon Islandicum — med rettelser der delvis er hentet i et håndskrift der ikke er baseret på den trykte udgave — foreligger nu i maskinlæsbar form hos Orðabók Háskólans, således at dens indhold let kan sammenholdes med andet ordbogsmateriale. Den bliver snart udgivet i instituttets ordbogsserie, *Orðfræðirit fyrri alda*.

3.3 Guðmundur Ólafsson

Fra 1600-tallet foreligger der en anden, ret stor, islandsk ordsamling med latinske forklaringer i et håndskrift i Kungliga biblioteket i Stockholm. Forfatteren er Guðmundur Ólafsson som fra 1681 til sin død i 1695 arbejdede for Antikvitetskollegiet i Sverige. En del af dette ordbogsarbejde er gået tabt, men ialt er der bevaret 16 bind, til dels i udkast og til dels i renskrift. Størstedelen af udkastet er bevaret — der mangler kun bogstaverne *e*, *g*, *p* og *r*, men af renskriften, som er på 7000 sider, er der kun bevaret den sidste del af bogstavet *b*, bogstaverne *c*, *d* og *e* og bogstavet *s*.

Tilsyneladende har forfatteren sat sig for at omfatte hele ordforrådet, og han gik meget langt med hensyn til at anføre lange rækker præfigerede og sammensatte ord. Han bruger 1200 sider på at opremse ord med præfikset *sí-* i betydningen 'altid', som f.eks. *sígóður*, *síplokkaður*, *síráðandi*, *síprúður*, *síráðalaus*, *síraður*, *síraulandi*, *síróandi*, *sísjóðandi*, *sísjúklegur*, og 842 sider med ord med præfikset *all-* i betydningen 'ret, temmelig', f.eks. *alldjúpur*, *alldökkur*, *allfagur*, *allgóður*, *allgrettur*, *allvondur*.

Sammensatte ord er karakteriserende for bogen, f.eks. har 40 ord *brúðar-* 'brude-' som første sammensætningsled, hvoraf Ordbogen kun har belæg på ca. halvdelen fra andre excerperede kilder. Blandt disse ord er *brúðarbekkur*, *brúðardans*, *brúðardiktur*, *brúðardragt*, *brúðarhár*, *brúðarhrós*, *brúðarhús*, *brúðarkyrtil*, *brúðarnafn*, *brúðarstóll* og *brúðarsæng*. Alle disse ord er gennemsigtige og let forståelige, således at de nok ikke alle ville blive optaget i en moderne historisk ordbog, men i betragtning af eksemplernes alder er det klart at forekomsten af flere af dem er af kulturhistorisk interesse, idet den hjælper til at kaste

lys over tidligere generationers sæder og skikke. Om mange af de sammensatte ord gælder at man kan tvivle på at de nogensinde har tilhørt almindelig sprogbrug, og det samme gælder en del suffigerede ord som det nok ikke er usandsynligt at Guðmundur selv har lavet under sit arbejde. Som eksempler kan man tage det store antal nomina actionis dannet af de fleste verber med afledningsendelsen *-an* (eller *-un*). Blandt sådanne ord kan man nævne *bröltan* af *brölta* 'sprælle, fægte', *brúan* af *brúa* 'bygge bro', *bruggan* (ved siden af *bruggun*) af *brugga* 'brygge', *brunan* af *bruna* 'bruse frem', *bryðjan* af *bryðja* 'knase mellem tænderne' og så videre.

En undersøgelse af dette store ordbogsmanuskript har dog vist at til trods for visse skavanker indeholder det et værdifuldt supplement til andre kilder til islandsk ordforråd på 1600-tallet og viser i mange tilfælde det første belæg på et stort antal ord som er velkendte fra yngre kilder.

Guðmundur Ólafssons store samling har aldrig været anvendt ved senere leksikografiske arbejder, men nu er man på Ordbogen i gang med intasteringen af dette værk, således at det i maskinlæsbar form lettere kan sammenlignes med andre kilder. Alle disse ordbogsværker fra 1600-tallet er det nødvendigt for Ordbogen at benytte for at udfylde huller med hensyn til de oplysninger andre sprogkilder fra denne periode kan give om det almindelige ordforråd.

4 Det 18. århundrede

4.1 Jón Ólafsson úr Grunnavík

Det ene af to hovedværker inden for islandsk leksikografi på 1700-tallet er den ordbog som Jón Ólafsson fra Grunnavík anlagde gennem næsten et halvt århundrede og samlede materiale til, medens han, det meste af tiden, syslede med mange andre projekter i den arnamagnæanske samling i København, hvor hans værk i 9 store foliobind foruden et stort antal sedler er bevaret og forventes her til landet i nær fremtid. De fleste af hans oversættelser og forklaringer er på latin, men i visse tilfælde på dansk eller islandsk. Professor Jón Helgason, som i sin disputats om Jón Ólafsson beskrev dette arbejde, skrev at når man gik i gang med at udarbejde den store ordbog som viste det islandske sprogs udvikling gennem århundrederne, ville det være nødvendigt at gennemgå Jón Ólafssons ordbog på en omhyggelig måde og sammenligne andre kilder med dens oplysninger; uden tvivl ville det vise sig at den indeholdt mange ord som man ikke havde ældre belæg på og at det i yderst mange tilfælde blev nødvendigt at henvise til dette upublicerede arbejde (Jón Helgason 1926:121).

Det har Jón Helgason haft ret i. På grundlag af fotografier af hele ordbogen — bortset fra de senere tilføjede sedler — blev den for nogle årtier siden excerperet af Ordbogens chef, Jakob Benediktsson (1994), og materialet har vist sig at være en yderst betydningsfuld kilde til islandsk sprog på 1700-tallet, som man bl. a. kan se af Ásgeir Blöndal Magnússons førnævnte etymologiske ordbog. I Jón Ólafssons ordbog finder vi ikke kun ord som han har hentet fra ældre og yngre skriftlige kilder, men også fra 1700-tallets talte sprog, ordforbindelser og ordsprog, således at vi ingen steder har en lige udførlig kilde til dagligsproget i denne periode. Dette materiale har i et vist omfang været brugt af leksikografer og andre forskere som har beskæftiget sig med historiske undersøgelser af islandsk sprog og litteratur. For nylig blev opslagsordene indtastet og de viste sig at være opimod 48 000. Denne ordliste bliver i nær fremtid koplet til Ordbogens øvrige maskinelt tilgængelige materiale.

4.2 Nucleus latinitatis

De leksikografiske arbejder fra tidligere århundreder, som jeg indtil nu har omtalt, har det til fælles at de indeholder islandske opslagsord med latinske oversættelser og forklaringer. Året 1738 udkom der på den anden side en latinsk-islandsk ordbog, *Nucleus latinitatis*, af biskop Jón Árnason på Skálholt. Fire år tidligere havde Jón Árnason udgivet en mindre ordsamling, *Lexidion Latino-Islandicum Grammaticale*, som skulle tjene en elementær tilegnelse af latinsk ordforråd. Begge disse ordbøger blev excerperet for Ordbogen, og det gav ca. 10 000 ordbogssedler. En senere sammenligning har vist at denne excerpering burde have været endnu mere omfattende.

Forbilledet for Jón Árnasons *Nucleus latinitatis* var en bog med samme titel, forfattet af Hans Gram, senere professor og rektor ved Københavns Universitet. Det nøjagtige årstal for førsteudgaven af Grams *Nucleus* har hidtil været ukendt, idet intet eksemplar af den vides at være bevaret, men af biskop Jón Árnasons brev til kongen 1732 fremgår det at han da har været i færd med at oversætte en *Nucleus latinitatis* som var udgivet 1712 i København, og det er formodentlig årstallet for førsteudgaven, idet andenudgave udkom 1722. Efter den latinsk-danske tekst følger der en dansk ordliste med henvisning til side og linje hvor disse ord fremkommer i ordbogsteksten.

Et hovedkendemærke ved Jón Árnasons *Nucleus* er hvor mange islandske oversættelser han ofte giver af et enkelt latinsk opslagsord. Han bestræber sig på i sine oversættelser at dække de latinske ords betydning ikke bare i guldalder- og sølvalderlatin men også i sin samtids humanistlatin. Som eksempel på Jón Árnasons rige islandske ordforråd i oversættelserne kan det nævnes at som oversættelser af *vulpus* 'ræv' anfører han ni islandske ord, som med én undtagelse er velkendte i islandsk nutidssprog.

Ordbogens islandske ordforråd består af 18 888 ord, når man har udeladt pronomener, de mest almindelige adverbier, konjunktioner og præpositioner. Når man undersøger dette ordforråd viser det sig at omfatte adskillige ord som ikke er dukket op i excerperede tekster. Nogle af disse ord har Jón Árnason sikkert dannet for at nå det latinske ords betydning. Ofte anfører han fremmede ord i den forklarende tekst, men sådanne ord markerer han ved kursivering. Seddelsamlingen indeholder ca. 1100 hapax-legomena fra *Nucleus latinitatis*, næsten udelukkende sammensatte ord. Kun 38 ord kan regnes for usammensatte.

Nomina agentis med endelsen *-ari* føjet til en verbalstamme er almindelige. Med sådanne ord oversætter Jón Árnason som regel latinske ord med endelserne *-arius* og *-tor*. Disse nomina agentis angiver som regel en persons natur eller hverv. *Nucleus* viste sig at indeholde ca. 350 sådanne ord, og heraf er der 114 hapax-legomena i Ordbogens seddelsamling.

Nomina actionis med endelsen *-ing* danner en anden stor gruppe i *Nucleus latinitatis*, ialt på godt 500 ord; i 150 tilfælde er der tale om Ordbogens ældste eksempler ifølge den maskinlæsbare liste over alle opslagsord i det seddelmateriale der er baseret på excerpering af sammenhængende tekster. De mest almindelige suffixer i de af Jón Árnason oversatte nomina actionis er *-io* og *-tio*.

I betragtning af *Nucleus*' værdi som leksikografisk kilde blev den genudgivet i *Orðfræðirit fyrri alda* sidste år. Efter optrykket af den gamle udgave blev der trykt en liste over de islandske ord som bruges i oversættelserne (Jón Árnason 1994).

Det skal her præciseres at udgivelsen af Ordbogens serie, *Orðfræðirit fyrri alda*, tjener et dobbelt formål, dels at gøre gamle og sjældne leksikografiske arbejder tilgængelige i trykt

form for interesserede, dels at gøre dem og de islandske ordlister, som er udarbejdet på grundlag af dem, tilgængelige for leksikografer og andre såvel i trykt som maskinlæsbar form.

4.3 Ordlistor i håndskrifter

Jón Árnason var en meget arbejdsom mand og han har bl.a. efterladt sig mange utrykte skrifter om forskelligartede emner. Et af disse skrifter er en islandsk-latinsk ordbog. Håndskriftet, der dateres til ca. 1740, er på 196 sider og uden større defekter; det mangler dog bogstaverne *h* og *i(j)*. Dette arbejde har sandsynligvis oprindeligt været tænkt som en ordliste, svarende til den danske ordliste i Grams *Nucleus*. Denne plan har Jón Árnason tilsyneladende opgivet og i stedet tænkt sig at udgive en selvstændig islandsk-latinsk ordbog. Der findes en noget afvigende version af dette værk i et andet håndskrift fra 1700-tallet med kun én mindre lakune, Gl. kgl. Saml. 2394 4to (nu i Stofnun Árna Magnússonar á Íslandi). Det sidstnævnte håndskrift indeholder et noget større ordforråd end Jón Árnasons egenhændige manuskript. Sammenligner man disse to håndskrifter med den islandske ordliste som er baseret på den islandske udgave af *Nucleus* og trykt med Ordbogens udgave af den, viser det sig at håndskrifterne indeholder en del ord som ikke findes i den trykte ordbog, således at de har en selvstændig værdi som kilder til islandsk ordforråd på 1700-tallet (Guðrún Kvaran 1994: XIV–XVI). Disse håndskrifter burde derfor gøres til genstand for en nærmere undersøgelse og deres tekster gøres maskinelt tilgængelige for at lette arbejdet med dem.

Der findes nogle andre håndskrevne ordlister fra 1700-tallet. Kun fire små samlinger er blevet udgivet af Jón Helgason i Bibliotheca Arnarnagæana XX (Jón Helgason 1960). Den femte ordliste, som Jón Helgason udgav der, er en samling som er udarbejdet af Rasmus Rask på Island; dette er en liste over udtryk hvis brug ifølge Rasks mening var dialektal.

Orðabók Háskólans har gennem årene excerperet en del uudgivne ordlister, medens andre ikke endnu er blevet benyttet, men de kan sikkert også give ny viden, bl.a. om udtryk fra talesproget, som det ikke er sikkert at man kan finde i andre skrevne kilder. Der kan ikke her gøres nærmere rede for disse ordlister, men erfaringen viser at de udfylder huller med hensyn til kendskabet til islandsk ordforråd på 1700-tallet (Guðrún Kvaran 1988).

Mod slutningen af 1700-tallet voksede islandsk bogudgivelse væsentlig i oplysningens ånd. Man udgav mange slags belærende skrifter som var beregnet til almenhedens åndelige udvikling og oplysning. Dermed bliver det ordforråd der foreligger i trykte bøger betydeligt bredere.

4.4 Lexicon Islandico-Latino-Danicum

Dette øgede ordforråd viser sig også i det sidste leksikografiske arbejde jeg her vil omtale. Det blev udarbejdet sidst på 1700-tallet af præsten Björn Halldórsson på Sauðlauksdalur på Vestfjordene. Fra Björn Halldórssons side var dette en islandsk-latinsk ordbog, men Björn døde inden den blev udgivet, og da Rask forberedte den til trykken forøgede han den med danske oversættelser. Den udkom i København i 1814 og blev genudgivet i *Orðfræðirit fyrri alda* for tre år siden (Björn Halldórsson 1992). Den maskinelt læsbare tekst gør det lettere at sammenligne dens ordforråd med det ordforråd som er indsamlet fra excerperede bøger. Den sammenligning giver til resultat at Björn Halldórssons ordbog har 2063 ord som ikke

er blevet excerperet fra sammenhængende tekster. Af disse 2063 ord kunne man finde 828 i ældre ordbøger og ordlister, trykte og utrykte, medens der ikke fandtes andre kilder til 1235 ord. Mange af disse ord er komposita, nogle af dem flersammensatte, og mange adjektiver med det privative præfiks *ó-* var iøjnefaldende. Det er imidlertid klart at denne ordbog, ligesom de ordbøger jeg har nævnt i det foregående, er uundværlig ved udarbejdelsen af en historisk ordbog over det islandske sprog. Desuden har Björn Halldórsson spillet en historisk rolle ved at være en vigtig kilde til flere af de ordbøger der er udgivet i dette århundrede.

5 Afslutning

Jeg har her givet en kort oversigt over nogle af de vigtigste leksikografiske arbejder på 1600- og 1700-tallet. Jeg har lagt vægt på disse værkers store betydning for indsamlingen af materiale til en historisk ordbog, idet de udgør meget vigtige supplementter til de oplysninger om tidligere århundreders ordforråd som vi kan nå frem til ved udelukkende at excerperer sammenhængende tekster. Ganske vist giver kun få af disse gamle ordbøger belæg på de enkelte ord i en kontekst, men hver for sig er disse ordbøger og ordlister kilder til islandsk ordforråd og de skal betragtes og benyttes som sådanne.

Bibliografi

- Ásgeir Blöndal Magnússon 1989: *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Björn Halldórsson 1992: *Orðabók. Íslensk-latesk-dönsk*. Eftir handriti í Stofnun Árna Magnússonar í Kaupmannahöfn. Ný útgáfa. Jón Aðalsteinn Jónsson sá um útgáfuna. Orðfræðirit fyrri alda II. Reykjavík: Orðabók Háskólans.
- Dücker, Joachim [Hrsgb.] 1987: *Das Grimmsche Wörterbuch. Untersuchungen zur lexicographischen Methodologie*. Leipzig: S. Hirzel Verlag.
- Grimm, Jacob und Wilhelm Grimm 1854: *Deutsches Wörterbuch*. Erster Band. Leipzig: Verlag von S. Hirzel.
- Faulkes, Anthony 1964: The Sources of Specimen Lexici Runic. I: *Lingua Islandica – Íslenzk tunga*. Tímarit um íslenzka og almenna málfraði 5, 30–138.
- Guðrún Kvaran 1988: Sérsöfn Orðabókar Háskólans. I: *Orð og tunga* 1, 151–164.
- Guðrún Kvaran 1994: Inngangur. I: Jón Árnason. *Nucleus latinitatis*. Ný útgáfa. Reykjavík: Orðabók Háskólans, ix–xxix.
- Gunnlaugur Ingólfsson 1993: Um orðabók Guðmundar Andréssonar. I: *OrðAForði heyjaður Guðrúnu Kvaran 21. júlí 1993*, 45–47.
- Jakob Benediktsson 1994: Glíman við orðabók Jóns Ólafssonar. I: *Hræringur úr ritum Grunnavíkur-Jóns*. Erindi flutt á málþingi um Jón Ólafsson úr Grunnavík laugardaginn 16. apríl 1994, 19–22.
- Jón Árnason 1994: *Nucleus latinitatis*. Ný útgáfa. Guðrún Kvaran og Friðrik Magnússon sáu um útgáfuna. Orðfræðirit fyrri alda III. Reykjavík: Orðabók Háskólans.
- Jón Helgason 1960: Fem islandske ordsamlinger fra 18. og 19. århundrede. I: *Bibliotheca Arnarnagana XX, Opuscula I*, 271–299.
- Jón Helgason 1926: *Jón Ólafsson frá Grunnavík*. Safn Fræðafjelagsins um Ísland og Íslendinga V. Kaupmannahöfn: Prentað hjá S. L. Möller

Landau, Sidney I 1989: *Dictionaries. The Art and Craft of Lexicography*. Cambridge/New York/etc.: Cambridge University Press.

Rask, Rasmus Kristian 1811: *Vejledning til det Islandske eller gamle Nordiske Sprog*. Kjøbenhavn: Trykt paa Hofboghandler Schubothes Forlag hos J. R. Thiele.

Plasshensyn i tospråklige ordbøker

The maximum of information in a minimum of space is a natural goal in a bilingual dictionary such as H. Haraldsson's Russian-Icelandic dictionary (to appear in early 1996). Gradually, various methods have been developed to reach this goal, some of which can be seen in V. Berkov's Russian-Norwegian dictionary of 1987 (rev. ed. 1994). In this paper, the author describes some of the methods used in the dictionary to save space and condense information. He shows that among other things, precious space can be saved by means of a careful analysis of the semantics, distribution and stylistic value of synonyms.

Å spare plass — å formidle mest mulig informasjon på minst mulig plass — er en viktig faktor i praktisk leksikografi. I en ny russisk-islandsk ordbok (RIO) som ventes å utkomme i første halvdel av 1996 vies dette problemet stor oppmerksomhet. Utstrakt bruk av koder og andre plassbesparende tegn med klart definert funksjon hører til ordbokstekstens mest karakteristiske trekk. I denne artikkelen vil jeg fokusere på to aspekter som har med fortetting av ordboksteksten å gjøre (jf. Wolski 1991:2837 ff.):

- Behandling av synonymrekker i bilingvale ordbøker
- Samleartikler med informasjon som angår flere parallelle lemmaer

1 Synonymrekker

1.1 Generelt om synonymer

Av flere grunner trenger synonymrekker spesiell oppmerksomhet og spesiell behandling ved utarbeidelsen av tospråklige ordbøker, bl.a. med hensyn til at to språk gjerne har liknende synonymrekker for en bestemt semantisk referanse. Norsk har f.eks. følgende rekke synonyme adjektiver: *behjertet, djerv, dristig, modig, tapper, fryktløs, uredd* o.fl.; *Norsk synonymordbok* fra 1974 (heretter NS) gir 17 synonymer til **modig**; i *Íslensk samheitaorðabók* fra 1985 (heretter IS) finner vi f.eks.: *hugrakkur, djarfur, frækinn, hugaður, hugprúður, kjarkmikill, ódeigur* — i alt gir IS 29 synonymer til **hugrakkur**; NS har bl.a. følgende synonymer til **kujon**: *feiging, knehøne, kryp, reddhare, svekling, tøffelheld, usselrygg*, i alt 21 synonymer. Under **gunga** gir IS: *hugleysingi, kveif, kveifa*, og det henvises ytterligere til *aumingi, bleyða, rola*; **aumingi** er oppført med 47 synonymer, **bleyða** med 24 og **rola** med 31 — riktignok med en god del overlapping. En russisk synonymordbok (*Slovar' sinonimov russkogo jazyka* 1968) gir 16 synonymer til **smelyj** 'modig', men bare ett til **trus** 'feiging', attpåtil nokså sjeldent (*zájač'ja dušá* 'haresjel').¹

¹Det er innlysende at ikke alle disse ordene er i daglig bruk. Av de 31 synonymene til **rola** i IS kjente undertegnede ikke 12, dvs. nesten 39%; min datter (som er primært norskstalende) kjente ikke 6 av 21 synonymer til det norske **kujon**, dvs. drøye 28%. I både mitt tilfelle og hennes ville de fleste likevel ikke bydd på vanskeligheter i en adekvat kontekst.

Kriteriene for hva som regnes som synonymer er vanskelig å fastslå definitivt og enda vanskeligere å følge konsekvent i praksis. Her vil en viss subjektivismen alltid gjøre seg gjeldende. Etter mitt skjønn er f.eks. begrepet synonymer tolket romsligere i IS enn i NS, men mer restriktivt i *Slovar' sinonimov* enn i NS; ifølge den russiske boken skal synonymer være "gjensidig utbyttbare i konkrete tekster". Forskjellige syn på spørsmålet kommer klart til syne i forordene til de tre bøkene. En ting har de alle felles: de er beregnet på morsmålsbrukere.

Synonymer med tilnærmedesvis samme semantikk kan være vidt forskjellige mht. stilistisk valør, spredning o.l. NS og *Slovar' sinonimov* finsorterer sine synonymer med utgangspunkt i at alle ord i en artikkel er synonymer til oppslagsordet, men ikke nødvendigvis til hverandre. Det forekommer meg at dette kriteriet ikke brukes i IS. — I NS brukes det ikke stilistiske markører; IS bruker to slike tegn: + (en hevet pluss) for å markere både "forældet", "sjeldent" og "regionalt", og ° (en hevet liten sirkel) for å markere slang, tvilsomme neologismer o.l. *Slovar' sinonimov* har en minutiøs differensiering mht. stilistiske nyanser.

En stor russisk tobinds synonymordbok, *Slovar' sinonimov russkogo jazyka* fra 1970, gir synonymrekker med detaljert karakterisering av forskjellene mellom de enkelte synonymene (både semantiske, stilistiske o.l.) og anfører tekstavsnitt der ordene forekommer i typisk funksjon i typiske omgivelser. Jeg mener dette er den optimale type synonymordbok, og slike verk må både i kildeSpråk og målspråk være et ønskemål for en som driver med tospråklig leksikografi.

1.2 Synonymer i tospråklige ordbøker

La oss se på hvordan synonymspørsmålet behandles i et utvalg tospråklige ordbøker. Jeg begrenser undersøkelsen til adjektiviske og substantiviske synonymrekker som alle er knyttet til begrepsfeltet MOT:

1.2.1 Adjektiv. Norsk dominant: *Modig*

- (1) NIO 1987:
behjertet (ritm.) djarfur, hugrakkur
djerv djarfur, hugaður
drístig djarfur, áræðinn
modig hugrakkur, kjarkmikill, djarfur, áræðinn; (*høyt.*) hugprúður;
(*handl.*) djarfmannlegur
tapper hugrakkur
fryktløs óttalaus, ókvíðinn
uredd (*person*) óhræddur, djarfur; (*opptreden*) djarflegur

Under (1) dreier det seg i alt om 11 ekvivalenter. *Djarfur* (5 ganger) og *hugrakkur* (3 ganger) har størst spredning, mens 7 ekvivalenter bare forekommer én gang. Referanse til hhv. handling og person belyses til en viss grad.

- (2) DIO 1992:
behjertet djarfur, hugaður, hugrakkur
djerv einarður, hispurslaus; djarfur, tvíræður
dristig 1. áræðinn, djarfur, djarflegur *en d. investering* áhættusöm fjárfesting; 2. frakkur
modig kjarkmikill, ókvíðinn, hugrakkur *en m. optræden*
tapper hugrakkur, hraustur
frygtløs óhræddur, kjarkmikill

Spredningen er: *djarfur* og *hugrakkur* 3 ganger hvert ord, *kjarkmikill* 2, de øvrige 9 ekvivalentene har én forekomst hver.

- (3) DIO 1957:
behjertet djarfur, hugrakkur
djerv djarfur, einarður, djarfmannlegur
dristig djarfur, áræðinn, djarflegur, ófeilinn, bíræfinn, djarflegur
modig hugrakkur, hugprúður, djarfur, áræðinn, einbeittur, djarflegur, ókvíðinn, stæltur, brattur, fjörugur
tapper hraustur, røskur, vaskur, frækinn; rösklegur, vasklegur
frygtløs óttalaus, óhræddur, ósmeykur, ókvíðinn, ófælinn

DIO 1957 oppviser stor variasjon i valg av ekvivalenter. Blant i alt 19 ekvivalenter forekommer 15 bare én gang. Også her er *djarfur* det sentrale ordet. Ordboken registrerer ikke hvorvidt referansen gjelder person, handling eller utseende (*djarfur* : *djarflegur* : *djarfmannlegur*).

- (4) EIO 1984:
courageous hugaður, hugrakkur, djarfur, ódeigur, áræðinn, hughraustur
brave hugaður, hugrakkur, hraustur, frækinn, djarfur
daring hugaður, djarfur, áræðinn
bold 1. hugaður, áræðinn, djarfur; 2. (um verknað) dirfskufullur; 3. frakkur, frekur
intrepid ótrauður, geiglaus, óhræddur, djarfur
fearless óhræddur, óttalaus; hugrakkur
undaunted óhræddur, óttalaus, ótrauður, sem lætur engan bilbug á sér finna

EIO 1984 oppviser stor spredning i valg av ekvivalenter: *hugaður* 4, *djarfur* 4, *hugrakkur* 3, *áræðinn* 3, *óhræddur* 3, *ótrauður* 2, *óttalaus* 2, *ódeigur* 1, *hughraustur* 1, *hraustur* 1, *frækinn* 1, *dirfskufullur* 1, *geiglaus* 1. Ingen ekvivalent dominerer helt. Bare i tilfellet **bold** gis det eksplisitt informasjon om hvilken ekvivalent kvalifiserer handling, ikke person (delbetydning 2).

1.2.2 Substantiv. Norsk dominant: *kujon*

- (5) NIO 1987:
kujon ragmenni, gunga, kújón
feiging hugleysingi, skræfa
reddhare lydda, hugleysingi
usling löðurmenni
dott: (1) ... (2) gunga, rola
kryster skræfa, lydda
kjerring (1) ... (2) (hversd. e. máll.) eiginkona; (humor) kerling (3) (*om mann*, hversd.) huglaus maður, raggeit

Eksemplarisk spredning. Intet av de norske ordene i synonymrekken oversettes med mer enn 2 islandske ekvivalenter; ingen av de 8 islandske ekvivalentene brukes mer enn to ganger.

- (6) EIO 1984:
coward raggeit, bleyða, heigull, kvefja, kveif, lydda, gunga, ragmenni, löðurmenni, skræfa
poltroon heigull, löðurmenni, raggeit, bleyða
craven heigull, raggeit, bleyða
dastard illmenni, þrælmenni, níðingur, varmenni

Tre av ekvivalentene, *raggeit*, *bleyða* og *heigull*, er knyttet til tre lemnaer.² Det mest sentrale oppslagsordet, *coward*, skiller seg ut ved å vise størst ekvivalentalternasjon, i alt 10 ekvivalenter. Typisk "embarras de richesse"!

- (7) CIED 1989:
raggeit coward
bleyða 1. (*heigull*) coward 2. (*læða*) female cat
heigull coward
kveif coward, milksop, weakling
gunga coward
skræfa coward

coward brukes 6 ganger og er nærmest den eneste engelske ekvivalenten. Av de 10 synonymene EIO 1984 gir for *coward*, utelblir 4 (*kvefja*, *lydda*, *ragmenni*, *löðurmenni*) i CIED 1989, der også *poltroon*, *craven* og *dastard* glimrer ved sitt fravær.

1.2.3 Synonymrekker i RIO

Under utarbeidelsen av RIO ble det stadig klarere hvor viktig behandlingen av synonymrekker er i en tospråklig ordbok. Dette førte til revisjon av en rekke synonymgrupper. Noen få eksempler:

²Webster 1978 definerer de engelske synonymene slik: a *coward* is, in a general sense, one who is afraid to meet danger, real or imaginary; a *poltroon* is a mean-spirited and contemptible coward; a *craven* is one who shrinks back at the approach of danger; a *dastard* is a sneaking, malicious coward

(8) Begrepet SKAM.

(a) Utkast:

позор^{1a} *m* skömm *f*⁹, smán *f*⁹, háðung *f*^{4c}, sneура *f*¹, hneisa *f*¹; óháefa *f*¹, svívirða *f*¹ (*неприличие*); како́й ~!, стыд и ~! þvílík smán {háðung}; э́то для него́ ~ *тж* þetta er honum til skammar {háðungar}; э́то стра́шный ~ *тж* það er háborin skömm að þessu; ~ им! ≈ skömm sé þeim! ◇ **выставить** {отда́ть} *кого-л. на* ~ ≈ setja¹ *e-n* í garastokkinn Δ навлéчь, покрýть ○; ~**ить**^{4a} *нсв <св о>* *V svívirða*^{2T},

срам^{1a} *I m 1.* (skömm *f*⁹, sneура *f*¹, svívirðing *f*⁴ **2.** *прост.* (половые органы) ≈ blygðun(*in*) *f*⁹ **II** *в знач. сказ.* það er (synd og) skömm ◇ **мёртвые** ~у **не ймут** *погов.* ≈ hetjur falla, háðung flýg Δ **стыд**

стыд^{1b} *m 1.* skömm *f*⁹, smán *f*⁹, sneура *f*¹; испýтывать ~ *см* стыдýться; чýвство ~á blygðunartilfinning, skömmusta *f*¹; сгорéть со ~á verða^Δ miður sín af skömm; у *кого-л.* нет ~á {ни ~á ни сóвести} *e-r kann*^Δ ekki að skammast sín {er (alveg) blygðunarlaus⁴}; э́то ~ и срам {позор} *тж* þetta er til háborinnar skammar **2.** (*позор*) háðung *f*^{4c}, hneisa *f*¹ **3.** *см* стыдлýвостъ ◇ к ~у *своему* ...

(b) Etter justering (bl.a. vha. *Slovar' sinonimov russkogo jazyka* 1970):

позор^{1a} *m* skömm *f*⁹, háðung *f*^{4c}; óháefa *f*¹, svívirða *f*¹ (*неприличие*); како́й ~!, стыд и ~! þvílík skömm {háðung}; э́то для него́ ~ *тж* þetta er honum til skammar {háðungar}; э́то стра́шный ~ *тж* það er háborin skömm að þessu; ~ им! ≈ skömm sé þeim! ◇ **выставить** {отда́ть} *кого-л. на* ~ ≈ setja¹ *e-n* í garastokkinn Δ навлéчь, покрýть ○

срам^{1a} *I m 1.* smán *f*⁹, hneisa *f*¹ **2.** *прост.* (половые органы) ≈ blygðun(*in*) *f*⁹ **II** *в знач. сказ.* það er (synd og) skömm ◇ **мёртвые** ~у **не ймут** *погов.* ≈ hetjur falla, háðung flýg Δ **стыд**

стыд^{1b} *m 1.* (чувство смущения, раскаяния) skömm *f*⁹, sneура *f*¹; испýтывать ~ *см* стыдýться; чýвство ~á blygðunartilfinning, skömmusta *f*¹; сгорéть со ~á verða^Δ miður sín af skömm; у *кого-л.* нет ~á {ни ~á ни сóвести} *e-r kann*^Δ ekki að skammast sín {er (alveg) blygðunarlaus⁴}; э́то ~ и срам {позор} *тж* þetta er til háborinnar skammar **2.** (*позор*) háðung *f*^{4c}, hneisa *f*¹ **3.** *см* стыдлýвостъ ◇ к ~у *своему* ...

Justeringen tar utgangspunkt i grundig vurdering av hvilke synonymer i hhv. russisk og islandsk svarer best til hverandre. Dette er enda viktigere enn plassbesparelsen (som i dette tilfellet ikke er iøynefallende).

Den viktigste erkjennelsen er denne: Når man lager en lemmabase (eventuelt — og fortrinnsvis — utstyrt med illustrerende brukseksempler) som bl.a. skal brukes som utgangspunkt for tospråklige ordbøker der vedkommende språk er kilde språk, er det viktig å definere nøye alle ledd i hver enkelt synonymrekke med hensyn til semantikk, bruksområde (distribusjon), hyppighet, stilistisk valør, bruk i faste uttrykk o.l. Dette er en nødvendig forutsetning for vellykket valg av ekvivalenter på målspråket, der man ideelt bør disponere over tilsvarende synonymutredning.

Min erfaring understreker og bekrefter den konklusjon V. Berkov har formulert omtrent slik: Alle "ensartede" lekskografiske oppgaver bør løses samlet for å sikre konsekvent behandling. Dette gjelder f.eks. konjunksjoner, preposisjoner, prefikser, suffikser og pronomen. Hit hører også egennavn, og disse bør først deles opp i temagrupper.

Dette er tatt til etterretning i arbeidet med en stor norsk-russisk ordbok under faglig ledelse av V. Berkov. De oppgavene jeg nettopp nevnte er tildelt visse personer og utelates foreløpig av de medarbeidere og/eller arbeidsgrupper som er i gang med de enkelte bokstaver.

2 Samleartikler og deres plassbesparende funksjon

Når leksikografen arbeider med to eller flere parallelle lemmaer som forekommer i samme type forbindelser, vil det være hensiktsmessig å redegjøre for de leksikalske forbindelsene bare én gang, dvs. i en samleartikkel, og henvide dit fra de andre lemmaene (jf. Svensén 1987:152–153).

Hos Berkov (1994:XXI) heter det i avsnittet om ordbokens oppbygning:

Dersom en hel rekke ord inngår i de samme type forbindelser, oppgis disse konstruksjonene bare ved det ordet som kommer først i alfabetet, mens det ved de øvrige bare gis henvisning etter oversettelsen. Denne henvisningen gis ved tegnet ⇒ (pil) og den glosen som leseren henvises til.

Som eksempel på dette angir Berkov oppslagsordet *četvérg* 'torsdag' med henvisning til *voskresén'e* 'søndag', der parallelle konstruksjoner for alle ukedager (som f.eks. *v četvérg* 'på torsdag') er samlet.

Hvilken enorm plassbesparelse dette kan gi, fremgår av følgende eksempler fra RIO:

2.1 Tallord

Her fungerer *pjat* 'fem' som grunnlemma. Under andre tallord markeres likheten med beskrivelsen i samleartikkelen med henvisende dobbeltpil:

пять числ. <ж^{8b}> fimm; ~ ма́льчиков {де́вочек, слов} fimm drengir {telpur, orð}; ~ю слова́ми í {теб} fimm orðum; им да́ли по ~ книг þeir fengu fimm bækur {sínar fimm bæurnar} hver {hvor o двух}, см тж по ④; ~ раз fimm sinnum, í fimm skipti; он при-ходил ~ раз hann kom fimm sinnum; че́рез ~ дней eftir fimm daga; ка́ждые ~ дней fimmta hvern dag, á fimm daga fresti; в пяти́ коломе́трах от сто́лицы fimm kílómetra frá höfuðborginni; к се́мь прибави́ть ~, {се́мь плус ~} бу́дет двена́дцать fimm við {plús} sjö eru tólf; из восьми́ вы́честь {от восьми́ отня́ть ~,} полúчитс́я три fimm frá átta eru þrjú; тепе́рь ~ часо́в klukkan er fimm; ро́вно в ~ (часо́в) klukkan (nákvæmlega) fimm, á slaginu fimm; он приде́т в ~ часо́в {часо́в в ~, о́коло пяти́ (часо́в)} hann kemur klukkan fimm {um fimtleutið}; по́сле пяти́ (часо́в) (einhvern tíma) eftir klukkan fimm; тепе́рь {бы́ло} без пяти́ (мину́т) ~ klukkuna vantar {vantaði} fimm mínútur í fimm; тепе́рь без два́дцати́ пяти́ (мину́т) ~ klukkan er fimm mínútur yfir hálf fimm; бы́ло без два́дцати́ (мину́т) ~ утра́ klukkan var tíu mínútur yfir hálf fimm um morguninn {að morgni (dags)}; он прише́л то́лько без че́тверти ~ (вече́ра) hann kom ekki fyrr en klukkuna vantaði korter í fimm (um kvöldið); он прише́л к пяти́ (часа́м) {до пяти́ (часо́в)} hann kom klukkan að verða fimm {fyrr klukkan fimm}; с пяти́ до восьми́ frá klukkan fimm til klukkan átta; у меня́ догово́ренность на ~ часо́в ég ætla að hitta mann {hef stefnumót} klukkan fimm; ему́ ~ лет hann er fimm ára (gamall^Δ); ей со́рок ~ лет hún er fjörutíu og fimm ára (gömul), hún er hálf fimmtug; ему́ лет ~ hann er á að giska fimm ára; де́вочка пяти́ лет fimm ára {vetra ust.} gömul telra; в ~ лет он бы́л ... þegar hann var fimm ára (gamall), var hann ...; ~ с полови́ной (ты́сячи) fimm og hálfur {hálf þúsund}, hálfur sjötti {hálf sjötta þúsund}; (в) ~ раз fimm sinnum, í fimm skipti; в ~ раз бо́льший fimm sinnum stærri {meiri, fleiri}; fimmfaldur⁵ {пятикратный}; увели́чить(ся) в ~ раз тж fimmfalda(st)³, ср тж вдво́е, раз I; ⇒ полови́на I 2. ж {оценка} (einkunnin f⁹) fimm, см пяте́рка 2

вóсемь числ. <ж^{8*bΔ}> átta ⇒ пять
шесть числ. <ж^{8b}> sex ⇒ пять; через (каждые) ~ месяцев *тж* á misserisfresti;
 ~деся́т^Δ **числ.** sex tíu ⇒ пятьдесят, пять; ~со́т^Δ **числ.** sex hundruð
 два^Δ **числ.** м/с <ж две> tveir ⇒ пять; две сестры *tvær systur*; в ~ ра́за бо́льше
 heimingi {hálfu, tvöfalt} meiri {fleiri}, *см тж* три
 три^Δ II **числ.** þrjú ⇒ пять; ~ бо́льших сто́л^á þrjú stór borð; ~ ста́рые кни́ги þrjár
 gamlar bækur; е́му (со́рок) ~ го́да hann er (fjögutíu og) þriggja ára (gamall); два~ го́да
 tvö-þrjú {tvö til þrjú} ár III *нескл.* *см* тройка 3; учи́тель поста́вил ей ~ ≈ kennarinn gaf
 henni þrjá
 четы́ре^Δ **числ.** fjórir ⇒ пять, три, ...◇ **играть в ~ ру́ки** spila³ fjórhent; **сиде́ть в
 четырёх стена́х** híma^{2D} innan dуга; **отпу́стить ко́го-л. на все ~ сто́роны** láta^Δ e-n
 fara frjálсан ferða sinna

2.2 Ordenstall

Grunnlemma *þjátýj* 'femte':

пяты́й порядк. fimmti; ~ая глава́ fimmti kafli; ~ая страница́ (blað)síða fimm, fimmta síða; (в) ~ый раз í fimmta skipti(ð) {sinn}; в ~ом ве́ке á fimmtu öld; се́годня ~ое ма́я í dag er fimmti maí; э́то случи́лось ~ого ию́ля það gerðist fimmta júlí; се́йчас {уже́} ~ый час klukkan er (farin) að ganga fimm, klukkan er yfir fjögur; се́йчас полови́на ~ого {полпя́того} klukkan er hálf fimm; в ~ом часу́ klukkan að ganga fimm, milli (klukkan) fjögur og fimm, eftir klukkan fjögur; те́перь де́сять мину́т ~ого klukkan er tíu mínútur gengin í fimm {yfir fjögur}; он звони́л в ше́сть мину́т ~ого hann hringdi (klukkan) sex mínútur yfir fjögur {gengin í fimm}; в нача́ле ~ого rétt eftir klukkan fjögur; е́му иде́т ~ый год hann er á fimmta ári(nu) {kominn⁷ á fimmta ári(ð)}; е́му поше́л ~ый де́сяток hann er (kominn) yfir fertugt, hann er á fimmtugsaldri {kominn á fimmtugsaldur}; *см тж* пя́тая, раз I, полови́на 2, час 1, 2 Δ колесó ◇, колóнна ◇, пункт ◇, спи́ца

пятна́дцатый порядк. fimmtándi ⇒ пяты́й
восемна́дцатый порядк. átjándi ⇒ пяты́й
восьмо́й порядк. áttundi ⇒ пяты́й ...; е́му поше́л ~ де́сяток hann er (kominn) yfir sjötugt
шестíой порядк. sjötti ⇒ пяты́й, ... ◇ ~о́е чу́вство sjötta skilningarvitidⁿ⁴

2.3 Árstíðir

Grunnlemma er *vesná* 'vår':

весна^{1*d} ж <мн вёсны> vor n^4 ; этой ~ой í vor; síðastliðið vor, í vor sem leið (*тж текущего года*); próšлая ~á síðastliðið vor; vorið í fyrra (*тж прошлого года*); próšлой ~ой í fyrravor ↑; síðastliðið vor, í vor sem leið, á síðastliðnu vori ↑; той ~ой það vor, á því vori, þá um vorið; б́удущей ~ой í vor, næsta vor, á vori komanda; в одну ~у eitt vor(ið), vor eitt; той же ~ой, в ту же ~у sama vor; р́анняя ~á, náçало ~ы vorbygjun; в náçале ~ы í vorbygjun, á öndverðu vori; кон́ец ~ы endað vor, vorlok n^4 р́í; в кон́це ~ы, п́оздней ~ой síðla vors, að áliðnu vori; seint um vorið (*тж той весной*); seint í vor (*этой весной*); seint um vor (*одной поздней весной*); бы́лая р́анняя ~á það var snemma vors; р́анней ~ой snemma vors; snemma um vorið ↑; в п́ервую половину ~ы framán af vori, á öndverðu vori; в середине ~ы á miðju vori, um mitt vor; (уж́е) наступ́ила ~á það er/var komið vor; наступ́ает ~á það er að koma vor; ~á наступ́ила ráno það voraði {vorið kom} snemma; ~á próšlá vorið er/var liðið; vorið leið; он ост́анется здесь на всю ~у hann verður hér í allt vor; всю ~у шли дожди það rigndi allt vorið {í allt vor *этой весной*}; ~ой 1997-го ѓода vorið 1997; в теч́ение ~ы с.м. веснóй Δ дру́жный 2, р́íзнак

зимá^{1*d} ж <В ег зиму> vetur m^{Δ} ⇒ весна́;

Det er viktig at uttrykk som er spesielle for VINTER, kommer etter pilen, f.eks.:

наступ́ает {наступ́ила} ~á тж vetur(inn) er að ganga {er genginn} í garð; среди ~ы, в середине ~ы тж um miðsvetrarleyti(ð), um miðsvetrarskeið...; в п́озднюю зиму тж á útmánuðum, á ofanverðum vetri; он п́робыл там всю {ц́елую, ту, э́ту} зиму тж hann var þar vetrarlangt...

2.4 Måneder

Grunnlemma ávgust 'august':

август^{1a} м ágúst $m^{4\Delta}$, ágústmánuður $m^{9\Delta}$; в ~е (м́есяце) í ágúst(mánuði); в náçале ~а í bygjun ágúst, í ágústbygjun; в с́амом náçале ~а fyrst í ágúst; в середине ~а um miðjan ágúst; в кон́це ~а í ágústlok, í lok ágúst; в ~е э́того ѓода í ágúst(mánuði); núna í ágúst *разг.*; í ágúst sem leið (*в минувшем*); в ~е б́удущего ѓода í ágúst (næsta ár), í ágúst að ári; в ~е п́рошлого ѓода í ágúst í fyrra; í ágúst sem leið ↑; ~овский *прл* ágúst

Igjen ser vi at det er viktig å plassere uttrykk som ikke har paralleller i grunnlemmaet, etter pilen:

апрéль^{2a} м apríl m^0 ⇒ август; п́ервого ~я с ним сы́грáли ш́утку hann hljóp {var láttinn hláupa} apríl; с п́ервым ~лем! fyrsti apríl!; ~льский *прл* apríl
дека́брь^{2b} м desember m^0 , desembermánuður $m^{9\Delta}$ ⇒ август; 31-ое декаб́рj тж gaml(a)ársdagur; ~ский *прл* desember

2.5 Etnonymer

En spesiell utfordring når det gjelder bilingvale ordbøker fra russisk til f.eks. nordiske språk består i at slaviske språk, inkl. russisk, danner og bruker konsekvent hunkjønnsvarianter til etnonymer: *norvežec* 'nordmann', *norvežka* 'norsk kvinne'; *datčanin* 'danske'; *datčanka* 'dansk kvinne', etc. Samtidig unngår russisk nøye å bruke adjektiver for å angi nasjonalitet, dvs. uttrykk av typen "Hun er norsk", "Han er tysk" osv. Dette redegjøres det grundig for i artikkelen *islandka* 'islandsk kvinne':

ислáндка^{3*a} ж íslendingur *m*^{5ab}, íslensk kona *f*^{1Δ} {stúla *f*^{1N} *молодая*; telra *f*^{1N} *девочка*}; oná ~ka hún er íslendingur {íslensk; frá Íslandi *из Исландии*}; emú hrávjatja ~ки hann er hrifinn af íslenskum stúlkum {konum}; ~ки добивáются равнопáвия konur á Íslandi {íslenskar konur} berjast fyrir jafnrétti; как тебе понрáвились ~ки? т ж hvernig leist þér á þær íslensku? *разг.*; см т ж íслáндец; ~ский *прл* íslenskur⁵; frónskur⁵ *поэт.*; ~ский язы́к íslenska *f*¹,

Det henvises fra andre kvinnelige etnonymer til *islandka*. Over 50% av alle henvisninger i RIO vha. ⇒ gjelder nettopp dette. Under bokstaven A finner vi f.eks.:

абисси́нка^{3*a} ж abessíníukona *f*^{1Δ} ⇒ íслáндка; ~ский *прл* abessínskur⁵; blá-lenskur⁵ *уст.*
абха́зка^{3*a} ж abkhazíukona *f*^{1Δ} ⇒ íслáндка
австрали́ец^{5*a} м ástralíumaður *m*^Δ, ástrali *m*^{1,3}; ~ийка^{3*a} ж ástrali *m*¹ ⇒ íслáндка;
 ~ийский *прл* ástralskur⁵
австри́ец^{5*a} м austurríkismaður *m*^Δ; ~ийка^{3*a} ж austurríkiskona *f*^{1Δ} ⇒ íслáндка
азербáйджа́нец^{5*a} м, ~ка^{3*a} ж aseri *m*¹, azerbajdzhani *m*¹ ⇒ íслáндка
алба́нка^{3*a} ж albönsk kona *f*^{1Δ} ⇒ íслáндка; ~ский *прл* albanskur⁵; ~ский язы́к albanska *f*¹
алжи́рец^{5*a} м, ~ка^{3*a} ж alsíri *m*¹, alsíringur *m*^{5ab} ⇒ íслáндка;
америка́нец^{5*a} м ameríkumaður *m*^Δ; amerikani *m*¹ (*преим. житель США*); bandaríkjamaður (*житель США*); kani *m*¹ *разг.*; ~ка^{3*a} ж amerísk {bandarísk ↑} kona *f*^{1Δ} ⇒ íслáндка; ~ский *прл*
англича́нин¹¹ м englendingur *m*^{5ab}; ~ка^{3*a} ж ensk kona *f*^{1Δ} ⇒ íслáндка

3 Sluttord

Jeg har begrenset denne redegjørelsen til to viktige faktorer som er med på å bestemme fremstillingen av ordboksteksten i en bifunksjonal tospråklig ordbok utfra den målsetning å spare plass og formidle den leksikografiske informasjon til brukeren på en presis og konsentrert måte. Men plassbesparing og fortetting kommer til uttrykk på flere måter i RIO, ikke minst vha. koder, strukturindikatorer og andre symboler som knytter seg til enkelte informasjonstyper. Dette er det ikke anledning til å drøfte nærmere i denne omgang. Berkov 1994 gir en god innføring i og oversikt over de fleste av disse og lignende løsninger.

Litteratur

Ordbøker

- Berkov, Valerij 1994: *Russisk-norsk ordbok*. 2. utgave. Oslo: Universitetsforlaget.
- DIO 1992 = *Dönsk-íslensk orðabók*. Hovedredaktører: Hrefna Arnalds/Ingibjörg Johannesen. Reykjavík: Ísafoldarprentsmiðja.
- EIO 1984 = Sören Sörenson o.fl.: *Ensk-íslensk orðabók með alfræðilegu ívafi*. Reykjavík: Örn og Örlygur.
- NIO 1987 = Hróbjartur Einarsson: *Norsk-ísländsk ordbok*. Oslo: Universitetsforlaget.
- DIO 1957 = Freysteinn Gunnarsson: *Dönsk-íslensk orðabók*. Reykjavík: Ísafoldarprentsmiðja.
- CIED 1989 = Sverrir Hólmarsson/Christopher Sanders/John Tucker: *Concise Icelandic-English Dictionary*. Reykjavík: Iðunn.
- IS = Svavar Sigmundsson (ritstj.): *Íslensk samheitaorðabók*. Reykjavík: Háskóli Íslands.
- NS = Dag Gundersen: *Norsk synonymordbok*. Oslo: Gyldendal Norsk forlag
- SIO 1982 = Gösta Holm/Aðalsteinn Davíðsson: *Svensk-ísländsk ordbok*. Lund/Reykjavík: Walter Ekstrand Bokförlag/Almenna bókafélagið.
- Slovar' sinonimov 1968 = Aleksandrova, Z. E.: *Slovar' sinonimov russkogo jazyka*. Moskva: «Sovetskaja Enciklopedija».
- Slovar' sinonimov russkogo jazyka* 1970. Tom I-II. Hovedredaktør A. P. Evgen'eva. Leningrad: «Nauka».
- Webster 1978 = *Webster's New Twentieth Century Dictionary of the English Language*. Unabridged. Second Edition. Collins World.

Sekundær litteratur

- Svensén, Bo 1987: *Handbok i lexikografi*. Stockholm: Esselte studium.
- Scando-Slavica 1980 = Instructions to Contributors. I: *Scando-Slavica*, Tomus 26, 199–201.
- Wolski, W. 1991: Formen der Textverdichtung im zweisprachigen Wörterbuch. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie* III. Red. Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta. Berlin/New York: de Gruyter, 2837–2854.

Berit Heer Henriksen

Korpusbaserede relationsoplysninger og lemmatisering af flerordsforbindelser

Of the approximately 100 000 entries selected for *The Danish Dictionary*, not all will be one-word entries. Some of them will be subentries consisting of word groups. These subentries are described in detail as if they were ordinary entries, but we also have other and shorter ways of presenting information about phraseology and valency. Bearing these different possibilities in mind, we examine the concordances and word distribution reports generated from a text corpus of 40 million running words — and make the selection of subentries for the dictionary.

Den Danske Ordbog bliver en stor, ny ordbog over nutidsdansk, og den udkommer i seks bind i 1998–99.

Nogle af ordbogens ca. 100 000 opslagsord er hovedlemmaer, andre er sublemmaer. Et hovedlemma er fx et ord som *båd*, der kan slås op på ordets alfabetiske plads i ordbogen. Et sublemma er derimod en gruppe af ord, fx *gå i bådene*. Det enkelte sublemma er placeret under et relevant hovedlemma.

Sublemmaerne i *Den Danske Ordbog* er kombinationer af ord, der altid optræder sammen. De danner begrebslige helheder med en konkret eller overført betydning. Det kan dreje sig om talemåder som fx *brændt barn skyr ilden*, eller det kan være idiomer som fx *være i samme båd*, eller det kan være andre kollokationer som fx *bryde ud, sætte i gang, til fods, tak for mad, både . . . og* — inklusive termer som fx *tungt artilleri*.

Men ud over sublemmaerne har ordbogen oplysninger om valensmønstre og om typiske samforekomster. Oplysningerne om syntaktiske mønstre er især relevante for verbernes vedkommende, mens vi om et ord som fx *båd* vil give den oplysning, at man typisk træffer samforekomsterne *lille båd og om bord i/på en båd*.

For ordbogsredaktøren kan oplysninger om valenser og en angivelse af typiske samforekomster somme tider fungere som pladsbesparende alternativer til lemmatisering af en ordgruppe. Dét forhold vil jeg beskæftige mig med i det følgende. Jeg vil gennemgå mulighederne for at medtage samforekomster og valenser i *Den Danske Ordbog* — herunder oprettelsen af sublemmaer. I den forbindelse vil jeg vise nogle af de udnyttelsesmuligheder, der ligger i vores store elektroniske tekstkorpus. Til sidst vil jeg pege på konkrete problemer, der kan opstå under selektionen af sublemmaer.

1 Sublemmaer, samforekomster og valenser

Vi har forskellige muligheder, når vi vil have en samforekomst eller en valens med i *Den Danske Ordbog*. Vi skelner mellem

- sublemma (fx *båd* [. . .] *være i samme båd*)

- valens (fx *spise* [...] [*NGN spiser NGT*])
- typisk samforekomst (fx *chef* [...]~ *for*)
- typisk kerne (fx *garvet* [...] *garvede politifolk*)
- typisk adled (fx *rød* [...] *blussende rød*)
- citat (fx *eg* [...] *som gave til mor havde han lavet et solidt sybord i eg*)

De forskellige oplysningstyper kan kombineres. Behandlingen af et sublemma kan fx godt indbefatte oplysninger om valenser og typiske samforekomster.

1.1 Sublemmaer i Den Danske Ordbog

Sublemmaerne i *Den Danske Ordbog* behandles lige så grundigt som hovedlemmaerne, så lemmatisering af en ordgruppe er en pladskrævende løsning.

Alle sublemmaer får en betydningsbeskrivelse, og i ordbogen fremstår sublemmaet med halvfed skrift e.l. Desuden angives trykforhold, hvis det er afgørende for betydningsdifferentieringen — fx er der forskel på '*skrive* *sammen* 'brevveksle' m.m. og *skrive* *sammen* 'sammenskrive'. Idiomer og talemåder lemmaklassemarkeres, men ikke kollokationer — således står der *talemåde* efter sublemmaet *man skal ikke smide barnet ud med badevandet*.

Et sublemma er som hovedregel placeret under første betydningsbærende substantiv. Fx placeres talemåden *brændt barn skyr ilden* under *barn*. I de tilfælde, fx ved partikelverber, hvor ordgruppen ikke indeholder et betydningsbærende substantiv, må vi bruge andre anbringelsesprincipper, men det vil føre for vidt at gennemgå dem her.

Vi håber, at ordbogens brugere efterhånden vil lære sig, hvor de skal lede efter sublemmaerne. Desværre læser ordbogsbrugere så godt som aldrig ordbøgers indledning, hvor principperne kan være gennemgået. Vi ved på forhånd, at brugere af ordbøger ikke altid forventer at finde sublemmaerne under første betydningsbærende substantiv eller substantivering. Det ved vi bl.a. fra en lille spørgeskemaundersøgelse, der er foretaget af Henrik Lorentzen, en af redaktørerne på *Den Danske Ordbog*.

I nogle tilfælde vælger vi at komme ordbogsbrugeren i møde ved at opføre et sublemma i flere artikler og så henvise fra den ene artikel til dén artikel, hvor sublemmaet er behandlet. Henvisningen kan foregå eksplicit eller via henvisningsmærke. I *brænde*-artiklen finder man således den implicitte henvisning *brændt* ▷ *barn skyr ilden*. I artiklen om *badevand* finder man imidlertid henvisningen *man skal ikke smide barnet ud med badevandet, se barn*, fordi *barn* ikke står i sin grundform i talemåden.

Alt i alt må man sige, at det er en pladskrævende løsning, vi vælger, når vi vælger en lemmatisering. Ikke mindst, hvis vi lemmatiserer udtrykket flere steder i ordbogen som en service til brugeren.

Men der er ofte også andre præsentationsmuligheder end lemmatisering, og det kan være vanskeligt at afgøre, hvornår en alternativ mulighed er at foretrække. Meget afhænger af den undersøgelse, vi foretager i vores store elektroniske tekstkorpus.

	mut.inf.	coocc.	mean-pos	dev-pos
fods	247.02	[10]	-1.00	0.00
ombord	122.30	[9]	-2.56	0.25
cup	57.65	[8]	-1.13	0.11
bord	56.66	[27]	-2.33	0.22
bunden	51.99	[16]	-2.06	0.06
vandet	13.26	[10]	-2.60	0.44
bil	11.85	[9]	-1.78	0.17
pr	10.50	[11]	-1.00	0.00
meter	9.81	[8]	-1.75	0.44
samme	9.10	[45]	-1.07	0.11
manden	8.01	[8]	-2.13	0.11
lille	6.62	[33]	-1.06	0.06
danske	5.83	[27]	-1.15	0.27
m	5.18	[11]	-1.73	0.38
ude	5.05	[10]	-2.20	0.16
ned	4.72	[25]	-2.20	0.16
egen	4.55	[9]	-1.33	0.44
sin	4.15	[45]	-1.36	0.45
anden	4.05	[22]	-1.36	0.60
en	3.79	[399]	-1.47	0.44
min	3.68	[26]	-1.23	0.18
ligger	3.46	[9]	-2.11	0.32
stor	3.42	[13]	-1.46	0.56
mand	3.22	[11]	-2.64	0.23
alle	3.15	[34]	-2.26	0.90
ny	3.13	[11]	-1.27	0.38
din	3.07	[8]	-1.13	0.11
i	2.70	[489]	-1.77	0.63
fra	2.68	[78]	-1.77	0.74
tilbage	2.59	[11]	-2.09	0.08
mens	2.58	[10]	-1.60	0.84
omkring	2.46	[8]	-1.38	0.23
op	2.27	[30]	-2.27	0.20
denne	2.25	[16]	-1.25	0.31
den	2.23	[171]	-2.01	0.29
kom	2.20	[15]	-2.47	0.38

Figur 1. Den mutuelle statistik er baseret på Den Danske Ordbogs elektroniske tekstkorpus. Statistikken viser, hvilke ord der overhyppigt optræder på første, anden og tredje plads til venstre for ordet båd.

2 Korpus som kildemateriale

Den Danske Ordbog bygger på mange kilder. En af de vigtigste er et elektronisk tekstkorpus, der indeholder ca. 40 mio. løbende ord, hentet fra danske aviser, bøger, blade, radio, tv, brochurer, reklamer, tegneserier, breve, skolestile, lejlighedsviser osv. Fra dette korpus kan vi hente konkordanslister og statistiske oplysninger, som vi bruger under artikelredigeringen.

2.1 Eksemplet *båd*

Da jeg skulle skrive ordbogsartiklen om ordet *båd*, lavede jeg først en mutuel statistik. En mutuel statistik er en undersøgelse af, hvilke ord der overhyppigt optræder sammen med et bestemt ord. Edb-programmet sammenligner sandsynligheden for at møde ordene hver for sig med sandsynligheden for at møde dem sammen. Mutuelle statistikker giver både et indtryk af et ords betydning/betydninger og af de samforekomster og valensmønstre, ordet indgår i. Statistikkerne laver vi kun på ord, der er relativt frekvente i korpus; der skal helst være mindst et par hundrede forekomster.

Corpus-Bench - BAADGAES.CBC		
File	New	Edit
View	Options	Help
CONC: båd (inf)		
<p>Ler fort under alle forhold. Og hvor er det dejligt at være om bord i en båd, hvor man stadig kan Ler n </fremhæv> forstærker følelsen af at være om bord i en sikker båd. Vindskærmen er ikke IFHZ nle. Få minutter senere hev Svend Kloster sin ven ombord i sin båd. Da var Hans Mortens IFHZ k Inga og Svend Kloster også egenhændigt bjærgt ombord i deres båd, men hans liv stod he Jediga dem og var ved at skylle deres garn. Og han vil om bord i en af bådene, den, der tilhørte Jediga han øje på to både. En af ligger ved søen. Jesus går om bord i en af bådene. Den tilhører fiske KmO den i den hensigt at tage dit eget liv ved at springe over bord fra båden i løbet af natten. E KmO ss havde virket fuldstændig naturlig, da han hjalp hende om bord i båden, han havde småsnak KmO var der ingen radiosender eller -modtager om bord i vor beskedne båd. Der var proviant til f KmO iam Bligh og hans "loyalister" havde de tvunget ombord i en lille båd og ladet dem tilbage i LanX Nykøbing Falster, når han besøger os, og springer adræt om bord i båden, siger Tom. - Han Mide r trukket dybt ned i Middelhavet. Der er altid en læge om bord på båden, når Angela og andr Ozvi bevidsthed. En søndag var jeg med min far og morbror ombord i båden, på kryds ud af Sor PmJ jazz og sejlbåde. En af hans mest prominente gæster ombord i sin båd var den brasilianske r PmJ lov til at gå til kajs i havnen ved Adriater-havet. Ombord i en af bådene var den jugoslavisk EOT nervagt, at han intet kunne huske. Kun, at han var gået ombord på båden, og at han havde va EOT chef for Nationaltheatret i Oslo. Og ideen, som han fik om bord på båden, var den helt oplagt EOT 20. august ved 6-tiden sejlede vi ud af Plymouth havn om bord på båden "Mistress" med sk EOT s 6 ounces og var helt tilbage fra 1976 og også sat ombord på en båd fra Plymouth ude i Ke EOT mthlvkvædet vise, og så taler de så myndigt. Specielt om bord på en båd, hvor der skal handles EOT ns motorbåd, og undrede sig over hvad Fingal ville ombord på den båd, hvor han jo intet hav EOT gte ved havnen, hvor postbåden netop havde lagt fra kaj. Om bord på båden befandt sig flere al EOT Da han vågnede, befandt Jeroen og Kurt sig om bord på Tobias' båd på vej ud af indseilin</p>		
Total words: 0	Node words: 0	Matches: 38
Report updated.		

Figur 2. Konkordanslisten afslører bl.a. en uenighed blandt sprogbrugerne mht., hvorvidt om bord skal skrives i et eller to ord.

Den mutuelle statistik i figur 1 er en undersøgelse af konteksten til venstre for ordet *båd*. Som man ser, står ordet *fods* øverst på listen i statistikken. Mellem *båd* og *fods* er der

altså relativt stor gensidig tiltrækningskraft; der er et højt tal i statistikens første kolonne (*mut. inf.*). Så længe dette tal er over 1, er der tale om en overhyppighed.

Corpus-Bench - BAADGAES.CBC	
File New Edit View Options Help	
CONC: båd (inf)	
IFCN	n det er vel et eller andet med, at du skal føle dig i samme båd som Stallone? — Nej, det er
IFCN	som Stallone? — Nej, det er ham, der skal føle sig i samme båd som mig (skraldgriner) ... b
19LE	hun blev udsat for i den forbindelse. Jeg har været i samme båd. Jeg var 17 år, da jeg blev
19LE	ne, holder vagt, o.s.v. Når det brænder på, er vi alle i samme båd, og det giver forståelse. —
12KE	t af det intime samvær, der var mellem mennesker i samme båd på sygehuset, men alligevel
12KE	nok." — Lone sukkede. Stakkels Jens. Han var altså i samme båd som viceværten. Jens kunne
KOLF	nsen — det navn der står på bl.a. visen "Vi er alle i samme båd". — Når far fik så mange opg
PALE	ge Kjærulf Schmidt </fremhæv> synge "Vi er alle i samme båd". — Det, denne artikel vil pre
PALE	at at være bange, for der var rigtig mange, der var i samme båd som os, fortsætter Merry. —
KWFI	ikke, vi har ikke råd, / midt om natten stod vi ud i samme båd / farligt vand at ligge og ko
KZ1V	ene op og binde sig sammen for at sige: vi er alle i samme båd. — Det var netop spontant. —
KZ1V	gumpe-tung hurtigleber! — Her får du to båd-typer i samme båd. — Som begge fungerer opti
100F	aja: Åhr hold kæft. — Elsebeth: Vi er jo alle sammen i samme båd, er vi ikke? — Maja: Hvorfor
100F	unden for det gamle ordsprog som siger: Vi er alle i samme båd. — </Tekst> </Korpusenhed>
01JK	bet det gynger og alle vi synger for vi er jo alle i samme båd. — 2. Her har vi Richard. — h
01JK	da det satte ind med storm. De fiskede også lidt fra samme båd, men det var jo ikke så far
01JK	ed forskellige typer, der i "rejsetiden" er, om ikke i samme båd, så i samme tog. — Her var
01JK	kan jeg i øvrigt tage Per Bregengaard fra VS med i samme båd, hvis de kunne forholde sig
01JK	hold mellem passagererne ud fra devisen: Vi er alle i samme båd! — Er myldretids-toggangen t
EACR	er samfundsforskere og politikere mere eller mindre i samme båd. — Men nu er det jo langt fra
EACR	teresser, som er placeret i flere europæiske lande, i samme båd. — Der resterer stadig 500 p
05JK	til nu uden held — og ved, at mange andre unge er i samme båd som jeg. Min generation lær
05JK	er Folketinget ved det kommende valg. — For Ninn er i samme båd som Schlüter og H.P. Claus
Total words: 0 Node words: 0 Matches: 43	
Report updated.	

Figur 3. I samme båd kan man være — konkret, men også overført.

Af næste kolonne (*coocc.*) kan vi se, at der er 10 samforekomster med *fods* og *båd* i korpus. At ordet *fods* står højt på listen betyder altså ikke, at der er mange samforekomster — kun at der er relativt mange. Der er fx mange flere samforekomster med *den* — hele 171 — men denne samforekomst rangerer meget lavere på listen, for *den* er ikke meget mere tiltrukket af ordet *båd* end af så mange andre substantiver.

Tredje kolonne (*mean-pos*) i den mutuelle statistik fortæller, at ordet *fods* optræder (gennemsnitligt) på den første plads til venstre for *båd* — ikke på den anden eller tredje plads. Den sidste kolonne (*dev-pos*) angiver, at afvigelsen fra denne position er nul, dvs. *fods* står hver gang umiddelbart til venstre for *båd*.

Et kig i korpus afslører, at statistikens oplysninger dækker over eksempler i stil med *en 50 fods båd*. *Fod* bliver altså overhyppigt brugt som længdemål om både, men dén oplysning vil jeg ikke bruge i min ordbogsartikel om *båd*, for jeg synes ikke, det er interessant nok.

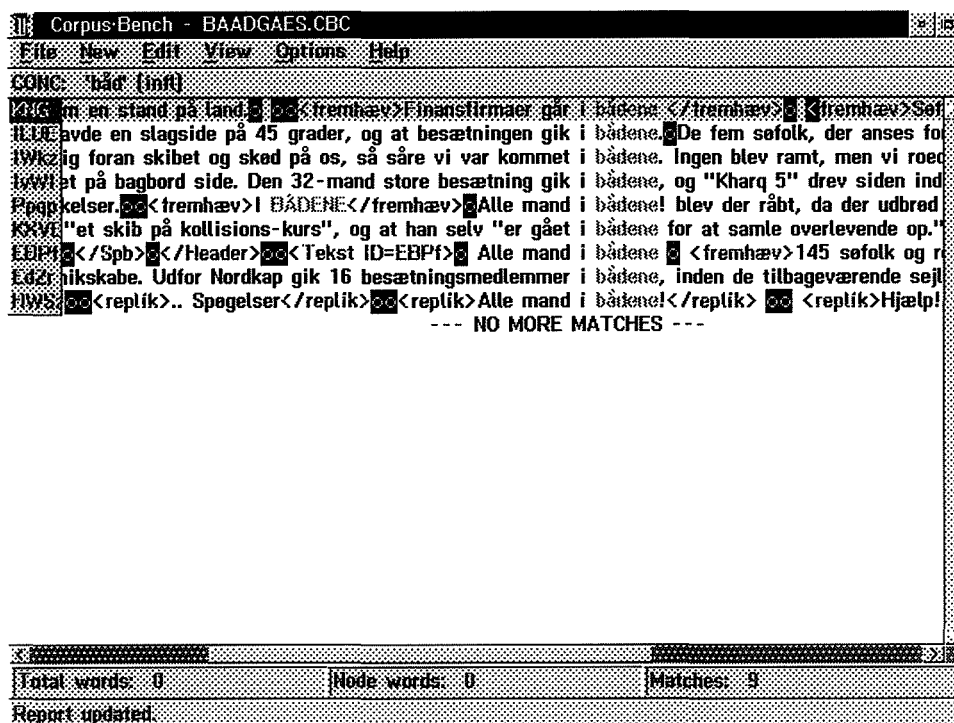
Mere interessante er de ord der står længere nede i den mutuelle statistik, men som samtidig repræsenterer mange samforekomster, bl.a. *bord*, *samme*, *lille* og *i*. Disse samforekomster undersøger jeg nu nærmere gennem konkordanslister. Jeg skal finde ud af, om samforekomsterne skal med i ordbogen — og i hvilken form.

Figur 2 viser en konkordansliste, hvor jeg har isoleret de linier, der har ordet *bord* eller *ombord* i *båds* kontekst. Vi kan se, at der tegner sig en samforekomst i retning af *være om*

bord i en båd eller *være om bord på en båd*. Det kan være interessant for især brugere med dansk som andetsprog el. fremmedsprog at vide, at man kan bruge begge præpositionerne, *i* og *på*, i forbindelse med en båd. På den anden side er det nok ikke nødvendigt ligefrem at lemmatisere samforekomsterne med *om bord*, for betydningen giver sig selv, hvis man kender betydningerne af henholdsvis *om bord* og *båd*. I den færdige artikel har jeg derfor valgt en angivelse af typisk samforekomst i stedet for oprettelsen af et sublemma.

Endnu et kig i korpus gør det klart, at ordet *lille* i ordbogsartiklen kan behandles som et typisk adled til *båd*, men at der så heller ikke er grund til at gøre mere ud af det.

Af den næste figur, figur 3, kan vi se baggrunden for, at ordet *samme* stod højt på listen i den mutuelle statistik. Der lader til at være en etableret brug af udtrykket *være i samme båd* i overført betydning. Konkordanslisten har flere eksempler på, at *alle* er i samme båd, men jeg opretter et sublemma, der lyder: *være i samme båd* og betragter det som en ordgruppe med overført betydning som et idiom.



Figur 4. Korpus viser variationer over sublemmatet *gå i bådene*. Her er fx konstruktioner med verberne *komme* og *sende*, og her er eksempler på kommandoen: *alle mand i bådene!*

Ordet *i* er der næsten 500 samforekomster med, viste den mutuelle statistik, og jeg er interesseret i at vide, om der er bestemte verber, man bevæger sig i både med. Min sprogfornemmelse siger mig, at der er noget, der hedder *gå i bådene*, hvor *bådene* betyder 'skibets redningsbåde'. Det bekræftes af konkordanslisten i figur 4, der viser samforekomster

med *båd* og *i*. Jeg opretter herefter et sublemma, der lyder: *gå i bådene*, og jeg betragter det som en kollokation, fordi betydningen er konkret.

Mit færdige udkast til artikel om *båd* kom til at se ud som i figur 5. Artiklen angiver et typisk adled, en typisk samforekomst og to sublemmaer. Begge sublemmaer har betydningsbeskrivelse og citater. Citatet under sublemmaet *være i samme båd* skal bl.a. illustrere, at man typisk taler om, at alle er i samme båd.

båd sb. -en, -e, -ene

1 mindre fartøj, der drives frem vha. årer, sejl el. motor typisk lille ~; om bord i/på en båd □ han [havde] allerede fået sportsvogn og ridehest og båd gå i bådene gå om bord i et (synkende) skibs redningsbåde el. -flåder □ coasteren havde en slagside på 45 grader, og .. besætningen gik i bådene være i samme båd ofø være i samme (problemfyldte) situation □ Vi var alle i samme båd, på godt og ondt. Det styrkede fællesskabet.

2 større skib, der sejler i fast rutefart typisk om bord i/på en båd □ fyren .. var blevet sat på englandsbåden på sin polterabend, og på båden møder han sit livs kærlighed.

3 trekantet stykke af rund frugt el. grøntsag, hvor stykkets ene side er frugtens buede overflade □ Skræl æblerne og skær dem i både.

SMS: til bet. 1 bådehavn, bådmotor; *motor-, *ro-, *sejlbåd

Figur 5. Et udkast til ordbogsartiklen om ordet *båd*.

3 Problemer

Men er det nu så indlysende, at artiklen om ordet *båd* skal se netop sådan ud? Man kunne vel tænke sig, at korpus' oplysninger om samforekomster havde været behandlet anderledes?

Fx kunne sublemmaet *gå i bådene* måske have været udeladt. I stedet kunne jeg have oprettet delbetydningen 'redningsbåd'. Jeg valgte en lemmatisering, fordi min sprogfornemmelse siger mig, at der er tale om et fast udtryk, bl.a. fordi der sker tryktab. Der er ikke tryk på verbet *gå*, men derimod på *bådene*, og dét er med til at cementere udtrykket. Desuden er *båd* i denne betydning altid i bestemt form pluralis, uanset hvor mange redningsbåde der er tale om. Var der kun én redningsbåd, ville man stadig råbe: *Alle mand i bådene! Kvinder og børn først!*

Man kan diskutere, om det var nødvendigt at lemmatisere *gå i bådene*, men man kunne modsat også hævde, at der burde have været flere sublemmaer i artiklen. Det gælder fx en ordgruppe, jeg har kategoriseret som en typisk samforekomst, *om bord i/på en båd*. Det gælder også samforekomsten *åben båd*, som man kunne vælge at opfatte som en term, der fortjener lemmatisering, men som jeg slet ikke har taget med.

Man kan også diskutere, om det var rimeligt at angive *lille båd* som en typisk samforekomst, når det nu gælder for vældig mange genstande, at de kan betegnes som små. Måske var det mere relevant at angive samforekomsterne med *fods*, selvom der var færre af dem.

Endelig kan man diskutere, om sublemmaerne i artiklen om *båd* i tilstrækkelig grad tager højde for de variationer, som korpus viste for udtrykkene *gå i bådene* og *være i samme båd*. Variationer behandler vi forskelligt alt efter, af hvilken art de er. Figur 6 viser nogle eksempler på, hvordan variationsproblemer kan løses.

Som vi har set, indgår der mange overvejelser i beslutningen om lemmatisering versus ikke-lemmatisering af en flerordsforbindelse.

Mutuelle statistikker skal holdes op mod konkordanslister, hvor de konkrete eksempler bag statistikkens tal findes — her afslører bl.a. homograferne sig. Konkordanslisterne skal så holdes op imod andre kilder, bl.a. andre ordbøger. Og ordbogsredaktøren må holde alle kilderne op mod sin egen sprogfornemmelse — og i sidste ende forsøge at forholde sig kritisk til samme sprogfornemmelse!

spurv	(..) skyde spurve/gråspurve med kanoner
jod	(..) rød(t) jod
hals	(..) få galt i halsen el. få i den gale hals
gelé	(..) have gelé i knæene (..) varierer i form <i>Hans sensuelle øjne og bløde basstemme gi'r pigerne gele i knæene</i>
alter	(..) ofre NGT på NGTs alter (..) <i>velfærdstanken .. er nu blevet ofret på overbudets og frådseriets alter</i>
skrive	(..) skrive 'af (..) [NGN skriver NGT af; NGN skriver af efter NGT/NGN; NGN skriver af fra NGT]
luft	(..) der er NGT i luften (..) <i>typisk der er forår i ~en, der er noget i ~en, der er romantik/øretæver i ~en □ Den 29-årige Esbjerg-spiller vendte ikke just den anden kind til, hvis der var lussinger i luften</i>

Figur 6. Et sublemma kan have variationer, og variationerne kan medtages i ordbogen på mere eller mindre pladsbesparende måder.

Vi har hele tiden ordbogens begrænsede omfang at tænke på, og lemmatisering er en pladskrævende løsning. I hvert enkelt tilfælde må vi tage stilling til, om der er tale om en begrebslig helhed, som brugeren skal kunne slå op — eller om der er tale om et valensmønster eller mere eller mindre tilfældige eller typiske samforekomster, der ikke behøver lemmatisering. Når vi tager stilling, må vi se på artiklen som helhed og vurdere de oplysninger, som artiklen i øvrigt giver om hovedlemmaet.

Vi skal finde den løsning, der giver størst mulig brugervenlighed på mindst mulig plads!

Ei ordliste trekt ut av tre ordbøker/ordlister

The article deals with the compilation of a spelling dictionary, *Nynorskordlista*, based on three electronically stored sources. Two of the sources are printed dictionaries. The types of data extracted from each source are described. The data were combined into the compilation format of *Nynorskordlista*. Some advantages of the method are discussed.

1 Innleiing

Dette innlegget er ein presentasjon av arbeidet med ei rettskrivingsordliste — *Nynorskordlista* — som er redigert på dataskjerm. Ordlista — som like gjerne kunne kallast rettskrivingsordbok — er laga på grunnlag av tre elektronisk lagra ordlister. I framstillinga ligg hovudvekta på dei meir tekniske sidene ved redigeringa.

2 Tre hovudkjelder

Nynorskordlista byggjer på desse tre hovudkjeldene:

- *Nynorskordboka*
Definisjons- og rettskrivingsordbok, 2. utg. 1993
90 000 oppslagsord, 755 sider
- *Bokmålsordlista*
Rettskrivningsordliste for skole, hjem og kontor, 1992
56 000 oppslagsord, 397 sider
- *IBM-morfologien*
Elektronisk ordliste med bøyingsmønster
111 000 oppslagsord

Ut frå desse hovudkjeldene — og frå mange andre aktuelle kjelder — har eg saman med Anne Torunn Engø redigert *Nynorskordlista*. Ordlista har fått om lag same omfanget som *Bokmålsordlista*. Utgjevinga skal vente til rettskrivingsvedtaka i Norsk språkråd i fireårsperioden 1992-95 er endeleg godkjende våren 1996.

Nynorskordlista er eit samarbeidsprosjekt mellom Avdeling for leksikografi ved Universitetet i Oslo, Norsk språkråd og Det Norske Samlaget. Initiativet til å utarbeide større rettskrivingsordlister for bokmål og nynorsk kom frå Språkrådet. Desse ordlistene er å rekne som offisielle rettskrivingsordlister på linje med *Retskrivningsordbogen* i Danmark og *Svenska Akademiens ordlista*.

3 Ordbokstype

Nynorskordlista blir ei større ordliste over det sentrale allmenne ordtilfanget i nynorsk. Talet på oppslagsord, bortimot 60 000, blir om lag dobbelt så stort som det vanlege skuleordlister har. Ordtilfanget byggjer særleg på *Nynorskordboka*, men opplegget for ordlista følgjer langt på veg *Bokmålsordlista*, og vi har bygd på den i stor grad.

I tillegg til grunnord og avleiingar, som har sin sjølvsgade plass i ordlista, er det med ein del samansetningar, ikkje minst for å vise samansetningsfuge. Ofte har vi vist samansetjingsmåten slik, utan å ta med døme:

religion [. . .] i sms.: **religions-**

Det er lagt vekt på å få med nye, aktuelle ord frå ymse område. Den offisielle norma i nynorsk har vore nokså atterhalden med å ta inn særleg dansk-tyske lånord, til dels også nyare engelske ord. Ordlista vår bryt ikkje med denne tradisjonen, men har opna litt for desse kategoriane av ord.

Mange namn er med som oppslagsord, særleg stadnamn, norske og utanlandske, i mange tilfelle med innbyggjarnemningar til. Ein del vanlege historiske og mytologiske namn er med, og til ein viss grad namn på statsinstitusjonar og internasjonale organisasjonar. Vanlege forkortingar og symbol er tekne med som oppslagsord.

Om kvart oppslagsord er det først og fremst teke med formelle opplysningar: rettskriving og bøying. Tydinga til orda er gjerne teken med for å skilje homografar frå kvarandre og ved meir ukjende ord. Uttaleopplysningar er med særleg ved uvanlege eller lite kjende ord.

4 Utval frå kjeldene og utforming av redigeringsformat

Dei tre hovudkjeldene er lagra elektronisk i nokså ulike format. I startfasen av arbeidet vart det valt ut aktuelle typar data frå kvar kjelde, og dei vart sette saman til eit redigeringsformat.

Utgangspunktet var *Nynorskordboka*, som vart skriven datamaskinelt i "Hjulstad-formatet", med namn etter Håvard Hjulstad, som laga edb-opplegget for ordboka. I Hjulstad-formatet har kvart oppslagsord ein **post** delt inn i **felt**, som her i dei to oppslagsorda *eple* og *eplekart* i *Nynorskordboka*:

```
=
NN001 eple
NN001a N2
TR006
..OPP #>$CepI@ n1
..ETY (norr $Bepli@; smh med $Bapal@)
..DEF $C1@ frukt av epletre
..UTR $Bplukke, hauste, sylte, mose e- !@
..UTR $Be-t fell ikkje langt frå stamma@
..FOR han, ho liknar far sin el. mor si !
..UTR $Bmåtte bite i det sure e-t@
..FOR måtte finne seg i noko ein ikkje liker
```

```

..DEF $C2@ potet, jordeple
..UTR $Bsetje, hyppe, ta opp e-@
..DEF $C3@ rundvoren del el. ting
..UTR $Badamse- !@
..UTR $Baugee-@
=
=
NN001 eplekart
TR006
..OPP $C-kart@
=

```

På trykk blir dette:

eple n1 (norr *epli*; smh med *apal*) **1** frukt av epletre *plukke, hauste, sylte, mose e- / e-t fell ikkje langt frå stamma* han, ho liknar far sin el. mor si / *måtte bite i det sure e-t* måtte finne seg i noko ein ikkje liker **2** potet, jordeple *setje, hyppe, ta opp e-* **3** rundvoren del el. ting *adamse- / augee- . . . -kart*

Først i kvar post er det linjer som ikkje er med i den trykte versjonen (topplinjene, NN. . . , og TR006), følgde av felt for oppslagsord med bøyning, etymologi, definisjonar og uttrykk. Til *Nynorskordlista* plukka vi ut oppslagsfeltet (. . . OPP), dessutan topplinjene, som inneheld fullt utskrivne oppslagsformer også for samansetningar, og dessutan bøyingskodar etter eit meir utbygd system enn trykkversjonen av boka.

Bokmålsordlista er redigert i eit tekstbehandlingsformat utan feltinndeling, berre med ei blank linje mellom avsnitt, som til dømes i oppslagsordet *eple* med samansetningar:

```
$Ceple@ $B-et@; $B-er@, $B-a@/$B-ene@ $C-kake-kart-slang@
```

Dette er på trykk:

```

eple -et; -er; -a/-ene-kake
-kart-slang

```

IBM-morfologien er ei elektronisk ordsamling med full bøyning for kvart ord. Nynorskversjonen, som vi brukte, har om lag 111 000 ord og berre hovudformene i offisiell rettskriving, ikkje sideformer (klammeformer). Ordsamlinga er utarbeidd av IBM Norge og kjøpt inn av Dokumentasjonsprosjektet. Frå IBM-morfologien vart det trekt ut ei oppslagsform for kvart ord, dessutan delar av bøyingsmønstra i høveleg rekkjefølgje.

Data frå dei tre kjeldene vart så samsorterte alfabetisk slik at det for kvart oppslagsord i råmanuset var samla dei utvalde opplysningane i éin artikkel. Formatet på råmanuset er ein omarbeidd variant av Hjulstad-formatet, og det skal nokså enkelt kunne konverterast til den så allmenne SGML-standarden.

To samsorterte ordartiklar i råmanuset med data frå alle tre kjeldene kan sjå slik ut:

```

IBMopp: eple
IBMbøy: -et; -e, -a

```

```

IBMtyp: NN (802) 17526
NN001 eple
NN001a N2
..OPP #>$Ceple@ n1
NOBløp: 15475
nold
BOL $Ceple@ $B-et@; $B-er@, $B-a@/$B-ene@ $C-kake-kart~slang@
BOLløp: 9247

IBMopp: eplekart
IBMbøy: -en; -ar, -ane
IBMtyp: MN (700) 17538
NN001 eplekart
..OPP $C-kart@
NOBløp: 15487
nold

```

Først i kvar av desse to postane kjem tre linjer frå IBM-morfologien, med oppslagsform, bøying og bøyingstype i IBM-typologien. Oppslagsorda fekk løpenummer i kvar kjelde. Deretter kjem linjene frå *Nynorskordboka* (NN. . . osv.). Tydingsfeltet *nold* er sett inn i redigeringsformatet. Til slutt kjem eventuelt linjer frå *Bokmålsordlista* (BOL).

5 Redigering

Redigeringa gjekk teknisk sett ut på å merkje av dei orda som skulle vere med som oppslagsord i ordlista, å kontrollere og eventuelt endre oppslagsform og bøying, og å setje til uttale- og tydingsopplysningar der det var ønskeleg. Dei to øvste linjene, altså *IBMopp*: og *IBMbøy*:, kunne ofte brukast direkte som dei var, til oppslagsform og bøying. Dei andre linjene fungerer som datagrunnlag, kontroll og referanse. Vi brukte redigeringsprogrammet **emacs** (GNU-emacs), som er kraftig og har mange funksjonar.

Dei same to artiklane som før, ferdig redigerte, ser slik ut:

```

#> #o
IBMopp: eple
nolb -et; -, -a
IBMbøy: -et; -e, -a
IBMtyp: NN (802) 17526
NN001 eple
NN001a N2
..OPP #>$Ceple@ n1
NOBløp: 15475
nold
BOL: $Ceple@ $B-et@; $B-er@, $B-a@/$B-ene@ $C-kake-kart~slang@
BOLløp: 9247
#o
nolo-kart

```


IBMopp: eplekart
nolb
IBMbøy: -en; -ar, -ane
IBMtyp: MN (700) 17538
NN001 eplekart
..OPP \$C-kart@
NOBløp: 15487
nold

Artiklar som skulle med i ordlista, markerte vi med visse teiknsekvensar over postane: #> #o står for ordartikkel i nytt avsnitt, #o står for påhengd artikkel utan nytt avsnitt.

I det første dømet ovanfor kunne oppslagsforma *eple* frå IBM-materialet brukast som ho stod. Bøyinga er ikkje utforma heilt som den som er vald for *Nynorskordlista*, og difor sette vi inn eit nytt bøyingsfelt (*nolb*) over IBM-bøyinga. Vi fastsette at det alltid er det øvste bøyingsfeltet som skal gjelde. Eit eventuelt uttalefelt kan setjast inn før bøyinga. Tydingsfeltet (*nold*) står tomt her.

I den påhengde artikkelen (*eplekart*) sette vi berre etterleddet med tilde framfor. Til det førte vi inn eit nytt oppslagsfelt (*nolo*) og eit tomt bøyingsfelt, fordi samansetningar i regelen ikkje har bøying oppført. Redigeringa gjekk i det heile ut på å leggje til, ikkje stryke.

Oppslagsord som ikkje stod i nokon av hovudkjeldene, vart redigerte i tilsvarende postar som dei over.

Det vart laga utskriftsprogram for to slags korrekturar. Dei fleste felte i redigeringsformatet vart strokne i korrekturutskriftene. Den endelege utskrifta blir om lag slik (temmeleg lik *Bokmålsordlista*):

eple -et; -, -a ~kart ~sider

6 Manglar ved samsorteringa

På to punkt fungerte ikkje samsorteringsprogrammet: ved homografar og ved teikn utanom bokstavane *a-z*. Homografar stod i klyngjer i råmanuset, programmet skilde dei ikkje frå kvarandre. Homografsepareringa måtte difor gjerast enkeltvis på skjermen i redigeringa. Ved ikkje-alfabetiske teikn som bindestrek og aksentar og til dels ved *æ*, *ø* og *å* slo også samsorteringa ofte feil. Kjeldene var ikkje alfabetiserte etter heilt dei same reglane. Det hadde truleg vore teknisk mogleg å lage eit program som greidde opp med både homografar og ulik alfabetisering, men det ville krevje så mykje meir programmeringsarbeid at det spørst om det ville vere verdt det.

7 Informatikarinnsats

Denne arbeidsmåten der så mykje blir gjort elektronisk, krev ein stor programmeringsinnsats, særleg i startfasen. Til samanstøypinga av råmanuset og utarbeidinga av redigeringsformatet fekk vi edb-hjelp frå Dokumentasjonsprosjektet. Det viste seg å bli ein heil prosess å komme fram til eit godt format og ikkje minst ein økonomisk arbeidsmåte, ein prosess med ein del prøving og feiling. Informatikaren frå Dokumentasjonsprosjektet var lydhør og hjelpsam,

og fann alltid tid til å løyse problem. På vegen fann vi fram til ein del tidssparande finessar, som er avgjerande for arbeidstempoet. Det er avgjort ein fordel å vere litt fascinert av kva datateknikken kan utrette. Arbeidet var òg lagt godt til rette med datautstyr på avdelinga, og det har vore god driftsstøtte på HF-fakultetet.

Eg nemner dette fordi vi slapp mange frustrasjonar som sikkert ville komme dersom utstyr og edb-eksperimente ikkje var så lett tilgjengeleg.

8 Vurdering av opplegg og arbeidsmåte

Ein stor fordel med å ha tre hovudkjelder elektronisk lagra og samsorterte er at ein slepp stadig slå opp i desse bøkene/kjeldene og jamføre. Ein er dessutan nokså godt sikra mot at ein overser aktuelle oppslagsord i desse kjeldene. På den andre sida er det ein fare for at ein kan bli for bunden av ordutvalet i akkurat desse kjeldene.

Det blir opplagt mindre skrivearbeid med denne arbeidsmåten enn med meir tradisjonelle måtar i og med at mykje av teksta alt er inne. Ein unngår ein del trivielle skrivefeil. Redigeringa går mykje ut på å flytte skrivemarkøren i posisjon, å kopiere, slette o.l., det vil ofte seie å bruke kontrolltastar, eventuelt mus, i staden for bokstavgastane. Det verkar som desse operasjonane er fysisk meir slitande enn vanleg tekstskriving.

Manuset til ordlista krev stor plass på dataskjermen. Med ein vanleg skjerm er det berre plass til to-tre oppslagsord på skjermiletet. Det ville vere ein fordel med ein større skjerm. I redigeringa var det nødvendig å kike i dei trykte bøkene for å vurdere kva ord som skulle veljast ut. Det var best å ta eit overblikk over trykksider for å vurdere korleis det tilsvarende utsnittet i ordlista skulle bli. Å redigere direkte på skjerm er lite oversiktleg.

Noko av meininga med å lagre ordlister og andre tekster elektronisk er at dei lett skal kunne rettast opp og forbetrast. Arbeidet med seinare utgåver av ordlista skulle såleis vere enklare enn før i tida. Feil som er funne i kjeldene mens arbeidet gjekk føre seg, er noterte nokså systematisk — for å betre kvaliteten på lagra data.

Sidan *Nynorskordlista* finst i elektronisk form, skulle ho liggje godt til rette for gjenbruk. Ordartiklane er delte inn i stort sett homogene felt, utan skriftkodar, og skulle kunne nyttast direkte om ein til dømes vil lage ei rein rettskrivingsordliste utan uttale- og tydingsfelt. Andre aktuelle gjenbruksprodukt er stavekontrollprogram og orddelingsprogram eller -ordbøker.

Ved sida av utskriftsprogram i to versjonar vart det laga enkle program for å plukke ut lister med typar av oppslagsord, til dømes ord merkte med fagområde eller stadnamn og andre namn. Redaktørane har lært å ta utskrifter og lage ymse lister sjølve, det er ein fordel ikkje minst i det avsluttande kontroll- og korrekturarbeidet.

Litteratur

Ordbøker

Bokmålsordlista 1992. Oslo: Universitetsforlaget.

Nynorskordboka 1993. 2. utg. Oslo: Det Norske Samlaget.

IBM-morfologien. Elektronisk ordliste utarbeidd av IBM Norge.

Annan litteratur

- Fjeld, Ruth Vatvedt (red.) 1992: *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden*. Oslo: NFL.
- Garde, Anna/Pia Jarvad (red.) 1994: *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden*. København: LEDA.
- Jörgen Pind 1989: *Computers, Typesetting, and Lexicography. I: Jörgen Pind/Eiríkur Rögnvaldsson (utg.): Papers from the Seventh Scandinavian Conference of Computational Linguistics*. Reykjavík: Institute of Lexicography/Institute of Linguistics.

Tove Jacobsen

Elektronisk snuing av Stor norsk–fransk ordbok

The Large Norwegian-French Dictionary is computer-readable and coded, and as such it is especially suited as a basis for other projects. By reversing the dictionary by computer, it serves as a base for a French-Norwegian dictionary, and, in this manner the material from an active dictionary is transposed to a passive one.

In this paper, the results of the reversing of the dictionary are described, and the effectiveness and quality obtained by such a method are discussed.

Moderne språkteknologi har gjort at gjenbruk av ordbøker blir mer og mer vanlig. En måte å utnytte en allerede eksisterende tospråklig ordbok på, er å snu den automatisk slik at det opprinnelige utgangsspråket blir målspråket. For at en slik elektronisk snuoperasjon skal være mulig og økonomisk lønnsom, må den eksisterende ordboken være lagret elektronisk. Den bør også være kodet på en enhetlig måte. Disse betingelsene var oppfylt for *Stor norsk-fransk ordbok* (heretter SNFO) da den kom ut i 1991 og senere i 1994. Man mente derfor at det ville være både lønnsomt og faglig interessant å snu SNFO, slik at resultatet kunne bli et råmanuskript for en ny stor fransk-norsk ordbok.

SNFO er lagret på tekstfiler i et postformat. Til den nye fransk-norske ordboken har man bestemt seg for å benytte et SGML-format. Det er flere grunner til det. SGML er et internasjonalt kodesystem som kan brukes både til utgivelser på elektroniske medier og på papir. SGML er også system- og maskinuavhengig. Ut fra gjenbruksperspektivet var det også viktig å ha et ordboksformat som kan brukes igjen ved utarbeidelse av nye ordbøker.

Selve snuoperasjonen ble utført ved hjelp av Pascal-programmering. Deretter foregikk redigeringen ved hjelp av Brief som viste seg å være et meget nyttig tillegg til Pascal-programmene.

Resultatet av en automatisk snuing vil naturligvis være avhengig av hva slags type ordbok man hadde i utgangspunktet. For SNFO's tilfelle ville det si at man hadde en ordbok som primært var laget for brukere med norsk som morsmål og som ville produsere tekster på fransk. Derfor inneholdt ordboksartiklene forholdsvis få oversettelsesekvivalenter, men desto flere eksempler hvor oppslagsordet ble oversatt i kontekst. For den nye ordboken som skal være både for forståelse og produksjon på norsk og fremdeles primært for norske brukere, vil behovet være annerledes. Oversettelsesekvivalentene blir det viktigste og antallet kontekster kan reduseres kraftig.

Eksempel (1) viser første del av en ordboksartikkel fra SNFO. Ved siden av å gi opplysninger om oversettelsesekvivalentene, har redaktørene også tatt med et stort antall eksempler for å vise leseren i hvilke kontekster de forskjellige oversettelsesforslagene kan brukes.

(1)

NB001	øde/legge
NB014	v
NB070	(1)
FR001	(<k>om fysisk skade</k>) casser, briser, détruire, (<k>delvis</k>) abîmer, endommager, (<k>fam</k>) esquinter, (<k>om mat</k>) gâter:
NB410	disse husene ble ødelagt under krigen
FR410	ces maisons ont été détruites pendant la guerre;
NB410	kjøtt ødelegges fort i varmen
FR410	la viande se gâte (<k>el</> s'abîme) vite par temps chaud;
NB410	denne gutten ødelegger alt
FR410	cet enfant casse (<k>el</k>brise) tout, cet enfant est un brise-fer;
NB410	hun fikk ødelagt ansiktet i en kollisjon
FR410	elle a été défigurée dans une collision;
NB410	regnet ødela avlingene
FR410	la pluie a endommagé les récoltes;
NB410	skarpe steiner ødela dekkene
FR410	des cailloux pointus ont endommagé (<k>el</k> abîmé, esquinaté) les pneus;
NB410	ø- sin helse
FR410	ruiner sa santé;
NB410	jeg har ødelagt en tann
FR410	je me suis cassé une dent
NB070	(2)

Etter snuoperasjonen vil hver oversettelsesekvivalent i FR001-feltet generere et nytt oppslagsord med en ny artikkel. For *ødelegge* (1) blir resultatet altså hele syv nye oppslagsord: *casser*, *briser*, *détruire*, *abîmer*, *endommager*, *esquinter* og *gâter*. Hvert av dem får med seg hele settet av opplysninger og eksempler. I (2) ses de første fire nye artiklene:

(2)

(a)

FR001	casser
XX000	<k>om fysisk skade</k>
NB014	v
NB001	øde/legge
FR410	ces maisons ont été détruites pendant la guerre
NB410	disse husene ble ødelagt under krigen
FR410	la viande se gâte (<k>el</> s'abîme) vite par temps chaud
NB410	kjøtt ødelegges fort i varmen
FR410	cet enfant casse (<k>el</k>brise) tout, cet enfant est un brise-fer
NB410	denne gutten ødelegger alt
FR410	elle a été défigurée dans une collision
NB410	hun fikk ødelagt ansiktet i en kollisjon
FR410	la pluie a endommagé les récoltes
NB410	regnet ødela avlingene
FR410	des cailloux pointus ont endommagé (<k>el</k> abîmé, esquinaté) les pneus

NB410 skarpe steiner ødela dekkene
 FR410 ruiner sa santé
 NB410 øde/legge sin helse
 FR410 je me suis cassé une dent
 NB410 jeg har ødelagt en tann

=

(b)

FR001 briser
 NB014 v
 NB001 øde/legge
 FR410 ces maisons ont été détruites pendant la guerre
 NB410 disse husene ble ødelagt under krigen
 FR410 la viande se gâte (<k>el</> s'abîme) vite par temps chaud
 NB410 kjøtt ødelegges fort i varmen
 FR410 cet enfant casse (<k>el</k>brise) tout, cet enfant est un brise-fer
 NB410 denne gutten ødelegger alt
 FR410 elle a été défigurée dans une collision
 NB410 hun fikk ødelagt ansiktet i en kollisjon
 FR410 la pluie a endommagé les récoltes
 NB410 regnet ødela avlingene
 FR410 des cailloux pointus ont endommagé (<k>el</k> abîmé, esquiné) les pneus
 NB410 skarpe steiner ødela dekkene
 FR410 ruiner sa santé
 NB410 øde/legge sin helse
 FR410 je me suis cassé une dent
 NB410 jeg har ødelagt en tann

=

(c)

FR001 détruire
 NB014 v
 NB001 øde/legge
 FR410 ces maisons ont été détruites pendant la guerre
 NB410 disse husene ble ødelagt under krigen
 FR410 la viande se gâte (<k>el</> s'abîme) vite par temps chaud
 NB410 kjøtt ødelegges fort i varmen
 FR410 cet enfant casse (<k>el</k>brise) tout, cet enfant est un brise-fer
 NB410 denne gutten ødelegger alt
 FR410 elle a été défigurée dans une collision
 NB410 hun fikk ødelagt ansiktet i en kollisjon
 FR410 la pluie a endommagé les récoltes
 NB410 regnet ødela avlingene
 FR410 des cailloux pointus ont endommagé (<k>el</k> abîmé, esquiné) les pneus
 NB410 skarpe steiner ødela dekkene
 FR410 ruiner sa santé
 NB410 øde/legge sin helse
 FR410 je me suis cassé une dent
 NB410 jeg har ødelagt en tann

=

(d)

FR001	abîmer
XX000	<k>delvis</k>
NB014	v
NB001	øde/legge
FR410	ces maisons ont été détruites pendant la guerre
NB410	disse husene ble ødelagt under krigen
FR410	la viande se gâte (<k>el</> s'abîme) vite par temps chaud
NB410	kjøtt ødelegges fort i varmen
FR410	cet enfant casse (<k>el</k>brise) tout, cet enfant est un brise-fer
NB410	denne gutten ødelegger alt
FR410	elle a été défigurée dans une collision
NB410	hun fikk ødelagt ansiktet i en kollisjon
FR410	la pluie a endommagé les récoltes
NB410	regnet ødela avlingene
FR410	des cailloux pointus ont endommagé (<k>el</k> abîmé, esquinaté) les pneus
NB410	skarpe steiner ødela dekkene
FR410	ruiner sa santé
NB410	øde/legge sin helse
FR410	je me suis cassé une dent
NB410	jeg har ødelagt en tann

=

Den snudde ordboken har på denne måten fått 35% flere oppslagsord enn den opprinnelige, og det er klart at man må foreta betydelige flytte- og strykeoperasjoner. I første omgang kan alle eksemplene hvor det nye oppslagsordet ikke er med strykes. Av de gjenblivne eksemplene er det sikkert også endel som bør fjernes fordi kontekster som tidligere nevnt er mindre interessante for den nye ordboken.

For (2c og 2d) vil det etter en første opprydning stå igjen (3 og 4)

(3)

FR001	détruire
NB014	v
NB001	øde/legge
FR410	ces maisons ont été détruites pendant la guerre
NB410	disse husene ble ødelagt under krigen

=

(4)

FR001	abîmer
XX000	<k>delvis</k>
NB014	v
NB001	øde/legge
FR410	la viande se gâte (<k>el</> s'abîme) vite par temps chaud
NB410	kjøtt ødelegges fort i varmen
FR410	des cailloux pointus ont endommagé (<k>el</k> abîmé, esquinaté) les pneus
NB410	skarpe steiner ødela dekkene

=

Spørsmålet blir deretter om disse eksemplene er interessante. Det er de antageligvis ikke, og fra en lang artikkel kan man sitte tilbake med bare oversettelsesekvivalenten *ødelegge*.

Til gjengjeld vil man ved mindre komplekse oppslagsord utrolig lett få en rekke nye oversettelsesekvivalenter fra ett og samme franske ord etter snuingen. Dette kommer tydelig fram i eksempel (5):

(5)

FR001 avalanche
 YY001 f
 NB014 s
 NB001 fonn
 NB015 -a, pl -er/fenner
 =

FR001 avalanche
 YY001 f
 NB014 s
 NB001 lavine
 NB015 -en
 =

FR001 avalanche
 YY001 f
 NB014 s
 NB001 ras
 NB015 -et
 =

FR001 avalanche
 YY001 f
 NB014 s
 NB001 skred
 NB015 -et
 =

FR001 avalanche
 YY001 f
 NB014 s
 NB001 snøskred
 =

Vesentlig er også at ordboksredaktørene automatisk vil få ut en del viktig leksikalsk informasjon om franske uttrykk som man ellers lett ville overse. Som et eksempel kan nevnes det faktum at mange norske adverb kan eller må oversettes til fransk med en bestemt type preposisjonssyntagme, nemlig *avec* + substantiv. Etter snuingen kommer denne kombinasjonen ut som eget oppslagsord, noe som gjør at man ikke kan unngå å bli oppmerksom på denne forskjellen mellom de to språkene. Erfaring har vist at når man skal oversette fra fransk til norsk blir man veldig bundet til den franske konstruksjonsmåten og man kommer ikke på at hele konstruksjonen faktisk tilsvarer et norsk adverb. I eksempel (6) sees en

del slike kombinasjoner (det finnes en hærsikare av dem) som har kommet fram som egne oppslagsord etter sruingen:

(6)

FR001 avec acharnement (<k>el</k> ténacité)
 NB014 adj
 FR070 adv
 NB001 iherdig
 =

FR001 avec admiration
 NB014 adj
 FR070 adv
 NB001 beundrende
 =

FR001 avec agilité
 NB014 adj
 FR070 adv
 NB001 behendig
 =

FR001 avec âme
 NB014 adj
 FR070 adv
 NB001 sjel/full
 =

FR001 avec amertume
 NB014 adj
 FR070 adv
 NB001 bitter
 NB015 pl bitre
 FR410 regretter qch amèrement
 NB410 angre bittert på noe
 FR410 dire qch avec amertume (<k>el</k> amérement)
 NB410 si noe bittert
 FR410 ce remède a un goût amer
 NB410 medisinen smaker bittert
 =

FR001 avec application
 NB014 adj
 FR070 adv
 NB001 flittig
 FR410 travailler avec application (<k>el</k> entrain)
 =

FR001	avec approbation
NB014	adj
FR070	adv
NB001	bifallende
=	
FR001	avec ardeur
NB014	adj
FR070	adv
NB001	idig
FR410	travailler avec ardeur
Nb410	arbeide idig
=	
FR001	avec attention
NB014	adj
FR070	adv
NB001	oppmerksom
NB015	-me
FR410	suivre qch avec attention
NB410	følge oppmerksomt med i noe
FR410	écouter attentivement, prêter une oreille attentive
NB410	høre oppmerksomt etter
=	

Disse kombinasjonene kan selvfølgelig ikke fortsette å stå som oppslagsord, men må flyttes for å stå under oppslagsordene *acharnement*, *admiration*, *agilité* osv. som da vil inneholde eksempler som *avec acharnement*, *avec admiration*.

Selv i større fransk-franske ordbøker som *Petit Robert* (1993) mangler ofte denne typen opplysning. *Avec approbation* og *avec amertume* finnes ikke, mens *avec admiration* finnes kun som definisjon på adverbet *admirativement*.

Slår man opp i en allerede eksisterende fransk-norsk ordbok (Kunnskapsforlagets) mangler alle av de ovenstående uttrykkene, noe som viser hvor nyttig den automatiske snuingen også kan være for kvaliteten på den nye ordboken. Mange opplysninger av denne typen burde gi denne ordboken et forsprang i forhold til ordbøker som er redigert på tradisjonell måte.

Bibliografi

Ordbøker

Grundt, L. O. 1991 og 1994: *Stor norsk-fransk ordbok*. Oslo: Universitetsforlaget.

Kunnskapsforlagets fransk-norske ordbok 1986. Oslo: Kunnskapsforlaget.

Petit Robert 1993 = Josette Rey-Debove/Alain Rey: *Le Nouveau Petit Robert*. Paris: Dictionnaires Paul Robert.

Annen litteratur

- Jacobsen, Tove/Randi Sæbøe 1996: The electronic conversion of a dictionary: from Norwegian-French to French-Norwegian. I: *Proceedings of the Sixth International Symposium on Lexicography* May 7-9, 1994. Tübingen: Niemayer. (under trykking).
- Massereeuw, P. C./I. Serail 1992: DictEdit: a computer program for dictionary data entry and editing. I: *Euralex '92 Proceedings I-II*. Tampere: University of Tampere, 237-247.
- Nygaard, Anders Drejer 1994: Datamatstøttet vending af ordbøger. I: A. Garde/P. Jarvad (red.): *Nordiske Studier i Leksikografi II*. Rapport fra Konference om Leksikografi i Norden 11.-14. maj 1993. København: Leda, 229-237.
- Honselaar, Wim/Marijke Elstrodt 1992: The electronic conversion of a dictionary: from Dutch-Russian to Russian-Dutch. I: *Euralex '92 Proceedings I-II*. Tampere: University of Tampere.

Kryptosammensætninger og skabsafledninger i virkeligheden og i ordbøgerne

In the paper I discuss the borderline between compounding and derivation exemplified by new words as *skabsbøsse* 'closet queen', *teflonpolitiker* 'teflon politician' and *eurokrat* 'Eurocrat', *økotrend* 'ecological trend'. I suggest the terms *kryptosammensætning* (crypto-compound) and *skabsafledning* (pseudo-derivation) for the results of a word formation created by highly productive elements which are traditionally regarded as affixes or roots but have a function as roots or affixes. The two types are described in present day vocabulary and compared with the treatment of them in two dictionaries.

1 Indledning

Titlen her er ikke helt gennemskuelig, og jeg vil derfor begynde med at definere de to nye termer som jeg opererer med, nemlig *kryptosammensætning* og *skabsafledning*. Jeg har i min bog *Nye ord — hvorfor og hvordan?* (Jarvad 1995), beskrevet kryptosammensætningerne og skabsafledningerne. Jeg vil her noget kortere karakterisere og beskrive disse i den virkelighed som kommer til udtryk i dannelsen af nye ord, hvorefter jeg vil se på behandlingen af dem i ordbøger.

I den traditionelle morfologi, som den fx kommer til udtryk i *Dæmonernes port* (Hansen 1977:10), i *Elementær dansk grammatik* (Diderichsen 1971:19f.), i *Ordlaging og ordelement i norsk* (Leira 1992:16ff.) og i *Att bilda ord* (Thorell 1981:6ff.), tages der udgangspunkt i morfemet som den bærende størrelse som er defineret som mindste enhed med semantisk indhold eller grammatisk betydning. Der skelnes mellem frie morfemer og bundne morfemer. Ud fra morfemet defineres rod, leksem, eller stamme som det også kaldes, bøjningsmorfem, afledning med brug af præfiks og suffiks, sammensætning med brug af rødder.

Selvom det ikke er nyt for læserne her, så vil jeg dog alligevel definere de grundliggende typer morfemer i orddannelsen:

Rod En rod kan ikke opdeles i mindre dele med selvstændig betydning, en rod er et frit morfem, kan stå alene, og en rod har en betydning som fx er angivet i en ordbog. Eksempler er *angst*, *rod*, *spring!*, *ikke*. Rødder er modtagelig for flektion og derivation, og med rødder som udgangspunkt kan der dannes komposition.

Affiks Et affiks (også kaldet afledningsmorfem) kan heller ikke opdeles i mindre dele med selvstændig betydning. Affikserne kan opdeles i to efter deres distribution: forrest i ordet er det **præfikser**, fx *ulykke* = præfikset *u* + roden *lykke*. Står de bagerst er det **suffikser**, *godhed* = roden *god* + suffikset *hed*. Et affiks er bundet til en rod eller et leksem, dvs. man kan ikke danne et nyt ord af to affikser, **uagtig*. Betydningen er af en anden karakter end betydningen af rødder, betydningen kan være at negere lekset,

ændre leksemets ordklasse eller lignende. Ved hjælp af præfikser og suffikser kan man danne nye ord.

Fleksiv Flexiver, bøjningsendelser, angiver det grammatiske indhold. Betydningen er mindre konkret, fx at *e* har indholdet flertal i *mure*, og *-te* angiver udsagnsord i datid i *spiste*. Ved at bøje ord danner man ikke nye ord, men nye ordformer af leksemet. Flexiverne er som affikserne bundne til et leksem, men adskiller sig yderligere fra affikserne ved at der er et begrænset, fast antal af dem, fx kan flertal angives ved *-e*: hus-huse; *-er*: hul-huller; *-r*: pige-piger; *-*: mus-mus; *-a*: stratum-strata; *-i*: konto-konti; *-s*: hotdog-hotdogs, men ikke flere.

Leksem (Stamme) Et leksem kan være en rod, en sammensætning (som består af mindst to rødder), og det kan være en afledning (som består af mindst en rod med mindst et præfiks eller suffiks). Et leksem og herunder en rod kan bøjes til forskellige ordformer, og et leksem og herunder en rod kan være udgangspunkt for at danne nye sammensætninger eller nye afledninger.

2 Overgangsformer

I det følgende vil jeg fokusere på overgangsformer mellem afledning og sammensætning.

Nye ord er fx *heterofil*, *biopat*, *femidom*, *elektrofoni*, *bifil*, *eurokrat*, *kronolekt*, *kvadrofoni*. I analysen viser de sig at være ord uden rødder, idet de er dannet af et præfiks og et suffiks; og det kan man ikke, fordi et af kendetegnene for præfiks og suffiks er at de er bundne morfemer, bundne til en rod.

En række ord som godt kan stå alene, får ved brug i sammensætning en anden betydning, fx *skabsbøsse*, *skabssocialist*, *numsevenlig*, *teflonpolitiker*, *grydeklar*, *herregårdsbil*, *voldspakke*. De har ved en umiddelbar analyse to rødder, men den ene af rødderne har i flere henseender karakter af et affiks.

Dannelser som *biopat* og *skabssocialist* har givet anledning til at diskutere grænsen mellem affikser og rødder for dermed at påvise at grænsen er svær at trække, fordi visse affikser i mange henseender fungerer som rødder og omvendt. Det sker hos fx Gudiksen (1994:13ff.) og Katamba (1993:292ff.).

I moderne orddannelse som den giver sig udtryk i dannelsen af nye ord, der etablerer sig, er en stor del dannet netop i grænselandet mellem affiks og rod.

Et vigtigt kriterium i den traditionelle analyse af ordene er om morfemet er frit eller bundet, mindre vigtigt er betydningen eller morfemernes produktivitet og distribution. Hvis man inddrager betydning, produktivitet og distribution mere, kan man måske beskrive den moderne orddannelse bedre.

Jeg vil derfor regne med to typer rødder og to typer affikser:

Rødder:

- a) Der er frie rødder, dvs. rødder som har en fast betydning som fx angivet i en ordbog, som kan stå alene, og som kan indgå i orddannelsen. Frie rødder er fx *hest*, *rock*, *wok*.
- b) Der er bundne rødder, dvs. rødder som har en betydning af samme faste karakter som den frie rod, som er bundet til et andet element der gør at de to tilsammen udgør et

leksem. De bundne rødder har en mulighed for at overgå til frie rødder over en tid, evt. i forbindelse med den enkelte bundne rods produktivitet. Distributionen i ordmassen er fri, dvs. kan både være forled og efterled. De kan substitueres med en fri rod, fx *neo-* og *nystalinisme*. Det er eksempler som disse: *krono-* og *-lekt* som i *kronolekt*, i princippet er der ikke noget til hinder for dannelsen af *lektografi* 'læren om sproglige varianter'. Disse rødder kalder jeg kryptorødder.

Affikser:

- a) Der er frie affikser, dvs. affikser som er stærkt produktive, som har en billedlig, indskrænket, degenereret eller vagere betydning end når de optræder som en fri rod, og betydningen angiver en retning eller tendens til det som leksemet betyder. Ofte er det ét enkelt ord der er udgangspunkt for analogidannelser. De er stærkt produktive, især med hinanden. De kan ikke have svagtryk som det bundne affiks kan have, og de har trykfordeling som sammensætningen. De kan enten være forled eller efterled. Dem kalder jeg skabsaffikser. Det er eksempler som *-venlig* i *numsevenlig*.
- b) Der er bundne affikser, dvs. affikser som har en ukonkret betydning, som angiver en retning eller tendens til den betydning som leksemet har, eller angiver ordklasse, og de kan ikke optræde uden en rod eller stamme. Bundne affikser overgår ikke til at blive frie affikser eller til at blive rødder. De er enten forled eller efterled. De kaldes affikser. Det er eksempler som *-lig* i *venlig* eller *be-* som i *bebrille*.

Der er altså fire elementer som kan benyttes i dannelsen af nye ord. De nye ord kan, alt efter hvad de er dannet af, karakteriseres som **sammensætninger**, der består af frie rødder, **afledninger** som er fri rod og affiks, **kryptosammensætninger** som er ord dannet med brug af kryptorødder, og **skabsafledninger** som er ord dannet med brug af skabsaffikser.

3 Kryptosammensætninger

Kryptosammensætninger er ord dannet med brug af bundne rødder, dvs. af rødder som har en betydning af samme faste karakter som den frie rod, men som er bundet til et andet element der gør at de to tilsammen udgør et leksem. Disse bundne rødder kaldes herefter kryptorødder. Orddelen *krypto* er af latinsk-græsk oprindelse, og det betyder 'hemmelig, i det skjulte', og disse ordsammenstillinger med brug af bundne rødder er netop i det skjulte at betragte som sammensætninger, selvom de er bundne til noget andet på samme måde som affikserne.

Det andet element som en kryptorod sammensættes med, kan også være en kryptorod, som fx *bio + naut*, *bio + pat*, og det kan være en fri rod som fx *bio + gas*, *bio + industri*. Kryptorødder kan i nogle tilfælde være forled eller efterled, fx er *fon* førsteled i *fonotek*, og efterled i *videofon*. Kryptorødder har mulighed for at overgå til frie rødder over en tid, evt. i forbindelse med den enkelte kryptorods produktivitet. Det er ved at ske med det højproduktive *øko-*.

Man kan skelne mellem tre slags kryptorødder:

3.1

Kryptorødder dannet ved forkortelse. Disse har samme faste betydning som de uforkortede ord, de er stærkt produktive, og de kan indgå i dannelser med egentlige rødder, med andre kryptorødder og med affikser. De kan være både førsteled og andetled, som fx *info*: *grøn-sagsinfo*, *infocenter*, *infofax*, *infomappe*, *infomøde*, *infotainment*, *infotek*, *infoteria*, *infotopia* (det sidste betyder 'informationsutopi'). Som selvstændigt ord er det registreret første gang i 1976.

Det højproduktive *øko*- betyder her 'økologisk', jeg medregner ikke *øko*- i betydningen 'økonomisk' her. Der er mange eksempler: *økobonde*, *økogris* 'gris fra økologisk drevet landbrug', *økokatastrofe*, 'økologisk katastrofe', *økolandbrug* 'landbrug som drives efter et økologisk princip og som kontrolleres herfor', *økolog* 'landmand som dyrker økologisk landbrug', *økomuseum* 'museum om sammenhæng mellem landskaber, dyr, planter, bygninger og redskaber og de mennesker, der har levet og lever i en egn', *økomælk*, *økosfære*, *økoteologi*, *økoturisme* 'økologisk turisme, det at holde ferie i uberørt natur, fx i regnskove og naturreservater'. Kryptoroden *øko*- er i sin betydning efterhånden så afsvækket at det bruges ikke sjældent blot som et plusord, som fx *økotrend*, som i annoncer om tøj blot betegner modetøj lavet af naturprodukter som bomuld, hør.

3.2

Den anden type kryptorødder er dannet med udgangspunkt i en slutdel af et andet ord. I ordet *femidom* 'præservativ for kvinder' er førsteledet *femi*- fra latin, og det betyder 'kvinde-', som man kender det i lidt ældre ord som *feminologi*, *feminologisk*, og fra et dansk ugeblad *Femina*. Andetledet *-dom* er taget fra ordet *kondom*, men i *kondom* har *-dom* ingen selvstændig betydning; *kondom* er if. Gyldendals Fremmedordbog dannet efter opfinderen af kondomet, englænderen Conton. Ved dannelsen af ordet *femidom* får *-dom* tillagt en betydning, nemlig 'præservativ'. Endelsen *-dom* har med denne proces mulighed for at danne nye ord, og *-dom* kan i første omgang anses for et nyt suffiks, og kan eventuelt senere overgå til kryptorod.

Ligeledes kan det være rimeligt at betragte *-arium* som en kryptorod. Den er også opstået ved analogi efter gamle ord som *akvarium*, *herbarium* 'samling af pressede blomster', og nye ord er *atomarium*, *delfinarium*, *dokumentarium*, det registrerede navn *Eksperimentarium*, *lapidarium* 'samling af bevaringsværdige gravsten', *planetarium*, *rosarium*, *solarium*, *sælarium*, *teletarium* 'center for æterbårne medier, som tv-programmer, lydbånd'. Tilfældige dannelser er fx *polarium* 'julemandens center ved Polarcirklen', *imaginarium* 'oplevelsescenter'. Produktiviteten og selvstændigheden af *-arium* og andre kryptorødder viser sig også ved at det kan kombineres ikke blot med ord og orddele af latinsk-græsk oprindelse, men også med danske. I alle ordene har *-arium* en fast leksikalisk betydning, nemlig 'sted, hvor det som førsteledet angiver, er samlet og evt. udstillet'.

I ordene kan *-arium* udskiftes med den frie rod *center*, fx *delfincenter*, *rosencenter*, og navnet (det registrerede varemærke *Eksperimentarium* ville nok have heddet Center for Eksperimenter, hvis det var opstået og navngivet i den periode hvor ordet center var mest

populært). Et ord med lignende betydning er *-land*, som i *badeland*, *chokoland*, *køkkenland*, *tæppeland*, *vandland*.

Med efterleddet *-tek* er der i de sidste tyve år dukket en del nye ord op. Efterleddet *-tek*, som i de gamle ord, der danner mønster for de nye, betyder 'samling', fx *bibliotek*, *glyptotek*, *kartotek*. Nyere ord er *mediatek* (1966), *fonotek* (1959), *videotek* (1971). Førsteleddet angiver hvad samlingen består af. Kryptoroden *-tek* er blevet stærkt produktivt og bruges ikke kun med fremmede førsteled, men også med hjemlige, fx *filmotek*, *helsetek*, *klippotek*, *legetek*, *spisetek*, *vinotek*, *whiskytek*. I mange af disse nye ord betegner *-tek* ikke kun samlinger af det som førsteleddet angiver, men også et sted, hvor det som førsteleddet angiver, foregår, fx *legetek* og *klippotek*.

3.3

Den tredje type kryptorødder er hvad der traditionelt kaldes præfikser og suffikser af græsk-latinsk materiale. Det er typen der kaldes „neo-classical compound“ hos fx Bauer (1988:38) og Katamba (1993:323), og altså regnes til sammensætning, mens det hos Thorell (1981:23) omtales som orddannelse på græsk-latinsk basis og regnet som afledning. En del af disse præ- og suffikser har som kryptorødder en friere funktion. De har samme faste betydning som en rod, kan i mange tilfælde udskiftes med en rod, fx *neo-* og *ny-kapitalisme*, og kan i visse tilfælde optræde alene, men de er oftest bundet til en anden kryptorod eller en egentlig rod. Her er et par eksempler: *etno-* betyder 'folke-, etnisk-, racemæssig'. Det betegner hvad der vedrører folkeslagene, fx *etnobort* 'bort på tøj med fremmedartet mønstre eller farver', *etnobotaniker*, *etnocentrisme*, *etnogruppe*, *etnokulturel*, *etnomagnetisme* 'det at være tiltrukket af etniske kvarterer', *etnopolitik*, *etnowestern*, *etnozoologi*, *etnoøkologi*. Det er på vej frem mod en stor produktivitet. Kryptoroden *-fil* angiver at det er en person som elsker eller ynder det som førsteleddet angiver, som i gamle ord som *frankofil* 'ynder af det franske', *anglofil* 'ynder af det engelske'. Kryptoroden *-fil* er i dag betydningsmæssigt belastet af at det har været brugt til betegnelser for seksuelt afvigere, som fx *bifil*, *homofil*, *pædoofil*, *nekrofil*, og kan i ellers neutrale ord som *amerofil* 'elsker af amerikansk kultur', *negrofil* 'elsker af negre og negerkultur' give 'frække' associationer. *Krypto-* som lægger navn til typen, betyder 'hemmelig' som i *kryptografer*, *kryptografering* 'det at sætte koder på noget så det forbliver hemmeligt'. Skautrup (1968:252) har det med under de ikke-produktive præfikser og oplyser at dannelserne er mere tilfældige. Det er nu på vej frem, med fx *kryptonazist*, *kryptoreligiøs*, *kryptovitalist*, *kryptokatolicisme*, og som man kunne høre i en tv-udsendelse i Danmarks Radio 18. 2. 1994: „kryptofascistiske kneb“.

Kryptoroden *meta-* udtrykker at andetleddet forholdes der til på en abstrakt, spekulativ, teoretisk måde som fx i ordene *metabevidsthed*, *metafiktion*, *metakommunikation*, *metakunst*, *metaleksikografi*, *metamagisk*, *metaniveau*, *metaplan*, *metaprodukt*, *metaroman*. I nye forbindelser angiver *meta-* ofte at andetleddet forholdes der til på en pseudoagtig måde, fx er „en metamorsom vittighed“ en 'søgt, kunstig vittighed', og ordet *metanyhed* er brugt om pressens evne til at skabe de begivenheder pressen omtaler.

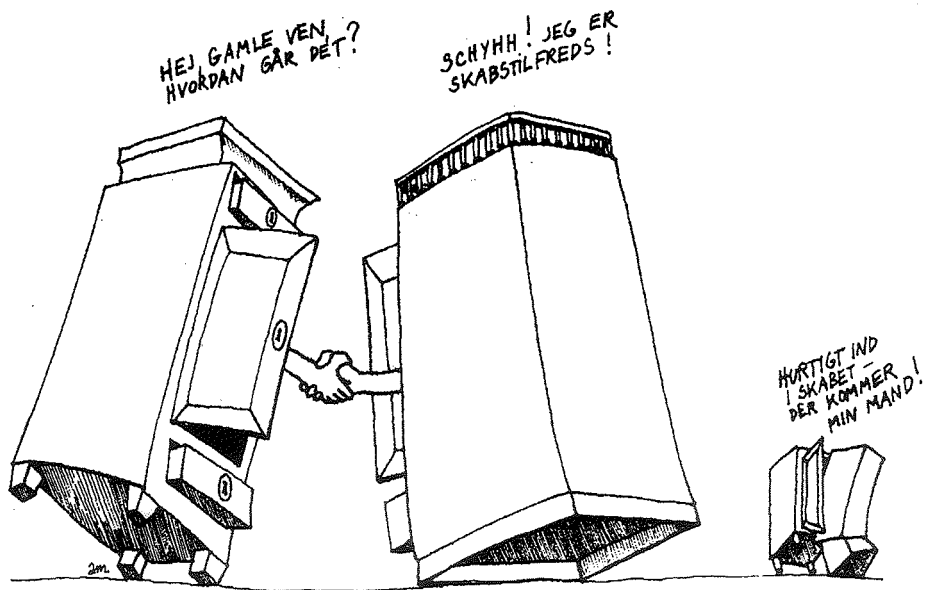
4 Skabsafledninger

Herefter går vi over til beskrivelsen af den anden type i grænselandet, nemlig det jeg kaldte skabsafledninger.

Orddelene *-syndrom* og *-effekt*, som jeg herefter kalder skabsaffikser, sammen med de ord som de sammensættes med, er ud fra en gængs betragtning resultat af en orddannelse bestående af to ord, to rødder, og dermed ligner de slående sammensætninger, men de er dog alligevel noget andet. Skabsaffikser er nemlig uhyre produktive, altså de er i stand til at danne nye lejlighedsdannelse i et overordentligt stort antal, og resultatet, de nye ord, som herefter følgeligt kaldes skabsafledninger, har ofte karakter af modeord.

Det er yderligere karakteristisk for dannelsen af disse nye ord at de har en „moder“, altså et udgangsord hvorfra den særlige betydning stammer, som skabsaffikset har eller får når det bliver produktivt. Denne moder er udgangspunktet for forståelsen af betydningen af nye skabsafledninger.

Skaberi



Grunden til at jeg har valgt ordet skabsafledning som betegnelse for det nye ord, og skabsaffiks som betegnelse for orddelen, er naturligvis at orddelen *skabs-* selv er et sådant og at det er et prominent eksempel: Brugen af *skabs-* som i *skabsaffiks* er nærmest eksploderet inden for de sidste år, fx *skabsgrundtvigianer*, *skabskommunist*, *skabskonservativ*, *skabsmor*, *skabsryger*, *skabsseer*, *skabssocialdemokrat*, *skabssocialist*, *skabsunionist*.

Alle ordene har en fælles moder, nemlig ordet *skabsbøsse* som kendes tilbage fra 1978, men som nok først i midten af 1980'erne bliver gængs. *Skabsbøsse* er en fordanskning af engelsk *closet homosexual* eller det mere slangprægede *closet queen*, og på engelsk betyder *closet* som navneord 'skab, kammer' udover at det betyder *wc*. Men som adjektiv betyder det 'hemmelig, privat', og det er den betydning der er overført til den danske orddel *skab* i ordet *skabsbøsse*; en *skabsbøsse* er således en bøsse som holder sin homoseksualitet hemmelig. Denne betydning går igen i brugen af *skabs-* som skabsaffiks. En *skabsunionist* er en person som i al hemmelighed er tilhænger af unionen, en *skabsseer* er en person som hemmeligt ser fjernsyn, og et *skabsaffiks* er en rod som i al hemmelighed er et affiks osv. Indholdsmæssigt har skabsaffikser en anden, enten udvidet eller indskrænket eller en billedlig betydning i forhold til ordet når det står alene. I visse tilfælde kan man tale om at skabsaffikset blot signalerer en vag betydning i en retning af grundordets betydning.

Det er også karakteristisk at forløbet fra skabsaffikset opstår til at det eksploderer er uhyre kort: for *skabs-* drejede det sig om at moderordet *skabsbøsse* dukker op i 1978 og bliver almindelig i midten af 1980'erne og de nye *skabs-*ord springer ud af skabet i slutningen af 1980'erne.

Nogle skabsaffikser kan kun være førsteled, andre kun andetled, og mange af dem kan både være førsteled og andetled. Skabsaffikser bruges gerne sammen, man kunne nærmest tale om indavl.

Skabsaffikserne er bundne således at de kræver at indgå i dannelse med en anden orddel, og får den for skabsaffikset særlige betydning. Det er disse karakteristika der gør at skabsaffikserne ligner affikser mere end de ligner rødder. Dette svarer også til hvad der er sket tidligere, nemlig en orddels overgang fra at være rod til at blive et affiks, og således er fx den gamle afledning *-dom*, som i dag næppe mere er produktiv, men som man kender i nutidige ord som *alderdom*, *mødom*, *sygdom*, det samme ord som grundordet *dom*.

Man kunne sige at karakteristikkene af en orddel som et skabsaffiks er en beskrivelse af orddelens nuværende status, på et senere tidspunkt kan det overgå til at være et egentligt affiks.

Her er nogle eksempler på skabsaffikser med udvidet eller indskrænket betydning i forhold til grundordet:

budget- bruges som skabsaffiks med betydningen 'billig-, lavpris-, discount-', og betydningen ligger langt fra grundordets betydning, som i en fremmedordbog er angivet som 'overslag over indtægter og udgifter i en periode' og som kendes i ordet *budgetkonto* 'konto til faste udgifter med det formål at udjævne sæsonmæssigt store betalinger'. Her er et par eksempler: *budgetklasse*, *budgetpriser*, *budgetferie*.

-kult betyder i ordbogssammenhæng enten 'religiøs dyrkelse af noget' eller 'ærbødig dyrkelse eller tilbedelse af en person eller et princip', men i ord som *efterretningskult*, *narkokult*, *nøgenkult*, *orgasmekult*, *personlighedskult* er betydningen udvidet til at omfatte også noget som ikke er positivt vurderet, og indskrænket til at ordet i bred almindelighed betyder blot 'dyrkelse'.

-pleje har tidligere nøje haft sammenhæng med mennesker, fx er der i *Nudansk Ordbog* (1992) disse eksempler på brugen af *pleje*: „patienten får en god pleje, sætte, være i pleje, pleje søges til 2 små børn“, men i dag er brugen af ordet udvidet

til således at bruges som i disse eksempler: *citypleje, debitorpleje, erhvervspleje, formuepleje, imagepleje, investeringspleje, kropspleje, kundepleje, landsskabspleje, medarbejderpleje, naturpleje, personalepleje, sorgpleje, talentpleje, tankepleje, vandløbspleje, vælgerpleje, økonomipleje.*

Andre skabsaffikser har billedlig og holdningsmæssig afvigende betydning af grundordet som fx *amatør-* hvor den glade amatør, der med samme sind taber eller vinder, er passé; som skabsaffiks betyder *amatør-* 'dårlig, ubehjælpsomt' som i ordene *amatørmarxisme, amatørpolitik, amatørsmugler, amatørvideo.*

Førsteleddet *barfods-* bruges om noget som er elementært, primitivt, i fx *barfodsforskning, barfodsjournalistik.*

Sidsteleddet *-bjerg* som skabsaffiks betyder 'større (an)samling af det som førsteleddet angiver', fx *affaldsbjerg, fiskebjerg, flæskebjerg, margarinebjerg, pulverbjerg, overskudsbjerg, smørbjerg, uddannelsesbjerg, æblebjerg, ægbjerg.* Synonymt hermed er *-pukkel*, fx *adjunkt-pukkel, arbejdsløshedspukkel, ansøgerpukkel, bogpukkel, boligpukkel, digterpukkel, fjerkræpukkel, fridagspukkel, godspukkel, kaffepukkel, mælkepukkel, svinepukkel, æblepukkel, ørredpukkel.* Et andet synonym *-sø*, bruges kun om samlinger i flydende form, fx *vinsø, mælkesø, skummetmælksø, whiskysø.*

Andre af samme type er *-bank*, fx *jobbank, jordbank, sædbank, videnbank* og *bur-* med betydningen 'som har dårlige livsvilkår', fx *burhøne, burbarn, burlilje.* Skabsaffikset *dåse-* bruges til at angive at andetleddet er færdiglavet, er kunstigt, er standard, som i *dåselatter, dåsemusik, dåsesoftware.* Skabsaffikset *-eksplosion* bruges i ord som *løneeksplosion, kundskabseksplosion, informationseksplosion, uddannelseseksplosion, videneeksplosion.* Førsteleddet *garage-* angiver at det som andetleddet betyder, er 'lille, evt. på amatørbasis, udført under primitive, vanskelige forhold', som i ordene *garageband, garageforlag.* Som skabsaffiks betyder *-jungle* 'uoverskuelig, uigennemskuelig samling af det som førsteleddet angiver', fx *skattejungle, fradragsjungle, cirkulærejungle, køkkenudstyrsjungle, lovjungle, uddannelsesjungle.* Som skabsaffiks betyder *-loft* 'øverste grænse for det som førsteleddet betyder', fx *skatteloft, akkordloft, udlånsloft, indtægtsloft, kreditloft, personaleloft, låneloft.* I politisk sammenhæng bruges *-pakke*, fx *helhedspakke, infopakke, krisepakke, påskepakke, ungdomspakke, værftspakke*, ligesom *-tank*, der også har politikernes særlige bevågenhed, fx *hjernebank, tænketank.*

Skabsaffikser kan have en afsvækket betydning, evt. blot med en vag hentydning til grundordet. Det gælder fx brugen af *-bevidst* i ord som *rygbevidst, klassebevidst, fodbevidst, kønsrollebevidst, forureningsbevidst, kvalitetsbevidst, kropsbevidst, boligbevidst, kvindebevidst, miljøbevidst, forbrugerbevidst, prisbevidst,* og *-ideologi* har afsvækket sin betydning til nærmest at svare til 'ide, samling af ideer', fx *masseideologi, saglighedsideologi, beatideologi, kønsideologi, statsideologi, samfunds-ideologi.* Andetleddet *-strategi* svarer omtrent til 'plan', fx *udviklingsstrategi, karrierestrategi, kulturstrategi, ledelsesstrategi, markedsføringsstrategi, nichestrategi, personlighedsstrategi, planlægningsstrategi, produktstrategi, salgsstrategi, virksomhedsstrategi.*

Andre af samme type er fx *-orienteret* som i *anvendelsesorienteret, billedorienteret, emneorienteret, forbrugerorienteret, fattigdomsorienteret, individorienteret, markedsorienteret, personorienteret, projektorienteret; -rettet: erhvervsrettet, fremtidsrettet, publikums-*

rettet; -rigtig: børmerigtig, brugsrigtig, energirigtig, ernæringsrigtig, fodrigtig, forbrugerrigtig, miljørigtig, prisrigtig; -venlig: allergivenlig, børmevenlig, handicapvenlig, kropsvenlig, miljøvenlig, naturvenlig.

5 Kryptosammensætninger og skabsafledninger i ordbøger

I gennemgangen af virkelighedens kryptosammensætninger og skabsafledninger har jeg nævnt de mere udbredte og prominente. Jeg vil nu se på hvordan de er behandlet i to ordbøger, nemlig *Nudansk Ordbog*, udgaven fra 1992, og Blinkenberg & Høybye: *Dansk-Fransk Ordbog* (1991).

Nudansk Ordbog er en ensproget modersmålsordbog, dens modtagergruppe er ifølge den selv bred, ja nærmest altfavnende, nemlig til alle der arbejder med det danske sprog i hjemmet, på ethvert kontor, i skoler og læreanstalter. I sit anlæg er den både en aktiv ordbog med henblik på produktion af dansk tekst og en passiv ordbog til forståelse af dansk tekst. If. Svensén (1987:10) er betydningsangivelser i modersmålsordbøger ved den passive brug af ordbogen central, mens den er af sekundær betydning for brugeren ved aktiv produktion. Da kryptosammensætninger og skabsafledninger repræsenterer et kerneområde i nyordsproduktionen, og de også i almindelighed bruges i en for grundordet afvigende betydning skulle der være grund til at antage at de også er fyldigt repræsenteret.

Dansk-Fransk Ordbog (1991:V) er en tosprogsordbog for studerende, korrespondenter, tolke, fagfolk, altså en knap så bred målgruppe med en større baggrund end *Nudansk Ordbog*. Den er primært for danske brugere som skal kommunikere på fransk, altså er den aktive brug af ordbogen i centrum, selv om det oplyses at den også er et hjælpemiddel for franskskyndige udlændinge som studerer dansk og som derved vil benytte den til passiv brug. Det primære i tosprogsordbogen er at give ækvivalenter på målsproget til kildesprogets ord og udtryk. Sandsynligvis har den enkelte skabsafledning og den enkelte kryptosammensætning som fx *rosarium*, *garageforlag* ikke helt samme betydning i en tosprogsordbog, derimod er det vitalt med en angivelse af brugen af forled og efterled når de benyttes på en afvigende måde i forhold til grundordet, med angivelse af den tilhørende ækvivalent. Det er jo ikke nødvendigvis sådan at selv om fransk har ækvivalenterne *aquarium*, *delphinarium*, *solarium*, *terrarium* til dansk *akvarium*, *delfinarium*, *solarium*, *terrarium* så er resten også parallelle, fx svarer *sælarium* til et længere udtryk, *bassin aux phoques*, *dokumentarium* til *programme documentaire*. I indledningen til ordbogen oplyses der således at der medtages opslagsord som er ord i gængs forstand, derudover optages præfikser, fx *gen-* som i *genhuse*, suffikser, fx *-løs* som i *barnløs*, førsteled i en sammensætning, som *krigs-*, sidsteled i en sammensætning, som *-bank*, fx *blodbank*.

5.1 Kryptosammensætninger

I *Nudansk Ordbog* er *-fil*, *-tek* og *krypto-* med som opslagsord. Det er angivet, at de har status som præfiks eller suffiks, en terminologisk oplysning som er i overensstemmelse med traditionen. Derudover gives oplysninger om brug og betydning efterfulgt af eksempler.

Kryptorødderne *meta-* og *øko-* er ikke med, men ordbogen har dog medtaget centrale kryptosammensætninger og behandlet dem som enkeltstående leksikaliserede sammensæt-

ninger på alfabetisk plads. Det drejer sig om *metakritik* 'kritik af kritik' og *økolog*, *økosystem*. Kryptorødderne *-arium*, *etno-* og *info-* er ikke med i ordbogen.

Det samme mønster ses i *Dansk-Fransk Ordbog*. Her er det *krypto-* og *øko-* der er med som opslagsord; de har status som præfikser og på alfabetisk plads er yderligere sammensætninger som *kryptokalvinist*, *kryptokatolik*, *økofeminisme*, *økofeminist*, *økopacifisme*, *økosystem*. Derimod er en række kryptosammensætninger med på alfabetisk plads, og dermed behandlet som enkeltstående leksikaliserede sammensætninger i *Dansk-Fransk Ordbog* som i *Nudansk Ordbog*. Således er *etno-* repræsenteret ved opslagsordet *etnolingvistik*, *meta-* med *metasprog*, *metasproglig*, *-arium* med *delfinarium*, *dokumentarium*, *solarium*, *sælarium* på alfabetisk plads, ligesom *-fil* er repræsenteret med *bifil*, *homofil* og *pædofil*.

5.2 Skabsafledninger

I *Nudansk Ordbog* er visse skabsafledninger angivet: Det drejer sig om opslagsordet *bank* hvor der oplyses: „[...] (overf.) om sted, hvor noget samles og distribueres, *blodbank*, *sædbank*“, og under opslagsordet *eksplosion*: „[...] (overf.) meget hurtig udvikling, voldsom forøgelse“, men her er ikke angivet eksempler på brugen. Under opslagsordet *jungle* står der: „[...] i sammensætn. om hvad der er svært at finde ud af, fx *asfaltjungle* om storby“ og ved opslagsordet *venlig*: „[...] som 2. led i sammensætn. i betydn. 'til fordel for', fx *fodvenlig* (om fodtøj), *forbrugervenlig*.“

I en række tilfælde er denne slags oplysninger ikke angivet, derimod er kun moderordet for det produktive skabsaffiks medtaget som opslagsord. Det drejer sig om *barfodslæge* 'sundhedsplejer i Kina, der tager sig af elementære sygdomme i landområderne', *burhøne* 'høne, der bliver holdt i et snævert bur', *dåselatter* 'latter optaget på lydband'.

Under opslagsordene *orienteret* og *strategi* antydes en mulighed for at brugeren kan indtolke en brug som skabsaffiks: „[...] (overf.) være godt orienteret i et emne; være amerikansk orienteret[...]]: have amerikanske sympatier“; „[...] planlægning af militære operationer; også i alm.: taktik“.

For skabsaffikserne *amatør-*, *-bevidst*, *-bjerg*, *budget-*, *-effekt*, *garage-*, *-ideologi*, *kult-*, *-land*, *-loft*, *-pakke*, *-pleje*, *-rigtig*, *skabs-*, *-syndrom*, *-tank* er der ingen oplysninger.

I *Dansk-Fransk Ordbog* er der i flere tilfælde angivet muligheden for at bruge et opslagsord som skabsaffiks. Det gælder *bank* hvor der står „[...] (produktivt 2. led i sammensætning) *blodbank*, *databank*, *sædbank*“; *bevidst* med betydningsangivelse på fransk og med eksemplerne *klasse-*, *krops-*, *pris-*; *-orienteret* med eksemplerne *forbrugorienteret*, *miljøorienteret*, *resultatorienteret* og *venlig* med eksemplerne *arbejder-*, *dansk-*, *kommunist-*, *Dreyfus-*, *tysk-*.

Også i *Dansk-Fransk Ordbog* er i en række tilfælde denne type oplysninger ikke angivet, derimod er kun moderordet for det produktive skabsaffiks medtaget som selvstændigt opslagsord. Det drejer sig om *barfodslæge*, *dåselatter*, *befolkningseksplosion* (som gemmer sig under opslagsordet *eksplosion*) og *skattejungle*. Under *loft* antydes en mulighed for at brugeren kan indtolke en brug som skabsaffiks ved oplysningen „fig“ med eksemplet „lægge loft over priserne“, og på alfabetisk plads finder man *skatteloft*, *kreditloft*, *udlånsloft*, men ikke fx *indtægtsloft*, *akkordloft*. Imidlertid er langt de fleste skabsaffikser ikke med, fx *budget-*, *bur-*, *garage-*, *kult-* eller *skabs-*.

Det fremgår således at i disse to ordbøger er der mangelfulde oplysninger om yderst

produktive nyere og nye orddele. Kun ganske få bliver således behandlet som det de er, og hovedparten af ord dannet ved brug af skabsafledninger og kryptosammensætninger opfattes som leksikaliserede enkeltord.

Litteratur

- Bauer, Laurie 1988: *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
- Blinkenberg, A./P. Høybye 1991: *Dansk-Fransk Ordbog*. København: Nyt Nordisk Forlag Arnold Busck.
- Brüel, S. m.fl. 1993: *Fremmedordbog*. 11. udg. København: Gyldendal.
- Diderichsen, Paul 1971: *Elementær dansk grammatik*. 3. udg. København: Gyldendal.
- Gudiksen, Asgerd 1993: Suffikset -vorn i de danske dialekter. I: *Danske Folkemål*, 1993, 3–248, København.
- Hansen, Erik 1977: *Dæmonernes port*. København: Hans Reitzel.
- Jarvad, Pia 1995: *Nye ord — hvorfor og hvordan?* København: Gyldendal.
- Katamba, F. 1993: *Morphology*. London: The Macmillan Press Ltd.
- Leira, V. 1992: *Ordlaging og ordelement i norsk*. Oslo: Det Norske Samlaget.
- Nudansk Ordbog* 1992. København: Politikens Forlag.
- Skautrup, P. 1968: *Det danske Sprogs Historie*, bd. 4. København: Gyldendal.
- Svensén, Bo 1987: *Handbok i lexikografi*. Stockholm: Esselte.
- Thorell, Olof 1981: *Svensk ordbildningslära*. Stockholm: Esselte.

Den færøske modersmålsordbog — en midtvejsrapport¹

This article is a half-way report from the Faroese monolingual dictionary, which is being prepared. It gives an outline of the edition of the dictionary which will be published in a year or two. The first part contains an outline of Faroese dictionaries historically and Faroese book publications during the last 173 years. The second part deals with the problems concerning the editing of the dictionary, especially the selection of the head words and, in this context, the conflict between purism and the spoken language. The general problems concerning the situation of Faroese as an intermediate form of language between West Scandinavian and East Scandinavian areas are discussed: where is Faroese placed typologically in this spectrum? What do we do with all the Danish loanwords and the neologisms?

1

1.1 Indledning

I mit indlæg vil jeg omtale færøsk-danske ordsamlinger og -bøger gennem tiderne samt seddelsamlingen, som arbejdet med modersmålsordbogen bygger på. Jeg vil starte med en oversigt over færøske bogudgivelser fra 1822–1994, hvilket kan give en indsigt i det samlede antal trykte tekster med tanke på excerpering. I dette tidsrum på 173 år er der udkommet 3 180 bøger på færøsk. Før århundredskiftet — et tidsrum på 78 år — udkom der 22 bøger, mens der fra fx 1974–1994 udkom over 2 000 bøger eller $\frac{2}{3}$ af det samlede antal.

<i>Tidsperiode</i>	<i>Udkommet på færøsk</i>	<i>Titler i gennemsnit pr. år</i>	<i>Befolkningstallet i tidsperioden</i>
1822–1899 (78 år)	22 bøger	0,28	6 200–15 000
1900–1939 (40 –)	197 –	4,90	15 000–28 000
1940–1949 (10 –)	124 –	12,40	28 000–31 800
1950–1959 (10 –)	173 –	17,30	31 000–34 600
1960–1969 (10 –)	274 –	27,40	34 600–38 600
1970–1979 (10 –)	686 –	68,60	38 600–43 200
1980–1989 (10 –)	1 074 –	107,40	43 200–47 834
1990–1994 (5 –)	630 –	126,00	47 834–43 678
1822–1994 (173 –)	3 180 bøger		

Tabel 1. *Færøske bogudgivelser 1822–1994*

¹Foredraget er delt op i to. Den første del står Anfinnur Johansen for og Jógvan í Lon Jacobsen står for den sidste.

1.2 Svabo og Mohr

De første færøske ordoptegnere er Jens Chr. Svabo (1746–1824) og Nicolai Mohr (1742–1790). De var barndomsvenner, begge tog studentereksamen fra latinskolen i Tórshavn i 1765 og begge rejste samme år til København for at studere de på daværende tidspunkt moderne fag, økonomi og naturhistorie, og de er dermed de første færinger, der bryder med den gamle færøske tradition at studere teologi. De tog begge examen philosophicum i 1769, men aldrig nogen afsluttende eksamen.

Jens Chr. Svabo begyndte tidligt at interessere sig for sit modersmål. Det fremgår bl.a. af, at et af manuskripterne til hans *Forsøg til en Ordbog eller Ordsamling i det færøske Sprog* er dateret på Regensen i København den 18. juni 1773. Det indeholder ca. 4 900 færøske ord. Og den interesse for færøsk, som dette *Forsøg* vidner om, bevarede han hele livet igennem. Ligeså lidt som vi ved, hvordan Svabo kom på den tanke at forfatte en færøsk-dansk-latinsk ordbog, ligeså lidt ved vi, hvilket forbillede han har haft ved anlægget af den. Ordbogen er blevet til uden for Færøerne, forfattet af en mand, der i en ung alder rejste hjemmefra og som i mange år næppe har haft anledning til at høre andre end studenter tale sproget, og som desuden måtte begynde på bar bund. På denne baggrund er den en imponerende bedrift. *Forsøget*, der nu kendes fra fem næsten identiske egenhændige håndskrifter, betragtede han selv kun som en fejlfuld og ufuldstændig redaktion, og på sin Færø-rejse 1781–82, som var en undersøgelsesrejse og hvis resultat blev det store værk *Indberetninger indhentede paa en allernaadigst befalet Reise i Færøe i Aarene 1781 og 1782*, fik han forøget sin ordsamling „næsten til det dobbelte“, som han selv siger. Nu eksisterer der ikke noget med Svabos håndskrevet ordbogsmanuskript fra 1780-årene, men vi tør formode, at samlingerne fra rejsen indgår i det af Nicolai Mohr skrevne ordbogshåndskrift, der også indeholder så godt som hele materialet fra *Forsøget*. Håndskriftet rummer ca. 7450 ord, deraf ca. 950 ord, som alene må tilskrives Mohr. Grunden til, at Nicolai Mohr har skrevet dette ordbogshåndskrift, må til dels tilskrives Svabos elendige tilstand i det pågældende tidsrum, da han var plaget af økonomiske sorger, mismod og sygdom. Men kun til dels, for Svabo besad ikke de grundige kundskaber i islandsk, som var nødvendige for at gøre hans ordbogsarbejde så værdifuldt for den lærde verden, som han ønskede. Det var oprindeligt Svabos tanke, foruden at meddele ordstoffet, også at give oplysninger om ordenes slægtskab med ord i andre sprog. I de manuskripter, som han selv har skrevet, finder man kun få sådanne etymologier, men man finder mange i Nicolai Mohrs håndskrift. Nicolai Mohr havde gode kundskaber til islandsk, dels fordi han i 1780–81 opholdt sig i Island, og dels fordi han i firserne havde gode kontakter med islændinge, der opholdt sig i København, og at han hos dem let kunne kontrollere sin viden. Mohrs etymologier beløber sig vel til et par tusinde eller mere, og de er i langt de fleste tilfælde så indlysende rigtige, selv der, hvor færøsk form afviger fra den islandske, at de ikke sjældent har kunnet udnyttes af senere forskere af det færøske sprog.

År 1800 rejste Svabo tilbage til Færøerne, hvor han levede resten af sit liv. Efter sin tilbagevenden tog han fat på en ny redaktion af sin ordbog, og denne arbejdede han på til sin død i 1824, idet han stadig indsamlede stof. Denne sidste redaktion — *Dictionarium med Appendices* — synes at være uafhængig såvel af *Forsøget* som af det i Mohrs manuskript optagne nye materiale. Dette er det største af Svabos ordbogshåndskrifter med tilsammen 8 400 opslagsord, hvoraf ca. 2 500 ord ikke findes i hans andre håndskrifter. Professor Chr. Matras udgav i 1966 alt Svabos ordbogsværk (indbefattet det af Nicolai Mohr skrevne) i

ét bind i en sådan form, at man let kan se, hvad der står i hvert manuskript. Bind II, der udkom i 1970, indeholder bl.a. et fuldstændigtregister i normalretskrivning (Svabo og Mohr anvendte en fonetisk ortografi) over alle opslagsord samt ord, der kun indgår i citaterne. Her skal tilføjes, at i 1780-årene skrev Svabo også et glossarium som et tillæg til sit store visehåndskrift. Det er det første forsøg på at forklare forældede og vanskelige ord og udtryk i det færøske visesprog. Det indeholder 840 ord og er kun i ringe udstrækning blevet optaget i hans andre ordbogshåndskrifter. Glossariet blev udgivet i 1943 af Chr. Matras.

1.3 Lexicon Færoense

Svabos og Mohrs leksikografiske arbejder blev ikke meget påagtet de mange påfølgende år. Bortset fra Rasmus Rask, der til brug for sin færøske grammatik skrev et uddrag af ord fra to af Svabos håndskrifter, er Svend Grundtvig og Jørgen Bloch de første, der fuldt ud udnytter Svabos og Mohrs arbejder til deres færøske ordbog, *Lexicon Færoense*, fra 1877–88. Dette store værk, der indeholder ca. 15 000 ord, er aldrig blevet trykt. Grundstammen i *Lexicon Færoense* er Svabos og Mohrs ordbogsarbejder, og ordmaterialet derfra beløber sig til ca. 11 000 ord. Af de ca. 4 000 resterende ord er 2 500 ord hentet fra *Føroya kvæði*, *Corpus Carminum Færoensium*, 550 fra oversættelsen af *Matteus evangelium* og *Færingesaga*, 400 fra to færøske grammatikker og ca. 650 fra forskellige prosaskrifter.

1.4 Færøsk Anthologi og dr. Jakob Jakobsen

V. U. Hammershaimbs *Færøsk Anthologi* udkom i København 1891 med et tekstbind og et ordsamlingsbind, forfattet af dr. Jakob Jakobsen. V. U. Hammershaimb og Ludvig Wimmer havde i 1886 opfordret dr. Jakobsen til at forfatte denne ordsamling. Han indvilligede men indså snart, at han blev nødt til at rejse til Færøerne for at foretage forskellige efterforskninger, bl.a. i dialekterne, og på denne rejse grundlagde han også sit betydningsfulde indsamlingsarbejde, som er en vigtig bestanddel af det ordmateriale, der findes i vores store ordsamling. Bind 2 af *Færøsk Anthologi* er den første trykte ordsamling eller ordbog på færøsk, dvs. færøsk-dansk ordbog. Den indeholder godt og vel 6 000 opslagsord. Dette ordmateriale går så at sige igen i alle senere udkomne færøske ordbøger.

Da dr. Jakobsen var ved at afslutte sin shetlandske ordbog, var hans hensigt at gå i gang med en stor videnskabelig færøsk ordbog. I en ansøgning til lagtinget, dateret 19. juni 1918, om økonomisk støtte meddelte han, at det færøske ordforråd må anses at være lige så stort som det norske eller islandske, dvs. ca. 40 000 ord. Samme dag som denne ansøgning var til behandling i lagtinget, den 15. august 1918, døde dr. Jakobsen, og dermed blev disse planer ikke til noget. Da første bind af Sigfús Blöndals islandsk-danske ordbog udkom i 1922 anmeldte professor Johannes Brøndum-Nielsen den i dagbladet *Politiken*. I slutningen af denne anmeldelse siger han, at det ville være ønskeligt, om der også snart udkom en „Søsterordbog: en Færøsk-Dansk“ (Brøndum-Nielsen 1922).

1.5 Færøsk-danske ordbøger i dette århundrede

På initiativ af Brøndum-Nielsen blev dette arbejde sat i gang under ledelse af Chr. Matras og M. A. Jacobsen. Denne færøsk-danske ordbog udkom i 1928. Den bygger på alle de tidligere eksisterende ordbogsarbejder foruden en hel del nyindsamlet materiale. Ved udarbejdelsen af denne ordbog blev grunden nemlig lagt for vores nuværende seddelsamling. Ordbogen indeholder knapt 20 000 opslagsord, og forfatterne siger i forordet, at det er knapt halvdelen af det færøske ordforråd — samme antal, som dr. Jakobsen angav i sin ansøgning. Af forordet fremgår også, at det exciperede materiale hovedsageligt er hentet fra bøger, tidsskrifter og aviser og indeholder derfor størstedelen af de nyfærøske skriftmålsord, men da mange sådanne ord ikke har fået en fast struktur, hverken i form eller betydning, har redaktørerne været forsigtige med at optage dem i ordbogen.

Da 1928-udgaven kun indeholdt ca. halvdelen af det ordmateriale, der fandtes i sproget, var den ret ufuldstændig, og behovet for at få en ny og forøget udgave voksede, både fordi det ret unge færøske skriftsprog efterhånden udviklede sig som litteratursprog, og fordi ordbogen efter en del år blev udsolgt. Excerpteringsarbejdet blev dog fortsat, men der skulle gå over 30 år, før en ny forøget udgave så dagens lys. Redigeringsarbejdet til denne udgave, der blev forestået af Chr. Matras i København, blev påbegyndt i 1959, og to år senere, i 1961, udkom bogen. Den er på 30 000 opslagsord. I tidsrummet mellem 1. og 2. udgave havde man fået skænket flere enkeltpersoners ordsamlinger. Desuden blev forskellige bogudgivelser exciperet, de færøske guldalderforfattere, dvs. fra slutningen af tyverne og nogle decennier frem, men især bøger med kulturhistorisk indhold, og materiale af denne art er i særdeleshed velrepræsenteret i seddelsamlingen. Dette materiale blev medtaget i ordbogen. Det talte sprog blev dog i stor udstrækning forfordelt.

I 1958 blev Málstovnur Føroya Fróðskaparfelags grundlagt, og en af dens første hovedopgaver var at købe et xerox-kopi af ordsamlingen, som Chr. Matras havde grundlagt og videreført alle de år, han var professor i København. Da Chr. Matras blev professor ved det nye Fróðskaparsetur Føroya i 1965, bragte han den originale seddelsamling med sig og sendte xerox-kopierne tilbage til Københavns Universitet. I 1968 blev Jóhan Hendrik W. Poulsen ansat på Fróðskaparsetur Føroya og gik da i gang med at udarbejde et supplementsbind til 1961-udgaven, som udkom i 1974. I tidsrummet imellem disse to udgaver voksede ordsamlingen meget, både på grund af excipering af bøger og andre skrifter og fra det talte sprog. Supplementsbindet indeholder ca. 15 000 opslagsord, hvoraf ca. 2 500 er tilføjelser til ordbogsartikler i hovedbindet. Her er også på sin plads at tilføje, at præsten A. C. Evensen begyndte i 1905 sin udgivelse, i hæfter, af en færøsk-dansk ordbog, som dog ikke blev tilendebragt — tiende og sidste hæfte udkom i 1909 og slutter med ordet *forfumma*. Ordbogen er på ialt 160 sider. A. C. Evensen har med sin ordbog ikke kun tænkt sig at belyse det færøske ordmateriale, sådan som det var, men har også tænkt på at udvide det, der hvor han syntes det var nødvendigt. Således har han hentet en hel del ord fra oldnordisk, og han har på en forstandig måde tillempt dem færøske sprogeregler. Desuden indeholder den en hel del nyoptegnede ord fra øen Sandoy, hvor han var præst. En stor del af hans materiale havde ligget ubenyttet hen både i 1. og 2. udgave men blev medtaget i supplementsbindet, med signaturen ACEv.

Under redigeringsarbejdet af supplementsbindet indså J. H. W. Poulsen, at en hel del af seddelsamlingens ældre ordmateriale var dårligt belyst både med hensyn til betydning, brug og udbredelse, og hvis det skulle med i en ordbog, var det nødvendigt at få mere

sikre og pålidelige oplysninger om dette ordmateriale. Med den islandske radioudsendelse *Íslenskt mál* som forbillede begyndte J. H. W. Poulsen efteråret 1969 en udsendelsesrække i den færøske radio med titlen *Orðabókin*. Udsendelserne, der er blevet sendt hver 14. dag vinterhalvåret, dog med enkelte års udebliven, har været af en halv times varighed, og hidtil er det blevet til 198 udsendelser. Hovedformålet har som nævnt været at spørge lytterne om oplysninger om dårligt belyste ord i seddelsamlingen, og gennem disse meget populære udsendelser har man dels fået mange sådanne værdifulde ordoplysninger, men især oplysninger om ord, der aldrig før er blevet optegnet, hovedsagelig dialektord og ord, der til dels er gået af brug. Det exciperede materiale beløber sig til ca. 20 000 sedler. Kun en lille del af dette materiale kom med i supplementsbindet fra 1974, så den resterende del bliver benyttet i udarbejdelsen af modersmålsordbogen. Disse radioudsendelser har tillige haft den store betydning, at venskabsbånd er blevet knyttet mellem ordbogen og informanter i hver eneste bygd, som de nuværende redaktører kan drage nytte af.

1.6 Seddelsamlingen

Vores nuværende seddelsamling består skønsmæssig af mellem 500 000 og 600 000 sedler med et samlet ordantal på ca. 200 000. Redigeringsarbejdet til modersmålsordbogen startede i 1989, og der er ikke foretaget nogen ny excerpering til den, bortset fra enkelte fagbøger af nyere dato. Den bygger derfor på ordstoffet i de tidligere udkomne ordbøger samt selvfølgelig også på den nuværende seddelsamling. Dels på grund af manglende bevillinger og dels på grund af mangel på kvalificeret arbejdskraft er der heller ikke blevet foretaget nogen systematisk excerpering de sidste mange år, og slet ingen de sidste 5–8 år, og som det fremgik af oversigten over færøske bogudgivelser fra 1822 til 1994 er $\frac{2}{3}$ af det samlede antal færøske bøger udkommet de sidste 20 år, dvs. at et stort antal tekster ligger derfor uexcerperet hen. For at råde bod på disse mangler foretog man en elektronisk tekstindsamling fra forskellige færøske trykkerier og aviser, også omfattende nyhedsteksterne fra den færøske TV-avis gennem 10 år. Hele denne indsamlede tekstmasse beløber sig til ca. 8 millioner løbende ord. Ud fra dette tekstkorpus lavde man en færøsk stavekontrol, der består af ca. 180 000 ord og ordformer, som er blevet indkorporeret i vores redigeringsprogram, som redaktørerne så kan benytte sig af, således at da et bogstav bliver redigeret, kan redaktørerne sideløbende slå op i stavekontrollen, hvor man ser ordet isoleret, uden kontekst, for at kontrollere, om det er med i ordbogen. På den måde har man fået en slags sikkerhedsventil over det nutidige sprog.

Apropos det sidste kan jeg tilføje, at der stort set aldrig er blevet foretaget en systematiseret excerpering af de færøske aviser. Og på Føroyamálsdeildin findes der også en stor båndsamling fra hele landet med bl.a. samtaler med ældre mennesker om gammel levevis osv., og dette delvist udskrevne materiale er kun i meget ringe udstrækning blevet exciperet. Forhåbentlig bliver ordbogsarbejdet i fremtiden prioriteret så meget, at disse huller kan blive udfyldt, i hvert fald så at 2. udgave af modersmålsordbogen kan blive suppleret med dette materiale.

I tabel 2 gives der en oversigt over opslagsordenes antal i de diverse ordsamlinger og ordbøger. Som det fremgår er opslagsordenes antal forøget med ca. 20 000 fra 2. udgave med supplementsbind til modersmålsordbogen:

J. C. Svabos „Forsøg“	4 900
N. Mohrs ordbogshåndskrift	7 450
J. C. Svabos „Dictionarium“ med „Appendices“	8 400
Lexicon Færoense	15 000
Færøsk Anthologi II. Ordsamling	6 000
Føroysk-donsk orðabók, 1. útgáva	20 000
Føroysk-donsk orðabók, 2. útgáva	30 000
Føroysk-donsk orðabók. Eykabind	15 000
Móðurmálsorðabókin	ca. 65 000

Tabel 2. Antal opslagsord i færøske ordbøger og ordsamlinger.

2

2.1 Indledning

Her vil lemmasektionen i forbindelse med modersmålsordbogen i korthed blive omtalt og status gjort, nu hvor godt og vel halvdelen af ordbogen er færdig redigeret. Denne fremstilling er derfor en midtvejsrapport fra slutningen af mai 1995, hvor en del leksikografiske tvivlsspørgsmål endnu er uafklaret.

Færøsk har de samme problemer, som Ásta Svavarsdóttir (1994) nævner i sin fremragende artikel om ordforrådet i den islandske modersmålsordbog i *LexicoNordica* 1. Udskifter man i hendes artikel ordet *islandsk* med *færøsk* fås indblik i problematikken med den færøske ordbog.

Nu (dvs. mai 1995) er der ca. 60 800 opslagsord i vores ordbog. En optælling fra slutningen af 1993 viser denne fordeling mellem ordklasserne:

66% substantiver
17% adjektiver
9% verber

Det fremgår af optællingen at to trediedele af opslagsordene er substantiver.

2.2 Færøsk i nordisk perspektiv

Lad mig indledningsvis slå fast, at færøsk er et selvstændigt nordisk sprog, og ingen „barbarisk bjergdialekt“. Men færøsk har i hvert fald ét stort uafklaret problem: Hvor befinder vi os i forholdet mellem øst- og vestnordisk? Foruden rent geografisk ligger vi også sprogligt ca. midt imellem det østnordiske område og Island. Vi står med et ben i hver lejr. Vi ved rent faktisk ikke, hvor vi står. Det historiske tilhørsforhold til Danmark har givet os en masse danske låneord, og sprogligt står dansk meget stærkt på Færøerne. På den anden side — i vest — ligger Island og lokker os med sin lange skrifttradition og puristiske sprogpolitik, som er helt anderledes end den mere liberale danske. Disse to diametrale modsætninger bombarderes vi med til dagligt. Vi ligger midt i hvirvelstrømmen. Strukturelt ligner færøsk meget islandsk med masser af bøjninger, tre genus, fire kasus, personbøjning af verber osv. Men ordforrådet har vi mere til fælles med dansk end med islandsk. Den færøske purisme benytter islandsk

som forbillede for sine nydannelser, samtidig med at en af dens opgaver er at være på vagt over for danismer. Dette må medføre et dilemma, og det er det vitterlig også. Danismerne er en organisk og naturlig del af nutidssproget, men en stor del af dem har ikke fået ordbogsstatus. Derfor er udvælgelsen af opslagsord til ordbogen et stort problem for os redaktører af den færøske ordbog. Hovedredaktøren, J. H. W. Poulsen, har udvalgt og indtastet alle opslagsordene på grundlag af seddelsamlingen og den færøsk-danske ordbog; men vi supplerer dog løbende, fordi seddelsamlingen er særdeles mangelfuld hvad angår nutidssproget.

Vi har ikke marked for specialiserede ordbøger og ordlister, som de større sprogområder har, og det medfører, at modersmålsordbogen foruden at være en almen ordbog også skal tjene forskellige mere eller mindre fagspecifikke behov. Vores dilemma er: Hvilke ord skal med i ordbogen? Hvor går grænsen mellem de ord, som kommer med og dem, som ikke kommer med? Årsagen til dette dilemma er interferensen mellem færøsk og dansk. Men hvis man vil, at ordbogen skal belyse det aktive daglige sprog, kommer man ikke uden om en meget stor del danismer. Islændingene har ikke hidtil helt kunnet afklare dette problem med sig selv, så en temmelig stor del opslagsord i den islandske modersmålsordbog er forsynet med spørgsmålstegn som signalerer et eller andet dubiøst. Men på det punkt er vores færøske ordbog mere konsekvent: Vi bruger ikke spørgsmålstegn! Enten er ordet med eller er det ikke med.

2.3 Brugsoplysninger

I stedet for at benytte et diffust og flertydigt spørgsmålstegn har vi i redigeringsprogrammet et felt med oplysninger om brugsområde angående opslagsordet, dvs. oplysninger om fx stilniveau, alder, udbredelse, talesprog/skriftsprog, fagsprog osv., således at ord, som af en eller anden grund har behov for brugsoplysning får en entydig markering. En søgning på brugskoder viser, at knap 11 000 ord er forsynet med en eller anden brugskode om stil, alder eller udbredelse, fx skriftsprog, talesprog, bibelsprog, lokalt sprog, kvadsprog, poetisk sprog, gammelt sprog osv. Ord med brugskoder udgør ca. 18%. Dvs. at en temmelig stor del af opslagsordene på en eller anden måde falder uden for det almindelige ordforråd. Den mest benyttede brugskode er „lokalt sprog“, dvs. dialektord, med 4,3% (ca. 2 600 ord), og hvis vi medregner brugskoden „sjældent sprog“ er vi oppe på 7–8%, som på en eller anden måde er lokalt ordforråd. Den mindst benyttede brugskode er „talesprog“ med 0,6% (ca. 350 ord), dvs. ord, som er almindelige i talesproget, men som ikke anbefales i skrift! Brugskoden „skriftsprog“ skal afdække, at ordet er indkommet gennem skriftsproget. Et enkelt belæg på et skriftmålsord er ofte nok til dens eksistensberettigelse i ordbogen, mens man er mere tilbageholden med optagelse af talemålsord. Talemålet er faktisk ikke godt nok repræsenteret. Ordbogen er mere en skriftmålsordbog. Vi lider af det, som Ásta Svavarsdóttir (1994) kalder for „samlerlidenskab“, dvs. at vi forbigår meget nødig ord og betydninger, selv om de er meget sjældne og har en meget begrænset udbredelse. Der kan være en vis overlappning mellem de nævnte grupper, da samme ord kan have mere end en brugskode; brugskoden kan også være anført som specifik for en særlig betydning af et opslagsord. Forholdet mellem talemålsord og skriftmålsordene, dvs. de egentlige skriftmålsord, kvadsprog, bibelsprog, gammelt sprog osv., er ca. 1:25.

De rent fagspecifikke ord udgør ca. 10% (knap 6 000 ord og betydninger). Det bedst dækkede fagområde er botanik med ca. 1 000 ord og zoologi med ca. 500; det skyldes, at vi

har en færøsk botanikbog helt tilbage fra 1910 og en zoologibog fra 1935, som begge har været brugt i skolerne.

botanik	1 020	kemi	138
zoologi	538	politik	125
medicin	514	fysik	121
båd og bådbyggeri	400	religion	118
fugl	315	håndværk	76
sømandssprog	280	elektricitet	48
fisk	279	teknologi	47
jura	267	EDB	44
fårefarver	230	håndarbejde	40
idræt	209	vejrterminologi	40
geologi	159		
<i>Ialt 5 806</i>			

Tabel 3. Antal opslagsord fra nogle fagområder i den færøske modersmålsordbog.

Alle disse ord, som er forsynet med en eller anden stilmæssig eller fagmæssig brugskode, udgør ca. 28% med de ovennævnte forbehold for overlapninger, men også med det forbehold, at enkelte brugskoder kan øges kraftigt i antal, da excerperingen ikke er færdig på nuværende tidspunkt.

2.4 Talemålsord i ordbøger

Vi kan i de nyeste ordbøger konstatere, at talemålsordene i større udstrækning end før er blevet accepteret som en del af vores ordforråd. Den engelsk-færøske ordbog fra 1992 har en del talemålsord i de færøske definitioner, fx *sirkul* (dansk *circle*) for det engelske *circle*, og *sivilisjón* står som synonym sammen med *mentan* og *siðmenning*. Den nye udgave af den dansk-færøske ordbog, som udkommer senere i år, er temmelig liberal på dette område. En skole- og retskrivningsordbog *Orðaskyn* fra 1994 har en hel del låneord med, som ikke findes i modersmålsordbogen: *analfabetur*, *anorakkur*, *automatiskur*, *batt*, *blits*, *dramatisera*, *dynamitt*, *elastikk*, *felga*, *fløti*, *fotografera*, *globus*, *kajakk*, *kapell*, *kettueyga*, *klips*, *klovnur*, *siffur*, *sirkul*, *sylindari*.

2.5 En lille leksikograf på 6 år

For sjov kan jeg nævne, at min lille søn, som er 6 år, sagde til mig forleden dag: „Far, et katteøje [færøsk *kettueyga*] er ikke øje på en kat, det er refleks på en cykel“. Han er leksikograf uden at vide det. Hans sproglige iagttagelse viste sig senere ikke at findes som opslagsord i modersmålsordbogen. Hvorfor ikke *kettueyga*, da vi fx har *kúgvæyga*, som ikke betyder 'øje på en ko', men 'et rundt lille vindue på et skib'? Men vi kan i denne midtvejsrapport konstatere, at dette ord sammen med en hel del andre ikke er med i modersmålsordbogen!

2.6 „Opsamlingsartikler“

Talemålsord med præfikset *be-* er samlet i en „opsamlingsartikel“, for at vise at de eksisterer, men deres status afdækkes ved manglende definitioner:

be- (herðingarveikt forfesti í tøkuroðum serst. í talumáli)
 t.d. *-draga, -døva, -gávaður, -givinheit, -grava, -gravilsi,*
-greinsa, -grípa, -grípiligur, -hagiligur, -halda, -handla,
-handling, -holdari, -høva(st), -kenna, -kentur, -kymra, -rømtur,
-skeð, -skeðin, -standa, -stemma, -stentur, -stilla, -stýrari,
-søk, -søkja, -taka, -tala, -tíðs, -týða, -vara, -vís, -vísa.

2.7 Modersmålsordbog med fremmede ækvivalenter

Feltet *fremmed ækvivalent* bliver brugt i 800 artikler, dvs. at 800 opslagsord er — foruden en færøsk definition — forsynet med et fremmed (mest dansk) synonym:

slitil -s slitlar *k* útbúnaður ið slítur streymsamband, avbrótari [da. kontakt].
siðalag *h* góðir siðir, hugburður um rætt og órætt, viðv. gerðum, atburði o.tíl.
 [da. moral].

Det drejer sig for det meste om sjældne skriftmålsord eller fagord, hvor det fremmede (dvs. danske) ord er bedre kendt end det tilsvarende færøske. Mange af disse nydannelser er fuldstændig ukendte! Vi ved, at det er lidt af en tilsnigelse, men ser det som en nødvendighed i denne fase. Men hvorfor ikke tage konsekvensen og optage disse ord som opslagsord i ordbogen, da de er bedre kendt end de færøske? Det er et åbent spørgsmål i denne midtvejsrapport. Muligvis bliver de mest almindelige af disse ord lemmatiseret i anden udgave, da de nu findes i ordbogen. Vi kan dog påpege enkelte sådanne ord, der har fået status som opslagsord, fx *pedalur*, men ordets lave status afsløres ved at definitionen skal søges under et af definitionsordene *ástig* eller *føtil*, som er helt ukendte nydannelser. Jeg mener at denne metode trods alt er et skridt i den rigtige retning, til at få indkorporeret de mest almindelige låneord i ordbøgerne, som fx *kontakt, spitari, moralur, parabolur, mikrofón* osv. En af fordelene vil være, at de færøske nydannelser på den måde bliver lanceret! Hvordan skal man ellers finde dem? Ved at skyde genvej gennem den dansk-færøske for dér muligvis at finde de færøske ord. Låneord med „ufærøsk“ struktur kan have sin berettigelse i ordbogen for at vise ordets færøske stavemåde. Hvordan skal fx ordet *pedagog* staves på færøsk (*pedagogur, pedagogur, pedakokur, pedakogur, petagogur, petagokur, petakokur* eller *petakogur*)? I enkelte tilfælde tillempes låneordet det færøske lydsystem, fx *lynlás* /ly:nlás/ til *lynlás* /luinlás/ og *planur* /pla:nur/ til *plánur* /pláanur/.

2.8 Purismen og det leksikografiske dilemma

Det tør siges, at purismen gør livet besværligt for mig som færøsk leksikograf, fordi den medfører, at en temmelig stor del af det aktive færøske ordforråd skal frasorteres som sprogligt ukrudt og kommer ikke med i ordbogen. Der er et eller andet skel mellem det som er færøsk set i diakronisk perspektiv og det sprog, som rent faktisk bliver talt på Færøerne i dag, som sætter mig i et gigantisk fagligt dilemma, fordi jeg giver ikke et sandt leksikografisk

billede af den færøske sproglige virkelighed. Vi tier en del bort. Vi søger at gengive et idealistisk billede i stedet for et sandt og troværdigt billede af sprogsituationen på Færøerne i halvfemserne. Danismerne kan sammenlignes med slumkvarterene i storbyerne: Man vil ikke acceptere at de findes, og man bestræber sig på at skjule dem — især for fremmede. Færøsk og islandsk er fælles om denne strengt selektive form for ordbogsarbejde. Islandsk har lang tradition for at danne islandske nydannelser for låneord, mens vores tradition er meget yngre. Jakob Jakobsen er vores første sprogenrer og med hans arbejder, særlig i begyndelsen af århundredet, får vi mange nydannelser, som er almindelige i færøsk i dag, fx *kollvelting*, *hugsjón*, *skjalasavn*, *trúbót* osv. Og denne tradition har man holdt fast ved lige siden med en masse nye ord for nye begreber — nogle kommer i brug og bliver en naturlig del af sproget, mens andre aldrig slår igennem.

2.9 Ordbogens funktion

Der er så stor forskel på talesprog og skriftsprog, at man vægrer sig ved at skrive talesproget endstige optage det i færøske ordbøger! Derfor må man gøre sig klart, hvilken funktion ordbogen skal have og hvad formålet skal være. Vi har ikke formuleret nogen overordnet målsætning med vores ordbog. Bare en færøsk modersmålsordbog. Da seddelsamlingen i det store og hele bygger på det litterære skriftsprog, bliver resultatet en ordbog over litteratursproget.

På nuværende tidspunkt ser det ud til at det historiske synspunkt er det dominerende i udvælgelsen af opslagsord, mens nutidssproget ikke er godt nok repræsenteret. Denne ubalance må der rettes op på.

Jeg vil afslutte med at referere Ásta Svavarsdóttir (1994:135) som siger, at den islandske modersmålsordbog præges af en sprogpolitik der ophøjer oldsproget, mens den er mistænksom over for nutidssproget. Og de samme tendenser kan denne midtvejsrapport desværre afdække.

Litteratur

- Ásta Svavarsdóttir 1994: Sprogsamfund, sprogpolitik og ordbøger. I: *LexicoNordica* 1, 219–237.
- Brøndum-Nielsen, Johannes 1922: Island og Danmark. I: *Politiken*, 30/11, 5.
- Ensk-føroysk orðabók* 1992. Red.: Anfinnur í Skála, Jonhard Mikkelsen og Zakarias Wang. Hoyvík.
- Evensen, A. C. 1905–09: *Føroysk orðabók*. Tórshavn: Bókastovn „Fram“s.
- Færøsk Anthologi* II 1891. Ordsamling og register udarbejdet af Jakob Jakobsen. København.
- Føroya landsbókasavn. Ársfrágreiðing 1994.
- Føroysk-donsk orðabók* 1927–28. Ved M. A. Jacobsen og Chr. Matras. Tórshavn: Felagið Varðin.
- Føroysk-donsk orðabók* 1961. 2. útgáva, nógv broytt og økt. Latið úr hondum hevur Chr. Matras. Tórshavn: Føroya fróðskaparfelag.
- Føroysk-donsk orðabók* 1974. Eykabind. Latið úr hondum hevur Jóhan Hendrik W. Poulsen. Tórshavn: Føroya fróðskaparfelag.
- Jacobsen, M. A. 1921: *Úr bókmentasøgu okkara*. Tórshavn: E. Thomsen.
- Lexicon Færoense* 1877–88. Udarbejdet af Jørgen Bloch. Utrykt. København.
- Matras, Chr. (ed.) 1939: *Svabos færøske visehaandskrifter*. København.

Matras, Chr. (ed.) 1943: *Svabos glossar til færøske visehaandskrifter*. København.

Orðaskyn. Mín fyrsta orðabók 1994. Red.: Ingrid Sondum, Petur Martin Danielsen, Niels Nattestad og Djóni Dalsgarð. Tórshavn: Føroya Skúlabókagrunnur.

Svabo, J. C. 1966: *Dictionarium Færoense. Færøsk-dansk-latinskordbog*. Udgivet efter håndskrifterne af Chr. Matras. I: Ordbogen. København: Munksgaard.

Svabo, J. C. 1970: *Dictionarium Færoense. Færøsk-dansk-latinskordbog*. Udgivet efter håndskrifterne af Chr. Matras. II: Indledning og registre. København: Munksgaard.

Jón Hilmar Jónsson

Nøkler til ordforrådet: Om lemmafunksjon, struktur og informasjonstyper i en ny kombinatorisk ordbok over islandsk

A new combinatorial dictionary of Icelandic, *Orðastaður*, includes a relatively few headwords (ca. 11 000), but these in turn give access to a great number of compounds (ca. 100 000, all told). The compounds under each headword are generally grouped according to meaning, which often gives the user access to a broad and varied vocabulary within a specific semantic field. This aspect is strengthened by cross-references between related headwords. The setup of the syntactic information pays special attention to the variety and variations of usage, as shown in the multiple choice of determiners and functional words in the collocations in which the headwords appear.

1 Innledning

En ordboksbruker har som regel problemer med å få klar oversikt over det ordforrådet som dekkes av en ordbok, og har i og for seg sjelden anledning til å overveie den indre sammenhengen innenfor dette ordforrådet eller tenke på hvordan leksikografen har avgrenset det ordforrådet som behandles i ordboka. Brukeren kan også sjelden hente noen utførlig informasjon om dette i selve ordboka utover det som sies om hvor mange oppslagsord ordboka inneholder. Med tanke på brukernes behov vil leksikografen heller ikke være særlig opptatt av å vise et helhetsbilde av det ordforrådet som blir beskrevet i ordboka. Det gjelder å sikre at brukeren kommer umiddelbart og problemfritt fram til hvert enkelt ord han ønsker å hente informasjon om. For å oppnå dette målet i en almann semasiologisk ordbok, vil det i prinsippet være hensiktsmessig å skille hvert enkelt ord ut som et selvstendig og isolert oppslagsord. Brukeren må altså regne med å stå overfor et temmelig oppsplittet og uoversiktlig ordforråd, selv om fravik fra en striktalfabetisk makrostruktur, særlig i form av en nestalfabetisk lemmaliste, og krysshenvisninger mellom enkelte lemmaer vil kunne imøtekomme brukerens behov og interesse for å se og oppdage innbyrdes sammenheng.

Men uansett hvilken vekt leksikografen har lagt på en klar og konsekvent avgrensning av det ordforråd som behandles i ordboka, vil en ferdiglaget ordbok med sin lemmaliste være et vitnesbyrd om at en slik avgrensning er blitt foretatt av leksikografen. Fremstillingen av lemmalisten og den funksjon lemmaene tildeles i forhold til de opplysninger de gir adgang til, vitner også om hvordan leksikografen har tenkt at brukeren kan komme frem til den leksikografiske informasjon ordboka inneholder. Og selve ordboksteksten er gjerne det eneste vitnesbyrd om denne viktige delen av ordboksarbeidet. En redegjørelse fra leksikografens side om lemmaseleksjon og makrostruktur er sjelden særlig utførlig, og en informerende drøftelse av den problematikken som man har med å gjøre i denne sammenheng, generelle forutsetninger for lemmaseleksjon osv., mangler som regel også (se Wiegand 1990:2129).

Leksikografen kan dog neppe gå ut ifra at en gjennomsnittlig bruker vil være klar over hvilken rolle faktorer som produktiv ordlaging og leksikalisering spiller for enkelte avgjørelser som har med begrensingen av lemmabestanden å gjøre.

Vi får vel innrømme at for leksikografen kan dette også være et nokså uoverskuelig område selv om han ikke vil kunne unngå å gjøre seg tanker om avgrensingen av det ordforrådet ordboka skal dekke, og om hvordan lemmalisten kan brukes som nøkkel til det. Her er vi naturligvis inne på forskjellen mellom ordbokstyper og det brukerbehov ordboka er beregnet på å dekke. Det forholdet gjenspeiles bl.a. i den status og funksjon lemmaene tildeles. Enten blir hele ordforrådet plassert på samme nivå som mer eller mindre likeverdige enheter, altså som oppslagsord eller lemmaer, eller så fungerer lemmaene også til en viss grad som nøkler til et større ordforråd, slik at ordforrådet som dekkes av ordboka, deles på to plan: en del er umiddelbart tilgjengelig som lemmaer, en del er plassert innenfor mikrostrukturen som informasjon til de enkelte lemmaer. Denne skilnaden står i en klar relasjon til skilnaden mellom en resepsjonsordbok, der leksikografen går ut ifra at brukeren er ute etter informasjon for å forstå ordene (som de står i en tekstsammenheng), og en produksjonsordbok, der fremstillingen er tilpasset en bruker som kjenner betydningen til lemmaet, men er ute etter nærmere informasjon med hensyn til tekstproduksjon eller bruker lemmaet som nøkkel til andre beslektede ord, gjerne for å sammenlikne alternativer. Denne bruken er typisk for onomasiologiske ordbøker, en synonymordbok f.eks., der brukeren gjerne slår opp på et ord han vet at ikke passer som uttrykk for det han skal uttrykke seg om, men utnytter sin kjennskap til det for å få adgang til beslektede ord som kunne passe bedre. Nyttan av en slik ordbok er til en stor grad avhengig av den innsikt brukeren har i oppslagsverdien til de enkelte lemmaer, for den tilsiktede informasjon kunne i prinsippet være tilgjengelig på forskjellige måter ut fra ulike lemmaer. Og en onomasiologisk ordbok kan også fremstilles som en tesaurus, for å imøtekomme de brukere som trenger en bred og detaljert kartlegging av et semantisk felt.

2 Begrenset lemmabestand dekker et bredt ordforråd

En slik fordeling av ordforrådet, der en del hører hjemme innenfor mikrostrukturen og lemmabestanden er forholdshvis begrenset, har også sin verdi i semasiologiske ordbøker. Her vil man ikke minst tenke på den informasjon som angår ordlaging, hvordan de enkelte lemmaer inngår i sammensetninger med andre ord og ordledd, eller hvordan enkelte prefikser eller suffikser realiseres som konkrete ord. Det formål å gi brukeren adgang til beslektede eller sammenhengende ord i tilknytning til og til en viss grad som en integrert del av den leksikografiske informasjon om de enkelte lemmaene, kan også være av grunnleggende betydning for struktureringen av ordboksartiklene.

Dette kan sies å være grunnmotivet for en ny ordbok over islandsk, *Orðastaður*, som kom i 1994. Som forfatter til denne ordboka vil jeg gi en grov skildring av ordbokas struktur, både makro- og mikrostruktur, og da spesielt med henblikk på de overveielser jeg nå har vært inne på.

Betegnelsen „kombinatorisk ordbok“ i overskriften på artikkelen skulle gi en grov antydning om hvor *Orðastaður* hører hjemme i ordbokstypologien. Men en sammenlikning med kjente ordbøker som vi forbinder med denne termen, som f.eks. BBI (1986) og Agricola (1962), er ikke helt dekkende. Det som utmerker *Orðastaður* utover det som sentralt hører til

ordbøker av denne typen (her kommer termene konstruksjonsordbok og kollokasjonsordbok også i betraktning), er den vekt det legges på de morfologiske kombinasjoner lemmaene opptrer i, og den paradigmatisk informasjon disse kombinasjonene formidler til brukeren. Det betyr bl.a. at *Orðastaður* til en viss grad vil kunne brukes som en onomasiologisk ordbok, der en alfabetisk makrostruktur gir adgang til et bredt og varierende ordforråd som knytter seg til et bestemt begrepsfelt. Forsåvidt kan den sies å være beslektet med analogiordbøker i fransk ordbokstradisjon, der den analogiske informasjon står i fokus (se Rey-Debove 1989).

Som det fremgår av statistikken i figur 1, inneholder *Orðastaður* forholdsvis få lemmaer i forhold til det store antall sammensetninger og ordforbindelser. Fordelingen av lemmabestanden på de enkelte ordklassene sier også noe om ordbokas formål og karakter. Vi legger merke til at det bare dreier seg om tre egentlige ordklasser, substantiver, adjektiver og verb, men i tillegg kommer en god del prefikser og andre forledd som er av interesse for de sammensetninger de opptrer i. På samme måte fins det blant substantivene og adjektivene en rekke suffikser og andre produktive bakledd som blir oppført som selvstendige lemmaer. Det er verdt å understreke at adverbene ikke opptrer som lemmaer. Det henger sammen med at med hensyn til den tilsiktede bruken, hører forbindelsene, spesielt de syntagmatiske forbindelser, primært hjemme under det ordet som utgjør basen eller kjernen i forbindelsen. Det betyr f.eks. at en forbindelse som består av verb + adverb, står under verbet, og det samme gjelder for adverb og adjektiver.

Oppslagsord	11 000	
Sammensetninger		
(knyttet til ordledd)	100 000	
Ordforbindelser	45 000	
Eksempler	45 000	
Substantiver	6 200	56,0%
Verb	2 400	21,0%
Adjektiver	2 200	19,5%
Forledd	400	3,5%

Figur 1: *Orðastaður*: noen statistiske opplysninger

3 Sublemmaer

I mange tilfeller deles ordboksartikkelen opp i to eller flere felt med en hierarkisk oppdeling, slik at visse kategorier opptrer som sublemmaer, enten konsekvent (som mediopassiv til verbene, jfr. figur 2) eller i de tilfeller der de anses å ha en (semantisk) særstilling overfor hovedlemmaet (som flertall til substantivene, jfr. figur 3). I beskrivelsen av verbene forekommer det tredje trinnet i dette hierarkiet, i og med at partikkelverb blir plassert (under hovedlemma eller sublemma), i innbyrdes alfabetisk ordning etter partikkelen (jfr. figur 2).

ráða so 1. hann hlustar ekki á það sem ég segi – því trúir ég vel, hann vill alltaf ráða, ráða <þessu; öllu, litlu, engu> þú mátt ráða því hvenær við byrjum, veðurfar ræður því að miklu leyti hvernig til tekst með heyskapinn, bankamir ráða að mestu leyti stefnunni í vaxtamálum; ráða <miklu, litlu, engu> um <þetta> ég ræð víst litlu um það hvernig peningum ríkisins er ráðstafað [. . .] 2. **ráðning** ráða <gátuna; drauminn; letrið, rúnirnar, táknið> 3. **rályktun** ráða af <ástandinu, viðbrögðum e–s, svipnum á henni> að–S ég ræð það af orðum hennar að hún sé óánægð með niðurstöðuna 4. **ráðning** ráða e–n/sig (<þangað, til að–NH>) hún var ráðin til að hafa umsjón með verkefninu, hann er nýbúinn að ráða sig á togara, ráða <fólk> til starfa, ráða e–n í <starfið, stöðuna>, ráða e–n/sig í vinnu 5. **ráðlegging** ráða e–m (<eindregið>) (til) að–NH ég ræð henni eindregið að leita læknis, ráða e–m (<eindregið>) frá að–NH ég réð henni eindregið frá að sleppa starfinu; ráða <honum, henni> heilt 6. **ofsi, ákefið** ráða sér ekki (fyrir <gleði, kæti, reiði>) ► **–ráða**: fast-, endur-, vist-, af- ★ **ráða e–ð af** <ákvörðun; sbr. afráða> við réðum það af að halda áfram og freista þess að ná heim í björtu ★ **ráða fram úr e–u** (<reyna að>) ráða fram úr <vandanum, vandamálinu> ★ **ráða fyrir e–u** ráða fyrir <landi; skipi, flota; búi; liði, her> ★ **ráða í e–ð** (<reyna að>) ráða í <þetta; skriftuna, letrið, orð e–s> ★ **ráða við e–ð/e–n** <sbr. viðráðanlegur> ráða (<vel, auðveldlega; illa, ekki>) við <þetta; vandann, starfið>/<hann, hana> heldur kennarinn ekki góðum aga? – nei, hann ræður ekkert við krakkana ★ **ráða e–ð við sig** <samráð, mat> þau eru að ráða það við sig hvað til bragðs skuli taka ★ **ráða yfir e–u/e–m**

yfirráð ráða yfir <honum, henni>/<landi, ríki; liði, vistum, birgðum> talið er að herinn ráði yfir öllum nýttiskuvopnum; hafa/eiga yfir að ráða <tækjum, búnaði; kunnáttu í fiskveiðum> við getum ekki mælt þetta nákvæmlega með þeim tækjum sem við höfum yfir að ráða ● **ráðast** miðm **framvinda, málaok** mér líst ekki á að þetta takist – við skulum sjá hvernig þetta ræðst, það er aldrei að vita; það ræðst á næstu dögum hvort af kaupnum verður, <þetta> ræðst af <aðstæðum> það ræðst ekki síst af verðinu hvaða þýrta verður keypt, það verður að ráðast (hv–S) hvenær ætlaðu að skila ritgerðinni? – það verður bara að ráðast, ég er ekki enn búinn með hana, láta <þetta> ráðast ★ **ráðast að e–m** ráðast að <honum, henni> (með <barsmið, skómmum, svívirðingum, látum>) ★ **ráðast á e–n/e–ð** <árás> ráðast á <hann, hana>/<landið, borgina>, ráðast á e–n <úr launsátri>; <gagnrýni> ritstjórnin ræðst harkalega á ráðherrann fyrir að hafa blekt þjóðina ★ **ráðast í e–ð** ráðast í <þetta; framkvæmdir, verk, fyrirtæki, byggingu, kaup> það þyrfti helst að skipta um alla glugga í húsinu – hvenær höfum við efni á að ráðast í það? ● **ráðinn** lh þt 1. vera ráðinn í að–NH mér skilst þau séu ráðin í að flytja til Reykjavíkur í haust; ráðinn og roskinn 2. **ásætningur** <gera þetta> af ráðnum hug hann viðurkennir að hafa barið hana en segist ekki hafa gert það af ráðnum hug ► **–ráðinn**: 1. auð-, fljót-, tor-, vand-; 6–2. fast-, laus-, ævi-; skip-, vist-, ný- ● **ráðandi** lh nt 1. vera (orðinn) sjálfs sín ráðandi, vera fjár síns ráðandi; ráðandi aðili 2. ráðandi öfl, ráðandi sjónarmið, ráðandi viðhorf, ráðandi stefna, ráðandi veðráttu, ráðandi vindátt ► **–ráðandi**: alls-, mikils- ⇒ –ræðni, ræður

Figur 2: Sublemmaer markeres med typografiske strukturindikatorer. I en verb-artikkel innledes mediopassiv og partisipp med symbolet ●, partikkelverb med ★.

beygla no kvk 1. það kom beygla á bílinn við árekssturinn 2. **um bíl; viðhorf** er þetta góður bíll? – nei, þetta er ótalleg beygla ► **beyglu**:- -legur ● **beyglur** flt

vandræði vera í (<hálfgerðum, bölvuðum>) beyglum (út af <þessu>)

Figur 3: I de substantiver der det foreligger en markant semantisk skilnad mellom entall og flertall, blir flertallsformen oppført som sublemma (innledes med ●).

I den morfologiske delen av artikkelen opptrer de aktuelle ordledd til lemmaet som sublemmatisk overskrifter til de enkelte avsnitt (jfr. figur 4). Den innbyrdes rekkefølge er konstant: først kommer sammensetninger med lemmaet som forledd (blant substantivene muligens i tre varianter: stammen uten endelse, med endelse for genitiv entall og med endelse for genitiv flertall), og deretter de sammensetninger der lemmaet står som bakledd.

fundur nokk fundur um <málið>, efna til fundar (með <samningamönnum, deiluaðilum>), kalla saman fund, skjóta á fundi, boða fund, halda fund; sækja fund(inn), <koma, mæta> á fundinn, <fjölmenna, fylkja liði> á fundinn[. . .] ► **fund-**: -vís, -vísí ► **fundar-**: 1. -boð, -boðun, -boðandi; -fært, -sókn, -fall; -höld, -seta; -maður 2. -tími, -dagur; -hlé; -byrjun, -lok 3. -staður, -herbergi, -salur 4. -form, -sköp, -stjórn, -stjóri 5. -efni; -ályktun, -samþykkt; -ritari, -gerð, -skýrsla; -bók 6. -laun ► **funda-**: -ferð, -höld; -frelsi; bók ► **-fundur**: 1. félags-, flokks-, skóla-; héraðs-, sýslu-, lands-, alþjóða-; fjölda-; þjóð-; úti-; mann-; aðal-, auka-; allsherjar-; árs-; skyndi-; morgun-, kvöld-, nætur-;

laun-, leyni- 2. þing-, þingflokks-, borgarstjórnar-, bæjarstjórnar-, miðstjórnar-, stjórnar-, fulltrúaráðs-, hreppsnefndar-, flokksstjórnar-, ríkisstjórnar-, ríkisráðs-; nefndar-, fulltrúa-, hluthafa-, kjörmanna-, sérfræðinga-, formanna-, leiðtoga-; borgara-, kennara-, foreldra-; safnaðar-; klúbb-; klíku-; blaðamanna-, fréttamanna-; miðils- 3. umræðu-, viðræðu-, samninga-, sátta-, samráðs-, rabb-; vinnu-; kynningar-, fræðslu-; mál-, stjórn mála-, þjóðmála-, þingmála-; framboðs-, kjör-, kosninga-; undirbúnings-, stofn-; framhalds- 4. [gjarna í fit] sam-; vina-; ásta-; gleði-, fagna-, [gjarna í fit] fagnaðar-, [fit] endur- 5. átaka-, æsinga- 6. fornleifa-, beina-, lík-; gull-, olíu-, landa-

Figur 4: Enkelte ordledd til lemmaet markeres med ► og fremheves i artikkelen som overskrifter til eksemplariske sammensetninger.

4 Valg og plassering av leksikalske forbindelser i forhold til lemmaene

Avgrensingen av lemmabestanden er i hovedsak avhengig av lemmaenes kombinatoriske verdi, syntaktisk og/eller morfologisk. Men dessuten styres lemmaseleksjonen av den funksjon de potensielle lemmaer anses å ha overfor brukeren som en logisk nøkkel til de forbindelser artikkelen inneholder. Når det gjelder konstruksjoner og kollokasjoner, betyr det at forbindelsene primært hører til beskrivelsen av det ordet som fungerer som base. Men det betyr også at typiske idiommer stort sett blir holdt utenfor, når de ikke inneholder et ledd som kan betraktes som base og har en klar semantisk sammenheng med forbindelsen som helhet.

Sammensetninger av lemmaet får som regel ikke adgang til den syntagmatiske delen, bortsett fra de forbindelser der ordet forekommer i sammensetninger med varierende forledd (jfr. figur 5). Men her er det verdt å nevne at sammensetninger kan forekomme i leksikaliserte forbindelser som utfra semantiske forhold godt kunne høre hjemme i beskrivelsen av det sentrale ledd i sammensetningen. Her dreier det seg om halvidiomatiske fraser som *taka við stjórnartaumunum* 'overta makten', *stíga á skipsfjöl* 'gå ombord' og *svifa inn í draumalandið* 'falle i søvn', der det første leddet i det sammensatte substantivet utgjør kjernen. Med hensyn til aktiv bruk ville slike forbindelser være best tilgjengelige under lemmaet til de respektive forleddene, altså *taka við stjórnartaumunum* under *stjórn*, *svifa inn í draumalandið* under *draumur* osv. Forbindelser av denne typen har ellers en forholdsvis uklar leksikografisk status. Betydningen er til den grad gjennomsiktig, ikke minst i en konkret tekstsammenheng, at den neppe ville bli anført i en passiv ordbok. Nyttan av å angi de sammensatte substantivene som lemmaer i en aktiv ordbok, er avhengig av at brukeren kjenner til forbindelsene på forhånd. I samsvar med de generelle prinsipper for lemmaseleksjonen i *Orðastaður* blir fraser av denne typen ikke anført i ordboka.

skyn nohv 1. skilningur, þekking bera (<gott; lítið, ekkert>) skyn á <þetta> *hann er útlendingur og ber lítið skyn á þær venjur sem hér gilda*, vita/kunna skyn á <þessu>; vít vera (alveg) skyni skroppinn *er hún svo skyni skroppin að hún geti ekki skilið svona einfalt atriði?* 2. gefa <það (að-S); ýmislegt> í skyn *hún sagði það ekki berum orðum en mér heyrðist hún vera að gefa í skyn að hún sækti um starfið* 3. <gera þetta> í <góðu, gustuka-, kurteis-, sparnaðar-> skyni *þau færðu henni blómavönd í þakklætisskyni*, <gera þetta> í því skyni að-NH *hann kom hingað í því skyni að leita eftir viðskiptum* ► **skyn-**: -færi, -fruma, -stöð, -taug; -mynd, -áhrif, -hrif, -reynd, -reynsla, -heim-

ur; -bragð, -næmi, -næmur, -bær; -rænn; -hyggja; -samur, -semd, -semi; um mann; viðhori -skiptingur; -lífill, -laus, -leysi, -leysing; -villa, -villtur ► **-skyn:** 1. sjón-, heymar-, lyktar-, ilm-, þef-, bragð-; rúm-, fjarlægðar-, form-; snerti-, þreifi-, hreyfi-, jafnvægis-; sársauka-, hita-, kulda-; raunveruleika-, veruleika-; siðferðis-, fegurðar-, ljóð-, náttúru-, skop-; tíma-; verð- 2. [þgf., sbr. 3] auðgunar-, ágóða-, áróðurs-, eiginhagsmuna-, greiða-, gróða-, gustuka-, hagnaðar-, háðungar-, hefndar-, heiðurs-, kurteis-, kveðju-, könnunar-, mótmæla-, refsingar-, samúðar-, sjálfsvarnar-, sparnaðar-, sæmdar-, varúðar-, viðurkenningar-, virðingar-, vísinda-, þakkar-, þakklætis-, öryggis-

Figur 5: Avsnitt 3 i den syntagmatiske delen av artikkelen inneholder eksemplariske ledd som forekommer i sammensetninger med lemmaet i en bestemt syntaktisk forbindelse. I avsnitt 2 i den morfologiske delen blir denne informasjonen utdypet gjennom flere morfologiske eksempler med henvisning til vedkommende avsnitt i den syntagmatiske delen.

5 Enkelte informasjonstyper

I hovedtrekk styres valget av de syntaktiske og morfologiske forbindelsene og fremstillingen av de leksikografiske eksempler av to motiver. De skal for det første vise typiske brukstilfeller (som kollokasjoner), og for det andre skal de gi uttrykk for den variasjon og det mangfold som karakteriserer bruken av de enkelte lemmaene. En kort gjennomgåing av den leksikografiske informasjon i ordboka ut fra konkrete eksempler vil belyse dette nærmere.

5.1 Syntagmatisk informasjon

I en ordbok av denne typen vil den syntagmatiske informasjon naturligvis innta en sentral posisjon. Den kommer først og fremst til uttrykk gjennom konkrete syntagmatiske forbindelser, enten det gjelder konstruksjoner, kollokasjoner eller friere forbindelser. Også grammatiske egenskaper som transitivitet og kasusreksjon kommer til uttrykk som implisitt informasjon i konkrete forbindelser, ut fra den tanken at brukeren har tilstrekkelige kunnskaper til å tolke det grammatiske innholdet i de enkelte former og bøyingsendelser. Grammatisk eller syntaktisk koder blir derimot brukt i meget begrenset omfang, faktisk bare til å betegne en leddsetning (med betegnelsen S) og et setningsledd som innledes av et verb i infinitiv (med betegnelsen NH for infinitivformen, foran infinitivsmerket *að*; jfr. f.eks. figur 2).

5.2 Morfologisk informasjon

I tillegg til de opplysninger om lemmaets opptreden i sammensetninger som vises i overskriftene til sammensetningslistene, gis en morfologisk informasjon i tilknytning til den syntaktiske delen, i form av en henvisende kommentar. Under verbene står det en henvisning til det tilsvarende nomen actionis, og under adjektivene henvises det gjerne til et

skera so [skurður] skera <bandið, taugina> (í sundur), skera brauð, skera kál, skera korn; skera mó, skera torf; skera hvannir, skera þang; skera hár <hans, hennar>/sitt, skera mön, skera <sneið, bita> af <kökunni, brauðinu>, skera af <kökunni, ostinum; endanum>, skera <skemmdina, meinið> burt, skera <ostinn> í bita, skera <efnið, pappírinn> til/í <lengjur, ræmur>[. . .] ★ **skera á e-ð** skera á <hnúttinn, bandið>, skera á kýlið/kýlinu ★ **skera e-ð niður 1.** skera <brauðið, kjötið> niður **2.** [niðurskurður] skera <útgjöld, framlög, fjárveitingu, kostnað> niður *útgjöldin hafa verið skorin niður um helming,*

gláður lo **1.** [glæði] vera/verða <barnslega; óumræðilega, ósegjanlega> gláður *ég man hvað við vorum alltaf glöð á jólnunum,* verða <mjög, harla> gláður við (er-S) *smalamaður varð harla gláður við er*

skera <þetta> niður við trog; skera niður (<féð, bústofninn>) ★ **skera e-n upp** [uppskurður] skera <sjúklinginn> upp (<við krabbameini>); skera upp lík ★ **skera úr (um e-ð)** [úrskurður] skera úr (um <þetta; málið, lögmæti e-rs>) (<fyrr dómstóli>) ★ **skera úr e-u** skera úr <þessu; málinu, deilunni>, (<reyna að, vilja>) fá úr <þessu> skorið *ég vil fá úr því skorið hvort þetta er heimilt* ★ **skera sig úr** *hann sker sig alveg úr; hann er svo miklu sterri en hinir strákarnir* ★ **skera (e-ð) út** [útskurður] skera út (<mynd, mynstur>) [. . .]

hann sá grilla í fjárhópinu í sortanum, verða gláður við <þessa sjón/frétt>; vera gláður í bragði; gláður og ánægður, gláður og reifur, gláður og kátur, sæll og gláður; gláður eða hryggur[. . .]

Figur 6: *Kommentarer av forskjellig slag blir angitt i konsentrert form innenfor en firkantet ramme i ordboksteksten. I en verbartikkel (som skera) henvises det til det tilsvarende nomen actionis. Adjektiv (som gláður) får en henvisning til et avledet substantiv.*

traust no hvk traust á <honum, henni>/<ráðstöfuninni>, sýna e-m (<mikið; fullt>) traust (með <þessu>) *mér þykir hann sýna henni traust með því að fela henni umsjón með verkinu,* setja traust sitt/allt sitt traust á <hann, hana> *ég hef ekkert vit á þessu, ég set bara allt mitt traust á þig,* <mér> er/þykir traust að <honum, henni>, bera (<fullt; ótakmarkað>) traust til <hans, hennar>, njóta (<mikils, óskoraðs>) trausts (<í flokknum>), hafa traust <hans, hennar> *það eru að vísu margir á*

móti henni en hitt vegur þyngra að hún hefur traust formannsins, verðskulda traustið, vera trausts(ins) verður; [sbr. traustvekjandi] vekja traust *ham er nota- legur í viðkyrningu og vekur strax traust, það er ekki hægt að segja að svona hringlandaháttur veki traust;* vera rúinn (öllu) trausti; <gera þetta> í trausti þess að-S *ég lána þér þetta í trausti þess að þú skilir því aftur;* [akvæði] verðugt traust; gagnkvæmt traust, <vera e-m, hafa e-n með sér> til halds og trausts ► **traust-** -vekjandi [. . .]

Figur 7: *Angivelsen i den firkantede rammen gjelder en sammensetning den etterfølgende kollokasjon står i ordlagingsrelasjon til.*

En selvstendig informasjon om relasjoner mellom morfologisk (og semantisk) beslektede ord gis også i direkte tilknytning til selve lemmaene, i form av en henvisning bakerst i artikkelen, som regel fra det mest sentrale ordet (jfr. figur 8).

5.3 Paradigmatisk informasjon

Den paradigmatiske informasjon står meget sentralt i den morfologiske delen. Her gis det i mange tilfeller adgang til en stor mengde ordledd som står i paradigmatiske relasjon til hverandre, og ved hjelp av en semantisk gruppering vil brukeren ofte få verdifull oversikt over nært beslektede ord, f.eks. synonymer (jfr. figur 8).

frost nohvk hann gengur í frost, það er (komið) frost, frost er/var á <um nóttina>, frostið harðnar/hefur harðnað (<með kvöldinu>), frostið fer harðnandi/vaxandi, hann herðir/er að herða frostið, frostið er að herða, það dregur/hefur dregið úr frostinu, hann linar/er að lina frostið; frostið bitur; <grösin> þola (ekki/illa) frost; (það er) <veggja> stiga frost; <vera úti> í (<miklu>) frosti, <rörið hefur sprungið> í frostinu, <grösin féllu> í fyrstu frostum; [ákvæði] <mikið, hart, grimmt, biturt, nístandi, brakandi; harðnandi, vaxandi; lítið, lint; minnkandi> frost; frost og kuldi, frost og stillur, snjór og frost, fjúk og frost ► **frost-**: 1. [mikið f.] -kuldi, -bál, -bitra, -bruni, -gaddur, -garður, -grimmd; -hörkur; -harður, -bitur, -grimmur, -kaldur 2. [lítið/minnkandi f.] -vægð, -lin, -lina, -leysa; -mildur, -vægur, -lítið, -laust 3. [njóð] -brak, -brestur, -dynk-

ur 4. [með úrkomu] -bylur, -él, -fjúk, -renningur, -hríð, -kafald, -snjór; -regn, -rigning, -skúr 5. [með vindi] -gjóla, -gola, -kul, -kylja, -kæla, -nepja, -næðingur, -stormur, -vindur 6. [stilla] -kyrra, -kyrrð, -kyrr 7. [ismyndun] -hella, -klammi, -skel, -skæni, -storka, -héla; -rós 8. [loft] -móða, -mistur, -reykur, -þoka 9. [tímabili] -nótt, -kaffi, -kast 10. [áhrif] -bit, -kali, -bitinn, -bólga, -bólgin; -springa, -sprenging, -sprengdur, -sprunga, -þensla, -lyfting, -veðrun; -hætta, -vörn, -lögur, -vökvi; -þol, -þolinn 11. [mæling] -mælir, -mark, -punktur ► **frosta-**: -kaffi, -vetur; -samur, -harður, -mikill; -vægur ► **frost**: 1. vetrar-, vor-, haust-; logn-; nátt-, nætur-; hrím-; þurra 2. [mikið f.] bál, bálhörku-, bruna-, gadd-, grimmdar-, heiftar-, hel(jar)-, hörku-, límings-, nístings-, stál- ⇒ freð-, freðinn, freri, frjósa, frysta, frysti-, frysting-, frystir

Figur 8: *Bakerst i en artikkel der lemmaet oppfattes som det sentrale ordet innenfor en gruppe morfologisk og semantisk beslektede ord, henvises det til beslektede lemmaer. — Den morfologiske delen av en ordboksartikkel kan som her gi adgang til (nær)synonymer som konkretiseres i sammensetninger til vedkommende lemma (som f.eks. i avsnitt 1, 4 og 5 under overskriften frost-).*

Men innenfor den syntaktiske delen gis det også en utstrakt paradigmatiske informasjon igjennom varierende bestemmelsesledd til de syntaktiske forbindelser. På den måten kan brukeren få en bred oversikt over alternerende kollokatorer som gjerne står i synonym- eller antonymrelasjon til hverandre (jfr. figur 9).

ganga so [. . .] 3. [göngulag, aðferð, hraði] ganga óstuddur, ganga haltur, ganga haltrandi, ganga við staf/hækjur; [ákvæði] ganga <hratt, greitt, hröðum skrefum, rösklega; hægt, í hægðum sínum, rólega, varlega, hjóðlega, hægum skrefum>, ganga snúðugt, ganga <álútur, silalega, álappalega, letilega, slyttislega; hnarreistur> [. . .] 5a. [framvinda, árangur] [ákvæði] <þetta; verkið> gengur <vel, fljótt, eftir áætlun, að óskum, glatt, greitt, greiðlega, átaka-

laust, skafið, liðugt, þegjandi og hljóðalaust, ekki amalega, eins og í sögu; sæmilega, bærilega, furðanlega, skaplega, upp og ofan, polanlega, þokkalega, vandræðalaust, stórslysalaut, svona la la, svona og svona, áfallalaust, snurðulaust, árekstralaust; illa, hægt, ekkert, andskotalega, ekki andskotalaut, helvítalega, erfiðlega, treglega, örðuglega, bölvánlega, báglega, ömurlega, afleitlega, böslulega, stirðlega, stírt, ekki prautalaust> [. . .]

Figur 9: *To avsnitt fra artikkelen **ganga** viser varierende kollokatorer til verbet knyttet til to sentrale betydninger. De alternerende adverbialer markeres med spissparentes.*

5.4 Semantisk informasjon

Den semantiske informasjon gis for en stor del implisitt ved at konteksten til lemmaet i enkelte forbindelser og eksempler er opplysende om betydningen. Som eksempel på den vekt konteksten tildeles i denne sammenheng, kan nevnes at adverbiale forbindelser blir som regel anført som en del av en større syntaktisk enhet der den andre delen skal belyse det semantiske innholdet (jfr. figur 10).

tími no kk[. . .] <þetta þótti fínt> á sínum tíma; <þurfa að fara á fætur> á ókristilegum tíma; <þetta breytist> með tíð og tíma/með tímanum, <viðhorfin hafa breyst> í tímans rás, <vonirnar ræstast> í fyllingu tímans, <hér vex upp skógur> er tímar líða (fram), <sárið verður gróíð> eftir <stuttan> tíma,

<þeir verða orðnir sáttir; hún kom aftur heim> að <nökkrum> tíma liðnum, <við lukum verkinu> á <skömmum, stuttum> tíma, <þetta hefur jafnað sig> á löngum tíma; <lánið er veitt> til <langs, skamms> tíma, <minningin lifir> um tíma og eilífið; [. . .]

Figur 10: *Et avsnitt fra artikkelen **tími** viser fremstillingen av adverbiale forbindelser til lemmaet. For å belyse adverbialelets funksjon og betydning blir det angitt i en større syntaktisk sammenheng, der den „belysende“ delen blir avgrenset med spissparentes.*

Her legges det også vekt på å trekke fram viktige semantiske egenskaper, f.eks. den konnotative betydningen, gjennom fyldige, illustrerende eksempler, gjerne i form av en dialog slik at en bred kontekst er med på å avsløre betydningen (jfr. figur 11).

þvæla so 1. **[þvælingur]** þvæla e–m <fram og aftur, fram og til baka, út í ófæru> *það er slæmt að þurfa að þvæla börnunum í pössun út um allan bæ* 2. **[ta; viðhorf]** *mér heyrst hann ekki hafa neitt vit á þessu, hann er bara eitthvað að þvæla, mér leiðist að hlusta á hann þvæla um trúmál* ● **þvælast** miðm 1. **[ferð, ráf; viðhorf]** þvælast <þangað, fram og til baka, um (landið)> *þau hafa ferðast víða? – það er nú líkast til, þau hafa þvælst um allan heim, (vera*

að) þvælast <niðri í bæ>; þvælast fyrir (e–m) geta strákarnir eitthvað hjálpað þér? – nei, þeir gera ekkert annað en að þvælast fyrir 2. *það er ljótt að sjá hvað kápan hefur þvælst, það verður að setja hana í hreinsun* ● **þvældur** lh þt 1. **[ú]** vera (orðinn) þvældur (eftir <ferðina>) 2. **[sli]** *ég get varla notað dúkinn svona, hann er svo þvældur; það nennir enginn að tala um þetta lengur; þetta er orðið svo þvælt; þvælt orðalag* ► **-þvældur:** marg-, út-, ó-

Figur 11: *Den semantiske informasjon blir gjerne utdypet gjennom fyldige, illustrerende eksempler (kursivert i teksten), der dialogformen ofte er utnyttet til å belyse (den konnotative) betydningen.*

En annen form for implisitt semantisk informasjon foreligger når de enkelte forbindelser eller sammensetninger blir gruppert på semantisk grunnlag og den semantiske skilnaden blir markert enten ved semikolon eller nummerering (jfr. f.eks. figur 2). Men den semantiske informasjon kan også gis eksplisitt som en identifiserende eller forklarende angivelse i ordboksteksten (jfr. figur 12).

fé no hvk 1. **[húsdýr]** (eiga) margt/fátt fé; reka féð (<á fjalli, af fjalli>), huska fénu <á undan sér>, hvuska við fénu, féð er á fjalli (<allt sumarið>), féð gengur sjálfala (<mestallt árið>) [. . .] 2. **[peningar]** spara

fé, safna fé, afla (sér) fjár, <honum, henni> verður/varð gott til fjár, græða fé, <honum, henni> græðist/græddist fé, **[viðhorf]** raka saman/moka saman fé; [. . .]

Figur 12: *Når det foreligger en klar semantisk skilnad, gjenspeiles den i struktureringen av ordboksartikkelen. I artikkelen **fé** blir den semantiske skilnaden markert ved hjelp av en identifiserende glosse som angis i en firkantede ramme etter betydningsnummeret.*

6 Sluttord

Jeg vil ikke gå nærmere inn på de enkelte informasjonstyper i ordboka og realiseringen av dem. Men jeg vil understreke hvor viktig det er at de konkrete forbindelser og eksempler

formidler den leksikografiske informasjon til brukeren. I samsvar med det skal leksikografen virke forholdsvis tilbakeholdende og mest mulig unngå å gripe inn med definisjoner og forklaringer som brukeren vil anse som forstyrrende eller overflødige. En kortfattet kommentar eller glossering er ofte tilstrekkelig til å formidle de opplysninger brukeren kan ha behov for.

Vi har sett at i en ordbok av denne typen får brukeren adgang til et meget omfattende ordforråd gjennom relativt få lemnaer. Leksikografen står altså overfor den oppgave å begrense lemmabestanden i samsvar med det genuine formålet og de forutsetninger han legger til grunn for bruken av ordboka, og det vil i mange tilfeller være nokså vanskelig. Det henger ikke minst sammen med at leksikografen har en tilbøyelighet til å regne med at brukeren reagerer sterkere på eventuelle lakuner, altså ord de synes mangler i lemmalista, enn på lemnaer som kan være overflødige. Her har vi med å gjøre en av de faktorer som kan motvirke en målbevisst og balansert avgrensning av lemmabestanden i ordbøker. Men vi kan vel heller ikke se bort fra at vurderingen av lemmaverdien til de enkelte ord, uansett ordbokstype, til en viss grad vil styres av ulike kriterier, slik at en fullt konsekvent og balansert avgrensning av lemmabestanden neppe vil kunne virkeliggjøres i praksis. Jeg kan bekrefte at *Orðastaður* ikke er noe unntak i så måte. En ytterligere begrensning eller utvidelse av lemmabestanden i *Orðastaður* ville bl.a. være avhengig av i hvilken grad illustrerende eksempler skal utnyttes for å formidle den leksikografiske informasjon. En annen avgjørelse henger sammen med lemmaverdien av produktive ordledd som prefikser og suffikser. Den tredje overveielser gjelder tilfeller der verdien av et potensielt lemma er begrenset til én isolert forbindelse (f.eks. av typen *svifa inn í draumalandið*). Men uansett hvilken virkning enkelte seleksjonskriterier har, vil hovedmotivet for lemmaseleksjonen og struktureringen av lemnaene være å gi brukeren de nøkler som passer best for å hente den mangfoldige informasjon som ligger i ordboka.

Litteratur

- Agricola, Erhard 1962: *Wörter und Wendungen*. Wörterbuch zum deutschen Sprachgebrauch. Leipzig: VEB Bibliographisches Institut.
- BBI = Benson/Benson/Ilson 1986: *The BBI Combinatory Dictionary of English*. A Guide to Word Combinations. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Orðastaður* = Jón Hilmar Jónsson 1994: *Orðastaður*. Orðabók um íslenska málnotkun. Reykjavík: Mál og menning.
- Rey-Debove, Josette 1989: Le traitement analogique dans le dictionnaire monolingue. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie* I. Red. Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta. Berlin/New York: de Gruyter, 635-640.
- Wiegand, Herbert Ernst 1990: Die deutsche Lexikographie der Gegenwart. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie* II. Red. Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta. Berlin/New York: de Gruyter, 2100-2246.

Kristín Bjarnadóttir

Lexicalization and the Selection of Compounds for a Bilingual Icelandic Dictionary Base

Artikkelen handler om de problemer som hefter ved lemmatisering av sammensetninger i en tospråklig ordbok der islandsk er kildepråket. Fordi enkelte ord kan vise varierende ordformer som førsteledd i sammensetninger, vil lemmaseleksjonen ikke utelukkende gjenspeile semantisk leksikalisering. Det må også tas hensyn til at leksikaliseringen i mange tilfeller er begrenset til en bestemt formvariant. Dette forholdet kompliseres ytterligere ved at sammensetninger som viser et produktivt ordlagingsmønster, kan inneholde polyseme ordledd, eller ved at ordleddene står i en flertydig relasjon til hverandre.

1 Introduction

The subject of this paper is the selection of vocabulary for a bilingual Icelandic dictionary base, in particular the selection of compounds. As expected, lexicalization proves to be the major factor in the choice of compounds but, due to the complexities of Icelandic word formation, I will maintain that lexicalization can be said to apply to formal aspects as well as to the more widely discussed semantic ones.

My interpretation of lexicalization, then, is a rather extensive one: In the following pages I will be assuming that lexicalization simply refers to all features which can not be inferred from the sum of the parts of any process of word formation, including ambiguity of form.

2 The Project

The project I am describing is the creation of an Icelandic base for a bilingual (i.e. Icelandic-Scandinavian) dictionary, being worked on at present at The Institute of Lexicography. The project is a joint venture of The Institute and Nordisk språksekretariat, and it is jointly funded by The Institute and Nordisk kulturfond. The intended size of the base is approximately 50 000–60 000 headwords. The work is to proceed in stages; the first stage, presently in progress, is the macrostructure; that is the word list itself, with information on inflections, word formation, idioms and other phrasal entries, etc.

The idea is to complete the work on the macrostructure as far as possible before turning to the description of individual headwords or the microstructure of the dictionary base.

At the outset a pilot word list containing just under 180 000 words was assembled from selected sources, and the observations made in this paper are drawn from the practical problems encountered while trying to reduce that number to the intended 50 000–60 000 words.

2.1 The Primary Sources

The main sources of the pilot word list are the archives of The Institute of Lexicography, a recent frequency dictionary of modern Icelandic (Jörgen Pind et al. 1991), compiled and published by The Institute, and the combinatorial dictionary *Orðastaður* described by the author Jón Hilmar Jónsson (1994) in his paper in this issue. As stated above, the basic word list arrived at from these three sources contains approximately 180 000 words.

As an indication of the scope of these sources the largest of the Institute's archives, *The Written Language Archive* (WLA), contains well over 600 000 headwords, the frequency dictionary contains just over 31 000 lemmas, and Jón Hilmar's book contains 11 000 headwords and over 100 000 compounds. These three sources are computerized so that producing the pilot list was relatively simple.

It should be noted that from the outset it was clear that some fields were underrepresented in the original data, which was only to be expected, given the nature of the sources. The historical aspect has always been in the foreground at the Institute, which in fact was founded just over fifty years ago for the purpose of producing a historical dictionary of Icelandic from the beginning of the age of printing in Iceland, i.e., 1540, to the present day. The WLA was created by excerpting for that purpose, and not intended as a base for a dictionary of the modern language (cf. Ásta Svavarsdóttir/Jón Hilmar Jónsson/Kristín Bjarnadóttir 1991). The combinatorial dictionary largely shows productive word formation, and the frequency dictionary, despite its rather hefty size, is based on a limited corpus of texts. The preliminary list of 180 000 words thus had to be reduced even further than down to 50 000–60 000, to leave room for new material. One of the sources of new material is the Institute's text corpus, and some excerpting in the traditional manner will also be done to supplement the original data. The bulk of the material will, however, come from the three sources mentioned above.

2.2 The Sorting of the Data

The sorting of the words in the original list was started in March 1994, and by now the whole alphabet has been sorted. Out of the 180 000 words in the original list, there are just over 138 000 compounds, and just over 35 000 derived or simple forms, i.e., non-compounds. Just over 4 500 "words" or items could not be analysed in this manner; these included abbreviations, some phrasal constructions, etc., as well as some items that can really only be classified as accidentals!

The sorting codes are kept as simple as possible, and the words are marked *included*, *omitted* and *under consideration*. Please note that *omitted* does not imply that the word is deleted from the lists, just that it does not get the status of a prospective headword.

Table 1 shows the proportions of base words, derived words, and compounds in the original word list, and the percentages in the sorted data for each type of word formation category, as they now stand in the lists:

	Base words		Derived words		Compounds	
To be included	8 276	44%	5 675	35%	27 247	20%
Under consideration	2 096	11%	2 290	14%	12 483	9%
To be omitted	8 555	45%	8 226	51%	98 495	71%
	18 927		16 191		138 225	

Table 1: *Proportions in the Pilot List*

The division between words to be included and those to be left out is not, of course, as clear-cut as these figures might imply. The compounds are in fact sorted into four categories, two of which contain candidates for our targeted 60 000 words. These categories were arrived at after quite some experimenting. Very simply put, the first category contains regularly formed compounds that are part of the core vocabulary, and common compounds in which meaning and form are not predictable, i.e., lexicalized words. The third category contains compounds which in fact do require explanation or listing, without being candidates for a base of the size intended here, as they are more peripheral in some manner, very often rare, specialized, old, dialectal, etc. The second category falls between the two, and the treatment of these words will be a matter of editorial policy in the future. The fourth category contains the words that do not really need explanations, being the products of fully productive and unambiguous word formation rules. Examples of this very simple outline of the classifications are given in (1), i.e., examples of compounds with *blóð* 'blood' as the first constituent.¹ All the examples seem to have the same morphological construction, and formally they are perfectly regular. The senses range from being fully lexicalized, as in (1a) *blóð-berg* 'thyme' (*timian*), to being fully productive as in (1f) *blóð-lykt* 'the smell of blood'.² The figures for each of the categories of compounds as they now stand in our files are given in parenthesis preceding the examples.

(1) *Examples of compound classification***Lexicalized compounds:**

INCLUDED (27 247):

- a *blóð-berg* 'Thymus arcticus'
- b *blóð-steinn* 'hematite'
- c *blóð-suga* 'blood sucker, vampire'

UNDER CONSIDERATION (12 483):

- d *blóð-arfi* 'Polygonum aviculare'
- e *blóð-miga* 'hematuria'
- f *blóð-rót* 'Potentilla erecta'

¹Please note that for reasons of clarity a hyphen has been inserted between the component parts of compounds in all examples, although this is not in accordance with Icelandic spelling. Please note also that the glosses are intended as rough guides to the meaning of the words, as it is sometimes impossible to give the exact meaning without giving lengthy definitions. Some words proved to be completely beyond translation except in context, and these are left without glosses.

²The ordinary user is not going to know that *blóð* in the first word is probably not the same as in the second. Cf. Ásgeir Blöndal Magnússon 1989.

OMITTED (13 529):

- g *blóð-björg* rarer name of *blóðberg*
- h *blóð-lýsa* 'leukemia'
- i *blóð-tala* 'red blood cell'

Productive compounds:

OMITTED (84 966):

- j *blóð-blettur* 'blood stain'
- k *blóð-bragð* 'taste of blood'
- l *blóð-lykt* 'smell of blood'

Independently of this very simple classification, the words are also sorted according to the field in which they are used, e.g., according to academic subject (*botany, zoology, medicine, etc.*), craft or occupation (*carpentry, sewing, agriculture, etc.*), and various other fields of diverse kinds (*literature, music, art, theatre; cars, traffic, flying; food, toys, etc.*) There are even fields for *sheep, cows* and *Icelandic national costume!* These fields will be used to make the classification more cohesive and consequent, thus giving a thematic key to the vocabulary.

2.3 The Criteria for the Sorting

The examples *blóð-berg* and *blóð-lykt* represent the two ends of the spectrum. One is fully lexicalized and the other is formed by a fully productive word formation rule. Lexicalizations of this kind are included in dictionaries or not, solely on the grounds of currency, i.e., whether the word is central enough in the vocabulary. As regards the productive part of compounding, the lexicographer is faced with the problem of deciding how far the prospective user can be expected to handle productivity, and how his needs can best be met.

The problem with sorting of the kind described here is of course that there are very many features that have to be taken into account when trying to determine what the prospective user may need. The problem is even greater when the prospective user is as remote as in this project. We are making a base for a bilingual dictionary which has to be useable for more than one set of languages. It is clear that the criteria are bound to be quite different according to the purpose of the end result, and we found it very important to try to construct the base in such a way that it could be used in as flexible a manner as possible. We also did not want the classification to be too complicated, for obvious reasons of time and money. We are therefore using a simple classification system to cope with very complex matters.

3 A Few Types of Compounds

Fully lexicalized words, like the examples I have been using, seem to have nothing whatsoever in common with their component parts. Fully productive compounds, on the other hand, seem to be nothing but the sum of their parts. Life would be relatively easy if that was all there was to it, but that is of course not the case.

3.1 Ambiguous Elements

Word formation rules are not unambiguous, and the ambiguity can appear both in the components themselves and the relation between them. Some regularly formed words are therefore not fully transparent.

In the examples in (2) below one of the component parts, *saumur*, can mean 'stitch', as in needlework, for example in *flat-saumur* 'flat stitch' or 'satin stitch', but *saumur* is also used in carpentry for some types of nails, such as *þak-saumur* and *pappa-saumur*.

- (2) a *flat-saumur* 'flat stitch', 'satin stitch'
 b *þak-saumur* 'roofing nail'
 c *pappa-saumur* 'nail used to fasten tar paper'

All homonyms are therefore a source of confusion when it comes to compounds.

The terminology in very many fields in Icelandic abounds with such words, as language policy dictates the preference of Icelandic neologisms over the use of loan words. "Ordinary" words are therefore very often used with specialized meanings, and the elements can thus very easily be semantically ambiguous.

The combining forms can be ambiguous as well, as shown in (3) and (4) below. The words *önd* 'duck' and *andi* 'spirit' can both have the combining form *anda-*. The accepted meaning of the compound in (4) *anda-læknir*, is 'spiritual healer', but I have seen it used in a police report to refer to 'a veterinarian specializing in treating ducks'.³

- (3) a *önd* fem. 'duck' Combining form: *anda-*
 b *andi* masc. 'spirit' Combining form: *anda-*
- (4) a *anda-læknir* 'spiritual healer'
 b *?anda-læknir* 'veterinarian specializing in ducks'

Although such examples can be humorous, they can also be the cause of very real and even quite serious misunderstandings.

3.2 Ambiguous relations of elements

An ambiguity which is perhaps not likely to cause confusion of a similar magnitude as the one mentioned above, is the ambivalence in the relation between the elements in a compound. This is demonstrated in the examples in (5) below.

- (5) a *lauk-baka* 'onion pie'
 b *sælkerabaka* 'gourmet pie'

The difference between 'onion pie' and 'gourmet pie' is obvious, but both words are regularly formed. Examples such as these are not likely to cause any difficulties for anyone, but not all cases are as obvious as these. In the sorting of our material we have found that it is necessary to be aware of these ambiguities in order not to overlook some less than obvious meanings.

³The story was that a sick duck was reported to the police. It got better on its own before the police found an *anda-læknir* for it.

3.3 The Choice of Lexical Items

The last but one of the types of compounds I will mention are words that are perfectly regularly formed, with perfectly transparent meaning, but they are lexicalized in the sense that the choice of components is not free. These are very often a source of mistakes for foreigners. To name an example (6), the words *verslun* and *búð* are synonyms for the word 'shop', although the first one is more formal than the second. Both *verslun* and *búð* can be used for places that sell books (as in (6a)) or flowers or cosmetics. But for an 'ice cream parlour' only *-búð* is acceptable, *ís-búð*.

- | | | | | |
|-----|---|-----------------|-----------------------|---------------------|
| (6) | a | <i>bóka-búð</i> | <i>bóka-verslun</i> | 'bookstore' |
| | b | <i>fisk-búð</i> | ? <i>fisk-verslun</i> | 'fishmonger's' |
| | c | <i>ís-búð</i> | * <i>ís-verslun</i> | 'ice cream parlour' |

The reaction of native speakers to mistakes made by foreigners in compounds such as these is usually "But you can't say that, actually, the Icelandic is ...". When pressed for an explanation, the only answer usually is: "Það er bara svona!" "That's just the way it is!"

3.4 Differences in Form

At the beginning of this paper I claimed that some formal aspects, such as differences in combining forms in Icelandic, have to be considered lexicalizations.

The most striking sets of combining forms do not in fact occur in compounds but in derivations (cf. (7)).⁴ As the glosses show, the meaning of each of the three words formed with the word *maður* 'man' and the affix *-legur* '-ly' is quite distinct from the other two:

- | | | | | |
|-----|---|---------------------|--------------------|-------------------------|
| (7) | a | <i>mann-legur</i> | <i>manns-legur</i> | <i>manna-legur</i> |
| | | 'human' | 'manly' | 'precocious, conceited' |
| | b | ? <i>barn-legur</i> | <i>barns-legur</i> | <i>barna-legur</i> |
| | | | 'childlike' | 'naive, childish' |
| | | | 'earnest, sincere' | |

Examples like these, where different combining forms are distinctive as to meaning, are very rare. More commonly only some of the possible combining forms are acceptable, even when the word formation is fully productive in all other respects. In order to explain this point a little excursion into Icelandic word formation is needed.

4 A Brief Word on Icelandic Word Formation

Contrary to what many recent publications on word formation maintain,⁵ Icelandic seems to contain word internal inflectional endings. A noun appearing as the first part of a compound can thus have the form of a stem or an oblique case, usually the genitive, or, more rarely, dative. Link phonemes do also occur. Adjectives, as first parts of compounds, contain inflectional endings as well, some of which are inflected for case inside the compounds. The

⁴Although the status of the affix *-legur* is a bit dubious on some formal grounds, at least it fulfills the requirement of being a bound form.

⁵I am mainly referring to discussions on Level Ordering.

verbal morphology in compounds is quite varied too. As noun+noun compounds are the most common and the most varied I will use these to demonstrate my point.

Examples (8) to (13) show some compounds with the nouns *bók* 'book' and *barn* 'child' as the first component part. Both these words commonly appear in compounds in the three forms shown in the examples. The choice of form for the first element in the compound seems to be largely arbitrary, as seen in (8a) and (8b). There is no semantic reason for the word 'book' to be in the plural in *bóka-búð* but not in *bóka-sala*, and the reason cannot be phonotactic either.

Noun+Noun Compounds:

	Stem:	Genitive plural:	
(8) a	<i>bók-sala</i>	<i>bóka-búð</i>	'bookstore'
	* <i>bók-búð</i>	* <i>bóka-sala</i>	
b	<i>bók-hlaða</i>	<i>bóka-safn</i>	'library'
	* <i>bók-safn</i>	* <i>bóka-hlaða</i>	
	Stem:	Genitive sg.:	Genitive pl.:
(9)	<i>bók-merki</i>	<i>bókar-merki</i>	<i>bóka-merki</i> 'book mark'

Words like the ones in (9) *bók-merki*, *bókar-merki*, *bóka-merki* 'book-mark' where more than one form is acceptable to convey the same meaning are common, as shown by the fact that The Institute gets a great number of phone calls from people asking about "the correct form".⁶ Part of the problem is perhaps that the difference can be quite difficult to hear, as in *bókar-merki* and *bóka-merki*. People have to be quite articulate for that distinction to be heard.

In examples (10) to (13) the reasons for the differences in forms cannot be semantic either. The difference between (13a) and (13b) is that *barns-meðlag* is 'child support' paid by a parent, whereas *barna-lífeyrir* is a part of the social security system. Both can be used to apply to one or more children.

(10) a	Stem:	<i>barn-fóstra</i>	'children's nurse'
b	Gen.sg.:	<i>barns-faðir</i>	'father of a child'
c	Gen.pl.:	<i>barna-pía</i>	'babysitter'
(11) a	Stem:	<i>barn-æska</i>	'childhood'
b	Gen.sg.:	<i>barns-aldur</i>	'childhood'
(12) a	Gen.sg.:	<i>barns-vagga</i>	'crib'
b	Gen.pl.:	<i>barna-rúm</i>	'baby's bed'
(13) a	Gen.sg.:	<i>barns-meðlag</i>	'child suport'
b	Gen.pl.:	<i>barna-lífeyrir</i>	'child support'

In an ideal system, the distribution of genitive singular and plural would obey the same rules as in syntax, where the feature *number* would be used consequently. Icelandic indeed has such words, as shown by the words for the son(s) and daughter(s) of farmer(s), brother(s) and king(s) below:

⁶A similar observation on Swedish is made in Malmgren 1994.

- | | | | | |
|------|---|-----------------------------|---|-----------------------------|
| (14) | 1 | $N_{gen.sing.} + N_{sing.}$ | 2 | $N_{gen.sing.} + N_{plur.}$ |
| | a | <i>bónda-sonur</i> | a | <i>bónda-synir</i> |
| | b | <i>bróður-sonur</i> | b | <i>bróður-synir</i> |
| | c | <i>konungs-dóttir</i> | c | <i>konungs-dætur</i> |
| | 3 | $N_{gen.plur.} + N_{sing.}$ | 4 | $N_{gen.plur.} + N_{plur.}$ |
| | a | ? <i>bænda-sonur</i> | a | <i>bænda-synir</i> |
| | b | * <i>bræðra-sonur</i> | b | <i>bræðra-synir</i> |
| | c | ? <i>konunga-dóttir</i> | c | <i>konunga-dætur</i> |

The words marked with a question mark, (14-3a) *bændasonur* ‘the son of farmers’ and (14-3c) *konungadóttir* ‘the daughter of kings’, are not very probable for semantic reasons, but not completely impossible. The first one is slightly better than the second one, as there are female farmers these days. The problem with the second one is that the word for *king* in Icelandic is not only a masculine noun but is used of males only, unlike the words *forseti* ‘president’ and *borgarstjóri* ‘mayor’ that are masculine nouns but readily used of women. The word marked with a star, (14-3b) *bræðra-sonur* ‘the son of brothers’, is obviously impossible for semantic reasons!

Words such as these are rare, whereas examples such as the ones in (8) to (13) are much more common. Usually the difference in form is just exactly that, an arbitrary difference. It is a matter of accepted forms and those that just sound wrong. Why they do is not a question that is easily answered.

4.1 The Distribution of Forms

The search for phonotactic, semantic, or formal explanations for the difference in combining forms only turns up isolated explanations. Some inflectional classes only seem to appear in certain morphological constructions. Feminine nouns ending in *-a* in the nominative singular only use the genitive singular as a combining form, even in nouns such as the one in (15a) *peru-tré* ‘pear-tree’ when the genitive plural would perhaps be expected semantically, as seen in (15b) *epla-tré* ‘apple tree’, where the apples appear in the plural.

- | | | | | |
|------|---|--------------------|----------------------|--------------|
| (15) | a | <i>peru-tré</i> | * <i>per(n)a-tré</i> | ‘pear tree’ |
| | | [fem.sing.] | [fem.plur.] | |
| | b | * <i>eplis-tré</i> | <i>epla-tré</i> | ‘apple tree’ |
| | | [neut.sing.] | [neut.plur.] | |

Another indication of a rule is that in multiple compounding the tendency seems to be to use the genitive form when the first part of the multiple compound is a compound itself (cf. Baldur Jónsson 1984, and Eiríkur Rögnvaldsson 1986), as demonstrated in (16a):

Multiple compounds as first component parts:

- | | | | | |
|------|---|--------------------|---|-------------------------------------|
| (16) | a | [[[N][N]]] | b | [[[N][N]] _{gen.sing.} [N]] |
| | | <i>borð-plata</i> | | <i>skrifborðs-plata</i> |
| | | ‘table top’ | | ‘writing table top’ |
| | | | | i.e. ‘desk top’ |
| (17) | a | <i>verð-hækkun</i> | b | <i>olíuverðs-hækkun</i> |
| | | ‘price rise’ | | ‘oil-price rise’ |

Even though there are some rules on a par with the two just mentioned, the problem is that very many factors play a role, and in the end the conclusion always seems to be: That's just the way this word is. That, I maintain, is a clear indication of lexicalization.

5 Conclusion

This very brief outline of some of the problems posed by the complexity of the formal aspects of Icelandic compounding is hopefully sufficient to show that the subject deserves careful attention.

As we are not producing a fully fledged dictionary, but rather a base from which dictionaries are to be produced in the future, our solution is to list all combining forms under the words they are derived from. The combining forms themselves are then graded for inclusion or omission in the base, in the same way as the compounds, or indeed all other words from the pilot word lists. All compounds are then placed under the proper combining form. The lists are therefore not in alphabetical order, but ordered morphologically, or rather indexed according to the morphology. This means that the word in (3a) *önd* 'duck' is not separated by the whole alphabet from its combining form *anda-*.

I should stress again that productively formed compounds (and derivations) are very much a part of the dictionary base. They are not deleted from the lists, just marked to show that they are not expected to be given the status of headwords.

It will be a matter of editorial policy in the future just how the base is developed and used. The relative weight of the factors I have mentioned will of course vary according to the aims of the editors, together with the question of how active or passive the prospective dictionary is supposed to be. For now we just have to make sure that the base is flexible enough to serve as many needs as possible.

Bibliography

- Ásgeir Blöndal Magnússon 1989: *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Ásta Svavarsdóttir/Jón Hilmar Jónsson/Kristín Bjarnadóttir 1992: Fra seddelsamling til database: Leksikografisk analyse af islandske verber. In: Fjeld, R. V. [ed.]. *Nordiske studier i leksikografi*, 390–402.
- Baldur Jónsson 1984: Samsett nafnorð með samsetta liði. Fáeinar athugasemdir. In: *Festskrift til Einar Lundeby*. 3. október 1984, 158–174.
- Eiríkur Rögnvaldsson 1986: *Íslensk orðhlutafraeði. Kennslukver handa nemendum á háskólastigi*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Jón Hilmar Jónsson 1994: *Orðastaður. Orðabók um íslenska málnotkun*. Reykjavík: Mál og menning.
- Jón Hilmar Jónsson 1995: Nøkler til ordforrádet: Om lemmafunksjon, struktur og informasjonstyper i en ny kombinatorisk ordbok over islandsk. [This volume, 245–254].
- Jörgen Pind [ed.]/Friðrik Magnússon/Stefán Briem 1991: *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.
- Malmgren, Sven Göran 1994: Sammensättningsmorfologi och lexikografi. In: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993*, 179–184.

Avledningar i Svenska Akademiens ordbok

The Germanic dictionaries on historical principles endeavour to describe relations between words, both formal and semantic relations within the lexicon of a special language as well as etymological relations to words in other languages. One of the described relations is the connection between a derivation and its base. The purpose of this paper is to give an account of this kind of description in The Swedish Academy Dictionary.

Det är en ambition hos de germanska, historiska ordböckerna att klargöra samband mellan orden. De påvisar formella och semantiska samband såväl inom det särspråkliga ordförrådet som förhållanden till ord i andra språk (genom redovisning av etymologin). Det är alltså en uppgift för ordböckerna att klargöra bl. a. det samband som råder mellan en avledning och dess bas. I artikeln beskrivs hur *Svenska Akademiens ordbok* (SAOB) fullgör denna uppgift. Det diskuteras också om lösningen kan motiveras, dels ur praktisk användbarhetssynpunkt, dels och framför allt ur språklig synpunkt.

1 Arrangemang i avledningsramsa

Både avledningar och sammansättningar presenteras i SAOB vanligen i en s. k. ramsa. Efter all betydelse­redovisning av ett huvudord kan det, under rubriken *Ssgr.* eller *Avledn.*, följa sammansättningar med det aktuella huvudordet som första led resp. avledningar av huvudordet. Den alfabetiska följd­en kan brytas av sammansättningar till en sammansättning eller avledning eller av avledningar till sammansatta eller redan avledda ord.

SNAPPA *snap*^{3a2}, v. -ade. [...]

Avledn.: **SNAPPARE**, om person m. | |ig., om djur m. l. r., om sak r. l. m. [jfr ä. d. *snappere* (d. *snapper*), t. *schnapper*] 1) till I 1: person som snappar (åt sig); i sht förr äv. närmande sig l. övergående i bet.: snattare l. tjuv l. [sannol. under påverkan av *snapphane*] plundrare l. rövare; nästan bl. i *ssgr*; jfr *penninge-*, *upp-snappare*. 2) till I 1 (a): djur som snappar (åt sig); numera nästan bl. i *ssgn* *flug-snappare* (l. elliptiskt härför). ÖSTERGREN (1942; o *m flugsnappare*). jfr *paradis-snappare*. 3) [möjl. utvidgad l. bildl. anv. av 1, 2] *urmak.* i slagverk hos ur: på mellersta bommen sittande liten hake som reglerar antalet slag gm att gripa tag i timtrappan o. lyfta denna hak för hak tills den är helt upplyft o. slåendet därmed stoppas.

HusgKamRSthm 1756—57, s. 491. Ssgr: **snappar-**, äv. **snappare-arm. urmak.** till **snappare 3**, i slagverk hos ur: från snapparhjulet utgående arm (se **arm**, sbst. **III 4**) som bildar snapparens huvuddel. ERICSSON *Ur.* 115 (1897). -**hjul.** [...] — **SNAPPERI**¹⁰⁴ [...]

REDUCERA [...]

Avledn.: **REDUCERBAR**, adj. [jfr t. *reduzierbar*] som kan reduceras; särsk. *a*) (i vetenskapligt fackspr., i sht *filos.*) till **3**; stundom: härledbar. WIKNER *Tidsex.* 229 (1888). En ej vidare reducerbar psykisk differens. SCHÉELE *Sjäsl.* 180 (1895). *b*) i sht *kem. o. metall.* till **8 (a o.) b**. FOCK *IFys.* 362 (1861). *Avledn.*: **reducerbarhet**, r. l. f. BJÖRKMAN (1889). särsk. (i sht *kem. o. metall.*) till **reducerbar b**. *JernkA* 1897, s. 280. — **REDUCERLIG**, adj.

Ramsorna tjänar naturligtvis syftet att ge överskådlighet och att vara utrymmesbesparande. Det som sparar utrymme är att orden i ramsan på ett enkelt sätt kan få sitt etymologiska samband med huvudordet redovisat. Sålunda anknys avledningar till betydelser hos avledningsbasen genom hänvisning till det moment på huvudordet där betydelsen beskrivs, t. ex. ”till **I I**” (se illustration ovan).

När en avledning självständigt har utvecklat en betydelse som har sin likhet hos avledningsbasen, t. ex. en bildlig betydelse, kan detta framhållas genom en hänvisning medelst ”motsvarande” (förkortat: motsv.) eller ”jämför” (förkortat: jfr).

2 Avledning eller sammansättning?

I ramsan gäller att ”avledning” uteslutande är suffixavledning. I övrigt omfattar begreppet i SAOB även kortavledning, dvs ordbildning åstadkommen genom avsaknad av suffix (se *nedan*). Företeelsen benämns emellertid inte alltid med termen kortavledning. Kombinationen av prefix och huvudord, ”prefixavledning” i moderna morfologiska beskrivningar, betraktas och behandlas däremot som sammansättning, vilket har junggrammatisk förankring (jfr Brugman 1906:4ff). Förleden benämns dock prefix och utgör oftast eget uppslagsord, som står i en divisartikel med sammansättningsramsa.

Det är förstaeligt med oklarhet om vad som är affix och vad som är stam. Ord på *-mässig*, t. ex. *pliktmässig*, behandlas som sammansättning men som avledning i många sammanhang utanför SAOB. Åsikten vad som är suffix har i ett fall hunnit ändra sig under utgivningstiden. I de första banden är ord på *-aktig* sammansättning, t. ex. *blyaktig*. I senare band beskrivs emellertid *aktig*-orden som avledning, t. ex. *silveraktig*.

3 Alternativ till presentation i ramsa

Det är inte alla avledningar som presenteras i ramsa, vanligtvis bara avledningar med inhemska och under den nysvenska perioden (från 1521) produktiva suffix. Några avledningar står som egna huvuduppslagsord, andra behandlas på det ord som utgör avledningsbas.

3.1 Eget huvudord

Det är inte ovanligt att en avledning har utvecklats till flera lexem som avviker från den betydelse som är mer given utifrån avledningsmorfemens normala användning och basens grundbetydelse. Sådana avledningar med mer eller mindre klar prägel av eget ord behandlas ofta i en huvudartikel. I den etymologiska parentesen upplyses om vilken bas avledningen har medelst formlerna "avledning av X" eller bara "(bildning) till X". Dessutom hänvisas genom att betydelser hos avledningen som har ursprung eller motsvarighet hos avledningsbasen markeras medelst "jämför" (jfr) eller "motsvarande" (motsv.).

Den lexikalisering eller annan omständighet som motiverar behandling som eget huvudord finns ingenstans beskriven. Den behöver naturligtvis inte heller anges. Någon praktisk olägenhet medför det knappast att ordboksanvändaren inte med säkerhet kan veta var en avledning står. Om den behandlas i ramsa hänvisas nämligen från den plats där den skulle ha stått som huvudord, och om den behandlas i huvudartikel hänvisas från den ramsa där den hör hemma. Det hänvisas också i aktuella fall från verbalsubstantiv som bara har nämnts i motsvarande verbs huvud (se nedan) till det ställe där avledningen beskrivs ytterligare. Allmänt gäller att som huvudord står avledningar med rik betydelseutveckling och avledningar med sådana suffix som inte är fornärvda i svenskan eller inte är övertygande produktiva i vårt språk, t. ex. romanska eller latinska suffix som *-or/-ör/-ator/-atör/-itör, -ant/-ent, -issa, -ist, -isera, -ik*.

3.1.1 Kortavledningar

Sådana verbalabstrakta bildningar som inte har någon avledningsändelse utan tvärtom utmärks av att sakna sådan och bara utgörs av stammen, s. k. kortavledningar, behandlas också i egna artiklar som huvudord. Men det hänvisas till dem från motsvarande verb, antingen på samma plats efter böjningsformerna där andra verbalsubstantiv nämns (se nedan), t. ex. GNÄGGA; jfr GNÄGG, eller i slutet av den etymologiska parentes, t. ex. SKICKA: . . . — Jfr SKICK. Det senare arrangemanget tillämpas när kortavledningen har mer utvecklad betydelse än den normala för verbalabstrakter.

3.2 Behandling tillsammans med avledningsbasen

Avledningar kan få sin betydelse redovisad tillsammans med sin bas, om denna utgör eget huvuduppslagsord. Här några typfall:

3.2.1 Verbalsubstantiv

Om ett verb bildar verbalsubstantiv med suffixen *-an, -andel/-ende, -else, -(n)ing, -eri* eller de personbetecknande *-are, -(ar)inna, -erska* framgår detta redan i verbartikelns inledning, i

det s. k. huvudet där bl. a. böjning och stavning redovisas. Avledningsändelserna uppräknas efter presentationen av moderna och döda eller mindre brukliga böjningsformer.

SKOTA *skw³ta²*, v.² -ade. vbalsbst. -ANDE, -NING;
-ARE (se *avledn.*). [till **SKOT**, sbst.², o. **SKOTA**, sbst.] 1)
skogsv. motsv. **SKOTA**, sbst. 1:

Ett *-ande* anges regelmässigt eftersom varje verb kan användas till sådan verbalabstrakt bildning. Avledningsändelsen *-(n)ing* anges om den förekommer till det aktuella verbet men ägnas, liksom fallet är med *-ande*, ingen beskrivning om avledningen inte har utvecklat någon mer specialiserad betydelse än den verbalabstrakta grundbetydelsen 'handlingen eller förhållandet att X-a, det att X-a'. De andra suffixen nämns om avledningen finns, och inom parentes efteråt anges om den får sin betydelse beskriven i ramsa, varvid det står "se *avledn.*" inom parentes, eller som eget huvudord ("se d. o." = detta ord). Verbalsubstantivet kan också behandlas i betydelsebeskrivningen av avledningsbasen, alltså det aktuella huvudordet. I det senare fallet, som bara kan gälla de sakbetecknande avledningarna på *-an*, *-andel-ende*, *-else* eller *-(n)ing*, anges ofta inom parentes i vilket moment betydelsen redovisas. I konsekvens med att dessa avledningar kan redovisas i verbartikeln kan avledningar till dem stå i ramsa efter verbet, t. ex. PÅVERKNINGSBAR i ramsa efter PÅVERKA.

3.2.2 Adverb av adjektiv

Den till nästan alla adjektiv förekommande adverbbildningen på *-t* redovisas på motsvarande adjektiv, omedelbart efter eventuell komparationsform och andra böjningsformer om dessa avviker från de regelmässiga, t. ex.

SPRAGGIG *sprag³ig²*, adj. -are. adv. -T. [sannol. *avledn.*
av **SPRAGGA**, v.]

Vanligtvis inskränks behandlingen av adverb på *-t* till detta nämnande av formen, men om betydelsen avviker från den normala, den som kan parafrastras "på X sätt", så kan adverbet beskrivas i eget moment på adjektivet. Så har skett på t. ex. **STILIG** och **SLARVIG**.

Adverbbildningar på *-liga*, *-ligen*, *-ligt* o. dyl. får oftare än dem på *-t* en behandling som eget huvudord. Skälet är att de är mindre förutsägbara, både i fråga om förekomst och betydelse.

3.2.3 Adjektiviska particip

Participbildningar behandlas inte som avledningar i SAOB. Däremot uppmärksammas i förekommande fall det speciella hos presens- och perfektparticip (förkortat p. pr. respektive p. pf.) som gör dem till annat än böjda former med verbal betydelse. Antingen står de som egna huvudord (SPÄNNANDE, INTRESSERAD), med ordklassbeteckningen *participiellt adjektiv* (p. adj.), eller så framhålls det i betydelsebeskrivningen av verbet att de förekommer i adjektivisk användning.

På sådana adjektiv som har en för sig själv i eget moment presenterad reflexiv och/eller deponentiell användning behandlas oftast också perfekt particip i eget moment. Anledningen är att perfektparticipet dels har en innebörd som motsvarar den reflexiva resp. deponentiella formen, dels en innebörd som motsvarar den transitiva eller intransitiva utan att röja detta i

sin form (*föräldrarna ska skilja sig — De är skilda; Man borde skilja näringsfrågorna från finansfrågorna. — Näringsdepartementet är skilt från finansdepartementet*).

Sådana bildningar som formellt överensstämmer med ett perfekt particip och kan knytas till ett verb behandlas oftast under detta även om ordet kan vara avlett av ett substantiv. Så återfinns t. ex. RYNKAD i eget undermoment av betydelsebeskrivningen av verbet RYNKA. Men såsom typiskt för ordboken framhålls det att ordet kan vara denominativt. Det står nämligen i etymologisk parentes att RYNKAD ”i vissa [. . .] användningar sannol. [är] att fatta ss. ett p. adj. avlett av RYNKA, sbst.” Ordklassbeteckningen *participiellt adjektiv* (p. adj.) har t.o.m. band 31 på ett tvivelaktigt sätt använts om liknande bildningar även när ordet inte är particip utan avlett av ett substantiv (jfr nedan om GLOS-ÖGD etc. under 4.1).

4 Några speciella avledningsförhållanden

Arrangemanget med avledning i ramsa efter ett huvudord passar särskilt bra när orden är avledda direkt av huvudordet. Eftersom så inte alltid är fallet nämns ofta avvikelser och oklar anknytning speciellt.

4.1 Avledda sammansättningar

Sådana avledningar som är gjorda samtidigt som de kommit att utgöra och ibland bara förekommer som senare sammansättningsled (s. k. avledd sammansättning) redovisas i sammansättningsramsa efter det huvudord som utgör första ledet i sammansättningen, t. ex. LÄG-LÄNT i ramsa efter LÄG, LÄNG-HÄRIG efter LÄNG och STOR-SÄLJARE efter STOR. Den senare är inte en avledning av STOR-SÄLJA; en sådan sammansättning finns inte i lexikaliserad form.

Det är ofta svårt att veta huruvida en sammansättning har efterled som är en självständigt bildad avledning (STOR+SÄLJARE) eller en efterled som med större säkerhet har gjorts i sammansättningsögonblicket. Det finns ju ett självständigt *säljare* i språket, men detta ord har knappast samma betydelse som *säljare* i sammansättning med *stor*, alltså ’något som säljs (i stor mängd)’, ’något som säljer bra’.

Ibland ligger en sammansättning i främmande språk till grund för den avledda sammansättningen i svenskan. Uppllysning om detta brukar då ges i etymologisk parentes, t. ex. på SJÄLA-SÖRJARE i SJÄLA-ramsan efter huvudordet SJÄL. Där hänvisas till medellågtyskans *seel(e)sorger* och tyskans *seel(en)sorger*, eftersom det svenska ordet förmodligen är bildat som översättningslån av ordet i medellågtyskan eller tyskan.

Många avledda sammansättningar liknar particip, t. ex. GLOS-ÖGD, GOD-ARTAD, HÅRD-HÄNT, HÖG-BARMAD, KO-BENT, LÄG-LÄNT, LÄNG-ARMAD, STOR-STILAD. Deras efterled förekommer inte som självständigt ord. I vissa fall skulle naturligtvis sammansättningen kunna vara avledd av ett annat sammansatt ord, ett substantiv. Sålunda kan GLOS-ÖGD möjligen vara avlett av GLOS-ÖGA men sannolikt inte HÅRD-HÄNT av *HÅRD-HAND. Dessutom torde i en del fall bildning ha skett i analogi med annan sammansättning; jfr sviten SID-LÄNT, HÖG-LÄNT, LÄG-LÄNT.

De participlika bildningarna GLOS-ÖGD osv. har hittills anmärkningsvärt nog fått ordklassbeteckningen *participiellt adjektiv* (p. adj.), trots att ordet inte är ett particip utan en bildning till ett substantiv. Skälet till benämningen är att samma morfem som bildar verbformen perfekt particip har använts, men man kommer inte ifrån att beskrivningen

”participiellt” är oegentlig i detta sammanhang. Ibland har dock ordklassbeteckningen adjektiv använts, t. ex. på HÖGLJUDD, som utgör eget huvuduppslagsord.

4.2 Sammansättningar med particip

Gränsen mellan ordböjning och avledning är inte alldeles klar. SAOB överensstämmer med gängse beskrivning genom att inte kalla particip med adjektivisk eller mindre verbal karaktär för avledning (jfr dock Andersson 1993:43), trots att adjektiviska particip påminner om avledningar, dvs medelst ändelsetillägg nybildade ord (jfr ovan under 3.2.3). De adjektiviska participen är förvisso böjda verbformer vars betydelse har utvecklats i adjektivisk riktning. Ett particip kan emellertid i kombination med förled bilda något som påminner om avledd sammansättning (se ovan). Likheten består i detta att det sammansatta participet saknar en motsvarande infinitiv. Ordboken redovisar sådana ordformer som egna uppslagsord med ordklassbeteckningen ”participiellt adjektiv”.

Sammansättningar med particip som har den likheten med avledda sammansättningar att de inte självklart går att knyta till en motsvarande sammansatt infinitiv är exempelvis BARN-LAGD, BLEK-LAGD, DJUP-KÄND, DJUP-GRUNDAD, FET-LAGD, GRUND-GÅENDE, LAND-LEGAD, SNAR-STUCKEN, SJÄLV-GÅENDE, SJÄLV-SKRIVEN, SPÅN-TÄCKT. De utgör alltså egna uppslagsord i sammansättningsramsa. I några fall finns en motsvarande infinitiv som också fått utgöra eget uppslagsord, men inte ens då har participet ansetts vara en form av detta motsvarande verb. Grunden för participet kan naturligtvis variera från fall till fall. Bortsett från analogier kan en sådan fras ligga bakom som inte nödvändigtvis innehåller en participform. Denna fras kan användas i en omskrivning av betydelsen, t. ex. *grund-gående* — *om fartyg: som kan gå på grunt vatten; fet-lagd* — *om person: som är lagd (beskaffad) så att han blir fet*.

Sammansättningarna med particip ovan överensstämmer med sådan particip som har en motsvarighet i ett partikelverb, i SAOB:s terminologi: särskild förbindelse, t. ex. LÄGGA IN. Participet till ett partikelverb är en sammansättning som kan vara bildad utan en motsvarande sammansatt infinitiv. Den redovisas oftast under en motsvarande sammansatt infinitiv, även om denna skulle vara mindre bruklig. T. ex. står INLAGD (om gurka o. dyl.) under INLÄGGA, visserligen med speciellt framhållande av den mer eller mindre adjektiviska participformen. Participet SAMMAN-HÄNGANDE har däremot fått sin självständighet mer markerad och står som eget uppslagsord, trots att det finns ett motsvarande SAMMAN-HÄNGA, vilket det hänvisas till när en betydelse hos participet har sin motsvarighet hos infinitiven. I sådana fall som SAMMAN-RINGLAD och PÅ-FLUGEN finns emellertid ingen motsvarande infinitiv (**samman-ringla*, **på-flyga*) redovisad som uppslagsord, vilket skulle ha varit missvisande. Participformerna är nämligen med största sannolikhet bildade till en förbindelse av verb och betonad partikel, *ringla sig samman*, *flyga på*, och har fått sin sammansatta form när participet bildades.

4.3 Retrogradbildningar

I en avledningsramsa förtecknas avledningar som har bildats till en bas genom att en ändelse lagts till. Varken kortavledningar eller retrogradbildningar är av detta slag. Som nämnts ovan är kortavledningar egna huvuduppslagsord. Det kan också retrogradbildningar bli,

t. ex. SPÄNST, som enligt upplysning i den etymologiska parentesen utgör retrograd bildning till SPÄNSTIG. Sådana retrogradbildningar som är senare led i sammansättning står ofta i sammansättningsramsa, men de hör inte självklart hemma där eftersom senare leden inte har lagts till förleden och därmed bildat sammansättning. Ordet har i stället utgått från en annan sammansättning och ombildats till ord med avledningsform som skenbart är ursprungligare. Därför är det missvisande ifall det inte i etymologisk parentes upplyses om det rätta förhållandet. Så görs i de senare banden, t. ex. SJÄLA-RINGA [retrograd bildning till **-ringning**], SKÄR-SLIPA [sannol. retrograd bildning till **-slipare**].

4.4 Osäkerhet ifråga om avledningsbas

Naturligtvis är det många gånger oklart hur en ordbildning har gått till. Här ett par sådana fall:

4.4.1 Avledning till sammansättning

Som redan nämnts kan en ramsa med sammansättningar eller avledningar också innehålla sammansättningar eller avledningar till något ord i ramsan. Det förekommer ganska sällan att en sammansättningsramsa redovisar avledningar till någon sammansättning. Det föreligger nämligen sällan säkra fall av sådan avledning. Ett ord som redovisats som avledning till en sammansättning är **skol-omogenhet** till SKOL-OMOGEN. Därmed visas att *omogenhet* sannolikt inte är en självständig avledning som utgör senare led i sammansättning med SKOLA. Avledningen är nog gjord av sammansättningen SKOL-OMOGEN.

I de flesta fall står alltså en avledning på samma nivå som övriga ord i sammansättningsramsan, t. ex. INRIKES-POLITISK och INRIKES-POLITIK. På sammansättningen med avledning som senare led hänvisas ofta till sammansättningen med icke avledd efterled. Därmed visas att avhängighet kan föreligga (STOR-POLITISK [jfr **-politik**]).

4.4.2 Avledning med beroende i annan avledning

En avledning kan utgå från annan avledning utan att vara en sekundär avledning med adderad ändelse av typen **reducerbarhet** till REDUCERBAR. Ett sådant fall är moverade bildningar på *-(ar)inna* och *-erska*. De på *-erska* kan visserligen vara avledda direkt av verbet men är ofta bildade till motsvarigheten på *-are*. BAGERSKA t. ex., som utgör eget huvuduppslagsord, uppges i etymologisk parentes vara femininbildning till BAGARE. SKOJERSKA har också fått anknytning till sin *are*-motsvarighet, SKOJARE, men genom att redovisas i avledningsramsa efter detta ord. På så sätt uppges SKOJERSKA vara bildat till SKOJARE, men för att något reservera sig och påvisa annan möjlighet hänvisar ordboken i etymologisk parentes till SKOJA; det är ju inte helt uteslutet att SKOJERSKA kan vara avlett direkt av verbet. Liknande är arrangemanget på SPELERSKA, som står i avledningsramsa efter SPELA och har etymologisk parentes där ordet uppges "delvis" vara bildat till SPELARE.

Ett annat fall av avledning med utgångspunkt i annan avledning är bildningar på *-eri*, som antingen med säkerhet utgår från en avledning på *-are* (*trädgårdsmästeri* till *trädgårdsmästare*) eller kanske gör det (*skojeri* till *skojare* eller till *skojja*). Jag bortser från de fall då ordet på *-eri* har inlånats direkt, t. ex. *mejeri*, *rytteri*, eller bildats i analogi,

Sven-Göran Malmgren

Nationalencyklopedins ordbok — några karakteristiska typer av information

In this paper, some characteristics of a new defining dictionary of Swedish, *Nationalencyklopedins ordbok* (NEO; The dictionary of the National Encyclopedia), are presented. The three-volume dictionary, comprising about 62 000 lemmas, will appear in 1995–96. Partly, it is based on another dictionary of Swedish, *Svensk ordbok* (SOB; A dictionary of Swedish). However, there are several new — or partly new — informational categories in the NEO, accounted for in the paper. These categories are (1) etymologies (for about half of the lemmas) and years of the first known occurrence (for every lemma and every numbered sense); (2) valency; (3) morphological analysis; (4) word formation; (5) musical accent, and (6) references to synonyms, antonyms, co-hyponyms, etc.

1 Inledning

Vid Institutionen för svenska språket vid Göteborgs universitet pågår sedan några år arbetet på en större svensk definitionsordbok, *Nationalencyklopedins ordbok* (NEO). Ordbokens tre band om sammanlagt ca 2000 sidor och med ca 62 000 lemman är planerade att utkomma hösten 1995, våren 1996 och hösten 1996. En CD-ROM-version av ordboken förbereds också. Ordbokens namn förklaras av att den ges ut i anslutning till den stora svenska *Nationalencyklopedin* (och av samma förlag, Bra Böcker); i bokhyllan kommer de tre ordboksbanden att se ut som lika många band av encyklopedin.

Arbetet på NEO utgår från den lexikala databas som utgjorde grunden för den 1986 utgivna *Svensk ordbok* (SOB). I synnerhet bygger betydelsebeskrivningarna på SOB:s betydelsebeskrivningar, om man bortser från några tusen nya ord i NEO (se vidare Allén & Malmgren 1993). NEO är emellertid inte en ny upplaga av SOB utan en ny ordbok, med bl.a. åtskilliga informationskategorier som inte fanns representerade i SOB och på något undantag när inte heller i någon annan allmän ordbok över nutidssvenskan. Nedan presenteras de viktigaste av dessa informationskategorier jämte några problem som vi konfronterats med vid utarbetandet av ordboken.¹

¹Arbetet på ordboken leds av professor Sture Allén med artikelförfattaren som biträdande projektledare. Övriga medarbetare är eller har varit Sture Berg, Bertil Engdahl, Birgitta Ernby, Carl von Garaguly, Sven Lövfors, Christian Sjögreen (systemering), Lennart Tegner och Maria Toporowska Gronostaj. Dessutom har ett stort antal experter engagerats för kontroll av olika typer av uppgifter i ordboken.

2 De nya informationskategorierna i NEO

De viktigaste nya informationskategorierna i NEO är:

- Etymologier; årtal för första kända belägg i svenskan
- Konstruktionsuppgifter vid alla verb och vid adjektiv och substantiv med bestämningar av valenskaraktär
- Ordled (≈morfologisk analys, av alla uppslagsord)
- Nomina actionis vid alla verb

Dessutom är informationen kraftigt utökad inom ytterligare två kategorier:

- Uttalet (bl.a. ges uppgifter om tonaccent)
- Hänvisningar mellan semantiskt besläktade ord, med åtskillnad mellan hyponymer, hypernymer och kohyponymer (samt, givetvis, synonymer och antonymer)

Av följande exempel ur det nu tryckfärdiga första bandet framgår hur artiklarna i NEO är disponerade, bl.a. hur de nya informationskategorierna är inordnade i artiklarna. Observera "rubrikerna" i kapital som är avsedda att leda läsaren rätt om han bara vill läsa delar av artikeln. (Etymologin och årtalet för första belägg står under rubriken "HIST." sist i artiklarna.)

abrupt adj., neutr. ~

ORDLED: *ab-rupt*

• plötsligt inledd eller avslutad särsk. om viljestyrt förlopp {→**hastig**, **tvär 2**}: *säga ett ~ adjö; ett ~ svar*

HIST.: sedan 1843; av lat. *abruptus* 'tvärt avbruten'; jfr **ruptur**

ackompanje'ra verb ~de ~t

ORDLED: *ac-komp-anj-er-ar*

SUBST.: **ackompanjerande**, **ackompanjer-
ing**; **ackompanjemang**

• spela ackompanjemang till viss huvudstämma etc.: *orkestern ~de kupletterna med säker stilkänsla*

BET.NYANS: överfört i fråga om tillkommande sinnesintryck e.d.: *vinerna som ~de var en torr sherry och ett halvtorr moselvin*

KONSTR.: ~ (*ngn* el. *ngt*) (*på ngt*)

HIST.: sedan 1779; av fra. *accompagner* 'ledsaga; ackompanjera'; jfr **kompani**

Fig. 1. Artiklarna abrupt och ackompanjera i NEO

Jag ska nu något kommentera de sex (helt eller delvis) nya informationskategorierna och samtidigt exemplifiera med hjälp av s.k. skärmskrifter från databasen. De sorteringar, i de flesta fall baklängessorteringar, som ligger till grund för exemplen kan när som helst göras av medarbetarna, och förhoppningen är att åtminstone vissa av dem blir möjliga att göra på CD-ROM-versionen av NEO.

2.1 Etymologier; årtal för första belägg

Etymologier ges i NEO för alla enkla ord och även för många avledda och sammansatta ord, då speciella skäl talar för detta. Oftast behöver endast (högst) en etymologi per lemma ges, men då lemmatiseringen i NEO görs på grundval av rent formella — icke-etymologiska — kriterier kan det förekomma att olika betydelsemoment inom ett och samma lemma har olika etymologier. Sammanlagt ges ca 27 000 etymologier.

Årtal för första kända belägg i svenskan ges för alla numrerade betydelsemoment, ca 68 000 stycken. Huvudkällan är *Svenska Akademiens ordbok* (SAOB) och, särskilt för den del av alfabetet som ännu inte täcks av SAOB (*t-ö* och en mindre del av *s*), SAOB:s samlingar (SAOS), som numera är mikrofilmade och direkt tillgängliga även i Göteborg. Andra viktiga källor är Söderwalls *Ordbok öfver svenska medeltidsspråket* och *Nyord i svenskan från 40-tal till 80-tal*. Dessa källor tillhandahåller tillsammans den stora majoriteten av uppgifterna om första belägg, men för flera tusen av uppslagsorden måste uppgifterna sökas på annat håll. Det gäller bl.a. många ord i början av alfabetet som är för nya för SAOB men för gamla för *Nyord i svenskan*, och det gäller ord som är för nya för *Nyord i svenskan*. I inte så fall har vi också funnit tidigare belägg än SAOB-beläggen med hjälp av institutionens (Språkbankens) egna korpusar, inte minst den stora Strindbergskorpusen. De ämnesexperter som engagerats för kontroll av artiklar om facktermer har också i många fall angett tidigare årtal för första belägg än någon av huvudkällorna.

Likväl har det hänt att vi i sökandet efter ett rimligt årtal för första belägg har sett alla tänkbara källor uttömda. Vi kanske har haft ett konkret belägg från 1980-talet, men vi har varit säkra på att ordet eller betydelsen funnits i svenskan sedan 1960-talet. I sådana (undantags)fall har vi, med stor tvekan, bestämt oss för att frångå kravet på ett konkret första belägg och tillåtit oss skrivningar som "trol(igen) sedan 1960-talet". I andra fall kan vi godta årtalet för det äldsta kända belägget men markera att ordet/betydelsen troligen är betydligt äldre med en skrivning av typen "åtm(instone) sedan 1971". Så gott som alla dessa problem uppstår vid ord eller betydelser som kommit in i svenskan relativt nyligen.

Fig. 2 visar årtal för första belägg och i ett par fall etymologier för ett antal ord med efterleden *-sång*.

minnesång 1/1	1848	till ty. *Minne 'kärlek'
ottesång 1/1	före 1520	fornsv. *ottosanger
vaggsång 1/1	1749	
folksång 1/1	1810	
kråksång 1/1	1843	

Fig. 2. Historiska uppgifter för ett urval ord med efterleden *-sång*

2.2 Konstruktionsuppgifter (valenser)

Alla verb i NEO — ca 9000 — har försetts med konstruktionsuppgifter av valenstyp. Det framgår alltså explicit om verben (eller de resp. betydelsemomenten av polysema verb) är intransitiva eller transitiva, och det ges även uppgift om ev. prepositionsobjekt, predikativ,

valensadverbial m.m. Om komplementen är fakultativa anges det med parentes. Det framgår också huruvida huvudorden i verbkomplementen är person- eller "sak"-betecknande ("ngn" resp. "ngt"), eller bådadera. Här nedan återges konstruktionsuppgifterna vid verben från *betunga* till *betyda*. (Frågan huruvida ett visst komplement hör till valensramen eller ej — med annan terminologi: om det är fyllnadsled eller tilläggsled — kan vara vanskelig att avgöra, men ett användbart testbatteri finns i Nikula 1986.)

betunga 1/1	& ngn/ngt (med ngt)
beltvinga 1/1	& ngn/ngt (med ngt)
betvivla 1/1	& ngt/att+SATS
betyda 1/1	& ngt/att+SATS
betyda 1/2	& ngt/att+SATS

Fig. 3. Konstruktionsuppgifter för verben *betunga*–*betyda*

Också adjektiven och substantiven har vid behov försetts med konstruktionsuppgifter. Här är problemet att dra gränsen mellan fyllnadsled och tilläggsled nog ännu svårare än vid verben. För adjektiven har dock en modell och ett par test utarbetats (se Malmgren 1987). Några av typerna kan exemplifieras med *artig* (*mot ngn*), *jämbördig* (*med ngn*) och *van* (*vid ngn/ngt*). En stor grupp substantiv som ofta får anses ta fyllnadsbestämningar är givetvis verbalsubstantiven, en annan grupp är "relationella" substantiv, t.ex. *bror* (*till ngn*). Av fig. 4 framgår konstruktionsuppgifterna vid ett antal adjektiv.

led 1/2	& (mot ngn)
ledlad 1/1	
ledbruten 1/1	
ledig 1/1	& (för ngn) (att + INF)
ledig 1/2	& (från ngt/ngn)
ledig 1/3	
ledsam 1/1	& (för ngn) (att + INF)

Fig. 4. Konstruktionsuppgifter för adjektiven *led*–*ledsam*

2.3 Ordled (morfologisk analys)

Uppslagsord med en flerstavig grundform eller böjningsform har försetts med en morfologisk analys, så när som på att morfem som inte innehåller en vokal inte beaktas. Denna analys innebär samtidigt en rekommendation om avstavning på konsekvent morfologisk grund i motsats till den blandning av mekanisk och morfologisk avstavning som hittills varit vanlig i

svenskan. (Ordboken är själv avstavad enligt denna morfologiska princip.) För att principen ska framgå så tydligt som möjligt analyseras vid substantiven och verben vanligen någon böjningsform som ofta har en stavelse mer än grundformen, i allmänhet bestämd form singular vid substantiven och presens vid verben. Sålunda t.ex. (i artiklarna **dag** resp. **jaga**): *dag-en* och *jag-ar* (eftersom *dag* och *jaga* inte skulle få några avstavningsmarkeringar).

En extra finess är att en stor morfemgräns anges vid sammansatta ord med hjälp av ett längre streck än divisen (t.ex. *buk-tal-ar-en*). Detta ger en extra upplysning om ordets struktur och samtidigt markeras därmed det bästa avstavningsstället. Att avgöra om sammansättning eller avledning föreligger, dvs. om långt och kort avstavningsstreck ska anges, är självfallet inte alltid lätt. I tveksamma fall har bl.a. ett betoningsskriterium använts. Sålunda har *geovetenskap* [jeˈ-] analyserats *geo-vet-en-skap-en* men *geografi* [-fiː] *geo-graf-in*.

Fig. 5 visar en skärmutskrift med ordledsanalyser av ett antal baklängessorterade ord (tecknet = står här för stor morfemgräns):

management 1	subst.	Ingen: böjning; n-genus:	manage · ment
logement 1	subst.	&et, plur. =, best. plur. &en	loge · ment · et
detachment 1	subst.	&et, plur. =, best. plur. &en	de · tache · ment · et
signalement 1	subst.	&et, plur. =, best. plur. &en	sign · ale · ment · et
element 1	subst.	&et, plur. =, best. plur. &en	ele · ment · et
bränsleelement 1	subst.	&et, plur. =, best. plur. &en	brän · sle=ele · ment · et
värmeelement 1	subst.	&et, plur. =, best. plur. &en	värme=ele · ment · et

Fig. 5. Ordledsanalyser av ett urval ord med slutleden -ment

2.4 Nomina actionis

Vid alla verb ges uppgift om möjliga nomina actionis, i första hand verbalsubstantiv på *-(n)ing* och *-ande/-ende*, men också på *-else*, *-tion* m.m. om dessa avledningar är vanliga. Ett viktigt krav är att avledningen fortfarande ska beteckna verbhandlingen. Har avledningen t.ex. fått en konkret betydelse tas den inte upp. Fig. 6 visar uppgifterna om verbalsubstantiv vid ett antal baklängessorterade verb.

mumifler\ a 1/1	-ade -at	-ande, -ing; mumifikation
teknifler\ a 1/1	-ade -at	-ande, -ing
saponifler\ a 1/1	-ade -at	-ande, -ing; saponifikation
personifler\ a 1/1	-ade -at	-ande, -ing; personifikation
unifler\ a 1/1	-ade -at	-ande, -ing
typifler\ a 1/1	-ade -at	-ande, -ing
verifler\ a 1/1	-ade -at	-ande, -ing; verifikation

Fig. 6. Uppgifter om nomina actionis av ett urval verb på -fiera

2.5 Uttal

Jämfört med SOB är uttalsangivelserna i NEO mycket explicita. Strävan har varit att förut-sätta enbart de allmännaste uttalsreglerna som bekanta för användaren (jfr dock strax nedan om bitryck). Den viktigaste egentligt nya uttalsinformationen i NEO gäller tonaccenten, om vilken uppgifter av någon anledning inte brukar ges i allmänna svenska ordböcker (undantag: inlärningsordboken *Svenska ord*). Här som i andra sammanhang gäller naturligtvis att det som ges är ett riksspråksuttal. Sålunda ges accent 2 som enda (riksspråks)uttal vid de flesta sammansättningar, trots att många dialekter kan ha accent 1-uttal. Däremot försöker vi täcka variation inom riksspråksuttalet, vilket bl.a. resulterat i att nära tusen ord fått både accent 1- och accent 2-markering.

Fig. 7 visar uttalsmarkeringarna i databasen för en grupp ord som ofta visar uttalsvariation (här är > markering för accent 1 och < för accent 2).

audlto<rlum	1	subst.			
dormlto<rlum	1	subst.			
kleresto<rlum	1	+subst.			
konserverlum	1	subst.	-to<r-	el.	-to>r-
konslstorlum	1	subst.	-to<r-	el.	-to>r-
konvltorlum	1	subst.	-to<r-	el.	-to>r-

Fig. 7. Uttalsuppgifter för ett antal ord på -orium

Viktigt är att ordboken redogör för oförutsägbara förändringar i uttalet från grundformen till böjningsformerna. Många ord har t.ex. accent 1 i grundformen och accent 2 i en eller flera av böjningsformerna. I de flesta fall går detta att beskriva med allmänna regler, men det finns en större undantagsgrupp: substantiv av 3:e deklinationen. Det vanligaste är här accent 2 i plural (*fil'mer*), men åtminstone i något hundratal fall förekommer accent 1 (*gre'ker*, *grän'der*). De senare fallen anges i NEO, och de allmänna reglerna ges i inledningen.

I ett avseende är uttalsangivelserna i NEO inte explicita: endast markering för huvudtryck ges, inte för bitryck. Inte heller ges alltid uttalsanvisningar för de stavelser som inte har huvudtryck och ev. är behäftade med uttalsvårigheter. Det är huvudsakligen sammansatta ord som på detta sätt inte får fullständiga uttalsmarkeringar. En anledning till detta var att det saknas en allmänt känd beteckning för bitryck i svenskan, en annan att fullständiga uttalsangivelser också skulle ha medfört krav på mycket vidlyftiga uppgifter om variantuttal (särskilt beträffande uttal av *o* i obetonad stavelse).

Med hjälp av ordbokens inledning och den morfologiska analysen bör dock även en icke-svensk användare av ordboken kunna komma fram till rätt uttal också av sammansatta ord. Antag att han slår upp artikeln **husbehov**, som har uttalsangivelsen [hu's-] och den morfologiska analysen hus-be-hov-et. Det långa strecket kan här fungera som en hänvisning till artikeln **behov**, och genom att summera uttalsuppgifterna på de båda artiklarna erhåller man — efter att ha förvandlat huvudtrycket på slutstavelsen i det självständiga ordet *behov* till bitryck på samma stavelse i efterleden *-behov* — det fullständiga uttalet av *husbehov*. Man måste dock också känna till regeln om accentflyttning i senare ledet av sammansättningar i

svenskan, men också denna regel är så gott som heltäckande (ord som *studiebidrag* uttalas t.ex. oftast med bitryck på sista stavelsen, trots att *bidrag* har huvudtryck på första stavelsen).

2.6 Hänvisningar mellan semantiskt besläktade ord

De flesta semantiska hänvisningarna är hänvisningar mellan kohyponymer (sammanlagt ca 30 000). Dessa hänvisningar markeras med en pil, riktad åt höger (se artikeln **abrupt** i fig. 1). Gränsen till vad som bedömts som synonymer är självfallet inte skarp, men allmänt kan sägas att vi valt att tolka synonymbegreppet ganska strängt; vi har alltså krävt mycket stor betydelselikhet för att ange synonymirelation. Motsvarande gäller för antonymirelationen.

Utom dessa relationer anges i många fall hyponymi- och hyperonymirelationer (med nedåtriktad resp. uppåtriktad pil). I artikeln **nederbörd** ges exempelvis hyponymerna *regn*, *snö* och *hagel*. I artiklarna om dessa ord ges visserligen inte motsvarande hänvisning till hyperonymen *nederbörd*, men detta framgår ändå indirekt genom att ordet *nederbörd* är huvudord i definitionerna. De totalt ca 35 000 hänvisningarna mellan semantiskt besläktade ord — i viss mån sålunda tillsammans med betydelsebeskrivningarna — torde sammantaget resultera i ett rätt finmaskigt semantiskt nätverk.

3 Avslutning: något om NEO:s målgrupp

De flesta av användarna av NEO kommer självfallet att ha svenska som modersmål. Men det är enligt min mening viktigt, inte minst med tanke på att Sverige numera är ett mångkulturellt land, att också se en stor enspråkig svensk ordbok som ett hjälpmedel vid avancerad svenskinläring. Det låter sig sägas, och sägs också ibland, att mycket av informationen i enspråkiga definitionsordböcker redan är känd för många av användarna, något som kan vara frustrerande för ordboksförfattarna (jfr Jonsson 1994). Om man, med en försiktig uppskattning, tänker sig åtminstone några tusen NEO-användare med annat modersmål än svenska, behöver man emellertid inte bekymra sig så mycket över eventuell redundans: då blir inte minst flera av de nya informationskategorierna (tonaccent, nomina actionis, ordledsanalyser, konstruktionsuppgifter) av särskilt värde. Då blir det också ytterst motiverat att skriva allmänna lexikonregler i inledningen — när man säger att ingen läser ordboksintroduktioner förbiser man rimligen denna grupp av användare. Och då kan man inte minst se etymologiuppgiften, i och för sig den uppgift som i många fall säkert har det största läsvärdet för svenska användare, som en ytterligare hjälp för icke-svenska användare att förstå ordets betydelse.

Litteratur

- Allén, Sture/Sven-Göran Malmgren 1993: Om Nationalencyklopedins ordbok. I: *Mänsklig mångfald*. Göteborg: Humanistiska utbildnings- och forskningsnämnden, Göteborgs universitet, 103–111.
- Jonsson, Hans 1994: Betydelsebeskrivningens förutsättningar i en enspråkig ordbok. I: *Lexico-Nordica* 1, 89–105.
- Malmgren, Sven-Göran 1987: *Om adjektiven i LPS-projektets lexikon*. Rapporter från Språkdata. 23. Göteborgs universitet: Institutionen för språkvetenskaplig databehandling.

NEO = *Nationalencyklopedins ordbok* 1–3. (utkommer 1995–96.) Höganäs: Bra Böcker.

Nikula, Henrik 1986: *Dependensgrammatik*. Malmö.

Nyord i svenskan = *Nyord i svenskan från 40-tal till 80-tal*. 1986. Utg. av Svenska språknämnden. Sthlm: Esselte Studium.

SOB = *Svensk ordbok*. 1986. Utarbetad vid Språkdata Göteborgs universitet. Sthlm: Esselte Studium.

Svenska ord = *Svenska ord — med uttal och förklaringar*. 1992. 2. uppl. Sthlm: Norstedts.

Söderwall, Knut Fredrik 1884–1918: *Ordbok öfver svenska medeltidsspråket* 1–3. Lund: Berlingska.

Nina Martola

Från enspråkig till tvåspråkig ordbok

Compiling a bilingual dictionary on the basis of a monolingual one means enormous advantages. There are, however, also problems attaching to it. When the languages involved are two as structurally different languages as Finnish and Swedish, there are probably more problems than if the languages are structurally close to each other. This article describes some of the problems we have had on the macrostructural level during our work on a new large Finnish-Swedish dictionary.

1 Inledning

Redan när det i mitten av 80-talet fattades beslut om att en ny finsk-svensk storordbok skulle utarbetas bestämdes det att det skulle bli en helt ny ordbok, inte en revision av den existerande och länge och flitigt använda storordboken av Cannelin, Hirvensalo & Hedlund 1976. Den nya ordboken skulle bygga på den enspråkiga finska ordboken *Suomen kielen perussanakirja* (Finsk basordbok), i fortsättningen *Perussanakirja*. Denna höll på att utarbetas vid Forskningscentralen för de inhemska språken och skulle enligt planerna omfatta 100.000 uppslagsord och bestå av tre band.

Arbetet med *Finsk-svensk ordbok* påbörjades vid Forskningscentralen som ett enmansprojekt, men efter det att förlagshuset WSOY gått med i projektet kunde arbetsstyrkan från och med 1990 utökas till fyra personer, varav två var anställda av Forskningscentralen och två av WSOY. Senare har ytterligare två deltidsanställda medarbetare tillkommit.

Nu, hösten 1995, är alla artiklar skrivna och vi håller på med slutredigeringen av manuskriptet. Vi kan alltså sammanfatta våra erfarenheter av att utarbeta en tvåspråkig ordbok utgående från en enspråkig. En del av våra erfarenheter äger förmodligen ganska stor allmängiltighet, andra är språkspecifika och vissa hänger naturligtvis samman med egenskaperna hos just den ordbok vi har haft som bas. Jag kommer i det följande att först kort ta upp vilka fördelarna har varit och sedan övergår jag till att redogöra för en del problem av makrostrukturell art som vi har haft.

2 Fördelar

Det innebär utan vidare enorma fördelar att ha en enspråkig ordbok som underlag när man utarbetar en tvåspråkig. Det är framför allt oerhört arbetsbesparande.

Lemmasektionen är gjord och huvudformer och biformer har fastslagits. Vi har dessutom haft förmånen att arbeta utgående ifrån en nyutkommen ordbok, vilket gör att lemmauppsättningen är aktuell (även om vissa nyord hunnit tillkomma och borde fås med).

Betydelseanalysen är gjord och idiomatiska uttryck, ordfogningar, talesätt, ordspråk, citat o.s.v. som är viktiga och relevanta i källspråket finns färdigt samlade. Stilvärde och fackområdestillhörighet finns färdigt bestämda.

En mycket handgriplig fördel för vår del var att *Perussanakirja* parsades och matades in i vår artikelstruktur, vilket gjorde att vi har sluppit att skriva in en massa finsk text. (Ett sådant tillvägagångssätt är naturligtvis det normala i våra dagar.)

Även om det finns ett beslut om att en enspråkig ordbok skall ligga till grund för en tvåspråkig är det naturligtvis skäl att fundera över hur bindande den färdiga lemmauppsättningen skall anses vara och vilka skäl det finns för eventuella justeringar.

3 Makrostrukturella problem och överväganden

3.1 Målgrupp och användningssyfte

Ett faktum som kan göra det befogat med vissa justeringar i lemmalistan är att målgruppen för den enspråkiga och den tvåspråkiga ordboken inte är densamma eller åtminstone bara delvis densamma. Användarens syfte när han slår i ordböckerna är knappast heller detsamma. Här talar jag om den eller de målgrupper som lexikografen har haft i åtanke under arbetet; att ordböckerna sedan kan användas av många andra och i andra syften säger sig självt.

I en enspråkig ordbok av *Perussanakirjas* typ slår användaren upp för att få information om sitt modersmål. Han eller hon vill kontrollera betydelser, formuleringar, användningsområde eller språkriktighet. *Perussanakirja* är både deskriptiv och normativ och den riktar sig både till yrkesskribenter och gemene man.

Finsk-svensk ordbok skall bli en produktionsordbok, och den riktar sig i första hand till finskspråkiga som skall producera text på svenska, antingen genom att de själva skriver på svenska eller för att de översätter egna eller andras texter från finska. I andra hand försöker vi beakta även svenskspråkiga översättares behov. Modern sakprosa är den texttyp som våra användare främst kan väntas producera, även om andra stilarter kan komma i fråga ibland. Skönlitterära översättare kommer däremot knappast att utnyttja ordboken i någon högre grad.

Det här betyder att t.ex. ålderdomliga ord, svåra (främmande) ord, ovanliga ord och onomatopoetiska ord i viss mån kan uteslutas till förmån för t.ex. sammansättningar. Mera om sammansättningarna längre fram.

Det som dominerar det dagliga lexikografiska arbetet är dock inte de målgruppsbaserade övervägandena utan de språkliga och kontrastiva frågorna.

3.2 Makrostrukturen i förhållande till mikrostrukturen

Lemmatiseringen måste naturligtvis ske i enlighet med den tilltänkta målgruppens sätt att tänka (i den mån man kan anse sig veta något om den saken), alltså utifrån hur användaren i första hand kan tänkas slå upp det ord han letar efter. Eftersom *Perussanakirja* är gjord av finskspråkiga för finskspråkiga har vi räknat med att den i grunden fungerar väl i det avseendet (även om enskildheter alltid kan ifrågasättas).

Men när det gäller makrostrukturen i en tvåspråkig ordbok kommer käll- och målspråk i en del fall i konflikt. Det som är en naturlig lemmatisering på källspråket fungerar inte alltid när ett annat språk, målspråket, skall beaktas. Det uppstår ibland systemluckor, som kan bli ganska påfallande. En strikt lemmatisering enligt källspråket kan också ge upphov till mindre naturliga eller teoretiska ekvivalenter på målspråket. Sådana kan möjligen i

viss mån accepteras om målspråket är användarens modersmål men borde definitivt inte få förekomma i en ordbok där målspråket är det främmande språket för användaren.

En principiell fråga blir därför, om man inte i en tvåspråkig ordbok kan fylla på med fler ingångar som faller sig naturliga ur målspråkets synvinkel — d.v.s. som ger så naturlig ekvivalens som möjligt — och sedan hänvisa till dessa uppslagsord från de källspråkliga huvuduppslagsorden.

3.2.1 Ordbildning, lemmatisering och lemmaluckor

Ett exempel på hur det uppstår luckor i lemmasystemet är particip II (ändelsen *-nut/-nyt*¹). Participformerna kan användas adjektiviskt liksom svenskans perfekt particip (som de i hög grad motsvarar) och de kan utgöra grundord för substantivavledning. I *Perussanakirja* behandlas participen, med undantag för de starkt lexikaliserade, som böjningsformer av verben och lemmatiseras alltså inte. Substantivavledning till participen lemmatiseras däremot alltid om de tas upp, eftersom ordklasskifte är ett kriterium för lemmatisering. Det här innebär att en substantivavledning kan finnas med som uppslagsord utan att grundordet finns med, något som verkar ganska påfallande i en tvåspråkig ordbok.

Tabell 1 nedan visar hur *Perussanakirja* har lemmatiserat ett antal substantivavledningar med tillhörande ordfamiljer. Kryssen i den första kolumnen anger om själva avledningen finns med som lemma, kryssen i den andra om participet, alltså grundordet till avledningen, finns med, kryssen i den tredje om infinitiven av verbet finns lemmatiserad och i den fjärde och femte kolumnen slutligen om motsvarande karitiva bildningar har lemmatiserats. Tecknet Ø anger lemmaluckor.

Orden är hämtade ur manuskriptet till *Finsk-svensk ordbok*. Det är ord som har använts som förklaringar till ekvivalenter i det material som har lästs av den finska språkgranskaren.

Vi ser att det är få av substantivavledningarna som är lemmatiserade i *Perussanakirja*. Förklaringen till det är att det i hög grad handlar om ord som är resultatet av produktiva ordbildningsregler och inte om lexikaliserade former.

Men användaren av en tvåspråkig ordbok skulle säkert undra varför substantivet *perehtyneisyys* 'förtrogenhet' finns med som uppslagsord men inte dess grundord *perehtynyt* 'bevandrad, insatt; förtrogen'. Det verkar också inkonsekvent att ge det karitiva adjektivet *turmeltumaton* 'oförstörd, ofördärvad' med därtill hörande substantivering *turmeltomattomuus* men däremot inte de icke-karitiva paret *turmeltunut* och *turmeltuneisuus*. Dessutom blir luckorna problematiska också såtillvida, att det i regel existerar betydligt fler och vanligare ekvivalenter till grundorden än till avledningarna.

I vissa fall ställer jag mig också tvivlande till om den lemmatisering tabellen åskådliggör verkligen följer graden av lexikalisering. Jag frågar mig till exempel varför inte *pettynyt* 'besviken' och *pettynisyys* 'besvikelse' finns med som lemmen om en gång *harjauntu-neisuus* 'vana; det att vara van' och *kiihottuneisuus* 'upphetsning' finns med. Möjligen skulle lemmatiseringen ha sett något annorlunda ut om *Perussanakirja*-redaktionen under hela sitt arbete hade haft tillgång till arbetsredskap som hade möjliggjort systematiska och strukturella sökningar av olika slag.

¹Växlingen *-nut/-nyt* och *-neisuus/-neisyys* beror på finskans vokalbalans.

		part.II -nut	inf.	karit. adj.	karit. subst.
hajaantuneisuus			x		
haltioituneisuus			x		
harjaantuneisuus	x	∅	x	x	
hermostuneisuus	x	x	x		
huolestuneisuus	x	∅	x		
hämmentyneisyys			x		
ihastuneisuus			x		
ikävyestyneisyys	x	x	x		
kiihottuneisuus	x	∅	x		
kiivastuneisuus			x		
kuihtuneisuus		x	x		
kyllästyneisyys			x		
lannistuneisuus			x		
masentuneisuus	x	x	x		
nöyrytyneisyys			x		
perehtyneisyys		x	∅	x	
pettyneisyys			x		
pingottuneisuus			x		
puutuneisuus			x		
päihtyneisyys		x	x		
repeytyneisyys			x		
ruskettuneisuus			x		
tottuneisuus	x	∅	x		
turmeltuneisuus	∅	∅	x	x	x
vaivautuneisuus			x		
väljähtyneisyys			x		

Tabell 1. Lemman i Perussanakirja.

Ett problem vid lemmatiseringen är naturligtvis också att det inte finns en otvetydig gräns när en participform har blivit så lexikaliserad att den kan anses utgöra ett eget ord, alltså när den har övergått från att vara en verbform till att vara ett adjektiv. I en enspråkig ordbok som är en beskrivning av språket är det naturligt att hålla samman sammanhörande former. I en tvåspråkig ordbok blir det klarare om man ger adjektiviska ekvivalenter (adjektiv och particip) i en egen artikel och inte tar upp dem under verbet. Dessutom är det inte ovanligt att det finns betydligt fler ekvivalenter till participet än till själva verbet.

Till det intransitiva verbet *sykähtää* 'slå' (om ett enstaka slag av hjärta eller puls) finns en kausativ bildning *sykähdyttää*, vars betydelse, om man ser till bildningen, är 'få (ngns hjärta) att slå ett slag'. Det finska verbet används dock enbart överfört och i *Finsk-svensk ordbok* ges ekvivalenterna *gripa* och *frappera*. I *Perussanakirjas* artikel finns det tre exempel som alla ger prov på användningen av particip I av verbet, alltså former av *sykähdyttävä*. I *Finsk-*

svensk ordbok har participet lemmatiserats.² De tre *Perussanakirja*-exemplen tas upp i den artikeln medan exempelmaterialet till verbartikeln har fått sökas på annat håll.

sykähdyttäv|ä¹⁰ gripande; hjärtgripande; **hurmaava** hänförande; bedårande; betagande; **komea** stilig; flott; tjusig; **näky oli** ~ synen var betagande (hänförande, bedårande); *se on -än kaunis* den är bedårande (betagande) vacker; den är så vacker att det gör ont; den är så vacker att man tappar andan (att det tar andan ur en); *ottelun -in maali* matchens snyggaste (stiligaste) mål

3.2.2 Onaturliga ekvivalenter

I finskan finns en stor grupp intransitiva verb på *-ua/-yä*. Verben anger i regel att en handling av något slag försiggår av sig själv. De motsvaras i vissa fall av mediala *s*-former i svenskan (Andersson 1993:156):

<i>muuttua</i>	förändras
<i>kuulua</i>	höras

De mediala *s*-verben är dock fåtaliga jämfört med den stora mängden finska verb och en betydligt vanligare ekvivalenttyp är *bli* + *particip* eller *bli* + *adjektiv*:

<i>ilahtua</i>	bli glad
----------------	----------

Men för många av de finska verben fungerar en konstruktion med *bli* otillfredsställande eller inte alls.

Verbet *tallentaa* 'spela in, banda; spara; lagra (adb); skriva ner osv.' har en intransitiv motsvarighet *tallentua*. Den artikeln ser i skrivande stund ut så här i *Finsk-svensk ordbok*:

tallentu|a^{52*J} **tulla tallennetuksi nauhalle** spelar¹ in, bli⁴ inspelad; bandas¹, bli⁴ bandad; **tulla kirjoitetuksi muistiin** tecknas¹ ner, bli⁴ nertecknad, tecknas¹ upp, bli⁴ upptecknad; **säilyä** bevaras¹; **atk: tulla tallennetuksi levykkeelle** sparas¹, bli⁴ sparad; **säilyä elektronisessa muodossa** lagras¹, bli⁴ lagrad; *tapahtumat ovat -neet filmiin* händelserna finns bevarade (inspelade) på film

Här har både artikelförfattaren och de som läst artikeln missat en viktig omständighet, nämligen att det finska verbet uttryckligen avser ett skeende som försiggår av sig självt. De föreslagna ekvivalenterna är däremot att betrakta som passiva former, där man underförstår en agent. De motsvarar med andra ord passiver av det finska verbet *tallentaa*. Schematiskt ser det ut så här:

Ääni ei tallentunut	≈ Ljudupptagningen misslyckades (p.g.a. ett tekniskt fel eller någon annan omständighet)
Ääntä ei tallennettu	= Ljudet spelades inte in (planenligt; man behövde bara bilderna)

Tabell 2.

²Kodsiffrorna som förekommer i artiklarna ur *Finsk-svensk ordbok* är dels böjningskoder för uppslagsorden och dels böjningskoder för svenska verb.

I vissa fall kan de svenska *s*-formerna fungera bra som ekvivalenter, trots att de i strikt bemärkelse är att betrakta som passiva former och inte som mediala *s*-former:

kokous lykkäntyi
mötet sköts upp

Eller med resultativa passivformer:

mötet är uppskjutet
mötet blev uppskjutet
(beroende på när utsagan görs)

Jag har inte undersökt vårt material i det här avseendet, men jag skulle tro att förklaringen till hur väl passiverna fungerar ligger i den grad av aktivitet som verbet uttrycker. Att skjuta upp ett möte kräver i regel mindre aktiva insatser än att göra en inspelning. Det finns säkert också individuella skillnader mellan olika talare för i vilken grad passiva former kan accepteras.

I en del fall är *s*-formen en så entydigt passiv att den definitivt inte kan komma i fråga som ekvivalent till ett *-ua/-yä*-verb. Ett ännu flagrantare fall än *tallentua* är *metsittyä*:

niitty on metsittynyt
ängen har vuxit igen
det har vuxit upp skog på ängen

Den finska satsen kan inte översättas med "ängen har beskogsat", eftersom det betyder att någon aktivt har planerat skogen.

Artikeln *tallentua* måste revideras. En elegant lösning ur svenskans synvinkel vore att behandla själva infinitivartikeln mycket kort och göra ett sublemma av participet där ekvivalenter ges:

tallentu|a ääni -i hyvin ljudupptagningen lyckades bra
tallentu|nut nauhalle inspelad; bandad; upptagen på band; **filmille** filmad;
muistiin kirjoitettu nertecknad, upptecknad; **säilynyt** bevarad; **atk: tiedostosta**
sparad; **informaatiosta** lagrad; **filmiin -neet tapahtumat** händelser som finns
bevarade (inspelade) på film; filmade händelser

De svenska ekvivalenterna är ju fortfarande former av aktiva, transitiva verb, det har vi inte kommit ifrån, men motsvarigheten är betydligt bättre, eftersom det nu är det resultativa som står i förgrunden och inte själva skeendet.

Den ovan föreslagna lösningmodellen har vi inte använt oss av i *Finsk-svensk ordbok*, vilket betyder att arbetet med denna typ av artiklar blir en ständig balansgång mellan önskan att förse användaren med svenska motsvarigheter och obenägenheten att ge ekvivalenter som är grammatiskt korrekta men som inte som sådana uppträder i svensk text.

Ett annat skäl till att lemmatisera participen i en finsk-svensk ordbok är att ekvivalenterna utgör stommen i både *vara*- och *bli*-konstruktioner och att man därigenom kan hänvisa till particippartikeln från andra med den sammanhängande artiklar och därigenom slippa en hel del upprepning. Dessutom är vissa av de svenska participen mycket allmännare än sina motsvarande infinitiver (t.ex. *förbluffad* vs. *förbluffa*).

Så här ser t.ex. gruppen kring *masentunut* ut:

- masennus** [...] *olla masennuksissa[an]* vara⁴
deprimerad **jne. ks. masentunut**
- masentaa** göra⁴ ngn deprimerad **jne. ks.**
masentunut
- masentua** bli⁴ deprimerad **jne. ks. masentunut**
- masentunut** deprimerad; **puhek.** deppig, deppad . . .
osv.

I stället för att upprepa alla ekvivalenterna i alla de fyra artiklarna ger man dem på ett ställe och hänvisar från de övriga. Förkortningen *jne.* betyder 'och så vidare'.

I *Perussanakirja* finns *masentunut* med som uppslagsord, därav denna uppställning. Min åsikt är att samma sak borde göras med många andra liknande grupper där *Perussanakirja* inte har lemmatiserat participet.

3.3 Förutsägbarhet

Traditionellt har arbetsfördelningen mellan grammatik och ordbok ansetts vara att ordboken skall redovisa det speciella medan det systematiska är grammatikens uppgift. Som tvåspråkslexikograf ställs man dock inför problemet att det inte råder något ett till ett-förhållande mellan det systematiska i de två språken.

3.3.1 Adverb

I finskan bildas adverb genom att ändelsen *-sti* fogas till adjektiv- eller participstammen. På svenska bildas adverb på motsvarande sätt genom att *-t* tillfogas.

Perussanakirja lemmatiserar adverb, medan *Finsk-svensk ordbok* tillämpar principen att adverb på *-sti* lemmatiseras bara om de får icke förutsägbara ekvivalenter, alltså inte bildningar på *-t* utan adverb av andra typer (*möjligen; naturligtvis; med fart* etc.). Det här är i och för sig en bra princip, men den kan vara nog så besvärlig att försöka efterleva i det praktiska arbetet.

För det första finns det en mycket väsentlig skillnad i funktion mellan finska adverb på *-sti* och svenska på *-t*, nämligen att de finska inte kan vara bestämningar till adjektiv eller till andra adverb, vilket de svenska kan, utan enbart till verb. Det är ju i och för sig också en systematisk skillnad, men det finns ändå återverkningar på lexikonets nivå. Man blir som lexikograf nämligen ofta tvungen att försöka hitta belägg för att en svensk *-t*-form faktiskt existerar som adverb och inte bara som neutrumform av adjektivet. Om adverbet existerar måste man också undersöka om det kan förekomma som bestämning till verb. Förekommer ett adverb bara som bestämning till adjektiv eller adverb kan det inte vara ekvivalent till ett adverb på *-sti*.

För det andra ställs man vid adjektiv-adverb-par ofta inför problemet att de flesta men inte alla ekvivalenterna till ett adjektiv har en motsvarande adverbform. I synonymraden *väldig, kolossal, enorm, jättelik* saknar det sista adjektivet adverbliell motsvarighet. Följer man den ovan anförda principen noggsamt borde sådana här fall redovisas, annars implikerar

man att det kan bildas adjektiv till samtliga adjektivekvivalenter. Med andra ord borde alltså adverbet lemmatiseras. I praktiken har vi nog tummat på principen, eftersom den kommer i konflikt med behovet att spara utrymme.

För det tredje finns det vissa verb som tar adverbiell bestämning på finska, medan deras svenska motsvarigheter tar predikativt attribut. Det finska adverbet *nuivasti* t.ex. motsvaras av svenska adverb som *kyligt*; *avvisande*; *kallsinnigt*. Men i exemplet *suhtautua nuivasti* (som är det exempel som *Perussanakirja* ger) skall det inte vara adverb i översättningen utan det heter *förhålla sig (ställa sig) kallsinnig*. Det mest ekonomiska vore att redogöra för denna satstyp en gång för alla i verbartiklarna, alltså *suhtautua* i detta fall, men å andra sidan är *förhålla sig kallsinnigt* en frekvent feltyp i översatta texter, och vi har därför strävat efter att ge exempel av detta slag, i all synnerhet som det är betydligt sannolikare att en användare slår upp *nuivasti* än att han slår upp *suhtautua*, som är ett ganska vanligt verb.

För det fjärde aktualiseras adverb ibland i finskan på grund av att de bestämmer attributivt använda verbformer i fall då svenskan har andra konstruktioner:

myrkyllisesti vaikuttava aine

*ett giftigt verkande ämne

ett giftigt ämne

vauhdikkaasti suoritettu esitys

?en fartfylld genomförd föreställning

en fartfylld föreställning

en föreställning med fart

3.3.2 Sammansättningar

Jag har vid rätt många tillfällen hört argumentet att självklara sammansättningar inte behöver tas med i en ordbok. I mindre ordböcker dikteras principen naturligtvis i hög grad av utrymmesbrist och är såtillvida förståelig, men särskilt användarvänlig är den inte i en produktionsordbok, därtill är andelen icke-förutsägbara sammansättningar alltför stor. Användaren kan ju aldrig veta hur det förhåller sig med den sammansättning han letar efter och inte hittar. Saknas den för att den är självklar eller för att ordboksredaktionen inte har råkat komma att tänka på den eller rent av för att den är så pass ovanlig att den inte kommit med av den anledningen, trots att den inte är förutsägbar?

Den språkmedvetna användaren har säkert upptäckt att det finns en mängd icke-förutsägbara sammansättningar. Hur kommer det sig till exempel att både *astmaanfall* och *astmaatack* är rätt medan *hjärtanfall* är fel? Varför heter det ibland *gat-* och ibland *gatu-* och när skall det vara *skog-* och när skall det vara *skogs-*?

En *älgskytt* jagar älg. En *tjuvskytt* jagar däremot inte tjuvar utan han jagar i smyg. Men vad gör en *tjuvjägare* på *tjuvjakt* egentligen?

Ännu mer komplicerat blir det då finska och svenska sammansättningar kontrasteras mot varandra.

Vi har *kirjoituspöytä* – *skrivbord*, *ruokapöytä* – *matbord*, *sohvapöytä* – *soffbord* men *tiskipöytä* – *diskbänk*.

Ordet *salaluukku* skall översättas med *fallucka* om det avser en lucka i golvet men med *lönnlucka* om det avser en lucka i väggen.

Sammansättningar på *meri-* motsvaras antingen av *sjö-* eller *havs-* och i vissa fall av en nominalfras med adjektivet *maritim*.

Förleden *moni-* motsvaras ibland av ekvivalenter på *fler-* och ibland på *mång-* och ibland av bägge två. I vissa fall skall det vara *multi-* och i vissa *poly-* och i en del fall skall det vara helt andra konstruktioner.

Ett urval av ord på *naaras-* visar ytterligare en typ av variation:

<i>naaras eläin</i>	hondjur
<i>-hirvi</i>	älgko
<i>-kissa</i>	honkatt; katta
<i>-koira</i>	tik
<i>-leijona</i>	lejonhona; lejoninna
<i>naaraspontti</i>	not, nåt (en typ av fog)

Det är naturligtvis omöjligt att redovisa alla de sammansättningar som kan tänkas förekomma i så sammansättningsrika språk som finskan och svenskan, men att redovisa en hel del sammansättningar, däribland också självklara, ingår enligt vår åsikt i en storordboks uppgifter.

4 De svenska ordböckerna som källor för målspråket

Jag skall avslutningsvis kort sammanfatta våra erfarenheter av de enspråkiga svenska ordböckerna som källor för beskrivningen av målspråket. Det finns vissa typer av information som vi ibland har letat förgäves efter och som antagligen saknas eftersom behovet av den aktualiseras när svenskan kontrasteras mot andra språk. En enspråkig ordbok skulle förmodligen vinna en hel del i explicitet om man gjorde kontrastiva jämförelser vid utarbetandet. Den information som skulle tillföras på så vis kunde säkert också modersmålsanvändarna ha nytta av.

Information som inte finns systematiskt redovisad i svenska ordböcker är:

- Om ett adverb kan bildas till ett adjektiv med hjälp av ändelsen *-t* (SAOL är bra men helt genomförda är angivelserna inte.)
- Om ett substantiv på *-het* kan bildas till ett adjektiv. För denna information har Språkdatas konkordanser i Göteborg varit den överlägset bästa källan.
- Om substantiv på *-het* anger enbart egenskap eller om de också kan avse det som så att säga är resultatet av egenskapen. Man kan *prata dumheter* och *ägna sig åt futtigheter* men kan man *putsa bort en solkighet*?
- Om en form på *-andel-ende* är enbart particip eller om den också är substantiv.
- Om ord på *-andel-ende* (och också andra verbalsubstantiv, t.ex. av typen *vrål, yl*) kan avse en punktuell handling eller om de har enbart oavslutad aktionsart.

- Om ett substantiv på *-[n]ing* är bildat till ett transitivt eller till ett intransitivt verb eller till båda (t.ex. *förändring*).
- Om avledningarna på *-[n]ing* är processuella eller resultativa eller både och. Om *limning* t.ex. kan betyda dels 'det att limma' och dels 'fog' hur är det med ord som t.ex. *klistring*, *sammansmältning* och *förtrollning*?
- För vilka av ett verbs eller ett adjektivs betydelser substantiveringar kan komma i fråga om det finns flera betydelsemoment.
- Om ett transitivt verb har en parallell medial *s*-form eller inte.
- Aktionsarten för *s*-former och reflexiva verb, alltså om handlingen är punktuell/avslutad eller durativt/iterativ. *Svensk ordbok* förklarar t.ex. *reta sig* under uppslagsordet *reta* med 'bli irriterad' och ger ett exempel *hon retade sig på hans sätt*. Den förklaringen tolkar jag som att *reta sig* avser en enstaka handling, något som strider mot min egen språkkänsla, som säger mig att *reta sig* används om upprepad eller ständig irritation.
- Om det vid en kombination av verb plus preposition handlar om ett enkelt verb som tar prepositionsobjekt eller om ett partikelverb med direkt objekt eller om båda alternativen rent av existerar. (Se närmare Cantell 1995.)
- Jag skulle helst se att det fanns valensuppgifter systematiskt angivna för både verb, adjektiv och substantiv. Om flera prepositioner är möjliga tillsammans med ett verb borde eventuella betydelskillnader beskrivas. Den inbördes ordningsföljden mellan olika synonyma prepositionsalternativ borde helst basera sig på något slags vanlighetskriterier. I en del fall går angivelserna i ordböckerna nu stick i stäv med antalet faktiska förekomster i konkordanserna.
- Stilvärdet kan ju inte mätas enligt någon entydig, objektiv skala och är därför vanskligt att fastslå, men jag tycker ändå att det är förvånansvärt ofta som de svenska källorna är oense. Vardagligt markerade ord i en källa står omarkerade i en annan och tvärtom. I en del fall har ett och samma ord etiketterats som vardagligt i en ordbok och som "något ålderdomligt", "något högtidligt" el.likn. i en annan (och jag talar nu om källor från samma tid, slutet av 80-talet).

En långt driven explicitet kan kanske rent av bli av ondo, eftersom språk som används ständigt lever och ändras och eftersom ett betydande mått av vaghet och subjektivitet ju hör till språkets natur. Det är väl det som gör det så intressant och utmanande — och samtidigt egentligen också omöjligt — att göra ordböcker. Men i och med att det utvecklas allt fler moderna och effektiva hjälpmedel i form av konkordanser, databaser och effektiva sökprogram blir det kanske litet mindre omöjligt i framtiden.

Litteratur

Ordböcker

Cannelin, Knut & Aulis/Lauri Hirvensalo/Nils Hedlund 1976: *Finsk-svensk storordbok*. Helsingfors: WSOY (4:e oförändr. upplagan 1986).

Perussanakirja = *Suomen kielen perussanakirja* [Finsk basordbok] 1990–94. I–III. Päätoimittaja [huvudredaktör]: Risto Haarala. Helsinki: Painatuskeskus.

SAOL = *Svenska Akademiens ordlista över svenska språket* 1986. Stockholm: Norstedts.

Svensk ordbok 1986. Utarbetad vid Språkdata, Göteborgs universitet. Stockholm: Esselte Studium.

Annat litteratur

Andersson, Erik 1993: *Grammatik från grunden*. Uppsala: Hallgren & Fallgren Studieförlag AB.

Cantell, Ilse 1995: Målspråkets verbkonstruktioner i en tvåspråkig produktionsordbok. I: *Lexico-Nordica* 2 (i tryck).

Ordbøker og folkelivsgransking (etnologi)

One of the fields interest for *ethnology* is the study of different aspects of labour: tools, equipment, and the work-process in its widest sense. Each field of work, each type of tool, and each method of operation has its own words or technical terms; and for the writer of dictionaries it is of importance to include lexical items, with appropriate semantic descriptions, from this field.

The present paper focuses on the following topics:

1. A brief attempt at drawing a distinction between every day vocabulary and technical terminology (following Coseriu 1973).
2. Examples of how Ivar Aasen (1850, 1873 and 1925) handled vocabulary of this type, illustrated by words/technical terms grouped under the heading "tools", "house", and "cleanliness/cleaning".
3. Aasen's lexical items and technical terms are compared to information from Norwegian ethnological literature, by Eilert Sundt, a contemporary with Aasen.
4. What can the writer of dictionaries learn from ethnological literature?

1 Innleiing

Folkelivsgransking — som ein tidlegare avgrensa det — er m a å granska arbeidsliv og alt som har med det å gjera: reiskap, utstyr, arbeidsgang og skikk og tru. Kvar arbeidsgrein og reiskap, kvar rådgjerd og åtgjerd har eigne ord eller nemne (Lid 1933). Ordbokskrivaren må få med slike ord og gjeva fullnøyande semantiske opplysningar om dei.

Dette innlegget tek opp fylgjande emne:

1. Freistnad på kort grenseoppgang mellom vanleg ordtilfang og terminologi (etter Coseriu 1973).
2. Døme på kva Ivar Aasen gjorde med slikt tilfang i ljøs av nokre ord eller termar / nemne for "reiskap" og "hus", "reinsemd" og "reinhold" — og so samanlikna desse nemna med tilsvarende tilfang i norske skrifter frå samtida om folkelivsgransking av Eilert Sundt. Berre i liten mon har eg nytta andre kjelder.

1.1 Kjelder

1.1.1 Ordbøker over nynorsk av Ivar Aasen

Den fyrste ordboka av Ivar Aasen (Aasen 1850), inneheld kring 25 000 oppslag; den andre (Aasen 1873) har kring 35 000 og dertil mange ufullnøyd opplysningar til ord som var med i 1850. I tillegg til desse to alfabetiske ordbøkene laga han ein tesaurus over norsk ordtilfang (Aasen 1925).

1.1.2 Skrifter av Eilert Sundt

Den fyrste norske etnologen heitte Eilert Sundt (1817–75). 1850–73 sende han ut innpå 20 bøker eller store utgreiingar, i alt kring 8000 prenta sider. Skriftene frå kystliv og fiske inneheld mange ord og nemne på båtar, båtdelar og fiskarliv. Andre skrifter har mange ord og nemne for reinsemdstall, hus og innbu (Sundt 1974–78 og Folkev).

1.2 Ordutval og oppsetjingsmåte

Her har eg berre teke med ord som har interesse for terminologisk gransking og for gjenstands-gransking. Utvalet byggjer på Aasen (1850, 1873 og 1925) — og på utskrivni ordtilfang or Sundt-skriftene i ArkivNO.

Oppslaga er attgjevne i nynorsk skriftform. I sitat er stor førebokstav i substantiv utbytt med liten og *aa* med *å*. Teiknet “—” framfor ei ordkjelde tyder at ordet, nemnet eller ordlaget vantar der. Eigne utfyllingar til å klårgjera samanhengen stend i hakeparentes: “[]”. Til å letta oversynet vert tilfanget stundom oppsett i spalter; definisjonar etter A stend til venstre, og utskurdar etter Sundt som kan tena til definisjonar stend til høgre. Døme og opplysningar frå Sundt og Aasen er oftast ordrett siterte — men stundom i samandregen form. Under einskilde døme gjev eg merknader og prøver å syna kva som sermerkjer Sundt, kva han har teke med, og kva som kan vanta hjå Aasen. Oftast har det vore nok å sitera definisjonar og uttydingar i Aasen (1873), stundom i Aasen (1850) au, og desse har eg samanlikna med Sundt. Etter å ha drøft eit nemne prøver eg stundom å forma ein definisjon på grunnlag av dei opplysningane som ligg i tilfanget.

Eg skriv nynorsk, og istadenfor “term” og “terminologi” nyttar eg helst *nemne*; for dansk “betegne”, tysk “bezeichnen” nyttar eg *merkja* og for “betegnelse” *merkjing*; for “betydning” nyttar eg *tyd* og *tyding*.

2 Ord og nemne

Det er vanleg å skilja mellom ord og nemne. Nemne eller termar høyrer i stor mon heime i fagmål, og slikt tilfang er knytt til fag eller emne. Nemna er ofte vanlege ord, nytta i ein slik samanheng at det å tolka dei krev meir innsikt i det emnet eller faget orda høyrer heime i enn kunnskap om språket i trongare meining. Om dette skreiv Coseriu (1973:26):

So dienen die Terminologien in der Regel zur Benennung von Verschieden, die nicht mehr durch die erste sprachliche Intuition, sondern erst durch wissenschaftliche Beobachtung festgestellt werden können.

Seinare har Coseriu (1988:270) peikt på det at denne sida av språket ber med seg ein tradisjon om sakkunnskap, og dette gjeld ikkje berre for vitskaplege og tekniske språk i akademisk samanheng, men òg for eit meir folkeleg eller populærvitskapleg og populærteknisk leksikon; slike nomenklaturar er gjerne knytte til fagmiljø, og folk utanfor slike miljø kan greia seg utan den innsikt som fagfolk må ha.

Å granska denne luten av ordtilfanget er kulturgransking òg Edward Sapir (1958:90–91) ordla det soleis i 1912:

It is the vocabulary of a language that most clearly reflects the physical and social environment of its speakers. The complete vocabulary of a language may indeed be looked upon as a complex inventory of all the ideas, interests, and occupations that take up the attention of the community, and were such a complete thesaurus of the language of a given tribe at our disposal, we might to a large extent infer the character of the physical environment and the characteristics of the culture of the people making use of it.

Dansken H. F. Feilberg (1910–14:II) tenkte i liknande lei:

Det var fra begyndelsen af min tanke ej blot at søge og muligst nøjagtig gengive ord og ordformer, men bruge dem i de enkelte artikler som et midtpunkt hvorom folkets tænkemåde gennem eksempler fandt et nogenlunde fyldigt udtryk.

Ivar Aasen (1873:XI) tykte det var innfløkt å uttyda slike ord.

Det nedarvede landssprog støtter sig til det fortroligste bekjendtskab med naturen og den simpleste skik i arbeide, huusholning og levemaade. [...] Almuens Tale vil altså have mange gode benævnelser for adskillige ting, som ere lidet bekjendte blandt dem, som mest befatte sig med læsning og skrivning [...] Derfor kommer man også idelig i forlegenhed, når man skal oversætte ord, som vedkomme arbeidet og den landlige indretningsmåde, så at man ofte her må give en lang forklaring over en meget simpel ting. Ved mange af disse ord vilde en liden tegning være mere oplysende end forklaringen;* og vistnok vilde det også være meget nyttigt, om ordbøgerne kunde forsynes med sådanne tegninger, som ofte ikke skulde behøve større rum, end at det kunde indvindes ved en afkortning i forklaringen.

*F. ex ved *grind, hesja, keip, kverva, leikinde, stabband, sula, truga, tvora, vidja*.

3 Døme med utgreiingar

3.1 Nokre reiskapsnemne

Aasen (1925) har med eit rikt ordtilfang for det me kan rekna under reiskap, verkty og utstyr av ymse slag; det ser me når me tel ord og nemne i dei einskilde bolkanen:

- 277 *førsleverk* og 280 *farkost* kring 500 ord;
- 278 *ilåt* og 279 *kjerald* 350–400 ord;
- 281 *reidskap*, 282 *vinneveler*, 283 *veideveler*, 284 *emningsreidskap*, 285 *husreidskap*, 286 *festeverk*, 287 *vevreidskap* og 290 *deilder av reidskap* kring 1500 ord;
- 288 *klædebunad* og 289 *småreide* [til klæde] kring 700 ord;

I det heile er dette vel 3000 ord og nemne for slike gjenstandar og arbeid.

3.1.1 Ard, plog og grind

Det fortelst at den norske etnologen Hilmar Stigum brukte å seia at *ard* ikkje var det same som *plog* endå om det kan sjå soleis ut i ordbøker. *Ard* vert gjerne definert med ‘einfeld treplog’ (Schjøtt 1914) og *plog* ‘reiskap til å pløya åker med’ (NOB). So einfelde uttydingar er ikkje fullnøyande. Ogso i ei vanleg ordbok må det gå fram at det er ein teknisk skilnad på desse reiskapane; arden kan ha *skjer*, men vantar ristil og veltefjøl og velter torv og mold til båe sider. Plogen har *skjer*, *ristil* og *veltefjøl* og velter torv og mold til ei side. Og slik reiskap vert dregen fram, i eldre tid av køyre dyr (hest) og seinare av motordrivne gogn (traktor) (jfr NRO og Rommetveit 1979).

Ein kunne kanskje definera orda soleis:

ard reiskap opph av tre (utan ristil og veltefjøl) til å løysa og smuldra jord med, velter torv og jord til båe sider, vert dregen fram av køyre dyr eller maskin (traktor)

plog pløyereiskap av jarn (med ristil, skjer og veltefjøl) til å skjera ut og venda (pløya) mark, voll (grassvord) og åker med, velter torv og mold til ei side, vert dregen fram av køyre dyr eller maskin (traktor).

Aasen drog fram *grind* som eit vanskeleg døme, men gav ein grei definisjon:

‘stengsel i eit gjerde, det har form som ein stige og samansett av *rimar* og *okar*’. (Aasen 1873: **rim** f “en af de stænger eller fjæle, som danne ribberne i en “grind”” — motsett **øke** m 2) “hjørnetræ i et led (s grind); et tykkere tværtræ, hvori enderne af ledets [d e *grind* (1)] stænger (rimar) ere indfældede”.)

Serleg i fl kan *grind* etter Aasen (1850) tyd 3 og Aasen (1873) tyd 5 vera ‘rekkje av samanbundne grindar nytta som kve’. Her kjem me innpå ein skikk, å *grinda* eller *grindgå* ‘setja opp grindar (kve, hegne, høye) til å ha buskapen i om natta, og då dei mjølka og stjorna’. Grindane vart oppsette på voll eller innmark, og ein flutte dei ikring so heile vollen eller marka fekk hevd. Skikken er gamal (nemnd i Hávamál 78: “fullar grindr sá ek fyr fitjungs sonum”) og har i ettertid vore kjend over mykje av Noreg; men ofte nytta dei andre ord eller nemne; termene *grind(ar)* og å *grinda* finn ein i øvre Numedal, Telemark, Agder, Hardanger og Nord-Noreg. Andre nemne var t d *kvea* eller *kvereka vollen* (Valdres) og *trø vollen* (Nord Østerdal) (se Reinton 1957:52–53 og Eidnes 1982).

3.1.2 Skavejarn, skavl og skjøve

Det er nemne for verkty som tresmeden (båtbyggjar, snikkar og tømremann) nytta. Sundt (1974–78 8:223–228) fortel koss ein brukar reiskapane, og opplysningane både om bruken og nemna hadde han frå øvre Hallingdal. Nemna er avleidde av verbet *skava* og det ein “skjev (av)” heiter

skav n.

A²: noget som er afskavet. [...]

Især om bark, som afskrabes til foder.

skava

A²: 1) skrabe, afskrabe, f. ex. bark af træer. *skava* åt *buskapen* skave bark til foderhjelp. 2) jævne ved skavning, afglatte eller polere træ.

S: man snitter barken af kviste og grene, ja tykke stammer, som er hugne [...] og af dette "skav" få kreaturene at æde (8:223; jfr 120). "skave suen" d. e. tildanne bådbordenes kant således, at de passe godt til hinanden i sammenføiningen (8:223).

Under 1) definerer Aasen med "skrabe, afskrabe", men Sundt har "snitter". Den som *skjev* nyttar eggreiskap, og med slik reiskap "skjer", "spikkar" eller "snitter" ein. Ein flink handverkar syter for å ha velskjerpte eggjer; det er berre med skjemd reiskap ein må "skrapa"; "snitter" er rettare. Under 2) gjeld det 'skjering', for Sundt opplyser at båtbyggjaren brukar høvel til slik arbeid, men når verbet er *skava* og ikkje *høvla*, er vel det "mindelse om en fortid, da man istedetfor høvl brugte et slags skave-jern" (8:223).

skavl m

A²: skavjern, redskab til at skave eller udhule med.

S: [reiskapen har skaft som sit på langs eller tvers for båe hendene] Den hallingdalske skavl ligger an mod træet, der skal skaves, ikke alene med eggen af jernet, men også med kanten af træverket (8:225).

skavl finn me ogso i ei samansetjing.

veggskavl

(vantar i alle ordbøker)

S: Når tømmermanden havde sat op en stuevæg, brugte han vægge-skavlen til at glatte stokkene med; med spidserne kunde han nå helt ind til mosefugerne. Som redskabet er tegnet, vender eggen frem- og nedover; man har grebet i skaftet med begge hender og halet til sig — til forskjel fra høvelen, som man skyder fra sig (8:227).

skjøve f

A²: et skavejern = skavl.

S: [skavereiskap] med skaft til een hånd (størelsen er netop så at den høire hånd får rum i skaftet) (8:225).

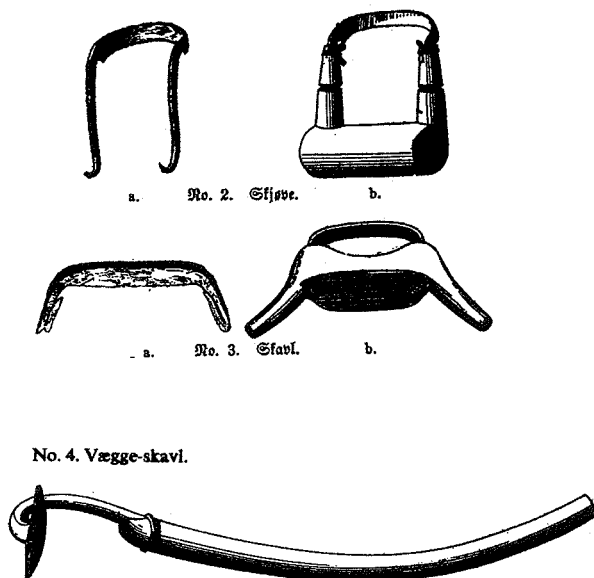
Skavl, *veggskavl* og *skjøve* var til å "jævne og glatte planker og stokke med". Fyrst nytta ein *skjøva*, og so finskov ein med *skavlen* (Sundt 1974–78 8:225). *Skavl* og *skjøve* er eggreiskapar som 'skjer' borken eller flisene av stokk eller anna treemne meir som ein høvel; ein 'skrapar' ikkje. Å definera *skjøve* med "skavl" er gale frå reiskapsynsstad. Sundt opplyste at *skavl* har 'skaft for båe hendene' men *skjøve* har 'skaft for ei hand'.

Ein kunne kanskje definera dei soleis:

skavl skavereiskap med langs- el tver-sitjande skaft for båe hendene.

veggskavl skavl med innkuva egg og tverrsitjande skaft for båe hendene til å glatta eller pussa tømmerstokkar i ein vegg med.

skjøve skavereiskap med handtak som sit på langs for ei hand.



Sundt 1974–78 8:224 og 227.

3.2 Husnemne

I Aasen (1925) under bolkane 272–276 *hus* osv finn me kring 900 ord eller nemne¹ for hus, husrom og alle slags delar på hus; Aasen skil mellom *hus*, *husryfte*, *husrom*, *husreide* og *stydjeverk*.

Husbyggjing og byggeskikk var emne Sundt kom til å arbeida mykje med. Hovudverket er *Om bygningsskikken på landet i Norge* (1974–78 6; jfr 1900), men i mange andre skrifter med kjem han innpå korleis hus er oppsette og inndelte.

Det var ofte mange hus på norske innlandsgardar; 10 til 15 bygningar av ymse slag var ikkje uvanleg. Sundt skilde mellom *udhus* på eine sida og *fremhus*, *forhus* eller *indhus* på hi. Det var *innhus* han granska, og til *innhus* rekna han (6:46):

- | | |
|--------------------------------------|-----------------|
| A første gruppe: | B anden gruppe: |
| 1. ildhus (nu for tiden bryggerhus), | 4. madbod, samt |
| 2. sommerstue, | 5. sengebod. |
| 3. vinterstue | |

Innhus vantar hjå Aasen, men står i NRO. ArkivNO har eit døme frå Vegårshei og nokre døme, mest or faglitteratur; nemnet er også kjent i dansk og svensk.

Til indhusene regnes *stuen* (d. e. stuebygningen) samt de af gårdens huse, der kunne betragtes som tilbehør til samme eller umiddelbart tjene til menneskenes brug og ikke til

¹I denne bolken (3.2) har eg fenge mange opplysningar av arkitekt Arne Berg.

kreaturenes, nemlig *ildhuset* (bryggerhuset), *stabburet* og *sengeboden*, bygninger som fra oldtiden af have været anseede for at høre til på en gård (Sundt 1974–78 9:80).

3.2.1 loft 'tvihøgda forrådshus'

Matbua og *sengebua* var 'forrådshus', og dei var utan eldstad (i motsetnad til *stova* som var oppvermd). I slike hus hadde dei rom til mat og klede, og dei var til å sova i, serleg om sumaren (Sundt 1974–78 6:6, 32–39 og 43; 9:260). Nemne for slike hus var *bur*, *stabbur* og *stolphus*.

Desse husa kunne vera einhøgda, men òg tvihøgda (stundom trehøgda). I det tvihøgda forrådshuset nytta dei fyrste høgda (ofte kalla *bu*) til matvarer, og andre høgda (kalla *loft*) til klede og til sengerom. Det siste nemnet, *loft*, kunne mangstads — som me skal sjå — vera nytta om heile bygningen au.

loft

A²: 1) rum med eget gulv under taget i et huus.

2) sal, overværelse, anden etage.

Hjå Sundt finn me ei tyding til:

S: 3) [forrådshus] med to etager, eller med bod under og loft ovenpå (9:260).

Denne tydinga står heller ikkje i eldre ordbøker (A, R, Schjøtt 1914 og NRO).

At tydinga 'tvihøgda forrådshus' vantar både i A og R, det er merkeleg, for på ferder kring landet såg dei mange slike hus; dei må ha høyrte kva dei vart kalla, og Aasen skriv om dei i ferdameldingane (1957–60 III:108 og 114).

Frå Setesdal i 1844:

Flere huse have to stokværk eller etager, hvoraf den øverste da er forsynet med en sval rundtomkring, hvilket giver bygningen et besynderligt udseende, da loftet derved bliver meget bredere end nederste stokværk.

Frå Seljord 1845:

De såkaldte "utloft" eller "bur" have her denne samme besynderlige skikkelse som i Sætersdalen og Øvre Råbygdelaget, idet den øverste høide eller etage er næsten dobbelt så bred som nederste.

I skildringa frå Setesdal nytta Aasen nemnet *loft* berre om 'øvste høgda', men frå Telemark nyttar han nemnet *utloft* om heile bygningen, dvs 'tvihøgda forrådshus'. Når me så slær opp på *utloft*, har A¹ og A² ulike definisjonar. A¹ er rettare enn A².

A¹: et eenligt huus med loft til forvaring af klæder og deslige

A²: loft på eenligstående buur eller forvaringshuus

I skildringa frå Seljord ser det og ut til at Aasen meiner *bur* og *utloft* er det same, men det er eit mistak.

Som døme på at *loft* tyder 'tvihøgda forrådshus' skal me nemna nokre opplysningar og andre kjelder: "Bur er eihægda, medan loft er bur med tvo hægdir," skreiv Rikard Berge

frå Telemark, og han skil systematisk mellom *bur* og *loft* (1925:576). “Det hende det stod tvo bur attåt loftet” (1973:66; jfr 44–84); frå Seljord skreiv Wille (1786:227): “loftet, som består af to etager, hvoraf den underste tiener til stabur at bevare alle madsager i, men den øverste til andet brug”. Tilsvarande nemningsbruk kjenner ein frå Hallingdal au; “ved sida av setustugu var det rettelege *loftet* [med bu under] viktigaste huset på garden. Andre høgda i loftet var mest som ei storstugu ...”; “*Stolphuset* står mest alltid ved sida av loftet” (Reinton 1938:374, 382 og 385). Hilmar Stigum skreiv m a (KHLNM, under *loft*):

[Loftet] består av et tømret rom i hver etage. Foran og på langsiden er en sval av reisverk, men det forekommer l[oft] med sval på baksiden. Trappa lå alltid utenfor den tømrede del og gikk opp i svala. L[oftet] var som regel det eneste hus på gården som hadde mer enn en etage. Det ble derfor lagt serlig vekt på vakker utførelse [...]

I norske innlandsbygder, serleg i Numedal og Telemark, var det ålment med tvihøgda forrådshus på gardane, og dei eldste er frå mellomalderen. Både “Staveloftet” (Ål) og “Finnesloftet” (Voss) er eldre enn 1350, og her ser me at *loft* er nemne for heile bygningen.

Sundt meiner det einhøgda forrådshuset (*bur*, *stolphus* eller *stabbur*) er eldst, og på norske gardar var det dette slags hus som fyrst fekk to høgder (1974–78 6:32–33; jfr 9:80):

det første hus på de norske gårde, som kom til at hæve sig i høiden, med et øvre stokverk, eller, som det endnu heder i landsmålet, et *loft* (i oldsproget *lopt*), var sikkerlig nettop denne bod.

Her gjer Sundt greie for ein tydingsvokster: *Loft* var nok i husterminologien opphavleg nytta om ‘rom under tak’ eller ‘rom i andre høgde’, men i tvihøgda forrådshus heldt dei andre høgda, *loftet*, gjævast; difor vart heile bygningen etterkvart kalla *loft*. Andre høgda

har været betragtet som den hovedsaglige del af bygningen, vigtigere end nedre stokværk, noget, som også kan sluttes af den talebrug på nogle steder, at man kalder hele bygningen slet og ret *loftet* (6:37).

Skilnaden mellom innhaldet i desse nemna, *loft* ‘tvihøgda (stundom trihøgda) forrådshus’ i motsetnad til *bur*, *stabbur* og *stolphus* ‘einhøgda forrådshus’ kjenner ein på Voss og i midlandet frå Hallingdal til indre Agder. Annanstad i landet, der slike hus var mindre vanlege, skil dei heller ikkje so grant, og då kan *bur*, *stabbur* og *stolphus* vera nytta som nemne på det ‘tvihøgda forrådshuset’ ogso.

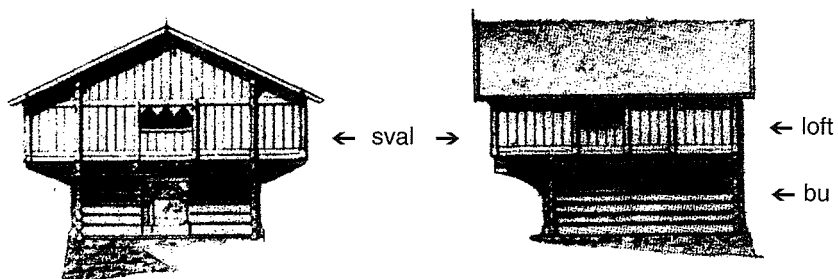
3.2.2 To nemne for avdelte eller tilbygde rom i stove og loft: forstove og sval

Stove var hovudnemne for det oppvermde bustadhuset på ein gard, og samstundes vart det nytta som nemne på det viktigaste opphaldsromet i ein slik bygning. Ein delte frå rom inne og bygde til rom utanfor, og når pipa avløyste ljoren, kunne ein nytta romet under raustet til å sova i. I stova kan ein skilja mellom ‘opphaldsromet’, ‘frådelt rom’, ‘tilbygde rom’ og ‘rom i andre høgde’, og dei har egne nemne.

Eit nemne som merkjer både ‘tilbygt og frådelt rom’, er

forstove

A²: lukket gang forved en stue.



Staveloftet frå Ål i Hallingdal. Bygt oml 1325.

Dette nemnet vanta i A¹, og det kan nok tenkjast A² har det frå Sundt, men Sundt (1974–78) har med meir om byggjemåten og kvar *forstova* er plassert.

1. *Forstova* er 'eit bordkledt tilbygg, sval, på utsida av laftebygget' (10:161–62; 6:53 og 164).
2. Stovebygningen kan vera delt i to; i den mandalske stova står *forstova* i midten mellom dei to stovene: *eldhuset* eller *gamlestova* som har ljore og er utan vindaugo, og *stova* eller *nyestova* som har pipe og vindaugo; denne bygningen "har forstue (fosstove) dannet af de to bindingsværks vægge, som forbinde ildhusets og stuens tømmerbygninger" (6:172; jfr 221).
3. Noko av ei stove kan vera frådelt med ein gavlvegg, og dette frådelte romet kan atter vera delt i to, "i en halv *kleve* og en *forstue* så hovedinngangsdøren ikke kommer til at stå på selve stuens væg" (6:103; jfr 127). Gjennom inngangsdøra kjem ein fyrst inn i gangen eller *forstova* (6:65) og derfrå inn i sjølve stova, og or stova er det "endelig [dør] ind i kleven eller det lille kammer som er afpanelet fra fordøren" (9:83; jfr 6:51).

Etter dei byggjetekniske opplysningar Sundt legg fram, er

forstove 1) bordkledt inngangsrom på framsida av stove, bislag. 2) bordkledt inngangsrom mellom to stover, tener samstundes til å binda dei saman. 3) (i eine enden av stove) frådelt inngangsrom attmed kleve.

Ovanfor høyrde me at *forstova* kunne vera eit 'tilbygg på utsida av laftebygget' men det (dvs *stova*) kunne ha eit tilbygg langs heile veggen au. Dette var ofte kalla

sval

A²: smal gang eller udbygning på siden
af et huus [...]
Også om en afsides gang med privet.

Attåt dette har Sundt ei viktug byggeteknisk opplysning: *svala* er bygd av bord eller plank i motsetnad til stova, buret eller loftet som er lafta.

Han nemner soleis

en svalgang med hel panelvæg (1974–78 6:4, jfr 89), en svalgang af bordvægge udenfor det tømrede hus (195). Svalgangene er lukkede med en heel bordvæg, og i dem ere afpanelede flere mindre rum (1900:299). Burets og burloftets brystsvæler vare før lukkede med stavværk [...] Sidesvælerne til buret og loftet ere af stavværk (369).

Elles har han fylgjande samansetjingar med *sval*: (på loft) *brystsvæl* (1900:369); (på bur) *busvæl* (1974–78 4:45); (på loft og stove) *dørsvæl* (6:50; 1900:77 og 100); (på stove) *langsvæl*: "ramstuen på Løkre har langsvæl foran" (1974–78 6:50, jfr 53); (på loft og tvihogda stove) "denne sval har to høider eller stokværk, nemlig *undersvæl* og *loftsvæl*" (5, jfr 35 og 50); (på stove) *midtsvæl* (1900:299 og 301) og (på bur eller loft) *sidesvæl* (369).

Aasen er for knapp og seier ingen ting om byggjemåten og for lite om kvar på ein bygning ei sval står. Om slikt tilbygg, *svðl*, på hus i Telemark skreiv Rikard Berge (1973:47–49):

me møter denne tilbygnaden i våre eldste trehus som stavkyrkjur, loft, stogehus, ei stolpereising innåt timberveggen. [...] Svaline minner um stavbyggnaden med di dei hev "stavar" som gjeng ned i ein gått i svilli og i ein liknande i slimsi uppe. Ingi svalir var timbra, anna det hender ein ser raustet i veslebursvøli av veggir. Stogune kunde hava svøl framme og på den eine sida, mest vanleg berre framme; den som gjekk på sida førde gjerne attetter til kamaren, so dei kalla ho "kamresvøl". På lofti låg jamnaden ei svøl framme både i fyrste og andre hægdi, i andre au svalir runn-ikring, so heile huset der låg innklædt av svøl, einskildstad vanta svøl bak. Dette kallast *loftsvøl*, medan svøli framme kallast *bursvøl*. Loftsvaline kvilte på *utskot*, d. v. s. på ein eller fleire stokkar øvst i fyrste hægdi på raustsidune som gjekk lenger utum veggjesalen. [...]

Svalklædningi kalla dei skjeldtile [...]

Skjeldtilet hadde alltid stavkonstruksjon, med hyrnestolpar, og i øvre svaline flate midtstolpar som batt ihop øvre og nedre slims (eller svill) til ei råde; millom stolpane stod då stavane eller "tiline". Jamnast stod stolpar og stavar i lodd [...] I skjeldtilet var skori glugg for luft og ljøs, helst små hol på ymis gjerd.

Arne Berg skriv at *sval* er "til vanleg ein overdekt gang el passasje langs yttervegg på bygning, ofte med opningar i den veggen som vender ut" (KHLNM, under *sval*). Ein ordboksdefinisjon bør vel òg ha med ei opplysning om at eit slikt tilbygg er "bordkledd", og likeins om kvar på dei ymse slags hus *svala* finst.

sval (på bygningar av ymse slag) overtekt og helst stavbygt el bordkledd tilbygg som ofte har gluggar el ljøsopningar øvst på utveggen; på framsida av bur og stove, *bursval*, *stovesval*, på sidene el rundt andre høg på loft, *loftsväl*, rundt kyrkje, *kyrkjesval*.

3.2.3 Reinsemd og reinhald: bada, lauga, badstove, gjeva på eller gjeva utor

Reinhald er i hovudsak kvinnearbeid, og om det emnet kan ordbøker seia lite eller vera ufullstendige. Menn har skrivi ordbøkene og samla opplysningane. Ein stor lut av heimelsfolka er ogso menn, og det er vel ikkje visst om dei hadde sansen for kvinne-verksemd, eller kanskje dei oversåg det (se Hageberg 1990; Aasen og Ross var sveinkallar).

Den fyrste som granska kvinnearbeid i Noreg, det var Eilert Sundt; og han skreiv mykje om reinsemd, reinhald og husstell. Han la seg mykje etter å få med ord og nemne knytt til slikt arbeid og likeins opplysningar om arbeidsgangen. Nedunder skal me drøfta nokre slike nemne.

bad

A²: især om varmt bad eller lægebad.

S: *finske* eller *russiske* bad eller *dampbad* (9:313), [det å vera] i hed damp (i et stuerom, en badstue (9:317).

bada v.

A²: 1) skylle i vand (= lauga). 2) varme, ophede (= baka); oftere v. n. give varme fra sig, være hed.

laug f, n.A¹: bad, badning.A²: vædske, vand, især til vaskning.S: 1) det bade- eller vaskevand, som var tillaget til menneskenes renselse (9:163): *ftru, vie tre lauger* (9:69).2) [vaskning] i et stort kar [...] så man kan dukke ned det hele legeme (9:317; jfr 1867b:20–21): *taka laug el lauga seg* (9:291 og 320).

I A¹ vantar tydinga ‘vatn til vasking’, og det er moglege A² kan ha den frå Sundt. Sundt har eit òg par samansetjingar med *-laug*:

sauelaug — ArkivNO

S: det lunkne vand (laug) som sauerne ere vaskede i før klippingen.

ull-laug — ArkivNO

S: [det lunkne vand] som man har vasket garn i når dette er spundet af uvasket uld (klippet af sauerne uden disses foregående vaskning) (9:280).

Frå Hadeland fortel Sundt at når dei hadde *lauga* sauene, “så pleiede gjerne flere af gårdens folk og navnlig børnene at bade sig i *sauelaugen*. Dette troede man skulde være brav for helbreden”. Sauelaugen kunne vera god til kledesvask òg (1974–78 9: 309–10 og 281).

lauga v.A¹: bade, vaske [...]A²: bade, vaske (ved neddyppelse i vand). [...] om at bade sig selv (*lauga seg*); sjeldnere om dyr eller noget andet: f. ex “uld”. *lauge ull*S: bade sig i vand (9:48); *lauge sig i elv eller sø* (9:320); [ein dukkar] sig i vand i et stort kar (9:317).

Jolekvelden heiter *laugarkvelden* sume stader, for då skulle alle *lauga seg* (Sundt 1867b:394). Sundt grunar òg på om det med *bada* “fra først af har været ment at vaske sig i et varmt rum eller at gjøre det varmt i et rum” (1974–78 9:313). Dersom tydinga ‘gjera det varmt i eit rom’ er opphavleg, kan det vera årsak til at nemnet *badstove* ‘oppvermt hus’ gjekk over til å merkja ‘hus til kornturking’, og i islandsk merkjer *badstofa* “selve [den oppvermde] dagligstuen i huset” (9:313, jfr 242 og Folkev 1858:109–10).

badstoveA²: tørrehus, hytte hvorri korn bliver tørret ved ild, førend det føres til møllen [...] Da huset ikke er egentlig bestemt for badning kunde man måske opfatte ordet som *barstova* af det gamle bar): korn [...], men alligevel er det dog rimeligere at forklare det af “bada” med betydningen: opvarme.

Hjå Sundt er det ei tyding i til; den svarar til NO tyd 1:

NO: (serl om e tilh) hus til å ta varmt bad
(serl dampbad) i (VTel og IAgd).

I samtida vart huset nytta "kun til sådanne ting som at tørke korn og [...] at rense klæder (*døyve klæder*) ved at udsætte dem for stærk hede" (Sundt 1974–78 9:313–14).

Tydinga *døyva klæde* 'reinsa klede i sterk hete' vantar i A. R4 har tyd 5) 'desinfisere, særlig klæder ved damp' (Setesdal; jfr NO *døyva* 6). Ogso Sundt (9:242) heimfester opplysningane til indre Agder.

Badstova var nok opphavleg bruka til kroppsvask; døme på det fann Sundt i norrøne og andre gamle kjelder. I Bygland var badstova enno i nytta soleis opp mot 1850, og i Åseral fortalde dei i 1866 koss badinga gjekk føre seg:

Dei elda opp *badstova* og tok med seg inn ei vassbytte og ein lauvdusk; for å få fram meir vasseim, skvette dei vatn på omnen eller omnsrøysa. "Så simpelt kan ovnen i en badstue være opløst eller opplagt af løse stene at den ser ud mest som en stenrøs." Dei la seg på ei hylle med eit klede for andletet og *baka seg* 'vermdde seg opp' i varmen og vasseimen (Sundt 1867b:407; jfr 402; 1974–78 9:314; jfr Skar 1903–15 IV:104f).

Badstoveskikken fann han òg på Finnskogen kring 1850:

På gangen igjennem den adspredt liggende husklynge hen til den fin, [...] kom jeg til en badstue. Badstuen var et sikkert tegn på, hvor vidt den finske nationalitet strækker sig ud over vore skovbygder. Den finske røghytte er nu ofte ombyttet med en norsk stuebygning; men hvor en fin bor, vil man over alt finde en badstue ved siden af. Det var lørdagsaften nu, rensesestiden, og her var livligt omkring den rygende bygning. Nogle kastede klæderne af sig på marken for at kvæge sig ved det forfriskende dampbad; andre hørtes derinde at pidske de svedende lemmer med ris af birkekviste; atter andre styrtede dampende ud og overøste sig med vand i den forbiflydende bæk, gamle og unge, mænd og kvinder i ugeneneret blanding. Dette sidste træk var vel altfor naturligt til, at det skulde kunne passe for vore forholde; men vel skulde jeg ønske, at der stod en badstue på hver norsk bondes gård, og at dens flittige brug kunde smykke hans person og hans hus med en sådan renlighed som den, der i høi grad udmærkede finnerne. Disse bad, forklarede finnerne mig, ere også overmåde tjenlige til sundhedens vedligeholdelse, og jeg troer, at jeg selv befriede mig for en begyndende forkjølelse, da jeg næste dag nogle øieblikke dukkede mig ind i den varme damp og derefter fik et styrtebad over mig (Sundt 1974–78 1:144).

I reinhaldsterminologien var det nok frå gamalt ein skilnad mellom *bad* — *bada* 'gjera seg rein i eit oppvermt rom med dampbading' og *laug* — *lauga* 'gjera seg rein med å støypa heile skrotten ned i (varmt) vatn'. Men når skikken å gå i badstova kom or bruk, gløymde ein den opphavlege skilnaden mellom *bada* og *lauga*.

For det å skvetta vatn på badstove-omnen hadde dei ordlaga:

gjeva på eller gjeva utor.

A: Vantar.

S: tage vand af bøtten og støbe, hælde, skvætte samme på glohede ovnstene, så det ved disses hede øieblikkelig steg op som damp (1867b:410; jfr 9:314 frå Åseral).

Sundt viser til ei badstove-skildring i Eyrbyggja cap. 28, og det norrøne ordlaget er *gefa á*: “síðan lét hann gefa útan á baðit í glugginn er yfir var ofninum”.

P. A. Munch omsette dette med “lægge i, fyre på”, men til det skriv Sundt (1867b:397): “han har [...] ikke tænkt sig badet som et dampbad, men da heller ikke fåt nogen klar forestilling om forholdet”. Fritzner (1886–96) definerer med “hælde, udgyde over noget”; Fritzner kan byggja på Sundt.

4 Avrunding

Alt då *Norsk ordbok* vart grunnlagt i 1931, la ein seg etter å få med den folkelege fagterminologien. Nils Lid sende ut tidskriftet *Ord og sed* og skipa “nemndi til gransking av norsk nemningsbruk”, seinare kalla “norsk etnologisk gransking”. Dei samla inn eit stort ordtilfang or norsk tradisjon, og det har vore til stort gagn for NO. Ved sida av dette innsamla tilfanget kan ein finna mykje folkelege fagord i skrifter om slike emne (Myhren 1988 og 1989).

Her har eg mest halde meg til nokre utvalde døme or tilfanget hjå Ivar Aasen som grunnla norsk talemålsgransking og ordgransking, og hjå Eilert Sundt som grunnla norsk folkelivsgransking. Meininga var å kasta ljøs over arbeidet deira frå ordbokskrivar-synstad. Men før me kan verdsetja desse livsverka i ein slik samanheng, må me ha i minnet at dei hadde kvar sine føremål. Aasen skulle laga oversynlege ordbøker til ålment bruk, og bøkene måtte ikkje verta for store. Aasen har likevel grunnkjerna av det norske ordfanget og hovudtydingane — det gjeld terminologi au, og sume stader er Sundt kjelde i Aasen (1873).

Sundt skreiv store bøker og artiklar om emna og hadde god plass til å drøfta ord og nemne. Han var den fyrste nordmannen som i nemnande mon dreiv med ord- og gjenstandsgransking, og som skreiv om norsk arbeidsliv og folkeliv. Sundt såg opp til Ivar Aasen, og dei kjende kvarandre tolleg godt, er det likt til (sjå Aasen 1957–1960). Like omhugsamt og med like stor innsikt som Aasen hadde granska norsk mål og ordfang, ville Sundt greia ut om norsk arbeidsliv og serleg “fremstille bøndernes egne betragtninger om tingene, de grunde, som fremføres, de sætninger og regler som ere blevne gjældende”. Målet var “et bidrag til *det norske arbeides historie*”, men det fekk han korkje arbeidsvilkår eller levedagar til å fullføra (Sundt 1974–78 3:149 og 7:183; sjå òg Refsum 1966 og Halvorsen 1901).

Både Aasen og Sundt bygde i stor mon på egne åtgåingar og for det meste på sjølvinn-samla tilfang. Utfrå kvar sine ståstader kjem dei med opplysningar som utfyller kvarandre. Sundt òg var ein næpen oppskrivar; han gjev att ord og nemne i god rettskriving og kjem med opplysningar om koss gjenstandar er laga og utforma; og han greier stundom ut om sjølve arbeidsgangen med pedagogisk innsikt. Dertil kunne han mykje gjenstandshistorie som han sameinte med kunnskapar i språkhistorie.

Endå om desse kjeldene er rike på tilfang, so kan det vel i visse høve seiast om båe at dei tek for lite med. Aasen kan stundom vera noko knapp og vel ålmenn i definisjonane; serleg

i terminologi kan han ha med for få semantiske drag, og soleis kan uttydingane verta for stutte og upresise eller falla noko på sida — sett frå fagmanns-synstad. Det ter ogso mange og store hol i hjå Sundt. Han skriv mykje om hus og båtar, men mindre om koss dei vart bygde.

Det eg har lagt fram her skulle likevel syna: i etnologisk litteratur kan ein finna opplysningar om gjenstandar, arbeid og arbeidsgang som kan hjelpa til å utvida kunnskapen om ordtilfanget — serleg om det traderte nemnetilfanget. Det kan tena til å avgrensa og utfylla ordboks-uttydingane og gjera dei meir visshøve og råkande.

Litteratur

A = A¹ og A².

A¹ = Aasen 1850, A² = Aasen 1873.

Aasen, Ivar 1850: *Ordbog over det norske folkesprog*. Christiania. [Stundom forkorta A¹].

Aasen, Ivar 1873: *Norsk Ordbog*. Christiania: B.T. Ballings Boghandel. [Stundom forkorta A².]

Aasen, Ivar 1925 [1875–77]: *Norsk målbunad*. Oslo: Norske samlaget.

Aasen, Ivar 1957–60: *Brev og dagbøker I–III*. Oslo: Norske samlaget.

ArkivNO = Setelarkiv for nynorsk mål i Avdeling for leksikografi ved Universitetet i Oslo.

Berge, Rikard 1925: *Bygningskunst*. I: H. H. Einung: *Tinn saga* I. Rjukan.

Berge, Rikard 1973: *Vinje og Rauland III*. Stavanger.

Coseriu, Eugenio 1973: *Probleme der strukturellen Semantik*. Tübingen Beiträge zur Linguistik 40. Tübingen.

Coseriu, Eugenio 1988: *Einführung in die allgemeine sprachwissenschaft*. Tübingen.

Eidnes, Asbjørn 1982: *Grindgang*. *Håøygminne* 64:224–264. Harstad.

Eyrbyggja = *Eyrbyggja saga*. *Grænendinga sögur*. 1935. Íslenzk fornrit IV. Reykjavík: Hið íslenzka fornritafélag.

Falk, Hjalmar/Alf Torp 1910–11: *Norwegisch-dänisches etymologisches Wörterbuch*. Heidelberg: C. Winter.

Feilberg, H. F. 1910–14: *Bidrag til en ordbog over det jyske almuesmål IV*: Tillæg. København: Thieles Bogtrykkeri.

Folkev = *Folkevennen*. 1855–67. Kristiania.

Fritzner, Johan 1886–96: *Ordbog over det gamle norske sprog I–III*. Kristiania: Norske forlagsforening [1. utg Chra 1867].

Hageberg, Arnbjørg 1990: Kvinner og menn og ordbøker. I: *Norsk lingvistisk tidsskrift* 8, 68–77.

Halvorsen, J. B. 1901: *Norsk forfatterleksikon V*. Kristiania.

KHLNM = *Kulturhistorisk leksikon for nordisk middelalder I–XXII*. 1956–78. Oslo: Gyldendal norsk forlag.

Lewandowski, Th. 1976: *Linguistisches Wörterbuch I–III*. Heidelberg: Quelle & Meyer.

Lid, Nils 1933: Um folkelivsgransking. I: *Helsing til Olav Midttun*. Oslo: Noregs boklag, 151–164.

Molbech, Christian 1833: *Dansk ordbog*. København.

Myhren, Magne 1988: Etnofolkloristiske og lokalhistoriske skrifter som kjeldor for ordtilfang. I: *Ord og mål*. Festskrift til Magne Rommetveit. Oslo: Kringkastingsringen, 128–144.

- Myhren, Magne 1989: Språk og ord som emne i Rikard Berges forfatterskap. I: *Fram då frendar*. Oslo: Norsk måldyrkingslag, 63–82.
- Myhren, Magne 1993: Eilert Sundt og norsk språkhistorie. I: *Språklig samling* 3, 9–13.
- NO = *Norsk ordbok*. 1950ff. Oslo: Det norske samlaget.
- NOB = *Nynorskordboka*. 1993. Oslo: Det norske samlaget.
- NRO = *Norsk riksmålsordbok*. 1937–57. Oslo: H. Aschehoug & Co.
- Ord og sed*. 1934–47. Oslo.
- Refsum, Helge 1966: Eilert Sundt. *Norsk biografisk leksikon* XV. Oslo: H. Aschehoug & Co., 276–295.
- Refsum, Helge 1941: Eilert Sundt, språksaken og bygdemålene. I: *Mål og minne*, 67–81.
- Reinton, Lars 1938: *Folk og fortid i Hol* I. Gjøvik: Hol herad.
- Reinton, Lars 1957: *Sæterbruket i Noreg* II. Oslo: Instituttet for sammenlignende kulturforskning.
- Rommetveit, Magne 1979: *Norsk landbruksordbok* I–II. Oslo: Det Norske Samlaget.
- R, R1–6 = Hans Ross 1895–1913: *Norsk ordbog*. Kristiania.
- S = Sundt 1974–78.
- Sapir, Edward 1958: *Selected Writings in Language, Culture and Personality*. Berkeley: University of California Press.
- Schjøtt, Steinar 1909: *Dansk-norsk Ordbog*. Kristiania: H. Aschehoug & Co.
- Schjøtt, Steinar 1914: *Norsk ordbok*. Kristiania: Norske samlaget.
- Skar, Johannes 1903–15: *Gamalt or Sætedal* I–VII. Kristiania: Olaf Norlis forlag.
- Sundt, Eilert 1974–78: *Verker i utvalg* 1–11. Oslo: Gyldendal norsk forlag.
- Sundt, Eilert 1900: *Bygningskikken på bygdene i Norge*. Kristiania: H. Aschehoug & Co.
- Sundt, Eilert 1867a: *Om skab-klåens udbredelse*. Christiania: J. Chr. Adelstedts forlag.
- Sundt, Eilert 1867b: Om badstuer. I: *Norden*. Et maanedsskrift, udgivet af J. Lieblein, 388–422
- Sundt, Eilert 1856–1866: [Artiklar i] *Folkevennen*. Kristiania: Selskabet for folkeopplysningens fremme. [forkorta *Folkev.*]
- Torp, Alf 1919: *Nynorsk etymologisk ordbok*. Kristiania: H. Aschehoug & Co.
- Wille, Hans J. 1786: *Beskrivelse over Sillejords præstegjeld*. København.

Exemplens funktion i ordböcker

It seems to be a generally accepted truth that examples have an important function in dictionaries. In spite of that there exists no general theory of dictionary examples. A sentence used as an example in a dictionary has a very different function from that of the same sentence used in normal communication. This means, for instance, that the problem of authenticity must be looked upon in a new way. A theory of dictionary examples could, as the German linguist Fritz Hermanns has shown, have the cooperative principle and the conversational maxims of Grice (1975) as one point of departure.

1 Inledning

Exemplens funktion i ordböcker är naturligtvis beroende av ordbokstypen. I detta bidrag rör det sig om vad man grovt kan beteckna som "praktiska ordböcker", t.ex. inlärningsordböcker och överhuvudtaget ordböcker som används för att förstå eller för att producera nya texter på det egna språket eller på ett främmande språk. Beaktade blir alltså inte historiska ordböcker, dialektordböcker o.d., dvs. ordböcker med rent deskriptiv-dokumentativ funktion. Distinktioner som allmänspråklig och fackspråklig ordbok, en- respektive tvåspråkig ordbok o.d. kan av utrymmesskäl inte beaktas, även om de har relevans i detta sammanhang. Vidare utgår jag utan närmare argumentation från att exemplen normalt utgörs av fullständiga satser, jfr Bergenholtz (1994: 422 ff.).

2 Autenticitet

En utgångspunkt får här utgöras av frågan om "äka", "autentiska" exempel dvs. belägg är "bättre" än konstruerade exempel. Min egen ståndpunkt är att konstruerade exempel ur en viss synvinkel rentav kan betraktas som mera äkta och naturliga, mera "autentiska" än exempel tagna ur korpora, jfr Nikula (1986). Denna ståndpunkt förefaller säkerligen mången ganska märklig och accepteras som sådan sannolikt inte av de flesta leksikografer. Den måste därför exemplifieras — med ett exempel.

Är satsen *Det regnar* ett belägg eller ett konstruerat exempel? Det är en sats som alla som levt i en svenskspråkig omgivning hört många gånger och den kunde sålunda vara ett belägg, ett äkta exempel, men jag har inte brytt mig om att leta efter satsen i någon korpus eller text, utan har konstruerat den för detta sammanhang.

Ett exempel som *Det regnar* används naturligtvis inte som exempel på sig självt utan för att exemplifiera någon företeelse i språket. Det kunde t.ex. vara ett exempel på en s.k. indirekt talakt i en lärobok i lingvistik. Låt oss anta att satsen är ett äkta belägg som yttrats i en bestämd situation i ett bestämt syfte. För recipienten var saken klar. Han eller hon tog paraplyet med sig innan hon gick ut. När vi läser satsen som exempel på en indirekt talakt i

en lärobok i lingvistik måste vi föreställa oss en motsvarande situation där vi är fiktiva hörare som hör en fiktiv talare — men vi reagerar naturligtvis inte enligt den fiktiva uppmaningen eftersom vi själva inte befinner oss i den fiktiva världen. Det är ju trots allt bara fråga om ett exempel!

Belägget, det äkta exemplet, har alltså en helt annan funktion än i den ursprungliga situationen, det är "ett citat som ryckts ur sitt sammanhang", jfr Hermanns (1988:166). Om jag däremot konstruerar en sats, t.ex. satsen *Det regnar*, som exempel på en indirekt talakt, är den exakt i det sammanhang som den formulerats för och är i den meningen mera autentisk eller äkta än ett belägg ur en "riktig" text skulle vara. I viss mening är dock, såsom Hermanns (1988:166) påpekar, även det konstruerade exemplet ett ur sitt sammanhang ryckt citat, eftersom "konstruktören" inte kan konstruera ett exempel utan att föreställa sig någon kommunikativ kontext.

Det hittills sagda innebär emellertid inte något ställningstagande till den intressanta frågan huruvida belägg är bättre eller sämre än konstruerade exempel. För att man överhuvudtaget ska kunna ta ställning till den frågan måste man närmare analysera exemplens funktion.

3 Exemplens funktion

Språkliga exempel i språkvetenskapliga arbeten, i grammatiker och i ordböcker har till en del olika funktioner. Exempelen i språkvetenskapliga arbeten kommer jag inte att beröra, eftersom deras funktion tydligt skiljer sig från exemplens funktion i grammatiker och ordböcker, medan exemplen i de senare har mycket gemensamt med varandra.

Exempelen i grammatiker och ordböcker kan användas för att exemplifiera grammatiska, semantiska, pragmatiska, stilistiska osv. egenskaper hos ord och andra språkliga uttryck. Exempelen kan vidare användas för att ge en uppfattning om sociala, historiska och kulturella aspekter på det eller de samhällen där språket i fråga talas. I SOB finner vi t.ex. under substantivet *pipa* exemplet "hon drömde om en lugn och trygg man som rökte pipa", vilket tyder på att synen på kvinnans ställning kanske inte förändrats så mycket som man gärna vill tro. Å andra sidan finner man exempel som "pappa ska mata lillan" under *mata* och "han lagar god mat" under *laga*, vilket ger en något annan uppfattning om det svenska samhället, eller åtminstone om ordboks författarna — eller kanske det rentav är så att författarna vill uppfostra ordboks användarna genom valet av exempel. Detta är inte en så långsökt tanke som det kanske låter. I t.ex. band 4 av DDR-ordboken *Wörterbuch der deutschen Gegenwartssprache* (WDG) i sex band sägs explicit att exemplet i detta och i de följande banden innehållsigt ska stå i överensstämmelse med den marxistiskt-leninistiska ideologin just i uppfostrande syfte. En intressant forskningsuppgift vore säkerligen att undersöka hur exemplen i grammatiker och ordböcker speglar det samhälle de härstammar från, jfr även t.ex. Bergenholtz/Tarp (1994:141), Grosse (1967), Pusch (1983), Römer (1973). Denna inte oväsentliga aspekt kan jag inte gå närmare in på här; exemplen i grammatiker och ordböcker används uppenbarligen huvudsakligen i andra syften.

I Ebba Lindbergs *Beskrivande svensk grammatik* finner vi följande regel (Lindberg 1976:158):

Topikalisering. Ett adverbial eller en nominalfras ur predikatsfrasen flyttas till satsens första plats, som i bassatsen upptas av subjektet. Som en följd av topikaliseringen måste subjektet flyttas åt höger, så att det i den transformerade satsen står efter det finita verbet.

Denna något abstrakta regel exemplifieras bl.a. av följande exempel (Lindberg 1976:158):

Jag har inte träffat honom än → *Honom* har jag inte träffat än.
Än har jag inte träffat honom.
Inte har jag träffat honom än.

Exemplen konkretiserar alltså en relativt abstrakt regel som utan exempel vore svår att förstå och använda. Egentligen borde vi inte tala om regel utan om regelformulering, dvs. exemplen tydliggör en regelformulering. Genom regelformuleringen försöker man naturligtvis beskriva en i språket självt inherent regel och i den meningen kan vi säga att exemplet exemplifierar en regel.

När det gäller grammatiker exemplifierar exemplen vanligen endast en regel åt gången. Exemplen i ordböcker exemplifierar däremot i normalfallet snarare flera än en regel, vilket dels har rent praktiska orsaker, dvs. utrymmesmässiga-ekonomiska, dels beror på att man slår upp i en ordbok av andra orsaker än i en grammatik. Grammatiken beskriver ju i princip allmänna regler, medan ordboken naturligtvis beskriver lexikaliska regler, dvs. regler som är förknippade med bestämda ord.

När man använder en grammatik slår man alltså upp en regel, när man använder en ordbok slår man upp ett ord för att få reda på en eller flera av de många mer eller mindre idiosynkratiska regler som gäller användningen av detta ord, varvid ett och samma exempel kan få belysa olika regler, indirekt kanske också sådana som det inte primärt är avsett att tydliggöra. Exemplet "han lagar god mat" anger samtidigt att *laga* är ett transitivt verb och att det utgör en betydelsevariant av detta verb ('tillreda, anrätta', inte 'reparera'), dvs. en syntaktisk och en semantisk regel. För ordboksanvändaren kan någondera eller båda reglerna samtidigt vara aktuella.

4 Exemplens textualitet

Det faktum att exemplen i grammatiker och ordböcker har som funktion att exemplifiera språkliga regler och inte att såsom språkliga uttryck i texter i allmänhet att förmedla vissa innehåll gör att vi måste ställa oss frågan hur ett exempel ska vara utformat utgående från helt andra synpunkter än när vi i normal kommunikation formulerar oss språkligt, jfr även Nikula (1988b).

En sats i en vanlig text är ju inte ett exempel på regler genom att utgöra en konkretisering av regelformuleringar utan är ett resultat av tillämpningen av ett stort antal regler i en viss kontext i syfte att förmedla ett visst innehåll. Detta faktum är viktigt att ta i beaktande när man tar ställning till användningen av autentiska respektive konstruerade exempel i ordböcker. Det är självklart att ordboksförfattaren bör använda sig av de korpora som står till buds, frågan är hur de används. När man med dagens datateknik kan behandla stora textmängder och man utgående från t.ex. frekvensundersökningar försöker sluta sig till språkliga regler

måste man beakta det faktum att belägg ur "normala" texter inte exemplifierar regler utan är resultaten av regel användningar och att sambandet mellan frekvens och regel aldrig kan spegla en regel direkt. Visserligen beror det faktum att t.ex. ett syntaktiskt objekt naturligtvis oftare förekommer vid transitiva än vid intransitiva verb på lexikaliska egenskaper hos dessa två verbklasser, men det att plats- och tidsadverbial ofta förekommer vid verb som *regna*, jfr även Karlsson (1983), beror inte omedelbart på samma sätt på lexikaliska egenskaper hos dessa verb utan snarare på kommunikativ relevans, jfr Nikula (1994). Detta innebär omvänt att den höga frekvensen av objekt vid transitiva verb inte kan uppfattas som en direkt indikator på någon lexikalisk regel såsom "transitivitet" utan att andra aspekter såsom t.ex. verbets semantik beaktas. När man i förordet till SOB säger att "*Svensk ordbok* är ett *tillrättalagt* [min kursiv.] val ur databasen" förefaller detta mig sålunda vara en klok strategi.

Exemplen måste naturligtvis väljas eller konstrueras utgående från sin funktion i ordböckerna och det har man naturligtvis alltid gjort, även om man inte alltid så tydligt formulerat uppgiften och framför allt kanske inte alltid så tydligt sett skillnaden mellan ett språkligt uttryck i normal kommunikativ funktion och som exempel på regler. Ett språkligt belägg är alltså inte ett exempel på en eller flera regler, även om man kan säga att det är ett exempel på att ett antal regler tillämpats vid ett visst tillfälle, dvs. på något som historiskt ligger mer eller mindre långt bakåt i tiden. Ordboksanvändaren är emellertid i fråga om praktiska ordböcker sällan intresserad av att någon vid något visst tillfälle sagt "Det regnar"; han vill genom exemplet få hjälp att förstå någon eller några av de regler som gäller användningen av verbet *regna*. Ordboksexempel har alltid en prospektiv karaktär, dvs. har funktionen att i någon framtida situation kunna ge ordboksanvändaren en uppfattning om en viss språklig regel. Belägget är däremot alltid i sin egenskap av belägg retrospektivt, dvs. pekar på något förflutet, men kan naturligtvis trots detta användas som ordboksexempel. Då är det emellertid inte mera retrospektivt utan prospektivt och också i den meningen "inautentiskt" sin autenticitet till trots.

5 "Maximer" för exempel

Ett språkligt uttryck, t.ex. en sats, i normal kommunikativ användning, befinner sig i en kommunikativ kontext där textproducenten vill förmedla någonting och textrecipienten är mer eller mindre intresserad av vad textproducenten vill förmedla genom texten. Den kommunikativa kontexten för ett lexikaliskt exempel består i sin vidare mening av ordboken och dess kontext, närmare bestämt emellertid av:

- a) de regelformuleringar exemplet ska förtydliga (t.ex. "itr. verb + 'betydelsedefinition'") i syfte att göra de språkliga reglerna gripbara, samt:
- b) den fiktiva kommunikativa kontext ordboksanvändaren måste föreställa sig för att tolka exemplet "som om det vore" ett språkligt uttryck i normal användning.

Den tyske germanisten Fritz Hermanns har utgående från Grice' bekanta kooperationsprincip och konversationsmaximer formulerat nya maximer som beaktar det lexikaliska exemplets särart, jfr även Bergenholtz (1994:429). Grice' fyra huvudmaximer var ju följande (Grice 1975:45 f.):

- 1) "Quantity", dvs. var så informativ som möjligt, var inte onödigt informativ.
- 2) "Quality", dvs. försök att göra ditt bidrag sådant att det är sant, säg ingenting som du håller för osant, säg ingenting, som du inte kan verifiera.
- 3) "Relation", dvs. var relevant.
- 4) "Manner", dvs. undvik dunkelhet i uttrycket, undvik flertydighet, uttryck dig kort, var följdriktig.

Grice (1975:46) tillägger: "And one might need others."

Hermanns' maximor för exempel lyder som följer (Hermanns 1988:177):

- 1) Ein Beispiel soll sprechend sein.
- 2) Ein Beispiel soll ansprechend sein.
- 3) Ein Beispiel soll echt sein.
- 4) Ein Beispiel soll kurz sein.

Maxim (1), ett exempel ska vara talande, innebär att exemplet verkligen ska tydliggöra de regler som är aktuella. Ett exempel som *Far läser under läsa* uppfyller inte ensamt detta krav bl.a. därför att det inte anger att verbet är transitivt. I kombination med andra exempel kan det naturligtvis fylla sin uppgift; bl.a. anger det att verbet utan särskild kontext kan stå utan objekt. Hermanns bifogar i likhet med Grice ett antal undermaximer.

- 1a) Ein Beispiel soll das semantisch Prototypische eines Lemmazeichens zeigen (Hermanns 1988:178). — T. ex. "Hunden sover" är inte semantiskt prototypiskt, mera prototypiskt och i den meningen bättre är "Hunden skäller, viftar på svansen, morrar".
- 1b) Ein Beispiel soll tautologisch sein (Hermanns 1988:178). — Ett tautologiskt exempel skapar en "stark kontext" (jfr Nikula 1988a:233). En icke stark kontext för *köpa* utgör "Han köpte en bok", medan "Han köpte en bok för 300 kronor hos sin bokhandlare" kan betraktas som en stark kontext. Beläggsatser som utgör starka kontexter är svåra att finna eftersom dels alltför mycket information pressas in i en enda sats, dels en del av denna information inte är kontextuellt relevant, jfr Grice' maxim (3) ovan, samt Nikula (1994).
- 1c) Ein Beispiel sollte multifunktional sein (Hermanns 1988:179). — Multifunktionella exempel som tydligt kan åskådliggöra flera regler på en gång är ofta svåra att finna i textkorpora.

Till maxim (2), ett exempel ska vara "tilltalande", fogar Hermanns följande undermaximer:

- 2a) Ein Beispiel soll interessant sein (Hermanns 1988:179). — Ett exempel som är intressant har svårt att uppfylla kraven i (1a, b, c).
- 2b) Ein Beispiel soll geistreich und witzig sein (Hermanns 1988:180). — Ett exempel som uppfyller kraven i (1a, b, c) är emellertid sällan särskilt spirituellt.
- 2c) Ein Beispiel soll anspruchsvoll sein.

- 2d) Ein Beispiel soll leicht verständlich sein (Hermanns 1988:180). — Ett exempel som är anspråksfullt (2c) är kanske inte alltid särskilt lättbegripligt (2d).
- 2e) Ein Beispiel sollte historisches Kolorit haben.
- 2f) Ein Beispiel sollte lokales und soziales Kolorit haben.
- 2g) Beispiele sollten auch in dem Sinn beispielhaft sein, daß sie besonders gelungene Formulierungen, namentlich aus der Literatur, präsentieren (Hermanns 1988:181). — Maximerna (2e, f, g) gäller kulturella aspekter, vilket talar för belägg; beträffande (2g) är belägg naturligtvis det enda möjliga.

Till maxim (3), ett exempel ska vara "äkta", fogar Hermanns följande undermaximer:

- 3a) Ein Beispiel soll authentisch sein (Hermanns 1988:181). — Ett exempel ska enligt Hermanns om detta är möjligt vara autentiskt i den meningen att det är ett belägg.
- 3b) Ein Beispiel soll seinem Zweck — dies oder jenes zu zeigen — optimal angepaßt sein oder werden (Hermanns 1988:181). — Ett exempel som är så väl lämpat som möjligt med tanke på sitt ändamål — att åskådliggöra regler genom att förtydliga regelformuleringer — finner man inte alltid så lätt ens i stora korpora, jfr även Hermanns (1988:182).
- 3c) Ein Beispiel soll glaubwürdig sein (Hermanns 1988:181). — Att exempel ska vara "trovärdiga" gäller även konstruerade sådana.

Till maxim (4), ett exempel ska vara kort, fogar Hermanns endast en undermaxim:

- 4a) Ein Beispiel soll lang sein (Hermanns 1988:181). — Bl.a. utrymmesskäl talar naturligtvis för att exemplen ska vara korta, medan långa exempel kanske lättare fyller kraven på naturlighet och också lättare bildar "starka" kontexter. Ett långt exempel kan emellertid genom sin större komplexitet vara otydligare i fråga om de regler det ska förtydliga, jfr även Nikula (1988a:234).

Det är uppenbarligen mycket motstridiga krav som ställs på exempel. Ett exempel ur verkligheten visar att ett exempel förefaller kunna vara bra just för att det är dåligt. Satsen *Die Heringe der Ostsee sind magerer als die der Nordsee* som exempel på användningen av determinativt pronomen som huvudord till genitivattribut tycks ha trängt djupt in i medvetanden hos alla svenskar som råkat ut för det i skolundervisningen som ett skräckexempel på ett tyskt grammatikexempel, jfr Sundqvist (1988), Freund/Sundqvist (1988:§488). Men ett exempel som man kommer så väl ihåg kan väl inte vara dåligt? Som grammatiskt exempel uppfyller satsen det centrala kravet att tydliggöra en regel genom att ställa pronomen och artikel bredvid varandra framför ett enkelt substantiv i slutet av satsen, "... die der Nordsee", vilket framhäver den från svenskan starkt avvikande strukturen. Exemplet uppfyller alltså väl det grundläggande krav som Hermanns formulerar i maxim (1). Genom att verka så uppenbart oäkta klarar det på ett elegant sätt av maxim (3) och därmed också maxim (2), vilket bekräftas av att man uppenbarligen har svårt att glömma exemplet. Eftersom det är lätt att komma ihåg är det knappast för långt varför det uppfyller kravet i maxim (4). Ett genialt exempel, som för övrigt ännu efter hundra år gör tjänst! (Jfr Sundqvist 1988).

6 Regel, regelformulering och exempel

Regelformuleringarna och exemplen måste vara så utformade att användaren utgående från sina språkliga och andra relevanta kunskaper får ut den information hon eller han söker. Detta innebär att de regler regelformuleringarna bygger på måste ha någon typ av överensstämmelse med ordboksanvändarens internaliserade regler. Dessa regler är inte direkt åtkomliga utan återspeglas endast indirekt i språkbrukarnas beteende, dvs. i deras produktion och reception av texter. Man måste dessutom alltid utgå från att det finns en viss interindividuell variation ifråga om de internaliserade reglerna utan att detta nämnvärt stör den språkliga kommunikationen, jfr Nikula (1995:avsn. 5.3).

Ett relativt tydligt exempel på svårigheterna utgör verben inom verbfältet "köpa och sälja". Inom tysk valensforskning, en deldisciplin av lexikologin med relevans för lexikografen, har dessa verb utförligt diskuterats och oenighet har rått och råder alltjämt om huruvida de ska betraktas som två-, tre- eller fyrställiga. Ska man t.ex. beträffande *köpa* (*kaufen*) förutom "köparen" och "varan" betrakta "priset" och/eller "säljaren" som valensbundna element, "fyllnadsled". Valet gäller frågan om vilka lexikaliska regler som ska antas gälla för verbet, vilket naturligtvis har konsekvenser för regelformuleringarna och därmed också för valet av exempel. Det är uppenbart att alla fyra element, "köpare", "vara", "pris" och "säljare", på något sätt har relevans för verbet och att olika valensbeskrivningar måste utgå från något olika uppfattningar om den centrala lexikaliska betydelsen hos verbet. Noteras kan att SOB av betydelsedefinitionen att döma, 'skaffa sig (ngt) som sin egendom mot ersättning' sannolikt utgår från att *köpa* är treställigt, "köpare", "vara" och "pris". I exemplen förekommer alla fyra element, "köpare", "vara", "pris" och "säljare".

Utgående från att alla fyra med verbet nära förknippade element, "köpare", "vara", "pris" och "säljare", är relevanta har man i den stora tyska verbvalensordboken *Verben in Feldern* (VIF) beskrivit *kaufen* 'köpa' som fyrställigt. Det är i detta sammanhang intressant att notera att det enda exemplet med fyra fyllnadsled utsatta på en gång är en konstruerad sats. Eftersom den korpus ordboken bygger på är elektroniskt läsbar är det enkelt att se att inga belägg med alla fyra fyllnadsled på en gång föreligger, att det överväldigande flertalet har "köpare" och "vara" utsatta, att endast ett mindre antal belägg med "köpare", "vara" och "pris" förekommer och att kombinationen "köpare", "vara" och "säljare" är ännu ovanligare, jfr Nikula (1994:205). Nämnas kan att inte heller SOB anför något exempel med alla fyra led, medan ÖSTERGREN anger ett belägg ur Bibeln: "Mat att äta skolen I k-a av dem för penningar" [min kurs.].

Som tidigare framhävts kan man inte direkt utgående från frekvensen dra slutsatser om språkliga regler. Utgående från andra överväganden som jag inte kan gå in på här kan man dock dra slutsatsen att förekomsten av lexikaliskt fyrställiga verb inte verkar särskilt sannolik (jfr Herslund 1988) och att det är fullt möjligt att beskriva verb som *kaufen* (och därmed också *köpa*) som ett lexikaliskt tvåställigt verb med "köpare" och "vara" som fyllnadsled, antaganden som åtminstone inte står i strid med korpusanalyserna.

Om emellertid *köpa* beskrivs som tvåställigt måste också "priset" och gärna också "säljaren" anges, eftersom dessa i motsats till t.ex. tids- och platsadverbial är intimt förknippade med just detta verb. En möjlighet är att göra det genom en kombination av väl valda exempelsatser så att exempel med "köpare", "vara" och "pris", samt "köpare", "vara" och "säljare" förekommer. Ytterligare kunde man liksom i VIF konstruera ett fyrställigt

exempel, ifall man inte finner ett lämpligt belägg, eftersom fyrställigheten skapar en stark kontext som bidrar till betydelseolkningen, alltså t.ex. "Han köpte en bok hos sin bokhandlare för trehundra kronor". En sådan sats är uppenbarligen inte särskilt sannolik i en normal text och inte heller speciellt "intressant", men kan fylla sin uppgift som exempel i en ordbok.

7 Sammanfattning

Ett exempel exemplifierar alltid en regel, ett ordboksexempel i allmänhet flera regler på en gång. Exemplet har som funktion att tydliggöra regelformuleringar så att ordboksanvändaren kan sluta sig till vilka regler det är fråga om. Exempelen måste alltså vara "genomskinliga" i förhållande till de regler de ska åskådliggöra, vilket tyder på att de helst borde vara korta och torftiga. Detta medför lätt att de verkar onaturliga. Mot kortheten och torftigheten talar det faktum att semantiskt "mättade" exempel, "starka kontexter", kan vara till fördel för att betydelsedefinitionen ska kunna tolkas riktigt. Också starka kontexter kan leda till "onaturligheter" eftersom sådant som i en normal text sannolikt vore kommunikativt irrelevant gärna ingår i en sådan kontext.

Exempelen i ordböcker bör vara enkla, korta och torftiga, de bör vara långa, semantiskt mättade och tautologiska, de bör vara intressanta osv. En mängd sinsemellan motstridiga krav ställs på exemplen och det enda sättet att klara sig ur dilemmat är att använda sig av rikligt med exempel, långa och korta, intressanta och torftiga, autentiska och konstruerade osv. Där sätter de praktiska, inte minst de ekonomiska realiteterna naturigtvis gränser. Eftersom exemplen helt uppenbart har en viktig funktion framstår det tydligt att en teori rörande det lexikografiska exemplet är nödvändig om man vill utveckla ordböckerna på denna punkt.

Litteratur

Ordböcker

SOB 1986 = *Svensk ordbok*. Utg. Sture Allén. Uppsala: Esselte.

VIF 1986 = *Verben in Feldern*. Utg. Helmut Schumacher. Berlin/New York: de Gruyter.

WDG 1970–1977 = *Wörterbuch der deutschen Gegenwartssprache*. (6 band). Utg. Ruth Klappenbach/Wolfgang Steinitz. Berlin: Akademie Verlag.

ÖSTERGREN 1981 = Olof Östergren: *Nusvensk ordbok*. (5 band). Stockholm: Wahlström & Widstrand.

Annan litteratur

Bergenholtz, Henning 1994: Beispiele in Fachwörterbüchern. I: Burkhard Schaefer/Henning Bergenholtz (utg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr, 421–439.

Bergenholtz, Henning/Sven Tarp 1994: *Manual i fagleksikografi. Udarbejdelser av fagordbøger — problemer og løsningsforslag*. Herning: forlaget systime als.

Freund, Folke/Birger Sundqvist 1988: *Tysk Grammatik*. Stockholm: Natur och Kultur.

Grice, Paul H. 1975: Logic and Conversation. I: Peter Cole/Jerry L. Morgan: *Syntax and semantics, Vol. 3, Speech Acts*. New York/San Francisco/London Academic Press, 41–48.

- Grosse, Siegfried 1967: Zur Aktualität in den Beispielen in der Schulgrammatik. I: Benno von Wiese/Rudolf Henß (utg.): *Nationalismus in Germanistik und Dichtung*. Berlin: Erich Schmidt Verlag, 261–272.
- Hermanns, Fritz 1988: Das lexikographische Beispiel. Ein Beitrag zu seiner Theorie. I: Gisela Harras (utg.): *Das Wörterbuch. Artikel und Verweisstrukturen (= Sprache der Gegenwart 74)*. Düsseldorf: Schwann/Bagel, 161–195.
- Herslund, Michael, 1988: On Valence and Grammatical Relations. I: Finn Sørensen (utg.): *Valency. Three Studies in the Linking Power of Verbs*. København: Nyt Nordisk Forlag Arnold Busck, 3–34.
- Karlsson, Fred 1983: Prototypes as Models for Linguistic Structure. I: Fred Karlsson (utg.): *Papers of the Seventh Scandinavian Conference of Linguistics II*. University of Helsinki, Department of General Linguistics, Publications No. 10. Helsingfors: Helsingfors universitet, 595–605.
- Lindberg, Ebba 1976: *Beskrivande svensk grammatik* (under medverkan av Ove Oskarsson). Stockholm: AWE/GEBERS.
- Nikula, Henrik 1986: Wörterbuch und Kontext. Ein Beitrag zur Theorie des lexikalischen Beispiels. I: Albrecht Schöne (utg.): *Akten des VII. Internationalen Germanisten-Kongresses, Göttingen 1985, Band 3*. Tübingen: Max Niemeyer Verlag, 187–192.
- Nikula, Henrik 1988a: Lexikon und Wörterbuch. I: *Fachsprachen und Übersetzungstheorie VIII*. Vasa: Vasa universitet, 227–239.
- Nikula, Henrik 1988b: Zur Frage der Textualität von Beispielsätzen. I: *Neuphilologische Mitteilungen* 89:4, 483–488.
- Nikula, Henrik 1989: Zur Beschreibung der Verbumbegungen in Wörterbüchern des Deutschen. I: *Fachsprachen und Übersetzungstheorie IX*. Vasa: Vasa universitet, 139–151.
- Nikula, Henrik 1994: Wörterbuch und Textkorpus. I: *Fackspråk och översättningsteori XIV*. Vasa: Vasa universitet, 196–209.
- Nikula, Henrik 1995: Valenz und Bedeutung. I: Ludwig M. Eichinger/Hans Werner Eroms (utg.): *Dependenz und Valenz*. Symposium an der Universität Passau, 18.–21. 3. 1992. Hamburg: Buske Verlag (under tryckning).
- Pusch, Luise F. 1983: "Sie sah zu ihm auf wie zu einem Gott." Das Duden Bedeutungswörterbuch als Trivialroman. I: *Der Sprachdienst* 27, 135–142.
- Römer, Ruth 1973: Grammatiken, fast lustig zu lesen. I: *Linguistische Berichte* 28, 71–79.
- Sundqvist, Birger 1988: Die Heringe der Ostsee. I: *Moderna Språk* 82: 4, 340–341.

Övriga källor

- MK = Mannheimer Korpus der geschriebenen deutschen Gegenwartssprache (MK1 ca 2,2 milj. o. MK2 ca 0,3 milj. ordformer). Institut für deutsche Sprache; Mannheim.

Partikelverb och lexikon: kriterier för att finna lexikaliserade partikelverb

This paper is an investigation of criteria which may be useful in deciding which phrasal verbs should be included in a dictionary. Most thoroughly scrutinized are syntactic and semantic criteria, and the conclusions are that semantic criteria combining the degree of localistic meaning and the specificity of valency roles in a phrasal verb, are the most useful ones. This is due to the possibility of more or less quantifying the degree of lexicalization in the verb by using these semantic criteria.

Ur lexikografiskt perspektiv är kanske den intressantaste frågan om partikelverb vilka som är lexikaliserade,¹ och därmed bör definieras i lexikon, respektive vilka som kan betraktas som en regelstyrd kombination av verb och ett slags adverb, och därmed kan lämnas åt den grammatiska beskrivningen av språkets regelsystem.

I den här uppsatsen kommer jag att behandla olika kriterier för lexikalisering av partikelverb. Men dessförinnan kan det vara på sin plats att ge en preliminär definition av vad som menas med partikelverb.

Det finns två former av partikelverb — löst sammanfogade och fast sammansatta. Ett löst sammanfogat partikelverb är en kombination av ett verb och ett oböjligt, adverbliknande ord som uppbär huvudtrycket i frasen. Ett fast sammansatt partikelverb är en kombination av samma slags adverbliknande ord med verb, men i omvänd ordning och utan ordmellanrum. Partikeln uppbär även i detta fall huvudbetoningen.

Med en sådan definition räknas uttryck som *hålla med, köra ut, höra upp, upphöra* etc. till partikelverben, och det är väl gott och väl. Men även fraser som *komma hem, skriva maskin* och *köra bil* kan emellertid räknas dit. I sådana fall vill vi kanske hellre tala om kombinationer av verb och reguljära adverb respektive s.k. nakna substantiv. Låt oss se om vi kan minska definitionens omfång med några fler kriterier. Jag börjar med dem som har med uttryckens form, syntaktiskt och ordbildningsmässigt att göra, och övergår sedan till semantiska kriterier.

1 Position i satsen

En typisk lös partikel har sin position omedelbart efter huvudverbet i ett satsschema av Diderichsentyp. Det betyder att den hamnar före objektet vid transitiva verb, eller före det egentliga subjektet, då man skriver om intransitiva verb i "presenteringskonstruktion".

Hon har kört ut barnvagnen på gården
Det har sprungit ut några älgar på vägen

¹Med lexikaliserad menar jag 'som inte kan genereras av regler'. Enligt andra definitioner kan lexikaliserad betyda exempelvis 'som utgör en idiomatisk enhet' o.d.

Samma konstruktion är möjlig för exempelvis adverb som *hem*

Det har kommit hem några grannar på middag

men inte för nakna substantiv

*Det har skrivit maskin en sekreterare

Positionskriteriet kan alltså användas för att skilja ut ett visst slags kombinationer av verb och adverb från verbfraser med liknande betoning. Men hur är det med lexikalisering och positionskriteriet?

Partikelverb med tydligt lexikaliserad betydelse, som *hålla med*, har endast möjlighet att placera partikeln omedelbart efter huvud verbet:

Jag har aldrig hållit *med* dig

Partikelverb med en betydelse som verkar mer regelbunden, som *köra ut*, har samma möjlighet:

Hon har kört *ut* barnvagnen på gården

Men vid sidan av denna möjlighet kan partikeln placeras som ett vanligt adverbial:

Hon har kört barnvagnen *ut* på gården

Den positionella skillnaden mellan lexikaliserade och semantiskt regelbundna partikelverb tycks alltså vara att de senare vid sin sida har en konstruktion med ett homonymt adverb. Det rör sig nästan uteslutande om adverb som anger ett urval allmänna positioner eller riktningar. Det kan också gälla en homonym preposition, med samma slags betydelse, som inleder ett adverbial, som i paret:

Hon la *i* tvätten

Hon la tvätten *i* maskinen

Vid ett lexikaliserat partikelverb som *hålla med* har partikeln ingen annan placeringsmöjlighet:

*Jag har aldrig hållit dig *med*

Det kan tolkas som att partikelns betydelse har "smält samman med verbet". Men det finns en viss grad av sammansmältning med verbet även hos de mer regelbundna partikelverben i jämförelse med deras alternativa konstruktioner. När partikeln står närmast verbet fokuseras dess betydelse mer än när det rör sig om ett homonymt adverb. Jämför

Hon har kört ut barnvagnen på gården

med

Hon har kört barnvagnen ut på gården

I den första meningen ligger fokus, både fonetiskt och sematiskt, på *ut*. Det är det positionella resultatet som är viktigast, inte sättet att utföra handlingen.

I det andra exemplet är det tvärtom. Här ligger tonvikten i verbhandlingen mer på själva körandet. Målet är sekundärt. Denna skillnad verkar regelmässigt uppträda mellan liknande par av meningar, så det är enligt min definition inte fråga om lexikalisering av partikelverben.

Vad beträffar lexikalisering tycks positionskriteriet i kombination med en undersökning av eventuella homonyma alternativkonstruktioner alltså kunna användas för att skilja mellan lexikaliserade och semantiskt regelbundna partikelverb.

Men det kan inte skilja mellan regelbundna partikelverb och kombinationer av verb och adverb med annan betydelse än att ange allmänna positioner eller riktningar. De senare har nämligen också alternativa positioner:

Det har kommit *hem* några grannar på middag
 Det har kommit några grannar *hem* på middag

Om vi vill skilja dessa uttryck från partikelverb kan vi behöva se oss om efter fler kriterier. Ett ytterligare skäl till vårt intresse är att partikelverb ofta är homonyma, dvs. det vi har kallat regelbundna partikelverb ger ofta upphov till lexikaliserade partikelverb, och det är därför viktigt att "hålla ögonen" på dem i semantiskt avseende. (Mer om detta längre fram.) Detsamma gäller inte för andra kombinationer av verb och adverb, troligen därför att adverbets egen betydelse hos dem är mer specialiserad och varierad.

2 Fundamentering

Fundamentering sägs normalt beröra en satsdel, och det är hela satsdelen som flyttas fram. Hypotesen skulle då vara att partikelverb kan fundamenteras därför att de utgör en semantisk enhet, men inte kombinationer av verb och adverb, som behåller var sin betydelse. Det tycks stämma:

Höll med gjorde hon sällan
Backade ut gjorde hon lite vårdslöst ur garaget

men knappast

?*Kom hem* gjorde några grannar (på middag)

Det kan vara intressant att lägga märke till att det ibland är hela verbfrasen, och inte bara verbet, som fundamenteras:

Kom hem på middag gjorde några grannar

Därför kan kriteriet inte användas för att skilja lexikaliserade partikelverb från semantiskt regelbundna eller från fraser bestående av bara ett verb och ett allmänt rörelseadverb som är homonymt med en partikel. I ett exempel som:

Backade ut gjorde hon aldrig

kan fundamentet vara ett lexikaliserat partikelverb: *hon* gav sig *aldrig*, ett regelbundet partikelverb: *hon backade aldrig* ut, (*däremot in*) eller en verbfras *hon backade aldrig ut* (*utan hon körde framåt*).

3 Fast sammansättning

Ibland används kriteriet ”ska kunna ingå i fast sammansättning” som kriterium på partikelverb. Man tänker då på par som *laga till* — *tillaga* osv. Tanken är väl att fast sammansättning är ett yttre tecken på semantisk sammansmältning. En följd av detta blir att man betraktar alla partikelverb som mer eller mindre betydelsemässigt enhetliga. Frågan är nu hur det egentligen förhåller sig med detta.

I det följande skiljer jag på tre typer av sammanfogning av partikelverb: alltid fast sammansatta, varierande sammansatta, dvs. sådana som med samma betydelse kan uppträda både i lös och fast form, och alltid löst sammansatta.

Att fast sammansättning skulle kunna vara ett indicium på lexikalisering stöds av att många partikelverb som alltid är fast sammansatta också är starkt lexikaliserade, t.ex. *avskriv* ’föra från dagordningen’, *tilltala* ’känns behaglig för’, *nedkomma* ’föda’, *pågå* ’äga rum’.

Mot ett fullständigt samband mellan alltid fast sammansättning och stark lexikalisering talar semantiskt mer regelbundna partikelverb som *tilltala* ’börja tala (med)’, *påminna* ’ge minnet åter’ m.fl. Även dessa uppträder endast i fast sammansättning.

Vad beträffar partikelverb som kan uppträda i varierande sammansättningsform är de oftast, mot hypotesen, semantiskt ganska regelbundna: *tillaga* — *laga till*, *bortspringa* — *springa bort*, *påbygga* — *bygga på*, *uppäta* — *äta upp*, *nedkomma* — *komma ned*, *påtända* — *tända på*, *avsalta* — *salta av*.

Det finns ett stort antal partikelverb av denna typ, men deras fast sammansatta alternativ är inte semantiskt mer enhetliga utan istället tecken på högre stilnivå, fackspråkighet o.d.

Några fall av mer lexikaliserade partikelverb med varierande sammansättning finns, men då gäller lexikaliseringen båda sammansättningsformerna: *tillkomma* — *komma till* ’börja existera’, *tillhöra* — *höra till* ’klassificeras som’, *tillgripa* — *gripa till* ’börja använda (ngt kraftfullt/olämpligt)’.

Vissa partikelverb som bara förekommer i lös förbindelse kan för all del ofta vara semantiskt regelbundna: *jogga på*, *sitta vid*, *ligga i (i vattnet)*, *skrika till*, *stå på (på spisen)*, men andra är lexikaliserade: *dra på* ’öka farten’, *hålla med* ’instämna’, *tycka om* ’gilla’.

Det finns alltså inga riktigt klara linjer vad beträffar sammansättningsstyp och lexikaliseringsgrad. Fasta sammansättningar kan ju bildas av andra skäl än att ange semantisk enhetlighet, t.ex. som ellipser för längre uttryck i texter eller som stilmarkör.² Och det är ju inte heller bara partikelverb som kan bilda sammansättning, utan alla slags fraser.

Sammanfattningsvis kan man om de mer formella kriterierna säga, att positionskriteriet i kombination med eventuell förekomst av parallella konstruktioner med adverbial kan skilja lexikaliserade partikelverb från mer regelbundna. Alltid fast sammansättning är oftast ett indicium på lexikalisering, alltid lös sammansättning kan vara det, men varierande sammansättning är tvärtom ett indicium på regelbunden betydelse. Fundamentering, slutligen, kan i viss mån användas för att skilja partikelverb och deras parallellkonstruktioner från andra typer av verb-adverbkonstruktioner.

²Det verkar dock finnas en tendens till semantisk differentiering på så sätt att alltid fast sammansatta partikelverb har alla slag av lexikaliserad betydelse, alltid lösa innefattar en grupp lexikaliserade verb med vissa slag av betydelse, och de varierande är mest regelbundna i sin betydelse. Dessa förhållanden tänker jag diskutera vid ett annat tillfälle.

4 Synonymer och ekvivalenter

Besläktat med sammansättningskriteriet är kriteriet "ett partikelverb ska kunna omskrivas med en synonym". Det måste förstås gälla lexikaliserade partikelverb, och fungerar naturligtvis bra i de fall då en synonym råkar finnas. (Det kan ju sällan bli tal om annat än ungefärliga ekvivalenter; äkta synonymi är ju ovanlig i naturliga språk.) Också förekomst av osammansatt ekvivalent på ett annat språk kan användas som kriterium för att urskilja lexikaliserade partikelverb i svenskan, men samma restriktioner gäller för dem — det är tur om de finns, och om man råkar känna till dem, men de kan inte utnyttjas systematiskt.

5 Semantiska kriterier

Frågan är vilka semantiska kriterier som bäst lämpar sig för att belysa skillnaderna mellan lexikaliserade och semantiskt regelbundna partikelverb.

I en analys av partikelverb, som utgår från deras sammansättningsform, har distinktionen mellan **konkret** och **abstrakt** betydelse använts (Johannisson 1954). Man har därefter allmänt menat att skillnaden mellan ett löst sammansatt *gå av* 'brytas itu' och ett fast sammansatt *avgå* 'lämna viss tjänst eller post' kan förklaras i dessa termer.

Med vår inriktning på lexikaliseringsfrågan kunde man tänka sig att ställa upp hypotesen "abstrakt betydelse är ett indicium på lexikalisering".

Men ett problem i sammanhanget är att definiera vad man menar med konkret och abstrakt betydelse när det gäller verb. För substantiv är distinktionen mycket klarare, och man kunde därför tänka sig att låta verbdefinitionen gälla verbens valensled (och kanske också tilläggsled), som ofta fylls av substantiviska referenter.

En definition av vad som avses med ett "konkret verb" kunde då se ut på följande sätt: ett konkret verb är ett verb där effekterna av verbhandlingen kan avläsas på minst en av innehavarna av valensleden med hjälp av minst ett av människans fem sinnen. Ett abstrakt verb har ingen sådan direkt iakttagbar verkan på valensrollsinnhavaren.

Ett exempel: man kan se och höra att någon springer, men man kan inte separera själva springandet från denna någon och från platsen där han eller hon springer.

Vissa verbhandlingar kan man dock direkt uppfatta med sinnena. Det gäller då främst verb som saknar valensled, och därför måste konstrueras med formellt subjekt (eller med innehållssubjekt), t.ex. *blåsa, snöa*. Dessa verb kan också kallas konkreta.

Är då ett verb som *avgå*, i t.ex. *hon avgick från ordförandeposten*, konkret eller abstrakt. Enligt vår definition kan det vara konkret, eftersom *hon* faktiskt rent fysiskt kan avstå från en viss sittplats i ett styrelserum, och enligt hypotesen borde det då vara semantiskt regelbundet. Är det det? Tja, det beror på hur bokstavligt eller metaforiskt *gå* tolkas och hur konkret rumsbetydelse *av* antas ha.

Som ett andra exempel kan vi ta *nedkomma* — *komma ned*. Det har den konkreta bokstavliga betydelsen 'komma (uppifrån och röra sig) nedåt'. (Den fasta sammansättningen har mycket hög stilnivå, som t.ex. i *änglarna nedkommo från himmelen*.) Det stämmer med hypotesen om att verb med konkret betydelse är semantiskt regelbundna. Men det lexikaliserade *nedkomma* 'föda barn' är också verkligt konkret, enligt vår definition. Vad som skiljer det från *komma ned* är inte konkretionsgraden utan föreställningen om en rörelseriktning.

Och då är vi tillbaka till det som nämnts redan tidigare i uppsatsen — typiskt för många verb-partikelkonstruktioner är att det ingår något slags **rumsföreställning** i dem. Det handlar oftast om rörelseriktning eller position. Partikeln anger rörelseriktning i partikelverb som *braka in* och *springa bort*. I det andra exemplet bidrar verbet också med en specifik rörelsebetydelse, medan i det första exemplet verbet enbart bidrar med att beskriva sättet för rörelsen.

I exemplet med ett lexikaliserat partikelverb som *nedkomma (med ett barn)* kan man också se att det är något speciellt med **valensrollerna** till detta verb, särskilt om man jämför det med det konkreta rumsliga *komma ner*. Vad som helst som befinner sig i ett högt läge kan komma ner, men *nedkomma* kan bara en gravid kvinna göra. Dessa två iakttagelser får bli utgångspunkt för två semantiska hypoteser om lexikalisering av partikelverb:

1. Partikelverb med riktning- eller positionsbetydelse är semantiskt regelbundna. Alla andra partikelverb är lexikaliserade, i olika grad beroende på om de har kvar rester av rumsbetydelse, har metaforisk rumsbetydelse eller helt saknar rumsbetydelse.
2. Partikelverb som specificerar sina valensrollsinnehavare mer än vad motsvarande simplexverb plus adverb gör, är lexikaliserade — mer ju större specifikationsgraden är.³

Tillägg eller minskning av antalet valensroller, liksom förändring av rolltyperna, är också indicier på lexikalisering. Det är detta som ibland hänvisas till som ”partikelns förmåga att transitivera”, men i själva verket handlar det om ny verbbetydelse, som sedan får konsekvenser för rolluppsättningen. Jämför t.ex. *hon arbetar* med *hon arbetar fram ett nytt förslag*. Att arbeta fram ett förslag är en väldigt specifik sorts aktivitet jämfört med att arbeta i allmänhet. Partiklar kan f.ö. också ”intransitivra”, av liknande semantiska skäl. Jämför t.ex. *höra (ngt)* med *hör upp!* Det är inte samma verbhandling som beskrivs.

Nu följer en genomgång av de partikelverb som redan nämnts i uppsatsen utifrån de semantiska lexikaliseringshypoteserna.

5.1 Partikelverb med konkret rumsbetydelse

Partikelverben *köra ut*, *springa bort*, *bygga på*, *ligga i* (i vattnet), *stå på* (på spisen), *komma ner*, *backa ut* och *braka in* har konkret rumsbetydelse. De är också alla semantiskt regelbundna, vilket stämmer med hypotesen.

5.2 Partikelverb med försvagad rumsbetydelse

Partikelverben *avsalta*, *tilltala* ’börja tala med’, *avgå* (om tåg o.d.) och *gå av* har kvar något slags rumsbetydelse, men andra element i verbhandlingen är lika viktiga. De har också något specialiserade valensroller, och får väl sägas vara något lexikaliserade.

³Inte alla partikelverb har något simplexverb plus adverb att jämföra med. Då får enbart hög specifikationsgrad för valensrollerna utgöra kriterium.

5.3 Partikelverb med metaforisk rumsbetydelse

I partikelverben *påminna*, *avskriva* 'föra från dagordningen', *arbeta fram*, *avgå* (från en post) är väl rumsföreställningen närmast metaforisk. De har, liksom gruppen ovan, något specificerade valensroller, och är väl lexikaliserade i ungefär lika hög grad.

5.4 Partikelverb med aktionsartsbetydelse

Om metaforiken i rumsbetydelsen abstraheras ännu ett steg hamnar vi i aktionsarternas sfär. Aktionsarter kan nämligen betraktas som olika slag av grammatikaliserad rumsbetydelse (se t.ex. Anderson 1971).

En enkel aktionsartsmodell för dynamiska verb, dvs. verb som beskriver handlingar, innehåller två huvudklasser — **gränsrelaterade verb** och **icke-gränsrelaterade verb**. Med begreppet *gräns* avses en "naturlig" gräns för handlingen. Om gränsen inte har uppnåtts kan handlingen inte sägas ha ägt rum, när gränsen har uppnåtts uppfattas handlingen som fullbordad (Andersson 1972).

De flesta verb är neutrala i förhållande till gränsrelatering. Först när de kombineras med ett visst objekt, adverbial eller subjekt kan man avgöra deras typ av gränsrelatering.⁴

Ett enkelt kriterium på om ett verb, en verbfras eller en sats är gränsrelaterad eller inte är kombinerbarhet med tidsadverbial som anger duration av två olika slag.

Icke-gränsrelaterade:

sova, sitta, bo, irra omkring, bygga hus + *i* flera år; * *på* två år

Gränsrelaterade:

somna, sätta sig, mogna, springa bort, bygga ett hus + *på* två timmar; * *i* flera timmar

Aktionsarterna kan, och har, finindelats på olika sätt. En sådan indelning har särskild relevans för partikelverben och ska därför nämnas här. Det gäller förekomsten av en föreställning om tidsåtgång eller ej för verbhandlingen. I det senare fallet kan verbet inte kombineras med något durationellt adverbial utan att få förändrad betydelse, som t.ex. *slockna*. Att *slockna* "tar ingen tid". Om man ändå kombinerar verbet med ett tidsadverbial, syftar detta på händelserna omedelbart före själva *slocknandet*, dvs. på en annan verbhandling: (*lampan började blinka och*) *slocknade på en minut*. Verb av typen *slockna* kallas ibland **momentana**, övriga (dynamiska) verb **durativa**. Verb kan också vara lexikalt neutrala mellan momentana och durativa.

Resten av de partikelverb som nämnts tidigare i uppsatsen kan delas upp efter sin aktionsart på följande sätt:

Icke-gränsrelaterade (durativa):

jogga på, sitta vid, hålla med, tycka om, pågå, tilltala 'vara njutbar för' ligga i 'arbeta hårt'.

Gränsrelaterade momentana:

⁴Också de verb som uppges ha en viss aktionsart kan få denna förändrad av kontexten, men det krävs då normalt att man går utanför meningsramen.

höra upp, tända på, tillgripa

Gränsrelaterade durativa:

avsluta, tillaga, äta upp, nedkomma 'föda', komma till 'skapas', dra på 'öka farten'.

Enligt hypoteserna om bevarad rumsföreställning borde partikelverben i den här gruppen vara mest lexikaliserade av alla, men av exemplet att döma gäller detta bara vissa av dem. Andra verkar mer eller mindre regelbundna. Vad är det då fråga om?

Jo, svenska partiklar verkar kunna användas grammatiskt helt regelbundet för att ge ett neutralt verb en viss aktionsart, eller för att förstärka aktionsarten hos andra. Exempel på detta ser vi när *jogga* kombineras med partikeln *på*, som fungerar som icke-gränsmarkör. *På*, liksom i viss mån *i* och *vid*, kan användas för icke-gränsmarkering tillsammans med många verb, t.ex. *jobba på*, *cykla på*, *prata på*. På liknande sätt kan partiklar regelbundet användas för att bestämma eller förstärka gränsrelatering. Så sker t.ex. med *av* i *avsluta*, *upp* i *äta upp*, *till* i *tillaga* och *på* i *tända på* (också initialt kan en gräns finnas).

Men det ligger ändå något i idén om att försvunnen rumsföreställning är ett indicium på lexikalisering. Visserligen har alla svenska partikelverb, till skillnad från simplexverben, en bestämd aktionsart, men när rumsföreställningen är helt borta, är de också lexikaliserade. Och den största "risken" att rumsföreställningen försvinner är när det bara är aktionsarten som uppbär den. Alltså är det i gruppen partikelverb med enbart aktionsartsbetydelse som vi kan vänta oss de flesta lexikaliseringarna. Bland exemplen ovan kan nämnas *hålla med*, *tycka om*, *pågå*, *tilltala* 'vara angenäm för', *nedkomma* 'föda', *komma till* 'skapas', *dra på* 'öka farten' och *säga till*. Deras valensroller är mycket specificerade.

Sammanfattningsvis kan man om de semantiska kriterierna "bevarad rumsbetydelse" och "specificerad valensroll" säga att de är bra indikatorer på lexikalisering. De kan t.o.m. användas för en försiktig gradering av lexikaliseringen — ju mer konkret rumsbetydelse och ju mindre specificerade valensroller, desto mer regelbunden betydelse. (Och samtidigt som man använder dessa kriterier klarnar definitionen av partikelverbet.)

Att de semantiska kriterierna kan användas för en gradering av lexikaliseringsgrad underlättar urvalet av partikelverb för ordböcker med olika omfång. Att åtminstone några partikelverb bör finnas med som stickord i en ordbok ser jag som en självklarhet.

Det finns dessutom grammatiska aktionsartspartiklar i svenskan, och de bör ges definitioner i lexikon, men i övrigt lämnas åt den grammatiska beskrivningen. Ett problem är att många av dem är homonyma, t.ex. kan ge både icke-gränsrelatering och gränsrelatering, som *på* ovan. De är också homonyma med en rumslig, adverbial, betydelse. Men lexikologer är ju vana att handskas med homonymi.

Litteratur

- Anderson, J. M. 1971: *The grammar of case. Towards a localistic theory*. Cambridge Studies in Linguistics 4. Cambridge: Cambridge University Press.
- Andersson, S.-G. 1972: Aktionslätät im Deutschen. Eine Untersuchung unter Vergleich mit dem russischen Aspektsystem. I: *Die Kategorien Aspekt und Aktionsart im Russischen und im Deutschen*. Studia Germanistica Uppsaliensia 10. Stockholm: Almqvist & Wiksell International.
- Johannisson, T. 1954: Om sammansatta verb i svenskan. I: *Nysvenska studier* 34, 162–181.

Korpus og seddelarkiv, fredelig sameksistens mellom det beste og det gode?

This paper focuses on how to save the information in a traditional citation archive for the computerized world. The sample archive is a citation archive at the Department of Lexicography at the University of Oslo, Norway. Originally the archive was to be keyed and given a SGML markup. Due to reasons discussed in the paper this approach was abandoned. The archive is now converted into a indexed facsimile data base which seems to be a better solution both from a scientific and from an economic point of view. The research, the program development and the actual conversion of the information in the archive forms a sub project of the Documentation project, which is a collaboration between the four Norwegian universities.

1 Innledning

Seddelarkivet er den tradisjonelle systematiske metoden for å samle belegg og opplysninger om ord til bruk i redigeringen av en ordbok. Korpuset og de tilhørende KWIC-konkordanser med ulike ordningskriterier er blitt muligjort i stor stil ved hjelp av datamaskinene. Den nye teknikken har skapt et skille i bruk av materialet i ordboksredigeringen. For noen står seddelarkivet som en tilfeldig samling opplysninger eller deler av en konkordans, og arkivet er kun et eksempel på hva gårdsdagens teknikk kunne produsere. For andre representerer seddelarkivet en skattekiste der hver seddel er valgt med omhu, mens det med et korpus kan være vanskelig å få med det spesielle og sjeldne språket. Etter forfatterens mening er korpusmetoden en åpenbart mer rasjonell måte å samle språklige data på.

I denne artikkelen vil det imidlertid bli fokusert på hvorledes en kan redde informasjonen i et tradisjonelt seddelarkiv over i den datamaskinelle verden. Det vil bli tatt utgangspunkt i Dokumentasjonsprosjektets arbeid med det 3,2 millioner store seddelarkivet til Norsk Ordbok ved Avdeling for leksikografi ved Universitet i Oslo. Jeg har tidligere gjort rede for prosjektet og for arbeidets gang i form av foredrag ved leksikografikonferansene i Oslo og København, Ore (1991) og Ore (1993).

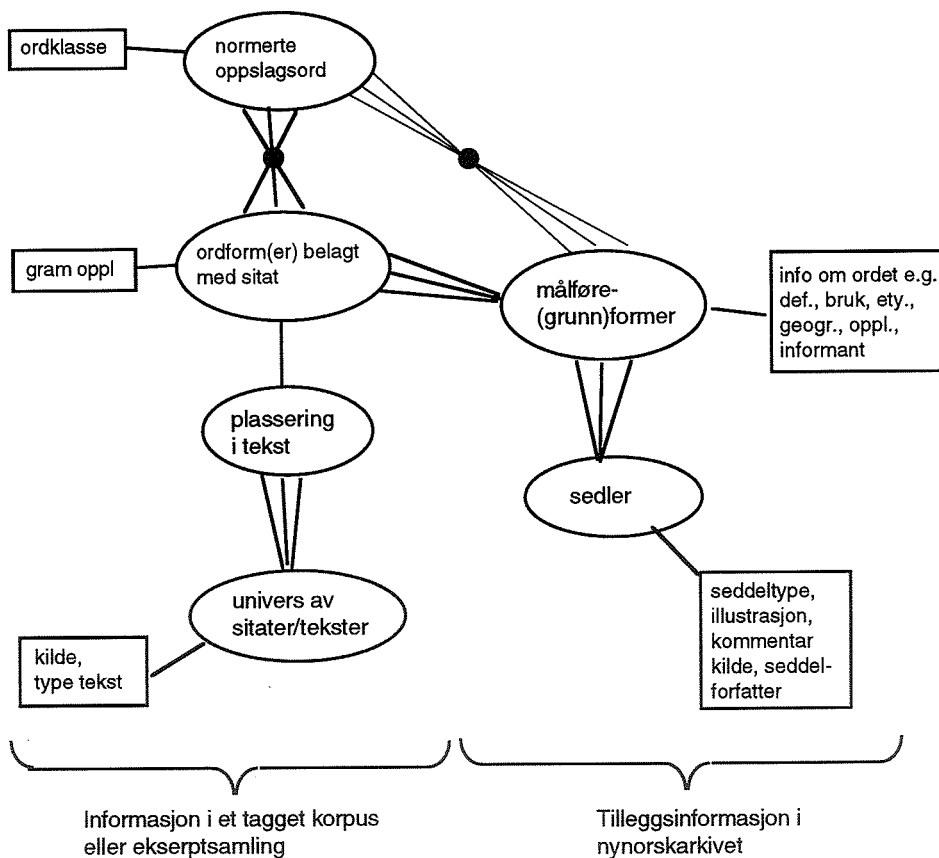
2 Det opprinnelige prosjektet

Informasjonen i et seddelarkiv kan på mange måter sammenlignes med den vi finner i et tagget tekstkorpus. I arkivet er det for et utvalg ord fra et utvalg tekstbrokker gitt opplysninger om bøyningsformen, ordets rot og annet. I et tagget tekstkorpus finner en gjerne de samme opplysningene knyttet til hvert enkelt av ordene i større tekstfragmenter (fra 25 sider løpende tekst og oppover). Denne observasjonen lå til grunn da vi planla konverteringen av nynorskarkivet. Vi ønsket å erstatte de rene ekserptsedlene med løpende elektronisk tekst :

De resterende sedlene skulle skrives inn og SGML-kodes (Sperberg-McQueen et al. 1994) slik at de kunne lastes inn i et databasesystem slik det er skissert i figur 1.

På bakgrunn av tidsstudier utført ved Nynorskavdelingen høsten 1989, ble avskrivingsarbeid og korrektur anslått til om lag 250 årsverk. Nynorskprosjektet ville dermed utgjøre en tredjedel av Dokumentasjonsprosjektet.

Konverteringsarbeidet begynte sommeren 1992 i Sigdal, Lier og Røyken, og i Mo i Rana i oktober samme år. Ved årsskiftet 1992/1993 var alt konverteringsarbeid flyttet til Mo i Rana. Konverteringsarbeidet ble gjort ved bruk av skannere og programmer for optisk lesning for de løpende tekstene. Sedlene ble skrevet av og kodet ved bruk av et vanlig tekstbehandlingsprogram.



Figur 1: Skisse av et kombinert seddelarkiv og tagget tekstkorpus

2.1 Problemer forbundet med den opprinnelige konverteringsmodellen

Det vil alltid være en avveining mellom hvilket tempo en ønsker, og kompleksiteten i arbeidet som skal gjøres. I 1993 erfarte vi at arbeidet gjort i Mo i Rana hadde en akseptabel

kvalitet, men at det gikk for langsomt. Hastigheten var om lag halvparten av den forventede.

Nynorskavdelingen tillot ikke at vi sendte originalsedlene til Mo i Rana. Av den grunn måtte sedlene fotokopieres. Fra innskrivingshold ble det hevdet at fotokopiene var så dårlige at det sinket arbeidet kraftig. Fra andre hold ble det hevdet at kodeskjemaet var for komplekst.

leta v 060605 "

leta (leta) v. tr. (a-a).

' farga, gjeva farge '

" leta gad'm herki letargam (letag a d'm)

1 "eg veit ikkje kor da er leta (um ah depr o. l.)

VossNLid NO

Figur 2: Faksimile av seddel

Det er klart at det var en god del å hente ved enten å bruke originalsedlene eller bedre kopier. Men kopieringen og nummerering av sedlene er ganske kostbar, om lag 50 øre pr. seddel eller kr. 1 600 000 for hele arkivet. Det var derfor lite tiltalende å investere mer i en ren produksjon av papirkopier av sedlene som i neste omgang skulle kastes.

Kodingen av kortene var i løpet av 1993 blitt forenklet, og det er nok mulig at tempoet kunne ha økt ved at innskriverne kodet mindre mens de faglærte assistentene kodet mer under korrekturlesingen. På den annen side viste det seg at innskriverne gjorde relativt lite kodefeil. En ytterligere forenkling av kodesystemet ville etter min mening ha undergravd hensikten med å konvertere sedlene til søkbar elektronisk tekst.

Det er neppe riktig å si at dårlige kopier og komplisert tagging var hele årsaken til den lave produksjonen. Man må også huske på at oppstarten av registreringen av nynorsk materialet falt sammen med oppstarten av en helt ny type sysselsettingstiltak. Det har altså vært satt i gang to helt nye prosjekter på en gang.

På bakgrunn av tidsstudier utført ved Nynorskavdelingen høsten 1989, ble avskrivingsarbeid og korrektur anslått til om lag 250 årsverk. Tatt i betraktning prosjektets størrelse var det umulig å skaffe ressurser til veie for å kunne fullføre det med halv hastighet. Det var derfor nødvendig å foreta en gjennomgang av prosjektet for å se hvordan det var mulig å få et akseptabelt resultat med de tilgjengelige ressurser.

3 En omlegging av registreringsarbeidet

3.1 Tekst og faksimile

Med elektronisk tekst menes det her den måten tekst lagres på i en datamaskin i for eksempel WordPerfect eller Microsoft Word. Hvert tegn i en tekst lagres som et tall på disken. Det er en kompakt og enkel lagringsmetode for vanlig tekst. Metoden gjør det mulig å søke etter bestemte tegnstrenger og å redigere teksten slik vi alle kjenner det fra vårt daglige arbeid med tekstbehandlere.

Hvis målet kun er å lese teksten, er det strengt tatt ikke nødvendig å lagre annet enn et bilde av teksten. Et bilde av en teksts side tar imidlertid mye mer lagringsplass enn en elektronisk tekstversjon, i beste fall om lag 100 ganger så stor plass. Men dersom selve innleggelsesprosessen for bildene er hurtig og rimeligere enn avskrift, kan kostnadene ved denne ekstra lagringsplassen lett forsvares.

Da Dokumentasjonsprosjektet ble startet, fantes det verken tilgjengelige høyhastighetsskannere eller tilstrekkelig rimelig lagringsplass. En inntasting av innholdet på sedlene var derfor det eneste mulige alternativet. I løpet av de seneste årene har imidlertid dette bildet endret seg totalt. Våren 1994 viste det seg at det var billigere å få skannet og lagret elektroniske faksimiler av alle sedlene enn å gjennomføre den nevnte utsorteringen og fotokopieringen av sedlene. Lagringsplass er fremdeles relativt kostbar, men vil antakelig innen prosjektets avslutning i 1997 være så rimelig at det er mulig å la alle de 3,2 millioner sedlene ligge „on line“ samtidig.

```

<NSET NR=60605>
<OPPF GRM=v>leta</OPPF>
<ORDF>lita<ORDF>léta</ORDF>
<KOM>tr</KOM><BFORM>a - a
<DEF>farga, gjeva farga
<SIT T=USET>
<ORDFS GR=inf>líta</ORDFS>gad'n<KOM>heril litargarn (lítagad'n) n.
</KOM></SIT>
<SIT T=SETN>eg veit ikje kor da èr <ORDFS GR=prp>líta</ORDFS><FORK>um eit
dyr o.l.</FORK></SIT></ORDF>
<KJEL>VossNLid NO

```

Figur 3: Kodet versjon av seddelen vist i figur 2

3.2 Tekstdatabase eller tekst/bildedatabase

Nynorsksedlene er laget av personer med og uten filologisk skolering over en 60 års periode og er svært forskjellige. Sedlenes eneste fellestrekk synes å være at de har et oppslag, en midtdel og en kilde. Det har vært hevdet at innskrivningen nettopp bør følge dette skjemaet. Det er lett å lære og enkelt å følge. Med dette skjemaet vil innholdet på sedlene bli tilgjengelig som søkbar elektronisk tekst. I tillegg vil oppslagsord, ordklasse og kilde være innganger til materialet.

Virksomheten i Nynorskprosjektet de to siste årene har imidlertid gitt en rekke erfaringer om problemer i registreringsarbeidet og om forhold i sedlenes utforming. På denne bakgrunn kan man spørre seg om det er verdt de ekstra kostnadene det er forbundet med å gjøre teksten på sedlene elektronisk tilgjengelig som elektronisk tekst i forhold til de lavere kostnadene ved

å lage en faksimiledatabase med oppslagsord, ordklasse og kilde som søkenøkler. Under er det listet opp seks sentrale punkter som taler mot en ren tekstbase med eller uten omfattende tagging:

1. Taggeskjemaet dekker innholdet på sedlene, men mange av sedlene er så lite systematisk satt opp, at en tagging krever en total innholdsfortolkning av hva som står på seddelen.
2. Det finnes flere lydskriftssystemer og mange ikke konsistente forsøk på slike, på sedlene.
3. En rekke av sedlene er ført med vanskelig leselig hånd. En avskrift blir dermed å likne med manuskripttranskripsjon.
4. Mange av sedlene i arkivet er det man kan kalle „huskelapp“-sedler. De inneholder kun informasjon om at det eksisterer et ord slik og sånn og hvor det er funnet brukt.
5. Om lag 25% av sedlene er basert på ekserpter fra ordlister, ordsamlinger og fra tett strekede bøker. I tillegg til det normerte oppslagsordet inneholder sedlene sjelden annen informasjon enn sitatet.
6. En rekke av sedlene har illustrasjoner.

Punktene 1–3 dreier seg om den formaltekniske kvaliteten på sedlene. Ved å registrere sedlene ved ord, ordklasse, ordform og kilde er man sikret at det alltid er mulig å hente frem originalsedlene og lese teksten slik den står. For „huskelapp“-sedlene vil man få registrert nettopp den informasjon de er ment å gi, nemlig koblingen mellom ord og kilde.

Det er liten hensikt i å lagre lydskrift som elektronisk tekst dersom det ikke finnes en konsistent definisjon av hvilke lyder tegnene angir. Det er vanskelig å formulere fornuftige søk i noe man ikke vet hva er. Det synes derfor som en bedre løsning å kunne hente bilder av de sedlene som inneholder lydskrift, eventuelt merke sedlene i basen slik at lydskriftsedler senere kan tas ut og behandles separat. På tilsvarende måte er det mulig å hente ut faksimiler av sedler etter en eller flere av inngangene oppslagsord, ordklasse, til dels også heimfesting, skriftlig kilde, tidspunkt for belegget og oppskriver (NO-medarbeider). Faksimilene av disse sedlene brukes så som punchegrunnlag for en utvidet registrering av seddelopplysningene der dette er ønskelig.

I de senere år har datateknikken muliggjort en mye mer effektiv oppbygging av den informasjonen som et seddelarkiv representerer. Elektronisk lesning av tekst (OCR), konkordansprogrammer og hjelpemidler for (halv-)automatisk markering av grammatisk informasjon på ord i løpende tekst kan nevnes. Det er derfor grunn til å vurdere om en avskrift av sitatene på sedlene nevnt under punkt 5, i noe fall ville vært formålstjenlig.

3.3 Registreringsarbeidet i dag

Beslutningen om å gå over til en faksimiledatabase ble fattet sommeren 1994. I løpet av det siste året er samtlige 3,2 millioner sedler sendt til Kodak Norge A/S for innskanning og mikrofilming.

I dag foregår registreringen av Nynorskarkivet ved hjelp av et spesiallaget program. Innskriverne får vist et seddelbilde på skjermen og skriver så inn oppslagsordet, ordklassen og kilden slik de klarer å tyde den. Deretter søker de i en kildeliste etter den „offisielle“ betegnelsen på kilden og kobler denne mot seddelbildet. Resultatet er at hver seddel får knyttet til seg et oppslagsord, en ordklasse og en standardisert kildereferanse og også en såkalt „NO-medarbeider“ (seddelskriver for Norsk Ordbok) der dette er oppgitt.

I forbindelse med fotokopieringen skulle disse rene ekserptsedlene uten tilleggsinformasjon og sedler basert på eldre ordbøker og ordlister bli frasortert. Underveis ble det klart at en slik utvalgsmetode har praktiske ulemper idet utvelgelsen av sedler tar lang tid. Å treffe riktige valg er vanskelig uten at arbeidet utføres av personer med inngående kunnskap om hele arkivet. En slik kunnskap er neppe å oppdrive for et arkiv bygd opp over 60 år av hundrevis av personer. Av den årsaken går nå alle sedlene gjennom innskrivningsprosessen. Det ville rett og slett koste for mye å foreta en sortering.

4 Konklusjon

4.1 Hva mister vi?

Faksimilemetoden gjør det umulig å søke i teksten på seddelen, da særlig etter ekstra informasjon om uttale, bruk, definisjon, bøyning og etter spesielle konstruksjoner og former i sitatene.

Det er intet å gjøre med sitatene. De kan kun nås gjennom kilder og oppslagsord. Men her kan man merke seg at mange av de litterære beleggene i arkivet er valgt ut etter mekaniske metoder. Enkelte leksikografer har valgt kun en forekomst fra hver bok og da gjerne den første regnet fra side 1. Beleggene er således ikke nødvendigvis de mest interessante. Videre har man ønsket å finne de ord som ikke forekommer ofte. Det kan her nevnes at under bokstaven A er det 25 sedler fra Arne Garborgs roman *Bondestudentar*. Om lag halvparten av disse er eksempler på ord som forekommer bare én gang i hele romanen. Det er videre i overkant av 120 lemma som begynner på A i denne romanen. Slik kan en muligens argumentere for at arkivet ikke kan betraktes som en sitatsamling som gir et representativt bilde av de omgivelser et ord blir brukt i.

Hva angår de andre kategoriene, viser det seg at en stor del av de sedlene som inneholder slik informasjon, kan gjenfinnes ved hjelp av kilde- og medarbeider-opplysningene. En grov måte å finne disse „interessante“ sedlene på, er ved å søke etter sedler uten litterær kilde og som ikke kommer fra visse deler av Norsk målførearkiv. Denne hypotesen har vi verifisert ved å bruke de 120 000 sedlene som ble skrevet av og fullkodet som en fasit. Denne fasiten viser oss også at stedfesting av ordopplysninger på sedlene i over 80% av tilfellene er lik seddelskriverens hjemtrakter. Det er klart at slike grove mål ikke er fullgodt med en total gjennomgang av sedlene.

4.2 Hva vi har

De erfaringene som er høstet de siste årene, peker entydig i retning av at det er riktig å nøye seg med å lage en indeksert faksimilebase av sedlene supplert med en fullstendig innskriving av sedlene til noen få utvalgte medarbeidere. Rett nok finnes det ikke søkemuligheter for

bildene. Men selv bare den enkle registreringen av oppslagsord, ordklasse og kilde for hver seddel gir mange nye innganger til materialet. Arkivet kan dessuten brukes uansett om man sitter i Volda, Tynset, Oslo, Kirkenes eller Minnesota. Det kan om ønskelig også publiseres som 75 CD-plater.

Dette bør etter forfatterens mening representere den endelige avslutningen av arbeidet med å vedlikeholde seddelarkivet slik vi kjenner det i dag. I fremtiden burde man konsentrere seg om å bygge opp leksikalske databaser basert på elektroniske tekster og om mulig en systematisk landsomfattende undersøkelse av talemålet. De frivillige medarbeidernes virksomhet som nær sagt trøffelsvin kan man vel aldri klare seg uten. Men jeg synes det er overordentlig viktig at en slik virksomhet foregår etter mest mulig standardiserte metoder som kan gjøre det mulig å etterprøve resultatene.

Litteratur

- Ore, C.-E. 1991: Dokumentasjonsprosjektet ved Det historisk-filosofiske fakultet, Universitet i Oslo. I: R. V. Fjeld (udg.): *Nordiske studier i leksikografi*. Rapport fra konferanse om leksikografi i Norden, 28.-31. mai 1991. Oslo: NFL, 403-408.
- Ore, C.-E. 1993: Blant fire millioner sedler, En situasjonsrapport fra Dokumentasjonsprosjektet. I: A. Garde/P. Jarvad (udg.): *Nordiske studier i leksikografi*. Rapport fra Konferanse om Leksikografi i Norden, 11.-14. maj 1993. København: NFL, 243-247.
- Sperberg-McQueen C. M./L. Bernard (eds.) 1994: *Guidelines for the Encoding and Interchange of Machine-Readable Texts (TEI P3)*. Chicago/Oxford.

Bo Ralph

Ordböcker som språkmodeller

It has sometimes been claimed that dictionaries which reflect the mental lexicon as faithfully as possible are likely to be more functional than dictionaries which do not. In this contribution, some fundamental properties of dictionaries which tend to create problems in this connexion are discussed. In particular, the way entry forms are traditionally selected in dictionaries and the lemmatization concept in itself give rise to asymmetries alien to the information network of the mental lexicon. However, lexicographers can contribute considerably when a realistic model of the lexicon is being established.

Det är normalt i vetenskapliga sammanhang att olika praktiska tillämpningar är sekundära i förhållande till (den primära) teoribildningen. Tillämpningarna är då inte bara logiskt sekundära, vilket är rimligt, utan också reellt, vilket inte är lika självklart. De kan alltså betraktas som senare uppkomna praktiska avknoppningar till redan existerande, vetenskapligt väletablerade forskningsdiscipliner. Detta är normalfallet inom exempelvis tekniska och medicinska vetenskapsgränar, och det kan nog sägas vara ett vetenskapligt ideal i den västerländska forskningstraditionen.

Ser man noga på saken framstår emellertid inte denna bild lika tydligt inom de humanistiska forskningsgrenarna. Många discipliner har där så småningom kommit att fördjupa en från början ganska svagt utvecklad teoretisk profil. Vår egen disciplin, lexikografen, uppvisar denna utvecklingsgång. Ett mera målmedvetet teoriarbete har där först efterhand utvecklats ur en redan väl fungerande praktisk verksamhet.

1 Lexikaliskt teoriarbete

Den teoribildning som är relevant för det lexikografiska arbetet är av två grundläggande slag: specifikt lexikografisk teori kring själva det lexikografiska arbetet å ena sidan och allmän språkvetenskaplig teori med direkt tillämpning inom den lexikografiska verksamheten å den andra. Teori kring lexikografi har naturligtvis lång tradition på det viset att praktiskt inriktade lexikografer teoretiserat i någon mening kring sitt arbete. Det kan gälla urvalsprinciper vid informationsuppbyggnaden, organisationen av denna information, utformningen i förhållande till tänkta målgrupper etc. Emellertid har de teoretiska övervägandena vanligen redovisats endast implicit, genom de konkreta resultat som de teoretiskt baserade besluten manifesterats i. Mera sällan har den teoretiska grunden för de underliggande resonemangen redovisats explicit i inledningar och förord. En monografisk framställning som Svensén (1987) med principiella lexikografiska överväganden var också närmast unik när den kom.

Lexikografiska ställningstaganden som en reflex av den språkvetenskapliga debatten är på det hela taget ett ganska nytt fenomen. Det är väl korrekt att säga att det är först när ordböcker på senare tid tillkommit i anslutning till akademiska miljöer som språkvetenskapliga argument mera uttryckligen lagts till grund för lexikografiska beslut, framför allt sedan den teoretiskt mycket genomarbetade *Longman's Dictionary of Contemporary English* (1978) visat vägen.

Dessförinnan har vi haft en relativt lång tradition av i första hand kommersiell ordboksutgivning, dvs. rena förlagsprojekt. Förord och inledningar i sådana ordböcker kan visserligen vara ganska utförliga men innehåller oftast praktiska anvisningar för användningen snarare än en redovisning av det teoretiska grundarbetet. Det som här sagts betyder givetvis inte att rent kommersiella projekt, som bedrivits vid exempelvis förlagens ordboksredaktioner, inte skulle ha tagit ställning till språkteoretiska och lexikografiteoretiska problem. Tvärtom — så har det säkert normalt varit, men det har oftast stannat vid enbart en *implicit* redovisning, som sällan lett till mer utbredda principiella diskussioner, helt enkelt därför att traditionen inte har inbjudit till det. Med tradition menas då ett ganska kortvarigt bruk i ”modern” tid; i äldre ordböcker kunde det ofta förekomma långa utläggningar av både grammatisk och encyklopedisk art, breda forskningsöversikter osv. Ett välkänt sådant exempel är Adelung (1774–86) på den internationella arenan och från svenskt område i viss mån redan Spiegel (1712) och Ihre (1769).

Det kan också vara på sin plats att understryka än en gång, att den distinktion som här görs mellan praktisk lexikografi och olika slags teori inte innebär någon värdering till nackdel för den ”praktiska lexikografin”. Vi vet ju att så kallad praktisk lexikografi inte alltid haft något högt prestigevärde i teoretiskt orienterade språkvetares världsbild, men man kommer inte ifrån att det är ”praktiska lexikografer” och inte ”teoretiska lingvister” vi har att tacka för de ovärderliga språkliga hjälpmedel som vi kallar ordböcker. (Se även mitt bidrag till volymen från Oslokonferensen (Fjeld 1992), särskilt s. 28–32!)

Som en tredje aspekt på relationen mellan teori och lexikografi vill jag nämna den rörelse som uttryckligen proklamerades i USA för ett femtontal år sedan, som kallades ”computational lexicology” av utövarna själva. Det hela startade med en uppmärksam doktorsavhandling av Robert Amsler vid University of Texas, Austin (Amsler 1980). Man syftade till att genom noggrann datalingvistisk analys av maskinläsbara ordböcker dra generella slutsatser av betydelse för den allmänna språkteorin. Särskilt avsåg man att utveckla språkmodellens lexikaliska komponent, som dittills blivit i stort sett ignorerad. Därvid utmanade man den tidigare allmänt omfattade språkteoretiska uppfattningen att syntaxen skulle vara språkmodellens helt centrala komponent. I ett för computational lexicology-rörelsen typiskt arbete citerar Raoul Smith (1981:1) Reiger (1978) och Mel’čuk (1978) i två mycket entydiga uttalanden:

There has been an increased interest within cognitive science in the status of the lexicon. In particular, the position of the lexicon as the central element in models of natural language users is being expressed clearly and strongly in the literature. Reiger, in his development of a story comprehender, argues:

Syntax [...] is secondary, something which accrues from a more primary language phenomenon. This phenomenon, we will propose, is the organization and selection of *word senses* [...] [Most] of the complexity in natural language, in fact most of its information content, lies in the individual word sense, rather than in any central or uniform system of rules [...] [W]ord sense discrimination is the key to language comprehension (1978:97–98).

Mel’čuk, in his study of a lexicographical model for text processing, agrees:

Text analysis and synthesis are conducted on the basis of stored information about words [...] and their combinations [...] Any text is built up of words having a definite meaning and combined in accordance with definite properties. To analyze or synthesize a text one must know word meanings and their combinatorial properties. The respective data are stored in the dictionary; therefore a dictionary is the most important component of any meaningful A[utomatic] T[ext] P[rocessing] system [...] Grammar rules are no more than a means of abbreviating a dictionary, an attempt to formulate useful generalizations about the dictionary, something that completely depends on the dictionary [...] Without a dictionary no ATP system is possible (1978:3–4).

Jag är osäker på vad som senare kommit ut av de planer företrädarna för ”computational lexicology” hade. Det anordnades ett internationellt symposium med ett mycket begränsat antal deltagare vid Stanford Research Institute International 1983, varvid en permanent organisation och en särskild tidskrift diskuterades, men det förefaller som om de drivande krafterna sedan dess absorberats av annan datalingsvistisk forskning eller olika kommersiella åtaganden. Hur som helst kan man konstatera att lexikaliska frågor generellt kommit att spela en betydligt större roll i språkvetenskapen under den följande tidsperioden, samtidigt som den lexikografiska verksamheten intensifierats och förfinats ytterligare. Praktisk lexikografi som allmän ram för ett snarast språkteoretiskt översiktsverk som Aitchison (1987) hade knappast varit tänkbart tidigare.

Under alla omständigheter innebär den inriktning som nämndes strax ovan, ”computational lexicology”, ett totalt perspektivskifte från den i bästa fall lätt nådiga attityd som språkteoretiker tidigare ofta intagit, med innebörden att den enda relation som råder mellan praktisk lexikografi och språk teori är den som består i att lexikografer kan sköta sitt uppdrag bättre om de är språkteoretiskt medvetna, eftersom de då kan göra språkvetenskapligt bättre genomarbetade ordböcker. Även om det är rimligt att anta att lingvistiskt väl skolade lexikografer åstadkommer bättre ordböcker än språkteoretiska ignoranter, kvarstår det fortfarande som ett obevisat påstående att ordböcker som troget speglar människans sätt att spontant hantera den lexikaliska sidan av språket är generellt mer ändamålsenliga än ordböcker som saknar sådana egenskaper. Om det påståendet emellertid vore sant skulle den idealiska ordboken också kunna fungera som något slags modell över språkbrukarnas lexikaliska förmåga. Det är med utgångspunkt i sådana frågeställningar som jag i det följande skulle vilja göra några opretentiösa iakttagelser och överväganden. Ett av skälen är att jag själv tidigare gjort mig till tolk för den antydda ståndpunkten men efter att ha stått vid sidan av praktiskt lexikografiarbete under en dryg tioårsperiod nu har fått många av mina tidigare perspektiv förskjutna. De frågor som tas upp nedan är välkända lexikografiska problem, men de behöver ånyo övervägas i just det här perspektivet.

2 Ordböcker och databaser

Ett av skälen till att relationen mellan lexikografi och språk teori ändrades i början av 1980-talet var den tekniska utvecklingen. Dittills hade man visserligen på sina håll ägnat

sig åt att konvertera traditionellt framtagna ordböcker till maskinläsbar form i samband med uppdatering och publicering av nya upplagor, men nu blev nya ordböcker omedelbart maskinläsbara i samband med att de etablerades direkt i de nya elektroniska medierna. Man fick således tillgång till stora datorläsbara lexikaliska material. Ett mål som emfatiskt framhölls inom den nya datorbaserade lexikografin var att omfattande, väl genomarbetade databaser skulle upprättas, varur olika specialiserade produkter skulle kunna genereras. Detta var ett av skälen till att åtskilliga aktiva språkforskare var ivriga att engagera sig i verksamheten. Det idealtillstånd som eftersträvades bestod i att man skulle få en funktionsuppdelning mellan databasen och de olika lexikografiska specialprodukterna. Databasen skulle vara en reservoar att ösa ur efter behag, när specifikt utformade ordböcker skulle sammanställas för olika ändamål. Visionen var nog i allmänhet att den grundläggande databasen snarare än de konkreta ordböckerna skulle ha karaktären av teoretiskt fullgånge språkmodell.

Den senare utvecklingen har väl visat vissa tendenser åt detta håll. För lexikografiska projekt som för andra mänskliga aktiviteter gäller emellertid, med en lätt modifiering av en rad hos den svenske diktaren Stagnelius, att "verklighetens hårda band oss trycka". De ofta starkt begränsade resurserna har nödvändiggjort att även de mest storslaget planerade databaser fått anpassas till de krav som ställts på de konkreta produkter som skulle extraheras ur dem. Den hett åtrådda kombinationen av databasperspektivet och ordboks-perspektivet i form av det *elektroniska lexikonet* — som alltså är något kvalitativt annorlunda än en datorläsbar ordbok och exempelvis skall kunna medge att informationen genomlysas från ett stort antal olika perspektiv — har veterligen inte förverkligats någonstans. Vad som tills vidare förekommer är *dels* elektroniskt tillgänglig lexikalisk information i mer eller mindre ordbokstrogen form, information som givetvis också kan ha ett allmänt språkvetenskapligt intresse av stor räckvidd, *dels* mera utpräglat modellliknande lexikaliska informationsmängder, antingen för teoretiska ändamål och då i allmänhet experimentellt utformade eller i starkt reducerad form, eller anpassade till mycket specialiserade ändamål av artificial intelligence-typ.

Vi kan bara konstatera att det ännu är långt kvar till en fungerande modell över människans totala lexikaliska förmåga. Lexikografer kan knappast ta något primärt ansvar för att en fullfjädrad sådan modell verkligen konstrueras. Frågan är om lexikografin över huvud taget har något att vinna på att ta del i dessa strävanden. En förutsättning är givetvis att man tror att det därigenom går att få bättre ordböcker.

Att en lexikonmodell syftar till att ge en trogen bild av människans lexikaliskt inriktade språkförmåga är självklart. Däremot kan ordböcker ha en hel rad olika funktioner. Ordböcker är i första hand praktiskt motiverade hjälpmedel, som följaktligen sammanställs i syfte att lösa en konkret uppgift. Syftet kan vara rätt generellt eller mer specifikt beroende på typen av ordbok. Frågan är då om det är relevant i de enskilda fallen att ordboken utformas i enlighet med en idé om det mentala lexikonets uppbyggnad. Det ligger kanske närmare till hands att ett sådant mål eftersträvas i enspråkiga definitionsordböcker än i tvåspråkiga allmänordböcker å ena sidan, olika specialordböcker å den andra. En enspråkig ordbok med en långtgående ambition att beskriva ett enskilt språks lexikaliska sida så utförligt som möjligt aktualiserar frågan om "psykologisk realitet" (i exempelvis Chomskys terminologi) på ett mer naturligt sätt än en specialordbok av något slag, låt säga en ordbok över flygterminologi. I det senare fallet kan det inte vara ett primärt önskemål att ordboken troget avspeglar en människas sätt

att mentalt lagra ifrågavarande ordförråd. Det viktiga är att den samlade termuppsättning som maximalt kan komma till användning inom det specificerade fackområdet på det aktuella språket finns väl beskriven i ordboken. På likartat sätt kan man givetvis resonera för de flesta ordböcker över specialområden. Ordbokens syfte i det enskilda fallet kan således vara avgörande för frågan huruvida en modell över det mentala lexikonet är ett eftersträvansvärt mönster eller ej.

Något svårare att besvara är möjligen frågan om önskvärheten av att tvåspråkiga ordböcker troget avspeglar formen för mänsklig lexikalisk kompetens. Syftet med dessa ordböcker är ju att underlätta översättning mellan språk. Att en tvåspråkig ordbok är utformad som en lexikonmodell skulle exempelvis kunna innebära att man försöker efterlikna en genuint tvåspråkig persons lexikaliska förmåga, vad nu detta skulle innebära konkret. Vad lexikografiteoretiker observerat på det här området är emellertid snarare, att tvåspråkiga ordböcker för att optimalt kunna fylla sin uppgift bör vara olika utformade med hänsyn till om de vänder sig till språkbrukare som primärt har det ena aktuella språket som modersmål eller det andra, om man har ett källspråks- eller målspråksperspektiv. Det är alltså snarast de i varje enskilt fall inblandade språkparens kontrastiva egenskaper som behöver fokuseras, och för varje språkpar skulle det i det idealiska fallet finnas åtta ordböcker: en produktions- och en perceptionsordbok på vardera språket med särskild utformning för talare av respektive språk. Detta fulla utbud brukar av praktiska skäl vara en omöjlighet att realisera, men en integrerad modell av en tvåspråkig talares kompetens — hur den nu skulle se ut — fyller under inga omständigheter samma funktion. De lexikografiskt kritiska frågorna gäller snarare informationsurval och informationens organisation.

3 Ordurval och lexikalisk systematik

En primär fråga inom lexikografien är hur orden i varje given ordbok skall väljas ut. Alla vet att mängden av uppslagsord är det kanske viktigaste försäljningsargumentet när en ordbok marknadsförs, oavsett hur sofistikerade lösningar på diverse andra problem man kan uppvisa. I praktiskt lexikografiska sammanhang blir därför ordurvalet ett optimeringsproblem, en fråga om hur man på effektivast möjliga sätt presenterar så mycket information som möjligt om så stor ordmängd som möjligt till en tänkt användargrupp. Olika skarpsinniga lösningar på diverse problem kan naturligtvis vara till nytta, men det är ändå vanligen de strikt *praktiska* bedömningarna som blir avgörande för resultatet.

Om ordboken skall reflektera en *lexikonmodell* blir redan ordurvalet en *teoretiskt* viktig fråga. Urvalet blir inte i första hand en utrymmesfråga, utan det skall också spegla någon form av *lexikalisk kompetens*. I själva verket skall det ordurval som man stannat för i ordboken återspegla ett *naturligt ordförråd* i någon mening. En sådan inställning kan leda till olika slutsatser. Om man vill inta en extrem ståndpunkt kan man säga att ingen specialordbok, t.ex. en terminologisk ordbok över ett visst fackområde, kan återge språksamhällets kollektiva kompetens. Dels är den lexikaliska informationen i en sådan ordbok så specialiserad att den svarar mot endast mycket få individers kompetens, dels saknas det mesta som bygger upp den språkliga kompetensen i allmänhet, nämligen större delen av det allmänspråkliga materialet. Här blir också valet av teoretisk grund avgörande. Konflikten mellan vad som ingår i individers lexikaliska språkförmåga och den kollektiva kompetensen är ett generellt deskriptivt problem. Specialordboken kan i och för sig mycket väl, liksom Chomskys

grammatikmodell, tillåtas spegla en i sammanhanget "idealisk språkbrukares" kompetens, nämligen den specialiserade fackmannens.

Det är emellertid ofrånkomligt att urvalskriterierna för ordmängden i praktiskt användbara respektive modellteoretiskt trogna ordböcker lätt råkar i konflikt med varandra. En lexikonmodell måste beskriva det reguljära allmänspråket, dvs. koncentrera sig på generaliteter i språket, vilket även en ordbok kan göra, men ordböcker behöver väl så ofta betona det speciella, det anmärkningsvärda: fackspråk, dialektformer, ålderdomlighet, olika stilregister osv. Vad av allt detta som inkluderas i en ordbok kan vara beroende av exempelvis nationell lexikografisk tradition. Exempelvis lämnar Blöndals isländsk-danska ordbok (1920–24) och Árni Böðvarssons *Íslensk orðabók* (1983), liksom den isländska lexikografiska traditionen i allmänhet, betydligt större utrymme åt ålderdomliga uttryck än vad som är vanligt i de flesta jämförbara skandinaviska ordböcker. Om dessa ordböcker över huvud taget skall betraktas som synkroniska beskrivningar av det isländska ordförrådet, får man förutsätta ett ovanligt generöst tilltaget synkroniskt stadium.

Botolv Helleland tog för några år sedan upp frågan om vilken plats egennamnen skall ha i en ordbok (Helleland 1992). Problemet är mer komplicerat än man kanske brukar tänka på. I enspråkiga ordböcker utelämnas ofta alla ord som man skriver med stor inledningsbokstav, medan sammansättningar med och avledningar av sådana ord genom sitt ordslut betraktas som principiellt icke-propriella och tas med. Däremot kan tvåspråkiga lexikon innehålla vissa egennamn som uppslagsord. Vanligen gäller det då bara sådana som har speciellt anmärkningsvärd form i ettdera av de språk lexikonet involverar, eller båda. Så kan i en engelsk-svensk ordbok *The Mediterranean (Sea)* med den svenska motsvarigheten *Medelhavet* återfinnas, vidare namn som *Prague* med uttalsangivelse och svensk motsvarighet *Prag*, medan urvalet av ort- och personnamn för den stora majoriteten av logiskt möjliga fall blir en bedömningsfråga; lexikografens föreställning om vad som är betydelsefullt för ordboksanvändaren blir avgörande.

Som jag påpekat i ett par andra sammanhang (t.ex. Ralph 1992:33) spelar namn en statistiskt ganska betydande roll i uppbyggnaden av texter, men framför allt intar de en mycket framskjuten plats i ordförrådet. I en realistisk lexikonmodell måste därför egennamnen beredas den plats som är motiverat med hänsyn till kravet på psykologisk relevans. När enspråkiga ordböcker återger uppslagsord som innehåller propriella morfem som element, medan själva de egennamn som ligger till grund för dessa utesluts, uppstår teoretiskt viktiga effekter. För att ta namn på stater och så kallade nationalitetsord som exempel, så bildar dessa vanligen naturliga lexikaliska grupper bestående av statsnamn (t.ex. *Island, Ungern*), inbyggarbeteckning (*islänning, ungrare*), nationalitetsadjektiv (*isländsk, ungersk*), och det är i många sammanhang också relevant att hit räkna beteckning på officiellt språk (*isländska, ungerska*; se även Sigurd (1965) och *Statsnavne og nationalitetsord* (1994)). Dessa grupper kan i ett givet språk, exempelvis svenskan, visa en ganska förutsägbar morfologisk form (som i de just nämnda fallen) eller en mer oförutsägbar, idiosynkratisk form (såsom *Danmark* : *dansk* : *dansk* : *danska*; *Norge* : *norrman* : *norsk* : *norska*). I och med att avledningarna bildas med hjälp av ett antal olika derivationsmorfem, måste varje enskild manifestationsform, dvs. den verkliga, konventionaliserade formen, preciseras i lexikonet, såväl människans mentala lexikon och den teoretiska modellen av detta som den fysiska ordboken.

Det finns emellertid trots allt en viss parallellitet mellan de fyra potentiella orden i varje grupp, ett slags partiellt återkommande systematik, även om det inte är helt förutsägbart

exakt vilka former som är gängse i varje enskilt fall. Denna tendens till systematik framgår av exemplen i tabell 1, där några olika typer samlats. Nu finns det i detta namnmaterial en del specialfall. Vi behöver bara gå till den återstående nordiska nationen, Finland, för att hitta ett. Som inbyggarbeteckning till statsnamnet *Finland* fungerar både *finländare* och *finne*, som nationaladjektiv på samma sätt *finländsk* eller *finsk*. Varianterna kan i vissa sammanhang användas distinkt (finlandssvenskar inkluderade eller ej), men de är i många sammanhang funktionellt synonyma. Som språkbeteckning saknas av naturliga skäl den ena morfologiskt möjliga varianten, dvs. **finländska*, medan den enda förekommande språkbeteckningen, *finska*, ansluter sig morfologiskt till *finne* : *finsk*.

Den lucka vi har i serien *Finland* : *finländare* : *finländsk* : — ser vi en motsvarighet till i en rad fall, där man saknar ett nationalspråk vars beteckning återgår på statsnamnet eller nationaladjektivet (utan att det finns en kompletterande serie som *Finland* : *finne* : *finsk* : *finska*): *Libanon* : *libanes* : *libanesisk* : —; *Liechtenstein* : *liechtensteinare* : *liechtensteinsk* : —; *Tanzania* : *tanzanier* : *tanzanisk* : —; osv. Det omvända förhållandet kan sägas råda i en del andra avledningsserier, såsom — : *same* : *samisk* : *samiska*, där något motsvarande statsnamn inte existerar av det enkla skälet att samerna inte bildat någon suverän stat. Beteckningen *Sameland* (*Sápmi*) har en något annorlunda status men kommer naturligtvis språkligt nära statsnamnen såsom beteckning på samernas traditionella bosättningsområde. Morfologiskt och semantiskt blir emellertid den ”icke-propriella” delen av serier som *same* : *samisk* : *samiska* och *finne* : *finsk* : *finska* fullständigt parallellt uppbyggd. Ännu lite annorlunda status har beteckningen *Kurdistan*, som i första hand har en etablerad geografisk innebörd. Området bebos huvudsakligen av den etniskt urskiljbara gruppen kurder, och även om Kurdistan inte är en suverän stat faller namnet således väl in i en serie *Kurdistan* : *kurd* : *kurdisk* : *kurdiska*.

Albanien	alban	albansk	albanska
Bulgarien	bulgar	bulgarisk	bulgariska
Danmark	dansk	dansk	danska
Frankrike	fransman	fransk	franska
Färöarna	färing	färöisk	färöiska
Grönland	grönländare	grönländsk	grönländska
Island	islänning	isländsk	isländska
Japan	japan	japansk	japanska
Norge	norrman	norsk	norska
Polen	polack	polsk	polska
Portugal	portugis	portugisisk	portugisiska
Sverige	svensk	svensk	svenska
Tjeckien	tjeck	tjeckisk	tjeckiska
Ungern	ungrare	ungersk	ungerska
Vietnam	vietnames	vietnamesisk	vietnamesiska

Tabell 1. Statsnamn och nationalitetsord. Symmetriska serier.

I en del andra fall med närmast etniska beteckningar saknas likaså allmänt accepterade beteckningar på landområdet. Så är det exempelvis för samojedernas del (om vi bortser från den skämtsamma beteckningen *Samojedien* för någon allmänt avlägsen plats). Vi får således

en ohjälpligt defekt serie, som i övrigt från språkliga utgångspunkter sett ansluter sig väl till de tidigare nämnda: — : *samojed* : *samojedisk* : *samojediska*. Ytterligare en variant, som kommer nära denna är — : *arab* : *arabisk* : *arabiska*, som likaså avser kulturförhållanden snarare än nationsförhållanden. Till denna defekta serie kan antingen territorialbeteckningen *Arabstaterna* läggas, som då är en sammanfattande beteckning på de stater som vi förknippar med ett i kulturellt avseende arabiskt livsmönster, eller *Arabien*, som får uppfattas som ett historiskt begrepp mer än ett politiskt. Den nutida motsvarighet i statshänseende som kommer närmast är väl *Saudiarabien* med avledningarna *saudiarab/saudier* respektive *saudi-arabisk/saudisk*, men man kan observera att beteckningen på det officiella nationalspråket är *arabiska*. Till denna svit av beteckningar som hänför sig till etniska grupper ansluter sig egentligen också på ett naturligt sätt *jude* : *judisk*, men här saknas såväl morfologiskt besläktad territorialbeteckning som språkbeteckning (om vi inte räknar *jiddisch*, vilket vore orimligt). De olika problemfall som behandlats här sammanfattas i tabell 2. De anförda exemplen är hämtade från svenskan men har uppenbarligen direkta motsvarigheter i andra språk.

Det är helt klart att man från språkliga utgångspunkter måste betrakta dessa etniska beteckningar som del av ett kontinuum, som mot den ena extremen rymmer de klara statsnamnen med deras respektive avledningar, mot den andra entydigt icke-propriella, rent appellativiska bildningar. Det är också rimligt att dessa nästan termliknande ord bildar en subkategori i språkbrukarnas mentala lexikon med väl utarbetade associationsbanor mellan de prototypiska fallen å ena sidan och till mer eller mindre avvikande fall å den andra. De mönster som delvis kan urskiljas i detta subsystem blir naturligtvis mindre tydliga om allt som liknar egennamn diskvalificeras, såsom faktiskt sker normalt i de flesta ordböcker. De asymmetriska förhållanden som råder mellan olika avledningsserier framstår också mindre tydligt. Detta är således ett exempel på hur ordböcker med nuvarande praxis omöjligt kan uppfattas som en trogen bild av människors lexikaliska kompetens.

Finland	finländare/ finne	finländsk/ finsk	— finska
Libanon	libanes	libanesisk	—
Liechtenstein	liechtensteinare	liechtensteinsk	—
Tanzania	tanzanier	tanzanisk	—
(Sameland)	same	samisk	samiska
Kurdistan	kurd	kurdisk	kurdiska
—	samojed	samojedisk	samojediska
(Arabstaterna/Arabien)	arab	arabisk	arabiska
Saudiarabien	saudiarab/ saudier	saudiarabisk/ saudisk	arabiska
—	jude	judisk	—

Tabell 2. *Asymmetriska serier.*

Man kan säga att egennamn som lexikaliska enheter vetter mot den encyklopediska kunskap som ordböcker sällan tar som sin uppgift att beskriva. Distinktionen mellan språklig och encyklopedisk information är emellertid ännu ett stående problem inom lexikografen. Kanske kan man för lexikonmodellens del påstå att en fullständigt realistisk sådan föreligger först när den är inordnad i en modell över människans totala kognitiva förmåga. I så fall

understryks diskrepansen mellan målet för lexikonmodellen och vad som kan vara ett rimligt mål för en ordbok ytterligare.

4 Den lexikaliska informationens lagringsform

Välkända är tidigare språkteoretiska beskrivningar av lexikonet som ”en lista över morfemen i ett språk”, ”över oregelbundenheter” eller ”oförutsägbara enheter som inte kan fångas med regelskrivning”. Den syn som kommer till uttryck i sådana formuleringar reducerar på ett avgörande och fatalt sätt lexikonets grundläggande mångdimensionella natur. I lexikonet kan ett stort antal subsystem av olika slag urskiljas. Egennamnen som nyss berördes kort bildar ett sådant system, och det går att urskilja ett stort antal liknande subsystem på olika grunder: grammatiska, semantiska, encyklopediska etc. Klasserna går sällan att skarpt avgränsa, utom i mycket speciella slutna system, som t.ex. beteckningar för militär rang. De måste också urskiljas i olika dimensioner, med följderna att en och samma enhet kan komma att ingå som medlem i olika klasser beroende på vilket perspektiv man anlägger.

Detta är ett synsätt som står i strid med grammatiktraditionens klassificering och därmed också ordbokstraditionens. Det skapar redovisningsproblem och aktualiserar en av de mest grundläggande frågorna i såväl arbetet på modellen över det mentala lexikonet som inom lexikografen: hur skall en lexikalisk enhet preciseras? I en ordbok som syftar till att efterlikna en lexikonmodell bör det rimligen vara god överensstämmelse mellan modellens lexikaliska enhet och ordbokens uppslagsord. Vad som skall kunna vara ett uppslagsord är naturligtvis också en lexikografiteoretiskt viktig fråga i sig.

Grundläggande för lexikonmodellens lexikaliska enhet är begreppet ”lexikalisering”, som vi ofta talar om, men som väl knappast är tillfredsställande definierat. Principiellt bör man av en lexikalisk enhet i det mentala lexikonet kräva att det är en konventionellt etablerad, fast bildning med specificerbar lexikalisk funktion. Liknande krav ställs ofta även på uppslagsord i ordböcker av teoretiskt intresserade lexikografer. Men frågan är hur enheternas grad av lexikalisering — i betydelsen ”konventionellt etablerade enheter” — skall uppfattas i de enskilda fallen. Eftersom ordböcker i första hand skall tjäna som praktisk hjälp, är det möjligen graden av förutsägbarhet som är det viktigaste urvalskriteriet, och det är inte riktigt detsamma som graden av lexikalisering. (För ett illustrativt fall i denna fråga se t.ex. Malmgren 1994.) Inte minst kan det just sagda gälla för tvåspråkiga ordböcker, där det inte bara rör sig om respektive enhets förutsägbarhet i respektive språk utan också förutsägbarheten av att två definierade motsvarigheter skall korrespondera till varandra i respektive språk, givet språkens regelsystem.

Uppslagsorden i ordböcker brukar representera en flexionsserie, ett morfologiskt sammanhållet paradigm. Den abstraktion som lemmatisering innebär och valet av kanonisk form som uppslagsord kan komma att dölja en skiftande verklighet. I en svensk flexionsserie *fort* : *fortare* : *fortast* kan möjligen formen *fort* uppfattas som logiskt primär framför övriga former, men hur förhåller sig komparativformerna *bättre*, *hellre*, *sämre* till superlativformerna *bäst*, *helst*, *sämst*, som saknar morfologiskt regelbundna motsvarigheter i positiv? Är någondera formen primär framför den andra? Normalt brukar dessa böjningsformer redovisas var för sig i ordböcker (alltså t.ex. *hellre* skilt från *helst*), med olika korsreferenser till varandra och till de positivformer som brukar suppleras (*bra*, *gärna*, *dålig*). Hur än redovisningsproblemet

blir löst kvarstår uppenbart en obalans för "suppletiva" böjningar såsom [*gärna*] : *hellre* : *helst* i förhållande till "regelbundna" flexionsserier som *fort* : *fortare* : *fortast*.

Total icke-förekomst av en böjningsform är bara ett extremfall av de frekvensförhållanden som kan råda inom paradigmerna. Vilken roll frekvens spelar för språkinlärning och lagring av språklig information är en omstridd och besvärlig fråga, liksom det är svårt att avgöra hur frekvensuppgifter skall kunna utnyttjas för att göra ordböcker mera ändamålsenliga. För att ta ett enda exempel förekommer ordet *resurs* i ett stort svenskt språkmateriale från 1965 (redovisat i Allén et al. 1970–80) endast i pluralis, medan singularformen tio år senare börjat uppträda och sedan dess blivit allt vanligare. Det som i det allmänna medvetandet snarast var ett defekt paradigm för en generation sedan är nu kompletterat, men i ordböckerna syns ingen skillnad, trots att själva uppslagsformen på 1960-talet bör ha varit en närmast konstruerad form. För andra ord kan det vara andra böjningsformer som är dominerande inom paradigmerna. Det står helt klart att de olika konkreta formerna inom ett paradigm har helt olika egenskaper med hänsyn till en rad faktorer, förutom frekvens även syntaktisk ram, fraseologi, semantiska restriktioner osv., med andra ord: skillnader som vi är noga med att notera för olika lemman! En modell över språkförmågan bör anvisa mekanismer för generering av icke ännu konventionaliserade men potentiella regelbundna former, samtidigt som den bör spegla det rådande tillståndet. I ännu högre grad är det viktigt att ordböcker ger besked om det faktiskt utnyttjade systemet.

Är det över huvud taget möjligt att fånga hela detta spektrum av lexikaliskt viktiga egenskaper hos böjda former under en och samma kanoniska form och hävda att denna har psykologisk relevans? I vissa avseenden verkar onekligen de enskilda formerna i ett paradigm vara de enda psykologiskt rimliga lexikaliska enheterna. Detta är knappast en lexikografiskt acceptabel tanke, men i olika teoretiska modeller har den föreslagits (t.ex. i flera arbeten av James McCawley), och den finner också språkpsykologiskt stöd i exempelvis språkinlärningens progression. Det har framhållits med ökande emfas att lexikonmodellen principiellt bör utformas som ett sofistikerat associativt nätverk (Aitchison 1987, särsk. s. 190 ff.), troligen så sofistikerat att det ännu är långt kvar till en förverkligad fungerande modell. Därmed sätter ordbokens fysiska förutsättningar vissa mycket bestämda gränser för hur nära den teoretiska lexikonmodellen man kan komma med en ordbok. Kraven på linearitet och val av en genomförd indelningsgrund i varje enskilt fall reser oöverkomliga hinder mot att lexikonmodellens lagringsform efterliknas i ordböcker trots aldrig så generösa korsreferenssystem. Naturligtvis kan man då återigen hävda att helt andra möjligheter öppnar sig med ett elektroniskt lexikon. Den frågan berörs något ytterligare avslutningsvis.

5 Framtiden

Mycket av det som här framförts kan tyda på att den gamla frågan om ordböckernas närmande till teoretiska språkmodeller för närvarande är en inaktuell frågeställning. Det finns emellertid ytterligare några aspekter som kan vara värda att poängtera. För det första går den tekniska utvecklingen entydigt i riktning mot att kraftfulla persondatorer blir tillgängliga för allt fler personer, och allt större och mer komplicerade datamängder hanteras med allt enklare medel. Därmed kan man fortfarande spekulera i möjligheten att elektroniska "ordböcker" för allmänt bruk kan bli verklighet i en någorlunda överskådlig framtid. Därmed kan flera av de tidigare nämnda problemen åter bli aktuella.

I väntan på denna utveckling är det kanske viktigare att konstatera att ingen har bättre förutsättningar att bidra till den teoretiska lexikonmodellens utveckling i många väsentliga avseenden än erfarna lexikografer. En bodelning mellan de båda verksamheterna skulle därför vara olycklig. Slutligen kvarstår åtskilliga olösta frågor som är av intresse för både lexikografiteori och lexikaliskt inriktad språkteori.

Lexikaliseringsproblemet är en sådan fråga som behöver mer belysning. Det är långt ifrån klart vad som skall betraktas som en lexikaliserad form och hur lexikaliseringsprocessen går till. Troligen måste man räkna med lexikalisering som ett skalfenomen snarare än ett binärt förhållande. Den nyss diskuterade frågan om uppslagsord och kanoniska former är av allmänt intresse, även om man skulle komma till olika resultat för ordböckernas respektive lexikonmodellens del. Det lexikaliska universets omfattning och en rad abstrakthetsfrågor behöver fortfarande preciseras såväl inom ordbokens som lexikonmodellens ram. Metaspråkets utformning är en högintressant fråga för både lexikografisk teori och allmän språkteori. Vi intresserar oss som språkvetare i allt högre grad för folktaxonomi, men vi avfärdar slentrianmässigt folketyologier såsom kontrafaktiva, trots att de tydligtvis representerar folkspråkliga kompetens och kan avslöja en del om deras språkliga slutledningsförmåga. En fråga som nästan inte alls uppmärksammas är folkdefinitioner, dvs. hur man spontant definierar ord. Kan dessa på ett intressant sätt utnyttjas teoretiskt och praktiskt inom den lexikaliska verksamheten såsom en spegling av den mänskliga ordsemantiska kompetensen?

Med detta hoppas jag ha argumenterat något för att lexikografer och språkteoretiker även fortsättningsvis bör ta del av varandras arbete med stor uppmärksamhet.

Bibliografi

- Adelung, Johann Christoph 1774–86: *Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der hochdeutschen Mundart* [...]. 1.5. Leipzig.
- Aitchison, Jean 1987: *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford/New York: Basil Blackwell.
- Allén, Sture et al. 1970–1980: *Nusvensk frekvensordbok baserad på tidningstext*. 1–4. Stockholm: Almqvist & Wiksell.
- Amsler, Robert A. 1980: *The Structure of Merriam-Webster Pocket Dictionary*. Austin, Texas: University of Austin, Texas.
- Blöndal se Sigfús Blöndal.
- Fjeld, Ruth Vatvedt (utg.) 1992: *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31 mai 1991*. Oslo: Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1.
- Garde, Anna/Pia Jarvad (utg.) 1994: *Nordiske studier i leksikografi II. Rapport fra Konferanse om Leksikografi i Norden 11. – 14. maj 1993*. Oslo: Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2. (Tr. Århus.)
- Helleland, Botolv 1992: *Proprium og leksikografi*. I: Fjeld 1992, 190–199.

- Ihre, Johan 1769: *Glossarium suiogothicum, in quo tam hodierno usu frequentata vocabula, quam in legum patriarum tabulis aliisque ævi medii scriptis obvia explicantur, et ex dialectis cognatis, moesogothica, anglosaxonica, alemannica, islandica ceterisque gothicae et celticae originis illustrantur*. Upsaliæ: Edmannianis.
- Íslensk orðabók handa skólum og almennungi* 1983. Önnur útgáfa, aukin og bætt. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Longman Dictionary of Contemporary English* 1978. Harlow/London: Longman Group Limited.
- Malmgren, Sven-Göran 1994: Sammansättningsmorfologi och lexikografi. I: Garde/Jarvad 1994, 179–184.
- Mel'čuk, I. A. 1978: A New Kind of Dictionary and Its Role as a Core Component of Automatic Text Processing Systems. I: T. A. *Informations* 2, 3–8.
- Ralph, Bo 1992: Ord på väg. I: Fjeld 1992, 28–39.
- Reiger, Charles J. 1978: Computational Linguistics. I: William Orr Dingwall (ed.): *A Survey of Linguistic Science*. Stamford, CT: Greylock Publishers.
- Sigfús Blöndal 1920–24: *Íslensk-dansk ordbog*. Reykjavík/København/Kristiania: Verslun Þórarins B. Þorlákssonar/H. Aschehoug & Co. (W. Nygaard).
- Sigurd, Bengt 1965: Morfologiska mönster i nationalitetsord. *Arkiv för nordisk filologi* 80, 231–276.
- Smith, Raoul N. 1981: *Conceptual Primitives in the English Lexicon*. Waltham, Massachusetts: GTE Laboratories Incorporated TR 81–401.1.
- Spegel, Haqvin 1712: *Glossarium Sveo-Gothicum eller Swensk Orda-book* [. . .]. Lund: A. Habereger, Gub. och Acad. Carol. Booktr.
- Statsnavne og nationalitetsord* 1994. Oslo: Nordisk spåksekretariat/Novus forlag.
- Svensén, Bo 1987: *Handbok i lexikografi. Principer och metoder i ordboksarbetet*. Stockholm: Esselte Studium/Tekniska nomenklaturcentralen.

Hur ser en användarvänlig ordbok ut?

User-friendly is a modern catchword, often used about appliances or systems. More and more frequently, it is also used about another type of advanced tool: the dictionary. When choosing a dictionary in earlier days, the most important criterion was size. It was assumed that as long as a dictionary contained vast amounts of information it was a good dictionary. It was up to the users to make friends with the dictionary — if they couldn't or wouldn't it was their problem and their fault. Present-day publishers realize that it's their job to get the information across. This paper discusses userfriendliness under headings such as *size, scope, contents, editing, typography, user instructions* and *front and end matter*.

1 Inledning

En ordbok är en sorts kunskapsmaskin, ett avancerat redskap. Det följer ofta med en detaljerad bruksanvisning. Problemet är bara att väldigt få av användarna tittar på bruksanvisningen — de kanske inte ens märker att den finns. I stället börjar han eller hon så att säga direkt trycka på alla knappar och dra i alla spakar och vrida på alla rattar för att få ut det som behövs.

Ordboksförfattare och förlag har väl i alla tider suckat över att användarna inte förstår att man ska läsa bruksanvisningen — att det med den här maskinen faktiskt är precis som med en ångmaskin, en motorgräsklippare eller en dator med CD-ROM-läsare! Traditionellt har man nog ofta tänkt att folk faktiskt får skylla sig själva om de inte begriper sig på mästerverket. Om de inte läser anvisningarna, inte förstår och inte kan sin grammatik så är det deras problem. En ordbok är som den är, och därmed basta.

Men det *är* inte användarnas problem, det är vårt. Det är kunden/användaren som bestämmer: köpa eller inte köpa, använda eller inte använda. Det är vi som gör ordböcker som måste se till att de är något att ha för användaren. Och hurdana ska de vara då? Det är minsann inte lätt att svara på den fråga jag satt som rubrik för det här föredraget, och jag har inte gjort några vetenskapliga undersökningar i fallet. Jag har inget fullständigt entydigt recept att komma med, för det finns nog inga absoluta sanningar. Men jag har en del funderingar kring utformning av ordböcker, sådant som man konfronteras med om man jobbar med att skapa ordböcker i den mera "kommersiella" verkligheten på ett bokförlag.

Om man ser sig om på ordboksmarknaden kan man konstatera att det numera finns rätt mycket att välja på, även i våra relativt små länder. Vi är många som vill dela på kakan. **Användarvänlig** är ett modernt honnörsord. Självklart vill alla göra anspråk på det för sina produkter. Men vad som är användarvänligt beror på vem användaren är. Ordboksanvändare har olika förutsättningar, olika språkfärdighet och inte minst olika ambitionsnivå. Än så länge kan man ju inte skraddarsy en ordbok för varje användare, utan måste försöka dela in användarna i kategorier, målgrupper. För att vara användarvänlig måste ordboken vara **målgruppsanpassad**. Den gamla standardmodellen duger inte längre.

Det finns mycket att tänka på när det gäller detta med att målgruppsanpassa. Utan att göra anspråk på fullständighet skulle jag vilja diskutera några aspekter som kommer in i bilden.

De har att göra med **storlek, innehåll, uttalsangivelser, redigeringsprinciper, typografi, och för- och eftertext** (dvs allt det som inte är själva ordbokstexten).

2 Storlek

Det är inte länge sedan det viktigaste, kanske enda, kriteriet vid val av ordbok var storleken. Bara den innehöll en mängd information var det en bra ordbok. Fortfarande väljer väldigt många ordbok efter antal uppslagsord — förutom efter priset förstås! För användaren är en bra ordbok en där man hittar det man söker. Så rent statistiskt har man väl större chans att finna den upplysning man letar efter ju större ordboken är. Men en användare med liten ordboksvana och begränsade språkkunskaper går lätt vilse i en för stor ordbok. En ordbok är ju i grunden en uppsättning rekommendationer. En liten ordbok kan och ska förstås inte försöka ge något som liknar en fullständig bild av språket. Ju mindre man kan desto viktigare är det att man får handfast hjälp, att man inte behöver fatta alla beslut själv: "Det här heter det här", punkt och slut. Att det sen finns andra översättningar eller rikliga möjligheter att variera sig kan man lära sig när man nått ett högre stadium. Man ska inte köpa en stor ordbok "att växa i". Det ger skavsår att mala sig fram mellan en massa information man inte kan tillgodogöra sig, på samma sätt som man får skavsår av att hasa sig fram i för stora skor!

En avancerad användare får naturligtvis lika mycket skavsår av "för små skor". Det är tröstlöst att leta efter saker som inte står i ordboken. Använder man språket på avancerad nivå, vill uttrycka sig väl och varierat, kanske med många fackuttryck, är det en stor ordbok man ska skaffa sig. Professor Jan Svartvik (1995) skriver, i en artikel om behovet av språkkunskaper, att man klarar sig förvånansvärt långt med ett engelskt basordförråd på omkring tretusen ord. Men därutöver måste man lära sig just de ord som behövs i olika speciella situationer. Så här skriver han: "3 000 ord täcker faktiskt in fyra av fem ord i en normal engelsk text. Men för att förstå de återstående 20 procenten behövs ett mycket större ordförråd, speciellt om det rör sig om olika facktexter. Ingen — inte heller de som har engelska som modersmål — behärskar alla engelska ord. Det är därför vi har stora ordböcker."

Det framstår nästan som en truism, men rätt ordbok för en nybörjare är en liten ordbok, för en avancerad användare är det en stor. För de allra flesta användarna är det följaktligen den medelstora ordboken som gäller. Den innehåller i huvudsak allmänspråk, men med tillsatser av fackspråk, och i synnerhet sådant som har med vardagsteknologi att göra.

Det här är viktigt! Och användarvänlighet är att tala om detta för användaren. Man måste med nödvändighet förenkla budskapet för att det ska bli tydligt. På mitt förlag använder vi sedan ett par år en schematisk uppställning, "Välj rätt ordbok", både i ordbokskatalogen och på eftersättsbladet i ordböckerna. Där visas ordböckernas inbördes förhållande. Det har varit uppskattat, inte minst i bokhandeln där man ska hjälpa kunden att välja. Men ibland har informationen uppfattats som för knapphändig. I vår nya katalog försöker vi åstadkomma ytterligare klarhet med ett enkelt kodsysteem som återkommer på olika sätt. Vi använder symbolen ruta för avancerad nivå, triangel för mellannivå och ring för den lägsta nivån. Vad symbolerna står för förklaras kortfattat på omslaget och något mera utförligt i katalogen.

Symbolerna finns sedan i katalogen intill alla de **allmänspråkliga** ordböckerna, dvs där det verkligen är fråga om nivåer.

3 Innehållets inriktning

Många små ordböcker, och till och med fickordböcker, innehåller ord som *abbot*, *abbedissa*, *karbonpapper* och *schellack* fast det faktiskt är ganska ovanliga företeelser idag. Det är viktigt att man strävar efter att anpassa innehållets inriktning, ordurvalet, efter den tilltänkta användaren. Ju mindre ordbok desto större blir andelen konkreta begrepp — det är ingen nyhet. Men det finns mycket att göra utöver det. En ordbok som i första hand är avsedd för skola och undervisning, ska spegla den värld de unga lever i. Den ska naturligtvis ge dem ett redskap för att uttrycka sig korrekt — ge viktiga verb och funktionsord och de vanliga, frekventa "tandborstorden". Men utöver det är det rimliga att den tar ett skutt ner på frekvensskalan och tar med vad som kanske egentligen är att beteckna som facktermer. Det gäller t ex sport, data och hästtermer (det sistnämnda därför att ridning är en av de absolut populäraste fritidssysselsättningarna bland svenska flickor).

En reseordbok ger mycket lite grammatikbetonad hjälp — syftet här är ju inte lika självklart att användaren ska kunna producera fullständiga satser. Ordförrådet är i mycket detsamma som i reseparlörerna, dvs uppbyggt för att vara relevant i vissa situationer. Den här ordboken används troligen för snabb kontroll av vad *plåster* eller *avgasrör* eller *välstekt* heter. I den situationen kan man tänka sig att en användare snarast skulle störas av en fylligare frasinformation i ordboken. Däremot kan ordboken gärna kompletteras med en liten parlördel, där den användare som vill kan få hjälp att säga "Jag skulle vilja ha en ask plåster, tack".

Som vi talade om nyss är det ju den större ordboken som innehåller mest fackspråk. Jag tror inte den tid är så förfärligt långt bort när ny teknik medger att man måttbeställer sitt ordförråd: allmänspråk plus medicin och biologi kanske. Men än så länge får man ta till den ordbok som innehåller mest för att få med just *sitt* fackspråk. Den stora ordboken är ju den som har den största bredden. Den ska spegla hela registret av språket, olika stilnivåer, olika specialområden, olika regionala varianter. Man brukar ju ofta säga att de stora ordböckerna är lättast att göra — det är i princip bara att ta med allt man kommer på! Det är väl en sanning med modifikation, men om den nu antas vara åtminstone delvis sann, är då de minsta ordböckerna svårast? Nej, jag tror inte det. Där kompletterar man det självskrivna grundordförrådet i den riktning man önskar tills man når den nivå man vill ha. Våra erfarenheter är nog snarare att det är de mellanstora ordböckerna som är svårast, och att det är där det finns stora möjligheter att hamna snett. Arbete kategorivis är ett sätt att hålla sin ordbok under kontroll. En bred, men balanserad ordbok är nog vad gemene man behöver. Användarvänlighet är att tillgodose kraven från den ganska diffusa målgruppen också.

Det finns många slag av information som man kan ta med i en ordbok eller lämna utanför. Många användare kan inte skilja på olika typer av referensverk utan vill helst ha allt just de behöver på ett och samma ställe. Ordböcker som innehåller encyklopedisk information, t ex Collins Plus 1989, är kanske något för dem. Många namn har bestämda konnotationer — och mången användare är betjänt av att få veta något om den speciella miljö man tänker på när man talar om *Mayfair* eller *Rive Gauche*, *Silicon Valley* eller *Grinzing*.

En annan aspekt som man kan fundera mycket över är hur mycket grammatik som egentligen ska finnas i en ordbok. Vill användaren ha information som liknar den i en grammatik eller egentligen bara en räkka bra exempel att pröva sig fram med och se om något passar? Hur mycket kan vi utgå ifrån att användaren vet? Hur mycket text orkar användaren ta sig igenom? Ja, som vanligt beror det på vem användaren är. Den avancerade användaren som förstår ett hierarkiskt system är kanske mest hjälpt av en tydlig typografi, men välkomnar med all säkerhet också sådant som ger ytterligare hjälp att hitta. Det är t ex viktigt att konstruktions- och prepositionsuppgifter ges i en standardiserad och lätt igenkännlig form. COBUILD 1995 ger den hjälpen genom sin numera välbekanta "extra column", som dessutom fungerar som ett slags register.

4 Uttalsangivelser

En aspekt som är värd att fundera över är uttalsangivelser i ordböcker där det främmande språket är källspråk. I Sverige har traditionen varit att ge fullständiga uttalsangivelser bara för engelska och på sin höjd lite betoningstips eller uttal "när det avviker" i fråga om andra språk. Till min glädje konstaterar jag att denna tradition är på väg att brytas och att man allt oftare ger uttalsbeteckningar, åtminstone i fråga om franska. Där har man brukat hävda att uttalet är så regelbundet att man inte behöver visa det. Men det är ju inte till hjälp för den användare som inte kan reglerna. Och även vi som en gång lärt oss franska kan minsann behöva kontrollera att vi minns rätt!

En svårknäckt nöt är att få in uttalsangivelser där användaren verkligen behöver dem: efter översättningen i ordböcker där det främmande språket är målspråk. I en liten enkel ordlista går det bra att göra så, men i en vanlig ordboksstruktur skulle det bli alldeles för oöverskådligt. Såvitt jag vet har ingen än kommit på en lösning, men jag hoppas att den som en dag gör det får Nobelpriset! Till dess får användaren använda hela ordboksparet vid produktion på främmande språk: svenska till X för översättning och X till svenska för uttalet. Men det systemet är egentligen inte särskilt vänligt mot användaren!

Med ny teknik kommer däremot något som borde kunna vara det allra bästa för de allra flesta: ordboken med ljud. I CD-ROM-tekniken är det ju inga tekniska problem att åstadkomma det. Inga uttalsangivelser i världen kan bli tydligare än att få höra de ord man undrar över. Här tror jag att det kommer att hända mycket — och att det kommer att ligga långt ifrån gamla tiders språklaboratorier och språkkurser på band!

5 Redigering

Det som utmärker en riktigt bra redigering är att den inte märks. Användaren hittar sin information utan att egentligen fundera över hur det går till och utan att snubbla över svårigheter som distraherar. En redigeringsprincip som jag uppfattar som svår och distraherande är att använda huvudstickord och understickord — det som på engelska kallas "nesting". Särskilt svårläst blir det om understickorden klipps av och bara syns som bindestreck och en orddel. För oss som dagligen sysslar med ord och ordbildning kan det vara en pedagogisk poäng att sammanhörande ord förs ihop. Men för ett språk som vårt, med alla fasta sammansättningar,

blir texten ofta till små korta ”räckor” som då och då bryts på ett sätt som är förbryllande för många användare. Det är verkligen inte lika självklart för alla varför det i en spalt står:

rot|blöta, -borste
rote
rotel
rotera
rot|fast, -frukt, -fyllning osv.

Om alla stickord i stället behandlas som likvärdiga och påträffas helt utskrivna i samma position i spalten, dvs i vänsterkant och med tydlig stil är de lätta att hitta. Det är viktigt att vi som gör ordböcker om och om igen påminner oss själva om att vi inte är några typiska ordboksanvändare, utan att vi tillhör en minoritet — även om vi ser många ut en sån här gång!

I äldre ordböcker tycks idealet ha varit att komprimera texten så mycket som möjligt, för att spara plats, men ibland kan man undra om det var en dygd i sig att göra texten svårgenomtränglig. Kanske något slags ”survival of the fittest”? Ett exempel: ”face-(skin-)lotion”, vilket skulle utläsas ”face-lotion eller skin-lotion”, men där ju skriftbilden blir ganska mystisk.

Något som egentligen borde vara självklart är att inte använda avstavningar i ordbokstext, utom möjligen i ord som ska innehålla bindestreck. Texten är svårläst nog ändå. Brist på utrymme brukar anges som skäl för att avstava i ordböckerna. Men enligt prov som vi gjort — visserligen på stort format och i trespalt — sparar man bara lite över 2% på att avstava. Det är bättre att kosta på sig några extra sidor för att erbjuda större läsbarhet.

Förkortningar eller beteckningar för ämnesområden har behandlats på olika sätt i olika ordböcker. En klassisk metod var att använda symboler: ankare för sjöterm, korslagda svärd för militärterm osv. De var ju tydliga på så sätt att de inte kunde förväxlas med den löpande ordbokstexten, men jag har inte sett exempel på att de använts som ersättning för mer än en handfull beteckningar. Medge att det skulle vara kul att se en ordbok med fler sådana symboler: kanske blomma för botanik, elefant för zoologi, kastrull för kokkonst!

I den nyutkomna *Cambridge International Dictionary of English* (CIDE 1995) använder man helt sonika pund- och dollartecken för att markera brittiskt respektive amerikanskt uttal. Det är djärvt och så vitt jag vet nytt. Det ska bli intressant att se hur det blir mottaget.

Det viktiga är att förkortningar och inprickningar avviker typografiskt så att de inte riskerar att läsas ihop med den övriga texten. CIDE använder versaler med ram omkring, andra använder kursiv eller liten stil. En till synes obetydlig detalj som en liten punkt gör att det tydligare framgår att man har att göra med en beteckning för ämnesområde — många av dessa bokstavsföljder bildar ord i sig och man kan få svårt med tolkningen. Det gäller inte bara *bank* och *sport* utan i ännu högre grad sådana som kan föra tankarna åt helt annat håll, som *anat* och *ekon* — i värsta fall kan de uppfattas som en semantisk förklaring.

6 Typografi

Mycket av användarvänligheten ligger i hur texten ser ut. En skicklig grafisk formgivare är nog trots allt användarens bästa vän. För oss vana ordboksanvändare går det väl att ta sig igenom en text som ser mer eller mindre lika ut hela tiden, nästan som en romantext, som

ju läses löpande. Men ovana, dvs vanliga, användare behöver den hjälp som god typografi ger. God typografi är inte detsamma som jättestor text, utan är en fråga om balans mellan **grad** och **kägel** — alltså mellan "svart" och "vitt" på sidorna. Det är också en harmonisk växling mellan olika grad, olika **stilsorter**, **rak** och **kursiv**, **fet** och **mager** så att de olika beståndsdelarna i informationen tydligt framgår. I god typografi är den växlingen en viktig signal till läsaren: "Se upp, här börjar något nytt".

När det gäller uppslagsorden har jag redan propagerat för placering i vänsterkant. Med ett bra typsnitt framstår de ännu tydligare. Det är inte en slump att de allra flesta moderna ordböcker har uppslagsorden i ett **linjärt typsnitt** (eller **grotesk** som man tidigare sa), en tung men tydlig stil med jämntjocka linjer. Däremot har man sällan själva ordbokstexten i linjär, utan väljer ett mjukare typsnitt som är mera lättläst i löpande text.

Det finns andra sätt att göra uppslagsorden mera framträdande. Särskilt i franska ordböcker står uppslagsorden ofta med versaler, stora bokstäver, vilket ju är helt förkastligt eftersom användaren här går miste om en del av informationen. Ett annat sätt är att använda färg, blå eller röd, i uppslagsorden. Det ser ju ganska trevligt ut, men det skulle vara intressant att veta hur färgen upplevs av synsvaga eller lässvaga användare.

Hur hjälper man bäst användaren att hitta rätt i ordboken? Kolumntitlar? Tumgrepp? Tonplattor med ledbokstäver i högerkanten? Kolumntitlar är klassiska, men utformningen varierar verkligen. Ibland står det en hel roman upptill på sidan, om man ska ha med "från" och "till" på varje sida och orden dessutom råkar vara långa. Men egentligen räcker det väl med första ordet på vänstersidan och sista ordet på högersidan? Tumgrepp är bra när de är nya och fräscha, men när de börjar fransa sig gör de det snarare svårare att bläddra. Tonplattor kan nog också vara bra, men tryckerierna har ofta lite svårt att hålla placeringen, så resultatet kan bli hoppigt och fult.

7 För- och eftertext

Jag nämnde i början de **användaranvisningar** som tyvärr inte många läser. Det var kanske det man tog konsekvenserna av när man skrev anvisningarna i Kjærrulf-Nielsens *Engelsk-Dansk Ordbog* (1977). Sjutton rader till en ordbok på närmare femtonhundra sidor — mer lapidariskt kan det knappast bli! I någon mån kan anvisningar annars ersättas av något slag av snabböversikt över hur ordboken är uppställd. Om en sådan översikt ska fungera bör den placeras så att man lätt hittar den, gärna på försättsbladet. Van Dale ger den dessutom på ett bokmärke, vilket ju har den stora fördelen att man kan flytta det med sig genom ordboken. För- och eftersättsbladen är ju egentligen idealiska för allt av översiktskaraktär: listor över fonetiska tecken, förkortningslistor, och sådant som både är användbart och dekorativt, t ex kartor och flaggor.

Man kan hitta alla de slag av information i supplement och inledande partier: allt från bilder av cricketplaner och fotbollsplaner, till mått och vikt och listor över vad berömda operor heter på originalspråket eller långa läsavsnitt om ordbildning. Det verkar inte finnas några gränser för kreativiteten, men även här är det viktigt att innehållet är anpassat till den tänkta användaren! I en ordbok för användare på elementär nivå kan det vara bra att ha en del inlärningsmaterial samlat: veckodagar, räkneord och ordningstal, klockan. Den som har kommit lite längre har kanske mer nytta av brev- och telefonfraser. Den avancerade användaren är förmodligen van att själv skaffa sig information, men skulle

kanske välkomna en del sådant som lättast låter sig visas tematiskt eller i tabellform — t ex kemiska beteckningar.

8 Avslutning

Redan idag finns ordböcker i många olika former, på disketter och CD-ROM-skivor. Det finns enkla ordlistor för stavningskontroll i datorn — från riktigt torftiga och dåliga till ganska bra (fast några ordböcker är de förstås inte). Man kan koppla upp sig on-line mot olika ordbaser via Internet. Man kommer troligen att kunna köpa ordlistor "on demand" och prenumerera på lösblad eller uppdateringsdisketter. Det kommer att finnas nya former som vi inte har en aning om idag. Förhoppningsvis kommer vi att få ha kvar ordböcker i bokform också. Och förhoppningsvis kommer innehållet att få fortsätta att vara det viktiga, inte den yttre formen. Och vad som än händer i morgondagens multi-media-samhälle kommer ordböcker att förbli avancerade redskap som man måste lära sig att använda. Det är vi ordboksmakare som måste se till att användaren vill ha dem, att användandet av ordböcker är mödan värt!

Litteratur

- CIDE 1995 = *Cambridge International Dictionary of English*. Editor in Chief: Paul Procter. London/Cambridge: Cambridge University Press.
- COBUILD 1995 = *Collins COBUILD English Language Dictionary*. New Edition. Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks. London/Glasgow: Collins.
- Collins Plus 1989 = *Collins Concise Dictionary Plus*. Chief Editor: Patrick Hanks, Managing Editor: Marian Makins. London/Glasgow: Collins.
- Kjærulf-Nielsen 1977. *Engelsk-dansk ordbog*. København: Gyldendal.
- Svartvik, Jan 1995: Vilka språkkunskaper behöver vi inom EU? *Norstedts ordbokskatalog 1995*. Stockholm, 18–20.

Leena Savolainen

Finnish Sign Language – Finnish dictionary: defining the head signs

Största delen av de teckenspråkiga ordböckerna har hittills snarare varit teckenlistor än egentliga ordböcker. Inom Finlands Dövas Förbund pågår för närvarande ett ordboksprojekt vars mål är att utarbeta en ordbok över det finländska teckenspråket med stöd i lingvistisk forskning. I denna artikel ger jag först en allmän presentation av själva projektet och dess målsättningar och arbetsmetoder. Därefter beskriver jag hur vi arbetar när vi lemmatiserar uppslagstecknen. Under det att arbetet har framskridit har vi kunnat konstatera att de lingvistiska fenomen och problem som vi stöter på med vissa undantag är desamma som författare till ordböcker över talade språk konfronteras med.

1 Finnish Sign Language dictionary project

The Finnish Sign Language – Finnish dictionary project was started in 1988 by the Finnish Association of the Deaf. At the moment the dictionary project team has six employees. Four are deaf, with Finnish Sign Language (FinSL) as their first language. Two are hearing, with Finnish as their first language; one of these is employed by the Research Institute for the Languages of Finland. The dictionary is planned to be published by the end of 1997.

In addition to the team of full-time employees there is a larger working group supporting the project. This group has both deaf and hearing members. The deaf are of different ages and from various parts of Finland, and both deaf and hearing represent various fields of sign language work and/or linguistic study. The group comes together twice a month on average to discuss the material we have produced.

Traditionally, sign language dictionaries have been more like lists of signs than like actual dictionaries. Each sign has been given only one spoken language equivalent (gloss), and the signs are represented by drawings or photographs, which are then listed according to the alphabetical order of the spoken language glosses. *Viittomakielen kuvasanakirja* (Pictorial Dictionary of Sign Language; 1973) — the dictionary currently in use in Finland — is an example of these traditional dictionaries.

A new kind of sign language dictionary is now being assembled by the dictionary project of the Finnish Association of the Deaf. In this dictionary FinSL is recognized as an independent language of equal status to the Finnish language, and the signs are arranged according to their structural components, not according to their spoken language equivalents. Comparable dictionaries have been made of e.g. the Thai, Australian and British Sign Languages.

The new FinSL dictionary will contain approximately 1 200 entries. Each entry includes:

- A head sign, which is represented by a picture with additional arrows showing movement
- The Finnish equivalents of the sign
- Translations of an average of four signed sentence examples into Finnish
- Grammatical information and other notes on the sign

In preparing an entry the following procedure is followed:

1. There is a list of glossed signs, which has been approved by the larger working group. The deaf team-members produce about four examples of sentences, with each sign used in different contexts. The sign along with its examples is videotaped.
2. The hearing members establish the Finnish equivalents of each sign and translate the signed examples into Finnish.
3. The deaf and hearing members together check the Finnish equivalents and translations, and also add grammatical information and other notes on the sign.
4. The larger working group verifies the accuracy of the entries and gives feedback to the project team.
5. Entries are then corrected as necessary.
6. The entries are videotaped again in their final form. The Finnish equivalents of the head signs, translations of the signed examples into Finnish and grammatical information and other notes on the sign are all checked once again.
7. For the illustrations the citation forms of the signs will all be videotaped from different angles. Suitable frames will be transferred (grabbed) to computer disc, and the pictures will receive their final finishing with arrows showing the movement in the signs.

In the next section I turn to describing the procedure we follow when defining the head signs. In other words I go into details of stage (1) of the procedure described above.

2 Defining the head signs

A preliminary list of about 1 200 signs was compiled in order to provide some basis for the more final decisions on which signs would be included in the dictionary and what exactly the forms of those signs would be. This list groups the signs under different headings, such as nature, home, question signs, numerals and time. The signs are noted in terms of Finnish-language glosses. Notes on the articulation of the signs are often added. The notes use partly a transcription system (Rissanen 1985:68–69) and partly explanations in Finnish. The 1 200 signs chosen for the book are basic signs of everyday vocabulary.

The 1 200 signs are discussed one at a time in the larger working group. During these discussions some signs are added to the list and some are dropped. The list supports the work, but it is not something we have to stick to.

With almost every sign we find individual variation between signers, or we get many variants from one signer. This variation can be dialectic, idiolectic or variation in FinSL which is commonly known by every deaf person in Finland. I would not like to call this commonly known FinSL a “standard FinSL”. The lexicon used differs from generation to generation, the difference being the biggest between the youngest and oldest generation of signers. Furthermore, nobody has ever standardized FinSL. It is not written. It has not been a subject of instruction at schools for the deaf (children were even totally forbidden to use it at schools for several decades in the current century). FinSL is a minority language which is only just beginning to gain appreciation.

The dictionary will omit regional dialectic and idiolectic variation. Our aim is to include only FinSL signs that are used or at least understood by all deaf people in Finland. Despite this, we know that the language used in the Helsinki area by fairly young people (25–35 years old) will be over-represented as our deaf team-members use mainly that particular form of the language.

3 Studying the sign more closely

To begin with I should mention that at this stage of the procedure we concentrate almost totally on FinSL in all discussions. The deaf people have lived in an atmosphere where FinSL was thought to be some kind of representation of Finnish and not an independent language in its own right. By concentrating solely on FinSL, we seek to minimize the influence of the Finnish language on the FinSL material produced. Only after an entry is videotaped do we begin to look for Finnish equivalents for the sign.

The deaf members of the larger working group make up sentences in which the sign is used in different contexts and — if the sign can be inflected — also in its different inflected forms. On this basis we discuss at least the following: What is the semantics of the sign? Is there a case of homonymy or polysemy connected to the sign? Which inflected forms of the sign really are inflections and which are actually derivative forms? Which category or categories of part of speech does the sign belong to? Which variations of the base form of the sign are dialectical, which idiolectical, and which are variants of the commonly known FinSL? In the following paragraphs I discuss each of these questions in turn.

To get a general picture of the meaning of the sign we list several different sentential contexts in which the sign can be used. During this listing we also look for cases of homonymy and polysemy, thus trying to draw a line between them.

Picture 1 gives an example of a case of **homonymy**. The sign in picture 1 can mean either ‘congregation, parish, parishioner, parishioners’ or ‘United States of America, America, American’. The latter sign is a loan from the American Sign Language and it happens to be of exactly the same form as the former FinSL sign the origin of which we do not know. In general in such cases the signs will be given two separate entries. But in this particular case the sign meaning ‘America’ etc. will be excluded, as we shall not include proper names in our dictionary.



- I) 'congregation, parish, parishioner, parishioners'
- II) 'United States of America, America, American'

Picture 1. *The picture is taken from the Pictorial Dictionary of Sign Language (p. 9).*

Picture 2 shows an example of **polysemy**. The sign in the picture has the meanings: 'a cross', 'a chemist shop', 'an ambulance' and 'Switzerland'. All these meanings will be put under one entry in the dictionary.



- 1. 'a cross, cross-shaped'
- 2. 'chemist shop, chemist, chemist's, pharmacy'
- 3. 'ambulance'
- 4. 'Switzerland, Swiss'

Picture 2. *The picture is taken from the Pictorial Dictionary of Sign Language (p. 268).*

These two cases — one of homonymy, the other of polysemy — are quite clear ones. Usually we do not know, for instance, what the etymology of a particular sign is. Thus we cannot use, as often as we would like to, etymology as a criterion for drawing the line between cases of homonymy and those of polysemy, not to mention the line between inflection and derivation.

Solving matters concerning **inflection** and **derivation** are often more troublesome than those of homonymy and polysemy. Some verb signs in particular can be very complicated, as there may be a group of base forms which all have something in common both in form and meaning. All of these base forms may then have their own inflectional paradigms and derivative forms. Unfortunately we do not have enough time to study these verbs thoroughly. However, we do our best to get them in a sensible order in the dictionary.

In the above list of questions the **part of speech** was also mentioned. Actually nobody has yet defined what the parts of speech of FinSL might be. It seems to share some categories with other languages known by project members (e.g. Finnish, Swedish, French, Chinese). Yet some signs behave in such a way that it is hard to say whether they can be classified in terms of part of speech similar to those of these other languages.

FinSL has, for instance, a class of nouns and verbs which share the same citation form. We put nouns and verbs of this kind in the same entry. This type of base form is very common in English, for example *loan*, *rule*, *part* and *dream*, just to mention a few.

The last stage in the procedure for defining the head signs is to establish the precise base form for the sign. The deaf members of the larger working group give the base form by intuition. But, as was mentioned above, there is a lot of variation in the base forms offered.

In the signs chosen for the dictionary the variation can be either phonemic¹ or allophonic. If there is allophonic variation in the sign, we decide on one allophone and use it in the citation form. If a sign has, for instance, two possible handshapes, and the meaning of the variants is exactly the same, we include both variants of the citation form in the entry. But note that we accept only variation in single phoneme, not variation in two or more phonemes.

Signs of the type I have been discussing in this article are sometimes called **non-polymorphemic signs** (see e.g. Engberg-Pedersen 1993). There exists also another type: **polymorphemic signs**. In a nonpolymorphemic sign the citation form consists of one morpheme (two or more in compound signs). In a polymorphemic sign the phonemes (or with some phonemes even different parts of the phoneme) have a meaning of their own, thus being morphemes. The problem with these polymorphemic signs for a lexicographer is that they do not have a base form, just a root. In the dictionary project we have not yet decided how to deal with this problem, and I will not discuss the matter further here.

4 In conclusion

Finnish Sign Language (FinSL) is still very little studied. When compiling the dictionary we cannot make much use of the Pictorial Dictionary of Sign Language — the FinSL dictionary currently in use — because it is more like a list of signs than like an actual dictionary, and it is not constructed on the kind of linguistic basis which we are working with now. This situation forces us to do a lot of basic research on FinSL during the process of compiling the dictionary. From a linguistic point of view, compiling a sign language dictionary is in many ways similar to compiling a spoken language dictionary. The lexicographic phenomena we are dealing with in FinSL are, with few exceptions, the same as those one faces while working with a spoken language.

References

Dictionaries

- Johnston, Trevor 1989: *AUSLAN Dictionary. A Dictionary of the Sign Language of the Australian Deaf Community*. Maryborough, Victoria: Australian Print Group.
- Brien, David (ed.) 1992: *Dictionary of British Sign Language / English*. London: Faber and Faber.
- Wrigley, Owen P. (ed.) 1990: *The Thai Sign Language Dictionary*. Revised and expanded edition. The National Association of the Deaf in Thailand. Bangkok: Thai Watana Panich Press.
- The Pictorial Dictionary of Sign Language = *Viittomakielen kuvasanakirja* 1973. Finnish Association of the Deaf. Helsinki: Kirjapaino Oy Surdus Ab.

¹The possible phoneme types in sign languages are: handshape, palm orientation, finger orientation, location (of the hands), movement of the hands, mouth movement, facial expression, head movement and body movement.

Other literature

- Engberg-Pedersen, Elisabeth 1993: *Space in Danish Sign Language. The Semantics and Morphosyntax of the Use of Space in a Visual Language*. International Studies on Sign Language and Communication of the Deaf, volume 19. Hamburg: Signum Press.
- Rissanen, Terhi 1985: *Viittomakielen perusrakenne* [Basic structure of sign language]. University of Helsinki, Department of General Linguistics, Publications No. 12. Helsinki: University Press.

Mer enn nok er for mye

This paper describes a research project at the Dutch publishing house Van Dale during 1994 and 1995. The main goal was to develop a practical working method for corpus lexicography. The second aim was to explore the borderline between general language and technical language. The title "More than enough is too much" hints at the idea that — for the sake of both lexicographer and budget — a corpus should be limited in size. During the project we built a small, but very rich **micro corpus** with medical texts. Selecting adequate texts, according to guidelines dictated by the profile of the dictionary, will become an important task for lexicographers.

Mitt foredrag "Mer enn nok er for mye" består av fem deler:

1. Van Dale og korpusleksikografi
2. Opplegg for undersøkelsen
3. Programvaren
4. Pilotundersøkelsen som er utført for det medisinske område
5. Konklusjoner

1 Introduksjon VDL & korpusleksikografi

Nesten alle eksisterende Van Dale-ordbøker er laget uten at en elektronisk tilgjengelig samling med tekster er ekserpert systematisk. Altså uten korpusleksikografi.

Våre ordbøker har et godt navn blant publikum, og selv er vi heller ikke misfornøyde med dem. Det er altså ikke absolutt nødvendig å endre den metoden som er brukt til nå. Men i dag *har* man muligheten til å utføre korpusleksikografi, og vi ville gjerne bli bedre kjent med metodens fordeler. Utgangspunktet måtte være å oppdatere og forbedre våre eksisterende ordbøker. (Fra nå av vil jeg snakke om "ordboken", og jeg henviser da til alle ordbøker hvor nederlandsk står på venstre siden. Dette er både moderne og diakroniske ordbøker, og både forklarende ordbøker og oversettelsesordbøker.)

Formålet var altså ikke å lage en ny leksikal beskrivelse av nederlandsk "from scratch", slik som COBUILD har gjort det for engelsk. Vi ønsket å bearbeide ordboken vår med utgangspunkt i noen av de fem nøkkelord som prof. Verkuyl (1994) nevnte i sitt foredrag under Euralex 1994 in Amsterdam: komplett, konsistent, korrekt, aktuell, sitater.

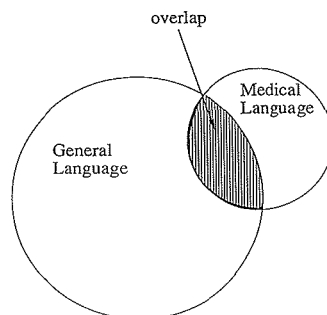
Vi vet at det er visse svakheter i behandlingen av enkelte fagområder i eksisterende ordbøker. Terminologien fra sjakk er ikke **komplett** hvis *springer* og *konge* er tatt med, men *bonde* mangler. Behandlingen er ikke **konsistent** hvis det i definisjonen for *løper* fortelles hvordan brikken beveges, og denne informasjonen mangler for *tårnet*. Informasjonen er ikke **korrekt** hvis det står at et *sjakkbrett* har 100 felt. Ordboken er ikke **aktuell** hvis det

kun er tatt med den gamle termen for en brikke, *kasteel*, istedenfor *toren*. Eksemplene er fiktive, men de viser hva fire av de fem nøkkelordene innebærer.

Hensikten med prosjektet som omtales i dette foredraget, var å undersøke hvordan bruken av tekstkorpus kunne føre til bedre kvalitet på ordbøker og større effektivitet under det dyre manuelle leksikografiske arbeidet.

2 Opplegg for undersøkelsen

Helt fra begynnelsen av hadde vi to målsetninger med undersøkelsen. Den ene var å utvikle en metode for innsamling av termer ved å bruke tekstkorpuser. Den andre var å avgrense fagområder i den generelle ordboken. I figur 1 skisseres hvordan fagspråkene overlapper det generelle språket – eller det utvalget som hører hjemme i en generell ordbok. Hvordan fastslås denne grensen? Hvilke termer fra fagspråkområdet hører hjemme i en vanlig ordbok?



Figur 1.

Vi har dannet en arbeidsgruppe av sakkyndige, med blant annet en terminolog, redaktøren for en enspråklig medisinsk ordbok og forfatteren av en tospråklig juridisk ordbok. Altså mennesker med kunnskap om både fagområde og leksikografi. Denne arbeidsgruppen har gitt oss råd om opplegget for undersøkelsen.

Arbeidsgruppen hadde følgende kommentarer:

1. Det finnes to typer korpusbruk. Den ene kaller vi **verifisering**, den andre **oppsporing**. Verifisering innebærer at man vet hvilket ord man vil vite mer om. For verifisering gjelder at et korpus nesten ikke kan være stort nok, og at det kan oppnås gode resultater med relativt enkel programvare. I vår undersøkelse ligger hovedvekten på **oppsporing**.
2. Det er vanskelig å formulere søkeoppdrag for ukjente termer. Så snart man ved at ordet X finnes i teksten, kan man undersøke den leksikale omgivelsen til X med ethvert computerprogram som har en søkefunksjon. Men først må man vite hvilke ord man ønsker å vite mer om. Hvilke ord er interessante? Hvilke ord er de ukjente X-ene? (Jeg kommer tilbake til svaret på disse spørsmålene.)
3. I forbindelse med **oppsporing** støter man ganske snart på fenomenet om begrenset ekstra utbytte. Jo mer tekst som er tilgjengelig, desto større blir arbeidet med å

finne noe interessant. Flaskehalsen i den leksikografiske prosessen er leksikografens arbeidstid. Gjør korpuset minst mulig, og sørg for en elektronisk "toolkit" som muliggjør hurtig søking i korpuset.

4. Gjør utvalget så bredt og produkt-uavhengig som mulig. Senere kan det foretas egne delutvalg for forskjellige ordbøker. Til en forklarende ordbok vil det bli valgt andre termer enn for en oversettelsesordbok. Et generelt kjennetegn for samlingen er: ikke-spesialistisk.
5. Velg tekstene slik at spørsmålet "skal denne termen tas med i ordboken?" allerede er besvart så snart termen finnes i en tekst. For det medisinske området er en populær legespalte i et ukeblad mer relevant enn et legevitenskapelig tidsskrift.
6. Ta hensyn til den relative, samfunnsmessige viktigheten av fenomener. Et "medisinsk" eksempel: Det faktum at det snakkes og skrives mye om en bestemt sykdom er viktigere enn antall personer som lider av den sykdommen.
7. Sørg for (transkribert) talespråk i korpusmaterialet.

Nå vil jeg så vidt gå tilbake til den andre kommentaren: For å kunne oppspore ukjente termer, har vi forsøkt å omskrive hva "interessant" innebærer innenfor rammen av vårt prosjekt. For hver kategori nevner jeg eksempler fra sjakkterminologien. Interessant er:

- et uvanlig ord som ikke hører til basisordforrådet, men som finnes i teksten (*matt*)
- et ord som finnes mye hyppigere i teksten enn i et korpus av generelt, ikke-spesialistisk, gjennomsnittlig språk (*hest, hvit*)
- faste kombinasjoner av ord som finnes utelukkende, eller oftere enn gjennomsnittlig i teksten (*fri bonde, slå en passant*)

3 Verktøy

En første betingelse for å kunne utføre korpusleksikografi er en **taggerlemmatiser**, et program som tilføyer ordtype og lemmaform til ordet. Det fantes ikke et slikt program for nederlandsk, så Van Dale har satt i gang arbeid med å utvikle det. Avdelingen for leksikologi ved Vrije Universiteit i Amsterdam har laget **D-tale** (Dutch-tagger-lemmatiser) for oss. På figur 2 kan man se hvordan programmet virker.¹ For hvert ord bestemmer det lemmaform, med et **kommersielt à-tegn** (@) foran. Etter **nummertegn** (#) kommer en tall/bokstavkode for ordtype og bøyningsform.

En annen forutsetning er en **toolkit**, som både kan brukes til verifisering og oppsporing. Hos Van Dale har vi utviklet programmet **Coco** til dette formålet.²

¹Eksemplet betyr: Er det kaldt i Reykjavik?

²Det eiendommelige navnet angir at programmet først og fremst hjelper til med søking av kollokasjoner (**collocations**) i et korpus (**corpus**).

D-tale (Dutch tagger/lemmatizer).

before D-tale:

Is het koud in Reykjavik?

after D-tale:

Is@zijn#3E het@het#7 koud@koud#2
in@in#6 Reykjavik@Reykjavik#1N
?@?#I

Figur 2.

Jeg begrenser meg til å nevne noen viktige funksjoner:

KWIC. D-tale og Coco arbeider med hele setninger. Ut-data avviker derfor fra det vanlige KWIC-formatet. På figur 3 kan man se hva søk etter ordet *operatie* (operasjon) i kombinasjon med et verb fører til.

Lagring av resultatet. Resultatet av et søkeoppdrag lagres automatisk. Materialet får den strukturen som brukes til lagring av leksikal informasjon hos Van Dale.

Frekvens. Frekvensen kan beregnes både for ordform og lemma. Spørsmålet om frekvensen av ordet *operatie* (operasjon) gir resultatet man kan se på figur 4.

Z-score. Den viser hvilke ord gjerne står i nærheten av hverandre. Eller mer vitenskapelig formulert: den relative grupperingsstyrke mellom søkeordet og dets leksikale omgivelse. Figur 5 viser resultatet av spørsmålet om preposisjoner fremfor ordet *operatie*.

in KWIC format

In een latere fase wordt de definitieve
[> operatie <] [> verricht <].

in Lemma format

<COMPO.> operatie
<GRAMT.> 1
<COMPO.> verrichten
<GRAMT.> 3
<STVRM.> een operatie verrichten
<CITAAT> In een latere fase wordt de
definitieve operatie verricht .
<BRONC.> Nigsys teksten
<BEWER.> rik

Figur 3.

```

Coco query result

Frequency of the lemma operatie

  1   Operatie@operatie#1E
  3   Operaties@operatie#1M
163   operatie@operatie#1E
 25   operaties@operatie#1M

Total 192

```

Figur 4.

```

Coco query result

Z-score for operatie, preceded by a
preposition

Expression A: operatie
Scope       : 3
Surr       : only before A
Collocator  : (B)
Minimum    : 2
Expr.Coll. : preposition
Totaal #Wrd : 444111

Score  F(B) F(A+B) preposition (B)
29.2802 587 30 na [after]
5.3387 265 4 tijdens [during]
4.2142 4146 18 bij [at]
2.0803 1943 7 door [result of]
1.8926 958 4 uit [from]

```

Figur 5.

Kompleks grupperingssøker. Denne kalkulerer de sterkeste grupperinger i en tekst uten at søketermen må angis. Eksempler fra et korpus om sjakk kunne være: *svart står sjakk; kort rokade.*

For å begrense regnetiden mens man bruker programmet, og for å ha en rettesnor mens man utfører valget, er fremstilling av hjelpelister en del av standardprosedyren ved forberedelsen av en tekst. Listene oppfyller det tidligere nevnte kriteriet "interessant".

Uvanlige ord. Ved å sammenligne ordet fra et korpus med to kilder med utelukkende "vanlige" ord, valgte vi det uvanlige ordet fra korpuset. Som kilde for vanlige ord brukte vi (1) Basisordbok fra Van Dale, en skoleordbok med 24 000 oppslagsord og (2) et normkorpus, basert på avistekster.

Siden det er valgt tekster med enkel språkbruk, er ordene i denne listen nøyaktig de ordene som er forbundet med tekstens tema, altså med det undersøkte fagområdet. Interferens oppstår særlig av sammensetninger, navn, akronymer, tall og skrivefeil.

Relativt høy frekvens. I denne listen sammenligner vi frekvensen til et ord i korpuset med den forventede frekvens på grunnlag av tellinger i normkorpuset. På denne måten håper vi å finne en spesiell betydning eller konnotasjon for det undersøkte fagområdet. Programmet virker slik at jo større avviket er fra det forventede, desto høyere score. Scoren er altså ikke direkte avhengig av ordets absolutte frekvens.

4 Pilotundersøkelse for det medisinske området

Vi har valgt det medisinske området for å utføre en pilotundersøkelse. Området er stort, det er i stadig utvikling og den eksisterende ordboken er ikke ajourført. Undersøkelsen består i å:

- fastslå hvilken type tekst innenfor området brukeren må kunne dekke/bearbeide med ordboken (tekster skrevet for ikke-medisinere, som når et stort publikum)
- velge tekster som oppfyller typen (legespalter i ukeblader, opplysningsbrosjyrer fra sykehus og pasientforeninger, samtaler mellom lege og pasient, en fjernsynsserie fra et sykehus)
- samle tekster og forarbeide dem elektronisk (be forlag eller forfattere om disketter; fjerne ikke relevante koder i tekstene; D-tale)
- ekserpere tekstene systematisk (redaksjonsarbeid)
- sammenligne utvalget med det som står i den eksisterende ordboken
- få en sakkyndig på området til å bedømme utvalget
- trekke konklusjoner. Er spørsmålene besvart, og er målene nådd?

De data som nevnes her, gjelder de medisinske *n*-ord, det vil si ord som begynner på en *n* og sammensatte uttrykk hvor ett av ordene begynner på en *n*:

Det finnes 199 *n*-oppslagsord med en <medisinsk> etikett i ordboken (det kan også være anatomi, patologi eller noe annet). Av de sammensatte forbindelsene som finnes med disse oppslagsordene, regner jeg 25 til det medisinske området. Totalt blir det 224. Det er for øvrig vanskelig å fastslå grensen. Under *nese* tar jeg også med *en tett nese* og *puste gjennom nesene* til det medisinske området, men ikke *pille nese*.

Det er valgt 443 *n*-termer fra mikrokorpusene (300 enkle som begynner på en *n*, 143 sammensatte hvor ett av ordene begynner på en *n*). Mange av disse termene står i ordboken uten <medisinsk> etikett. Det forklarer de forskjellige tallene i feltet "+/++" på figur 6.

in dictionary	In μ -corpus of relevant domainrelated texts	
	+ 443 N	-
+	+/+	+/-
224 N <med.>	64 (30%)	160 (70%)
without label	148 (33%)	
-	-/+	-/-
	295 (67%)	

Figur 6.

Forskjellene kan gjengis i følgende skjema:

- +/+ Gir en bekreftelse på at utvalget allerede er gjort tidligere av en leksikograf i forbindelse med ordboken.
- +/- I korpuset er det ikke funnet noen bekreftelse på ord som er tatt med i ordboken. Bedømmelse av en sakkyndig gir svar på om den eksisterende ordboken inneholder noe som ikke hører hjemme i den, eller om utvalget av tekstene eller av termene fra de tekstene ikke var det riktige.

- /+ Ordboken inneholder åpne felt. Også i dette tilfellet må en sakkyndig bedømme om det dreier seg om nyttige tillegg, eller om det skyldes et mindre omhyggelig valg av tekster, eller av termer fra de samme tekstene.
- /- Alle ikke-eksisterende ord kommer i denne kategorien. Men vi forblir nysgjerrige etter de interessante termene som vi åpenbart ikke har valgt de riktige tekstene til.

4.1 Bedømmelse av kategori +/– (finnes i ordboken, ikke i μ -korpus)

Som forventet er mange termer modne for en kritisk bedømmelse. Flere vil bli sløyfet, eller blir merket som <gammel>. De har for lett blitt tatt med fra tidligere utgaver, selv om termen eller fenomenet som blir angitt med termen, er gått ut av bruk. Enkelte burde kanskje aldri ha vært tatt med, fordi de var altfor spesielle for en generell ordbok, f.eks. *net* (fettrik hinne mellom buk og innvoller), *netelkoorts* (feber som forårsakes av irritasjon av huden med brennesle ol.), *nijdnagel* (der hvor huden er revnet langs en negl), *nootgewricht* (kuleledd, ledd som kan beveges i alle retninger), *natriëmie* (å forebygge natrium i blodet), *negeponder* (baby som veier ni pund ved fødselen).

I andre tilfeller dreier det seg om sammensetninger hvor det er tvilsomt om de er relevante for ordboken. Betydningen av ordene er ganske åpenbar, de forekommer ikke hyppig og de utgjør ikke grunnlaget for lengre sammensetninger, f.eks. *neusklier* (nesekjertel), *neusheelkunde* (nese medisinske), *nierader* (nyreåre), *nierkanker* (nyrekreft). Det er ikke nødvendig å fjerne dem fra databanken, men det er ikke særlig sannsynlig at de tas med i det utvalg som ethvert leksikografisk produkt er.

En annen, mindre del av ordene i denne kategorien hører imidlertid absolutt hjemme i ordboken. Det at de ikke finnes igjen i μ -korpussen, skyldes som regel den vage begrensningen av området. De "medisinske" etiketter i ordboken er mer omfattende enn de kriteriene som tekstene for korpussen ble valgt etter. Anatomiske termer er et eksempel på dette, særlig når de i større grad gjelder dyr enn mennesker (*nekhaar* (nakkehår), *nekvel* (nakkeskinn)).

4.2 Bedømmelse av kategori –/+ (finnes i μ -korpus, ikke i ordboken)

Siden ekserperingsarbeidet er utført med en instruks om at termer skal velges i tilfelle tvil, er samlingen stor, større enn det som får plass i en gjennomsnittlig generell ordbok av normalt omfang. Det er selvsagt ingen ulempe, men det er viktig å være klar over det ved bedømmelsen av materialet.

Antall tillegg i forhold til ordboken er imponerende. Helt vanlige ord, som også mange lekfolk på det medisinske området umiddelbart vil gjenkjenne, viste seg til vår forundring å mangle i ordboken: *nepmiddel* (placebo), *neurotransmitter* (transmitter; definisjon i *Bokmålsordboka*: "stoff i kroppen som overfører impuls fra nerve til muskel"), *natriumbepert dietet* (diett med nedsatt natriuminnhold), *natuurlijke bevalling* (naturlig fødsel), *bijkomen uit de narcose* (våkne/komme seg etter narkosen).

Det er også noe interferens i materialet.

1. I denne gruppen finnes det dubletter som kun avviker i stavemåte og et ganske stort antall semi-dubletter av typen *negatieve uitslag* (negativt resultat), *negatieve testuitslag* (negativt prøveresultat), *negatieve aids-test* (negativ aids-test).
2. Det finnes et temmelig stort antall forbindelser og sammensetninger hvor ordet (orddelen) som begynner på bokstaven *n*, ikke tilhører det medisinske området, slik som *ta* (pulver, medisin, tablett, paracetamol) og *ikke* (ikke gravid, ikke arvelig, ikke medfødt).

Til nå har jeg utelukkende snakket om innsamling av termer og ikke om hvordan disse termene må behandles leksikografisk. Noen av de valgte tekstene er av opplysende art. Det innebærer at mange termer blir forklart i teksten. *Coco* lagrer en valgt term automatisk, sammen med hele setningen som termen brukes i. Slik oppstår det i tillegg til ordene en samling med sitater som kan spille en viktig rolle ved definisjonen av termene. Noen eksempler:

nekrose: lokal død i vev (for eksempel ved koldbrann eller forbrenning).

neseendoskopi: innvendig undersøkelse av nesen med endoskop.

5 Konklusjoner

5.1 Pilotprosjektet (medisinske termer)

Selv et svært begrenset korpus gir mye brukbart materiale for oppgradering og supplering av den eksisterende ordboken.

- Mange av termene som er tatt med i ordboken, kan fjernes eller revurderes ved hjelp av materialet i μ -korpuserne.
- Materialet gir en sammenhengende, relativt komplett samling av termer, hvorav mange utgjør et viktig supplement til den eksisterende ordboken.

5.2 Den valgte fremgangsmåten

- Omfattende ekserpering av et ganske lite korpus kan føre til en avbalansert samling av termer, forutsatt at korpuset er fremstilt svært nøye. Kvaliteten på utvalget er helt avhengig av valget av tekstene. Den nøyaktige begrensningen av området er av avgjørende betydning.
- Hjelp fra område-ekspertise — altså en lege i vårt eksempel — kan begrenses til rådgivning om valget av tekster og bedømmelse av resultatene. En stor del av arbeidet kan utføres av computere og ikke-eksperter. Det reduserer utgiftene.
- En videre automasjon av prosessen, særlig ved indre kontroll av ordboken, ser ut til å være mulig og vil gjøre prosessen enda mindre arbeidskrevende.

- Den undersøkte metoden begrenser seg ikke nødvendigvis til fagspråk. Også til oppsporing av neologismer eller ungdomsspråk kan det arbeides med et lite korpus av nøye utvalgte tekster.

Litteratur

- Johansson, Stig 1991: Times change, and so do corpora. I: Karin Aijmer/Bengt Altenberg (eds.): *English corpus linguistics*. London: Longman.
- Meijs, Willem 1994: Computerized lexicons and theoretical models. I: Nelleke Oostdijk (ed.): *Corpus-based research into language*. Amsterdam: Rodopi.
- Summers, Della 1993: Longman/Lancaster English Language Corpus — Criteria and Design. I: *International Journal of Lexicography* 6, 3, 181–207.
- Verkuyl, Henk 1994: *Knowledge Representation in Dictionaries*. Keynote lecture delivered at the 6th Euralex International Congress 1994. Amsterdam.

Multimedialeksikografi: nye utfordringer for leksikografisk teori

Når vi bruker ordet bok, har vi alle en noenlunde klar forestilling om hva vi mener. Det er en gjenstand som består av trykte sider, bundet sammen i ryggen og gjerne med et omslag eller innbinding. Vi sier at en forfatter har skrevet ei bok, og mener da at han har skrevet en tekst som blir presentert i trykk mellom to permer. Vi er så vant med det halvtusenårgamle ekteskapet mellom boktrykkerbransjen og ordets folk at vi knapt tenker over at våre forestillinger er farget av teknologien. Ordet bok betydde opprinnelig ei treplate av bøk som ble brukt til å risse runer inn i. I Bibelen betydde ei bok en papyrusrull med skrift på. De gamle islandske mestre skrev sine bøker på skinn. De eldste bøkene vi kjenner fantes antakelig først bare som muntlige overleveringer fra forteller til forteller, før de endelig ble nedtegnet i skriftlig form. Forestillinga om boka er altså i høy grad et produkt av teknologi.

Leksikografien er en bestemt disiplin innen vitenskapene om ord. Den har til de grader vært bundet til den trykte boka at den har tatt preg av det. Den trykte bokas omfang og begrensninger har virket tilbake på leksikografien og leksikografene. Nå som den elektroniske boka er på full fart inn i hverdagen, vil vi om litt bli konfrontert med dette.

1 Den trykte bokas begrensninger

Den trykte boka har satt klare grenser for

- omfanget av stoffet
- systematikken
- rekkefølgen
- presentasjonsformen

La oss se på disse begrensningene:

1.1 Omfang

Den norske *Bokmålsordboka* har ca. 5000 tegn pr. side på i alt 697 sider, dvs ca. 3,5 mill. tegn. I forhold til en standard roman på kanskje vel en halv million tegn, er jo ordboka reine mursteinen. Men i forhold til kapasiteten på en CD-ROM, blir mursteinen redusert til en grusbit, eller nærmere bestemt en halv prosent av de 700 millioner tegn ei slik plate kan romme. Til sammenlikning er det som om det bare skulle vært trykk på de første 3,5 sidene i *Bokmålsordboka*, mens de resterende 693 sidene var blanke.

Ønsket om å klemme maksimalt med ordboksinformasjon inn i ordbøkene har skapt et eget ordbokspråk med forkortelser, komprimerte setninger, koder og eventuelt også symboler. I *Bokmålsordboka* kan f.eks. en artikkel se slik ut:

I bo subst (Ity *behof*, sm o s *behov*) i uttr *ha bo* trenge, ha bruk for

I denne korte artikkelen er det seks forkortelser og en kode. Hadde ikke ordboka vært skrevet på denne måten, ville den kanskje krevd tre ganger så mange sider. På CD-ROM forsvinner dette argumentet. Vi ville fortsatt ha hatt 98,5% av plata til disposisjon.

1.2 Presentasjonsform

Den trykte bokas fysiske begrensninger påvirker også presentasjonsformen. Hvor mye grammatikk bør ei ordbok inneholde? Hvor mange eksempler? Hvordan skal sammensatte ord framstilles?

Vi har vendt oss til å bli avspist med kode- og telegramspråk. Bare den aller nødtørftigste grammatikk får plass og da gjerne i form av koder som *m1*, *v2* og *a5*. Eksempelene er sjelden belagt med formuleringer fra det virkelige liv, hvis vi da ser bort fra svære verker som *The Oxford English Dictionary* i nærmere 20 bind. Vil vi ha etymologi ut over det rudimentære, må vi gå til egne etymologiske ordbøker. I norske ordbøker settes gjerne sammensatte ord inn under en hovedartikkel, slik som

ord n1 (norr. . .) . . . **-bok** alfabetisk ordnet samling av ord fra et språk

Men med den plassen vi får til rådighet på en CD-ROM kan vi gjøre dette radikalt annerledes.

I en del ordbøker og de fleste leksika brukes illustrasjoner for å forklare teksten bedre. Den første som utstyrte et oppslagsverk med illustrasjoner skal ha vært den hellige Isidore fra Sevilla. Denis Diderot la meget stor vekt på illustrasjonens betydning i den store franske encyclopedien. Tekst og illustrasjon skulle bidra til å sprengte laugenes kunnskapsmonopol og bringe kunnskapen ut til folket. Slik fikk illustrasjonen en viktig rolle i encyclopediske verk. Men fortsatt har den trykte boka tvunget oss til å holde oss til skriftlig framstilling av lyd og bevegelse. Med elektroniske leksika stiller saken seg ganske annerledes.

Jeg skal komme tilbake til noen forslag.

1.3 Rekkefølge

Den vanligste rekkefølgen i ei ordbok er alfabetisk orden. Men vi kjenner mange avvik fra dette. I kryssordbøkene er betydning og antall bokstaver som regel overordnet alfabetet. I rimordbøker er slutten av ordet viktigere enn begynnelsen. I tesauriske oppstillinger kan fagområde være første nøkkel.

Men felles for alle mulige rekkefølger i den trykte boka: *når rekkefølgen er valgt, ligger den fast*. Den trykte boka er et *lineært* medium. Det er ikke den elektroniske boka. Her er alle ønskelige rekkefølger mulige. Utfordringa for leksikografen blir da delvis å ha tenkt gjennom hvilke alternative sorteringsmuligheter som skal kunne tilbys og delvis å sørge for at innholdet i den elektroniske boka er slik at den alternative sorteringa gir mening. Hvis alle oppslagsord er forsynt med ordklasse, kan f.eks. brukeren sortere dem først etter ordklasse og dernest omvendt alfabetisk, så har hun en rimordbok over adjektiver!

Mange ordbøker og de fleste leksika bruker henvisning til andre del av samme verk. For brukeren kan dette bli tungvint, så vi prøver derfor å redusere slik kryssreferanse til det absolutt nødvendige. I elektroniske verk stiller dette seg også helt annerledes. Intelligent

krysshenvisning, såkalt *hypertekst* blir både enkelt og morsomt i bruk. Men det stiller nye krav til leksikografen.

1.4 Systematikk

I encyclopediske verk er systematikken helt sentral. Den kommer ofte ikke fram i presentasjonen, men ligger bak kunnskapsmengden som en skjult struktur. Den encyclopediske tradisjonen går tilbake til middelalderens *Summae*, og graver man lenge nok, vil man finne spor av denne arven i encyclopedienes skjulte struktur. Så lenge encyclopedien var en sidedisiplin til teologien, finner man teologenes verdensoppfatning meget klart uttrykt i systematikken. Vincent de Beauvais delte sin *Speculum majus* opp i tre deler: *Naturale* — Gud, skapelsen og mennesket; *Doctrinale* — språk, etikk, fagene, medisin; *Historiale* — verdenshistorie. Francis Bacon tok skrittet ut av denne teologiske tradisjonen da han delte opp sin *Instauratio magna* i følgende tre deler: Den *ytre naturen* (omfattet bl.a. astronomi, geografi, meteorologi), *mennesket* (anatomi, fysiologi osv.) og *menneskets påvirkning av naturen* (bl.a. medisin, kjemi, arkitektur og transport og en mengde andre emner). Den store franske encyclopedisten Denis Diderot overtok og videreførte Bacons systematikk og gjennom ham har denne arven blitt ført videre til oss. Denne systematikken har sine utvilsomme styrker, men den har også innebygd i seg en del svakheter. Det jeg særlig vil legge vekt på her er faren for å utelukke tverrfaglighet. Verden splintres opp i enkeltvitenskaper for at den skal bli lettere å katalogisere og beskrive, men den virkelige verden eksisterer bare i sammenheng og samhandling. Hvis vi griper den nye teknologien på en riktig måte, skal vi kunne gjøre interessante forbedringer av oppslagsverkene fra en slik synsvinkel.

2 Noen forslag til hva leksikografer kan bruke CD-ROM til

Jeg begynte å arbeide med leksikon på CD-ROM i 1986, samme år som teknologien ble lansert internasjonalt og fikk æren av å lansere Europas første leksikon på CD-ROM i april 1987. I disse dager lanserer vi det tredje CD-ROM-leksikonet som jeg er redaktør for. Nå med lyd, interaktive kart, massevis av bilder, video, animasjoner og lyd. Disse erfaringene har lært meg en del om hvilke muligheter dette mediet inneholder for oss leksikografer. Henning Bergenholtz (1995) sa i sitt foredrag at elektroniske ordbøker til nå bare har vært elektroniske versjoner av de trykte bøkene. Dette kan jeg skrive under på. Ingen har ennå tatt inn over seg til fulle de endrede vilkårene for leksikografien som de nye mediene gir. Derfor vil jeg benytte denne anledningen til å utfordre kremen av nordiske leksikografer til å være med å forme en leksikografisk teori og framstillingsform som tar elektronikken på alvor. Jeg vil skissere noen av dem og peke på de teoretiske utfordringene som følger i kjølvannet av disse mulighetene. For å sette fantasien i sving har jeg tillatt meg å beskrive ei stor norsk encyclopedisk ordbok som ikke er utgitt eller planlagt. Vi forestiller oss at dette er både ei ordbok og et leksikon, slik at vi kan få nytte av alle de mulighetene som mediet gir.

2.1 Omfang

Vi begynner med omfanget:

Siden vi har så mye plass, bør vi legge opp til å bruke den. Vi begynner derfor med å sløyfe forkortelser og koder. Vi skriver inn den fulle bøyningstabellen for alle oppslagsord. Dette gjør vi ikke bare for å bedre innholdet i den enkelte artikkelen. Men dette gjør at alle former av ordet vil være søkbar i databasen. Et søk i fri tekst etter ordet *ordbøker* vil gi treff selv om oppslagsordet er *ordbok*. Av samme grunn må vi slutte å føre opp sammensatte ord på den tradisjonelle kortformen: *ord, ... -bok*. Allerede her vil vi merke at formen på artikkelen vil endre seg.

Så bestemmer vi oss for å legge inn et par norske grammatikker. Vi kople dem intelligent til hver enkelt ordboksartikkel, slik at brukeren kan få øyeblikkelig grammatikalsk hjelp. Her må leksikografen leve seg inn i brukerens situasjon og tenke ut hva slags støtte hun vil ønske.

Berit Heer Henriksen (1995) ga en lærerik framstilling av problemene knyttet til lemmatisering av avledete begrep, sublemma osv. Spørsmålet er hvor mye av dette som knytter seg til papiret som begrensning. Plassen hindrer oss ikke lenger i å lage det antallet sublemmaer vi måtte ha bruk for. Og det stående uttrykket kan utmerket godt brukes som sublemma for så mange ord i uttrykket vi måtte ha behov for. Søker du på *barn*, finner du *brent barn skyr ilden*, blant mange andre eksempler. Men du kan også finne det dersom du har søkt på *ild*. Leksikografens problem er hvordan disse mulighetene skal tilrettelegges for å gi brukeren mest mulig nytte og glede. For eksempel bør det finnes en mulighet for å hente fram alle idiomer, talemåter og stående uttrykk, samlet og hver for seg.

Siden plass ikke er noe problem for oss, så legger vi inn både nynorsk og bokmål på samme plata. I neste versjon bestemmer vi oss for at vi også skal legge inn ei samisk-norsk ordliste, slik at man kan slå opp betydninga av enkeltord begge veier. For å være forberedt på dette, har vi alt i første versjonen tenkt ut hvordan dette skal gjøres.

Ennå har vi ikke brukt 5% av plata, så vi kan fortsatt boltre oss en masse. Vi legger inn ei svær eksempelsamling fra skjønnlitteratur, aviser og dagligtale, komplett med årstall og kilde. I tillegg til det arbeidet språkfolk allerede har gjort på dette området, vil vi ha muligheten til å bruke de store avisdatabasene for å finne gode eksempler på bruk av bestemte ord. På denne måten blir den generelle ordboka også ei sitatordbok og en språkhistorisk dokumentasjon. Kanskje vi også ønsker at brukerne skal få tilgang til svære korpussamlinger. Da legger vi inn en valgmulighet for å gå fra den elektroniske ordboka og over til en stor database via Internett.

Noen leksikografer, eller kanskje mest encyclopedister, hvis vi får bruke et slikt ord, har eksperimentert med å føre inn lydlike oppslagsord med henvisning til riktig skrivemåte. Det ble f.eks. gjort i *Damms store leksikon* i Norge, men har ikke slått gjennom. I vår digitale ordbok kan vi åpne muligheten for såkalt *soundex*-søk, dvs. at vi søker på lydlikhet. Hvis vi taster inn *fono*, vil vi også finne *phono*. I motsetning til i den trykte boka, er dette ikke noe plassspørsmål lenger, det er et programmeringsspørsmål. Apropos programmering: når vi først er i gang, bør vi sørge for at boka er programmert slik at man kan få tilgang til den fra et tastatur for blindeskrift. Da vil vi plutselig ha ei svær encyclopedisk ordbok for blinde samtidig som for seende.

2.2 Systematikk

Så har vi systematikken:

Ordklassene har vi sjølsagt med. Dette vil glede for eksempel kryssordløserne, fordi de nå lett kan finne "preposisjon på syv bokstaver der fjerde bokstav er *n*". Men vi bør også være nøye med å få med hvor faguttrykkene hører hjemme. På denne måten kan vi for eksempel finne fram alle maritime faguttrykk og glede oss over hvor mange av dem som kommer fra nederlandsk og plattysk. I leksikondelen av denne ordboka vil vi gå enda lenger i oppdeling av stoffet i fagområder. Der vil jeg også foreslå at vi tar med en tesaurus med synonymer og relevante emneord.

Referanse til andre artikler legges inn som hypertekst, slik at brukeren bare kan peke og klikke seg fram. Hvis vi indekserer hele teksten på riktig måte, vil brukeren ganske enkelt kunne peke på et hvilket som helst ord og finne hovedartikkelen for dette ordet.

Jeg kan ikke forlate dette emnet uten å ta med min favorittsystematikk. Den er hentet fra et kinesisk leksikon fra keisertida og inneholder følgende inndeling av dyreartene:

- dyr som tilhører keiseren
- balsamerte dyr
- tamme dyr
- pattegriser
- havfruer
- fabeldyr
- frittgående hunder
- dyr som beveger seg som sinnsyke
- utallige dyr
- dyr som er tegnet med tynn kamelhårspensel
- dyr som nettopp har unngått fare
- dyr som ligner fluer på avstand

2.3 Søking i fri tekst

Vi legger sjølsagt opp til at brukeren skal kunne søke både i fri tekst og etter fagområde, ordkombinasjon og ordklasse. Kanskje bør det også være mulig å spesifisere antall tegn og kanskje også antall stavelser. Søking med synonymer er også en mulighet vi bør vurdere. Siden vi har lagt inn ei synonymordliste, er det enkelt å programmere denne funksjonen.

2.4 Lyd

Først nå beveger vi oss fra tekst til lyd:

Nå er det ikke så få leksikografer som har fylt ordbøker med lyd, problemet er at det har vært skriftelig lyd, nemlig lydskrift. Det er jo en bleik erstatning for virkelig lyd. Vi har også det problemet at jo riktigere lydskriften er, fra en faglig synsvinkel, jo ferre er det som har kompetansen til å forstå den; og jo lettere den er å forstå, jo mer omtrentlig er den. Så nå som vi har muligheten til å gi brukerne våre den virkelige lyden i stedet for den skriftlige forflatninga av den, bør vi gi dem det. Tenk bare på hvor mye enklere det blir å forklare tonelagsforskjeller eller palatalisering, eller klikkelydene i xhosaspråket. Dialektprøver blir

også mer interessante på denne måten. Vi bestemmer oss altså for å la alle oppslagsord ha uttale og det samme med en del av eksempelmaterialiet.

Men lyd har også sin egen forklaringsverdi. I *Cappelens store musikkleksikon* ser definisjonen på en tritonus slik ut:

det intervall som dannes av tre heltonetrinn og som i den diatoniske skala av stamtoner representeres av intervallet f-(g-a-)h, dvs. en forstørret kvart.

Dette er forståelig for alle som ikke trenger å få begrepet forklart og uforståelig for alle som trenger det. Litt lyd her ville vært av det gode. Men dette er porten inn til en hel verden av lyd for ordbokfolket. I CAPLEX, Cappelens ettbinds leksikon, er fagotten forklart slik:

fagott, treblåseinstrument med dyp tone og nasal klang.

Fra nå av venter brukerne at vi også lar dem få høre hvordan instrumentet lyder. Hvis ikke leksikografen nå har skjønnt at ordboka er i ferd med å bli noe helt annet enn vi er vant til, vil det gå opp for ham når vi bestemmer oss for også å inkludere bevegelse som enda en ny dimensjon i ordboka.

2.5 Illustrasjonen må med

I følge et gammelt kinesisk ordtak som oftest feilsiteres, er det bedre å se en ting en gang, enn å høre om den hundre ganger. (Kineserne ville aldri ha satt bildet opp mot ordet, både på grunn av sin respekt for ordet, og fordi deres ord er piktogrammer.) En teknisk innretning kan av og til beskrives tilfredsstillende ved hjelp av bare ord. Men de fleste har innsett at det blir enklere, hvis man også har en illustrasjon å støtte seg på.

Riksmålsordboka definerer en sjakkell slik:

sjakkell, bøile av metall som i begge ender har hull (øie) for en bolt som den lukkes med, brukt til å forbinde med hverandre (særl.) gjenstander som er forsynt med (ell. ender i) ring ell. bøile.

Vi kan formelig føle hvordan leksikografen sliter med å presse virkeligheten inn i språket. En liten skisse i tillegg ville gjort det ganske mye enklere. Dette var jo noe vår læremester Diderot innså til fulle, og i vår visuelle tid er det ikke blitt mindre viktig. Nå kan vi heller ikke skyldes på plassproblemer. Strekillustrasjoner av typen sjakkell kan vi antakelig legge inn ti tusen av uten å oppleve plassproblemer. Problemet blir i stedet leksikografens. Hun blir plutselig bilderedaktør også. Kanskje kan det være fornuftig å lage samleillustrasjoner, slik man gjør i illustrerte ordbøker, men med den forskjellen at vi kan indeksere hver enkelt bildetekst slik at den blir søkbar og lagt inn i kryssreferansetabellen.

2.6 Ordbok med bevegelse

Bevegelse i ordbok eller leksikon, helligbrøde, jåleri, vulgært. Hva skal vi med det? Det skal innrømmes at mye av det vi har sett av video på multimedia oppslagsverk har vært heller tvilsomt fra en faglig synsvinkel. Men det er nå en gang slik at virkeligheten ikke foregår bare i to eller tre dimensjoner. Den fjerde, tidsdimensjonen, er i høyeste grad en

realitet, noe i det minste våre grå hår burde minne oss om. Den reine ordlista har sjølsagt ikke særlig behov for å kunne vise bevegelser. Men med en gang vi beveger oss over til definisjonsordboka, for ikke å snakke om leksikonet, så ser vi at vi svært ofte bruker tidsdimensjonen i skriftlig framstilling. Vi simulerer tid gjennom skrift, vi lager skriftlig bevegelse. Dette kan vi ha behov for når vi skal forklare så forskjellige begreper som platetektonikk, sluse, firetaksmotor, matematiske funksjoner, histogrammer og frontenes utvikling under anden verdenskrig. Man skulle kunne tenke seg at dette ville være en fin måte å følge opp det arbeidet Leena Savolainen (1995) fortalte om i forbindelse med ordbok for tegnspråk. Hun har imidlertid fortalt meg at dagens CD-ROM teknologi gir for lite plass til video. Men om fem år er det problemet løst, så vi kan like gjerne begynne å tenke i de banene.

3 Utfordringene blir altså

Til nå har jeg snakket om muligheter med denne nye teknologien. Hva har det med leksikografisk teori å gjøre? En av kollegene var så ufin at han spurte om det finnes en leksikografisk teori. Men la oss gå ut fra at det gjør det. De utfordringene jeg ser er at vi må begynne å utvikle begreper, framstillingsmåte, kriterier og metodikk som tar helt og fullt inn over seg at vi skal publisere for et elektronisk medium.

For at den elektroniske ordboka skal bli god, er god og hensiktsmessig formatering eller koding alt.

Spørsmål 1: Hva er god og hensiktsmessig formatering i denne sammenhengen?

Spørsmål 2: Gitt at vi har all denne plassen til rådighet, hvordan bruker vi den? Hvor går grensa mellom god og rikholdig informasjon og overflødig eller hemmende informasjon?

Spørsmål 3: Skal vi tilby brukerne alternative fagkodesystemer? Hvilke er disse og hvordan skal de presenteres?

Spørsmål 4: Hvilke andre måter kan vi utnytte ikke-lineariteten på? Synonymordbok, enderimsordbok, sitatordbok, fagordbok osv. i samme verk. Er det mulig å gjøre dette på en god måte, og i tilfelle hvordan?

Spørsmål 5: Det finnes muligheter for å tilby brukeren forslag til søkestrategier, som hun kan bygge videre på. Hvilke bør brukes og hvordan skal de tilrettelegges. Hva med hjelpeverktøyenes rolle? Skal brukeren få muligheten til å bruke konkordansverktøyer i ordbok/leksikonet vårt?

Spørsmål 6: Hvordan leksikaliserer vi lyd, bilde, animasjon og video? I billedarkiver lagres bilder med en beskrivelse av bildet, f.eks. om det er fra Island og viser en leksikograf som drikker "Svarti dauði" foran en vulkansk fjellvegg. Skal de nye mediene beskrives på denne måten, slik at innholdselementene i dem blir søkbare og hvordan skal det gjøres? Når blir slike medier verdifulle tilskudd til verket og når blir de hype? Skal det legges inn tabellarisk informasjon og skal den kunne manipuleres?

(Skal for eksempel folketellingsdata kunne tas inn i et regneark og legges ut som grafer?)

Så velbekomme, kjære kolleger leksikografer. En ny generasjon brukere av ordbøker og leksika er kjent med hvilke muligheter de nye mediene gir. De venter av oss at vi skal bruke disse mulighetene.

Litteratur

- Bergenholtz, Henning 1995: Leksikografi. Hvad er det? [Denne rapporten, 37–49.]
- Bokmålsordboka = M. I. Landrø/B. Wangensteen 1993: *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. Oslo: Universitetsforlaget.
- Cappelens musikkleksikon* 1978–80. Oslo: Cappelen.
- CAPLEX. Cappelens ettbinds leksikon* 1991. Oslo: Cappelen.
- Damms store leksikon* 1985–90. Oslo: N. W. Damm & søn.
- Henriksen, Berit Heer 1995: Korpusbaserte relationsopplysninger og lemmatisering af flerordsforbindelser. [Denne rapporten, 195–203.]
- Riksmålsordboka = *Norsk riksmålsordbok* 1937–57. Oslo: H. Aschehoug & Co.
- Savolainen, Leena 1995: Finnish Sign Language-Finnish dictionary: defining the head signs. [Denne rapporten, 359–364.]
- The Oxford English Dictionary* 1989. I–XX. Second Edition. Oxford: Clarendon Press.

Nordisk lexikografisk ordbok: artiklarnas struktur och innehåll

The project *Nordic Lexicographical Dictionary* (NLO) aims at producing a dictionary of Nordic lexicographical terminology, with definitions in Norwegian bokmål, and with term equivalents in five other Nordic languages as well as in English, German and French. After completing data collection and preliminary data selection, and after assigning to each concept a place in the classification scheme specially developed for NLO, the project has now reached a stage where the main task consists of making the final selection of concepts and terms and compiling the dictionary entries. This contribution presents and discusses, from various points of view, some of the entries compiled by the Swedish member of the project group.

1 Bakgrund

När Nordiska föreningen för lexikografi stiftades i Oslo 1991, beslöts att föreningen som en av sina första uppgifter skulle låta utarbeta en ordbok över lexikografins terminologi. I slutet av året bildades en arbetsgrupp, som startade sin verksamhet i början av 1992. Gruppen enades ganska snart om följande beskrivning av den tänkta produkten, som fått arbetsnamnet *Nordisk lexikografisk ordbok* (NLO).

Ordboken skulle ha ett sjufaldigt syfte. Tre funktioner skulle vara primära, nämligen (i prioritetsordning): (1) produktion på modersmålet, (2) reception på modersmålet, (3) inhämtande av kunskaper om fackområdet. De fyra sekundära funktionerna, som är svårare att ordna prioritetsmässigt, skulle vara reception på främmande språk, översättning från främmande språk, produktion på främmande språk och översättning till främmande språk.

Beskrivningsspråket, och lemmalistans språk, skulle vara norskt bokmål. Termekvivalenter skulle ges på danska, finska, isländska, nynorska, svenska, engelska, tyska och franska.

Grundprincipen för urvalet skulle vara att ordboken i första hand skulle behandla sådana begrepp och termer som är specifika för lexikografen eller åtminstone har en särskild betydelse eller användning inom lexikografen. Ordboken skulle vara förhållandevis normativ när det gällde valet av termer på de nordiska språken men snarare deskriptiv när det gällde betydelsebeskrivningen.

Arbetet har hittills haft följande förlopp.

Potentiella engelska, tyska och franska termer utvaldes ur registret till det stora trebandsverket *Wörterbücher*, och nordiska termer excerperades ur existerande metalexikografiska arbeten på nordiska språk. Dessutom excerperades inledningar och bruksanvisningar till ett antal ordböcker på de ingående språken. För en mer detaljerad presentation av selektionsprinciperna, se Mikkelsen (1994).

De utvalda engelska, tyska och franska termerna försågs i förekommande fall med nordiska ekvivalenter. För en presentation av ekvivalentproblematiken, se Fjeld (1994).

Parallellt med detta undersöktes ämnessystematiken, eftersom den är av stor betydelse för urvalsprocessen. För olika överväganden beträffande systematiken, se Svensén (1994a) och Svensén (1994b). Ämnessystematiken står i centrum också för den systematiska inledning till ämnet som skall ingå i ordboken. För en preliminär version av en sådan inledning, se Bergenholtz & Svensén (1994). På grundval av ämnessystematiken genomfördes sedan en preliminär klassning av samtliga termer i materialet.

Klassningen låg till grund för fördelningen mellan projektmedarbetarna av de olika delområdena inför det fortsatta arbetet, som gick ut på att var och en inom det tilldelade området skulle dels göra det slutliga urvalet av termer och begrepp, dels författa själva ordboksartiklarna. Denna fas i arbetet pågår nu för fullt, och syftet med föreliggande bidrag är att presentera och kommentera ett urval av de artiklar som det har fallit på den svenske projektmedarbetarens lott att utarbeta.

2 Artiklarnas struktur

Först något om ordbokens övergripande struktur.

Ordbokens **megastruktur** består av tre huvudavsnitt. Det centrala avsnittet är förstås själva **lemmalistan**, alltså hela följderna av artiklar. Lemmalistan föregås av den så kallade **förtexten**, som i NLO utgörs av ett förord, en bruksanvisning och den systematiska inledningen. Efter lemmalistan följer **eftertexten**, där man återfinner ett alfabetiskt termgregister för vart och ett av de 8 ekvivalentspråken.

NLO:s **makrostruktur** kan beskrivas mycket kort. Den är glattalfabetisk — utan nischer och nästen — och grundar sig på de norska bokmålstermernas stavning.

Ett typiskt exempel på en fullständig **mikrostruktur** i NLO är artikeln **receptionsordbok** [exempel 1]. Jag vill dock genast göra den reservationen att artikeln i två avseenden inte är typisk, och det gäller f.ö. alla exempel som kommer att visas i det följande: (a) Beskrivningsspråket, liksom lemmats språk, är ännu så länge artikelförfattarens modersmål; översättning till norskt bokmål blir nästa steg. (b) Utöver bokmålslemmat skall ju NLO ge termekvivalenter på ett antal språk; dessa ekvivalenter är inte medtagna här.

(1) Artikeln inleds förstås med lemmat. (2) Efter lemmat följer en definition av helt traditionellt slag. (3) Definitionen följs av en mer sammanhängande text (av oss i projektgruppen något lösligt kallad "förklaring"), som här inleds med ett tankstreck, och den kan ha mycket olika omfång, uppläggning och innehåll. (4) Därefter anges eventuell synonym. (5) Efter synonymen ges eventuella hänvisningar till andra artiklar. (6) Den hänvisning till "§ 11" som följer därefter avser paragraf i den systematiska inledningen. (7) Artikeln avslutas i sin nuvarande form av litteraturreferenser.

En informationstyp som inte finns med här och vars vara eller icke vara vi ännu inte har fattat beslut om är klassningskoderna. Det kan ju vara till en viss hjälp om det i varje artikel ges upplysning om begreppets plats i ämnessystematiken.

Alla informationstyperna är inte obligatoriska. Det kan självfallet inträffa att det inte finns några synonymer, några lämpliga besläktade termer eller någon paragraf i den systematiska inledningen att hänvisa till, eller att det inte finns någon relevant litteratur att referera till.

receptionsordbok

[1]

ordbok avsedd för förståelse av text

— En receptionsordbok kan vara enspråkig och är då i regel avsedd för förståelse av text på användarens modersmål. Receptionsordboken kan också vara tvåspråkig, vilket är det normala när det gäller förståelse av text på ett för användaren främmande språk. Ingenting hindrar emellertid att även en enspråkig ordbok på ett främmande språk används för förståelse av text på detta språk. Det senare användningssättet ställer större krav på användaren, eftersom han måste ha en sådan kompetens på det främmande språket att han förstår de ordförklaringar och övriga upplysningar som ges. — Åtminstone i teorin är det möjligt att skilja mellan tvåspråkiga receptionsordböcker och passiva översättningsordböcker, som är avsedda som hjälpmedel vid överflyttning av en given text från ett för användaren främmande språk till hans modersmål. I verkligheten är det dock inte särskilt vanligt att en ordbok entydigt är avsedd att användas för det ena och inte det andra av dessa båda ändamål; förmodligen är det bara en fråga om att ordboken i praktiken används för ett visst syfte.

= avkodningsordbok

<—> produktionsordbok

—> passiv översättningsordbok

§ 11

Litt.: Hausmann, F.J. & Werner, R.O.: Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher. I: *HSK* 5:2729–2769 (särskilt avsnitt 9) — Kromann, H.-P., Riiber, Th. & Rosbach, P.: Principles of bilingual lexicography. I: *HSK* 5:2711–2728 (särskilt avsnitt 3–4) — Kühn, P.: Typologie der Wörterbücher nach Benutzungsmöglichkeiten. I: *HSK* 5:111–127

När det gäller betydelsebeskrivningen och vad därtill hör finns det som vi sett två olika informationstyper, och inte heller de är obligatoriska. De finns båda med i den artikel som nyss visades [1]. Men det finns också många artiklar som inte har någon definition utan enbart en resonerande förklaring [2]. Det är också fullt tänkbart — fast i mitt material mindre vanligt — att en artikel har definition men saknar resonerande förklaring [3].

eponymordbok

[2]

— En eponymordbok behandlar egennamn som, med utnyttjande av de konnotationer som förknippas med en viss bärare av namnet (eponymen), antingen själva har övergått till att (även) användas som appellativer (*giljotin*) eller givit upphov till sammansättningar (*judaskyss*), avledning (*köpenickiad*) eller syntagm (*Kolumbi ägg*) som anspelar på referenten i fråga. Av exemplen framgår, att egennamnet inte nödvändigtvis behöver vara ett personnamn. — Lemmalistan kan bestå antingen av de ursprungliga egennamnen eller av de appellativer etc. som de har givit upphov till; den senare typen är den vanligaste. Information ges om egennamnet och dess ursprungliga bärare, om det språkbruk som namnet givit upphov till och om hur detta har gått till. Därigenom kan en eponymordbok också innehålla åtskillig encyklopedisk information.

—> namnordbok

§ 30

Litt.: Bernet, C.: Le dictionnaire d'éponymes. I: *HSK* 5:1258–1260

deskriptiv ordbok

[3]

ordbok vars syfte är att beskriva språket, inte att påverka användarnas språkbruk

<—> normativ ordbok

§ 4, 9

Litt.: Ripfel, M.: Die normative Wirkung deskriptiver Wörterbücher. I: *HSK* 5:189–207

En fråga som naturligen inställer sig är: Vad är det som avgör om en artikel skall ha både definition och förklaring eller enbart definition eller enbart förklaring? Normalfallet är att båda delarna förekommer. Ofta kan man behöva kommentera definitionen; ännu vanligare är att förklaringen ger tilläggsinformation, ofta av mer encyklopediskt slag. Jag återkommer till innehållet i förklaringarna.

Det förekommer också ganska ofta att definition saknas och det bara ges en förklaring. Skälen till detta kan variera.

Det vanligaste fallet är att termen är så pass genomskinlig att en definition blir överflödigt och t.o.m. skulle te sig lite löjlig, t.ex. typen **XYZ-ordbok**, definierad som 'ordbok över XYZ'. I stället ges en mer encyklopediskt inriktad beskrivning [4]. Ännu mer påtagligt blir det när det har varit nödvändigt att konstruera en nordisk term. En konstruerad term skall ju helst vara genomskinlig till betydelsen, och termen **ordbok för främmandespråksinläring**, som alltså skall motsvara engelskans *dictionary for foreign learners*, är i praktiken sin egen definition, och man kan gå direkt på den encyklopediska förklaringen [5].

förkortningsordbok

[4]

— I en typisk förkortningsordbok utgörs lemmalistan av förkortningar, och i mikrostrukturen ges information om dels den oförkortade formen, dels dennas betydelse. Det finns dock även förkortningsordböcker som utgår från de fullständiga formerna och i artiklarna anger motsvarande förkortningar; sådana listor kan också förekomma som komplement till den förstnämnda typen av förkortningsordbok. [...]

§ 30

Litt.: Menzel, H.-B.: Das Abkürzungswörterbuch. I: *HSK 5:1261–1266*

ordbok för främmandespråksinläring

[5]

— De flesta ordböcker som benämns ordböcker för främmandespråksinläring är enspråkiga. De mest kända av dem förekommer inom den anglosaxiska världen och kallas där *dictionaries for foreign learners*, t.ex. *Oxford Advanced Learner's Dictionary of Current English* (OALDCE) och *Longman Language Activator*. Denna typ av ordbok är i regel avsedd att fungera som både receptionsordbok och produktionsordbok. [...]

Av de hittills visade exemplen på artiklar framgår att det är de ordbokstypologiska avsnitten som har fallit på den svenske projektmedarbetarens lott. Det handlar om två partier i systematiken, dels gruppen 41000 "Allmän ordbokstypologi", dels gruppen 81000 "Typologi" hörande till huvudområdet 80000 "Specialordböcker" (Svensén 1994b:233 ff). Det visar sig här — kanske inte helt oväntat — att behovet av regelrätta definitioner är större när det gäller den allmänna ordbokstypologin än när det gäller olika typer av specialordböcker.

3 Articklarnas innehåll

3.1 Lemma

Om lemmat som informationstyp är inte mycket att säga. Lemmat representeras i NLO av bokmålsformen, i princip utan några särskilda uppgifter om böjning, ordklass etc.

3.2 Definition

Om definitionerna är inte heller särskilt mycket att säga. I början av arbetet diskuterade vi inom projektgruppen att eventuellt skriva definitioner i något slags COBUILD-format men enades slutligen om systemet med både en definition av konventionellt slag och en mer resonerande "förklaring", där man för varje artikel skulle kunna välja att ha med endera eller båda.

Arbetet har inte delats upp efter alfabetet utan efter den klassning av hela materialet som gjorts på grundval av den preliminära ämnessystematiken. Att samma person har behandlat samtliga begrepp inom en viss begreppsklass ökar förutsättningarna för att besläktade begrepp skall definieras på likartat sätt [6–7].

syntagmatisk ordbok [6]
specialordbok som beskriver lexikaliska enheter i relation till andra enheter med vilka de kan samförekomma

paradigmatisk ordbok [7]
specialordbok som beskriver lexikaliska enheter i relation till andra enheter mot vilka de kan utbytas

3.3 Förklaring

Som tidigare påpekats, kan förklaringsdelen ge information av mycket olika slag. Här följer en liten provkarta.

3.3.1 Explicit kommentar till definitionen

Förklaringen utgör i många fall en explicit kommentar till definitionen. Den kan exempelvis ge närmare information om något som nämns i definitionen; i artikeln **lista** [8] får man i förklaringen veta vad som i definitionen menas med termen "textelement". I andra fall ges en motivering till definitionens utformning; under **enspråkig ordbok** [9] får man veta varför det i definitionen står "är identiskt med eller utgör en varietet av beskrivningsspråket".

lista [8]
efter vissa principer utvald och ordnad samling av självständiga textelement
— Textelementen i en lista kan bestå av alfabetiska tecken, siffror, logogram etc. eller av kombinationer av dessa typer. Inom lexikografin utgörs de i regel av språkliga enheter (ord, orddelar eller ordgrupper). De kan vara utvalda antingen i sin egenskap av språkliga enheter (som i de flesta ordböcker och ordlistor), såsom representanter för företeelser i världen utanför språket (som i encyklopedier, kataloger, person- och sakregister etc.) eller i båda egenskaperna (som i encyklopediska ordböcker). Textelementen kan vara ordnade efter olika principer, t.ex. alfabetiskt, systematiskt eller slumpmässigt.
= förteckning

enspråkig ordbok [9]
ordbok som har endast ett objektspråk och där objektspråket är identiskt med eller utgör en varietet av beskrivningsspråket
— Att objektspråket kan vara en varietet av beskrivningsspråket innebär alltså, att exempelvis slangordböcker och dialektordböcker med standardspråkliga förklaringar i princip räknas som enspråkiga ordböcker, likaså ordböcker över historiska varieteter med förklaringar på nutida språk. [...]

Man kan vidare kommentera definitionen genom att ange vad den avser att utesluta, alltså vad som *inte* ingår i begreppet [10]. Det förekommer också att den valda definitionens lämplighet diskuteras, eventuellt också att utgångspunkter för alternativa definitioner anges [11].

normativ ordbok [10]

ordbok i vars syfte ingår att påverka användarnas språkbruk
— [...] Av definitionen framgår, att en ordbok räknas som normativ endast om den har ett normativt syfte. Till de normativa ordböckerna räknas alltså inte sådana ordböcker som, genom den höga status ordböcker åtnjuter hos allmänheten, kommit att i något avseende bli oavsiktligt normerande.

vetenskaplig ordbok [11]

ordbok utarbetad efter vetenskapliga principer
— Termen **vetenskaplig ordbok** används på flera sätt, och den valda definitionen är en bland flera möjliga. (Ett alternativ hade exempelvis varit att ta fasta på ordbokens användningsområde och låta termen stå för en ordbok som är avsedd att användas i samband med vetenskapligt arbete, ett annat att använda termen för att beteckna motsatsen till en kommersiell ordbok.)

3.3.2 Kommentar till sådant som är speciellt för NLO

Man kan även använda förklaringen till att kommentera sådant som är speciellt för NLO, alltså när ordboken avviker från gängse språkbruk och försöker vara innovativ [12].

kumulativ ordbok [12]

ordbok som enbart räknar upp de utvalda lexikaliska enheterna utan att kommentera dem individuellt
— [...] Termen **kumulativ ordbok** har här givits en mer generell innebörd och användning än i *Wörterbücher*, där den enbart används om vissa typer av synonymordböcker och citatordböcker och då med vissa betydelskillnader. Ovanstående definition ansluter i huvudsak till den i *Wörterbücher* gjorda distinktionen mellan olika typer av citatordböcker.

3.3.3 Diskussion av begreppets existensberättigande

Förklaringsdelen kan också vara ifrågasättande. Man kan t.ex. föra en diskussion om ett begrepps existensberättigande över huvud taget [13].

allmän ordbok [13]

ordbok som inte är specialiserad i fråga om behandlade språkvarieteter, texter, lemmatyper eller informationstyper, inte heller när det gäller presentationssätt, avsedda användarkategorier etc.
— [...] Ordet "allmän" skall härvid inte fattas som synonymt med "allmänspråklig": i den meningen är en allmän ordbok i själva verket inte särskilt "allmän", eftersom de flesta allmänna ordböcker har inslag av olika språkvarieteter såsom fackspråk samt ålderdomligt, dialektalt och litterärt språk etc. Man kan därför i viss mån ifrågasätta både termen och begreppet **allmänspråklig ordbok**.

Någon gång motiveras begreppets existens av att det behövs som motsats till ett annat. Ett sådant fall är det som på engelska kallas *dictionary for native speakers* och som på svenska har fått en — jag betonar det — mycket tentativt ekvivalent [14]. På liknande sätt motiveras ett begrepp genom att det kan utgöra överbegrepp till en grupp andra, t.ex. det som på engelska kallas *interlingual dictionary*, på svenska preliminärt **mellanspråklig ordbok**. Definitionen blir därefter [15].

ordbok för modersmålsbrukare [14]

ordbok vars utgångsspråk är identiskt med de tilltänkta användarnas modersmål
— Termen används främst i samband med enspråkiga ordböcker. Dess huvudsakliga roll är att utgöra motsatsen till den mycket vanligare termen **ordbok för främmandespråksbrukare** och framför allt dennas pedagogiska variant **ordbok för främmandespråksinläring**. [. . .]

mellanspråklig ordbok [15]

tvåspråkig ordbok eller flerspråkig ordbok
— Begreppet har sitt berättigande enbart som hyperonym.

I sådana sammanhang kommer man lätt in på förhållandet mellan teori och verklighet. När det gäller ordböckers funktion har NLO nämligen ett begreppssystem där man, åtminstone teoretiskt, skiljer mellan tre kategorier: **produktionsordböcker**, **receptionsordböcker** och **översättningsordböcker**:

- 1 ordbok
 - 1.1 produktionsordbok
 - 1.1.1 enspråkig produktionsordbok
 - 1.1.2 tvåspråkig produktionsordbok
 - 1.2 receptionsordbok
 - 1.2.1 enspråkig receptionsordbok
 - 1.2.2 tvåspråkig receptionsordbok
 - 1.3 översättningsordbok
 - 1.3.1 aktiv översättningsordbok
 - 1.3.2 passiv översättningsordbok

Huvudskälet till detta är att all översättning, såväl till som från modersmålet, involverar både ett receptions- och ett produktionsmoment. Om man däremot ser till den rent praktiska sidan av saken, kan det finnas skäl att diskutera vad det i verkliga livet är för skillnad mellan exempelvis den tvåspråkiga receptionsordboken och den passiva översättningsordboken [16]. (Inom parentes sagt vore det en lättnad om vi på de nordiska språken kunde komma ifrån adjektiven *aktiv* och *passiv*, som egentligen ger fel associationer. Tyskarna är i den avundsvärda situationen att de förfogar över prefixen *her-* och *hin-*, varigenom de har kunnat mynta termerna *Herübersetzungswörterbuch* och *Hinübersetzungswörterbuch*.)

passiv översättningsordbok [16]

översättningsordbok avsedd för översättning av text från ett eller flera främmande språk till modersmålet
— Ätminstone i teorin är det möjligt att skilja mellan passiva översättningsordböcker och tvåspråkiga receptionsordböcker (som är avsedda för förståelse av text på ett för användaren främmande språk). I verkligheten är det dock inte särskilt vanligt att en ordbok entydigt är avsedd att användas för det ena och inte det andra av dessa båda ändamål; förmodligen är det bara en fråga om att ordboken i *praktiken används* för ett visst syfte.

3.3.4 Begreppssystematik

Liksom i den systematiska inledningen sätts i artikelns förklaringsdel ett begrepp ofta in i sitt begreppsliga sammanhang. Oftast blir det då fråga om en ytterligare underindelning [17]. Ibland har det t.o.m. varit lämpligt att sammanföra alla definitioner inom ett mindre begreppssystem till en enda artikel, nämligen under det överordnade begreppet [18].

specialordbok [17]

ordbok som har en särskild inriktning när det gäller behandlade språkvarieteter, texter, lemmatyper, informationstyper, presentationssätt eller avsedda användarkategorier — Självfallet finns det en enorm rikedom av ordbokstyper inom kategorin specialordböcker, och flera olika typologier har presenterats i olika sammanhang. [...] Den kategorisering som tillämpas i NLO innefattar tre huvudgrupper: **syntagmatiska specialordböcker**, **paradigmatiska specialordböcker** och **avgränsade ordböcker**. Därefter görs ytterligare uppdelningar i flera steg, särskilt i den sistnämnda gruppen.

normativ ordbok [18]

ordbok i vars syfte ingår att påverka användarnas språkbruk — En normativ ordbok kan vara **öppet normativ** eller **förtäckt normativ**. I den öppet normativa ordboken redovisas det normativa syftet explicit, vilket kan ske dels genom generella programförklaringar i förordet eller ordboksgrammatiken, dels genom normativa angivelser inom de enskilda ordboksartiklarna. I den förtäckt normativa ordboken har urvalet av uppslagsord och exempel, böjnings- och uttalsvarianter etc. skett i enlighet med utgivarens språkliga normer; det normativa syftet redovisas inte explicit, utan ordboken ser på ytan ut som en deskriptiv ordbok. — Den normativa ordboken kan vara **normskapande** eller **normbeskrivande**. Den förra typen utgör ett redskap för språkplanering och används vid introduktion av standardskriftspråk där sådana saknas, t.ex. i utvecklingsländer. Den senare typen utgör ett redskap för språkvård inom ramen för väl etablerade standardspråk. [...]

3.3.5 Diskussion av terminologisk oreda

Man kan också komma in på motsatsen till begreppssystematik och kommentera fall av terminologisk oreda. När det gäller den besvärliga termen **ordlista**, framgår den i NLO valda betydelsen av definitionen, medan övriga betydelser måste kommenteras [19].

ordlista [19]

lexikografiskt uppslagsverk bestående av en lista över ord i vilken endast summariska ordförklaringar ges eller där ordförklaringar förekommer endast sporadiskt eller helt saknas

— Med **ordlista** avses i huvudsak den alfabetiska allmänspråkliga ordlistan av enklare slag. Inom terminografin används också termen **ordlista** men avser då i regel en alfabetiskt eller systematiskt ordnad förteckning över ett visst fackområdes termer (**terminologisk ordlista**), med utförliga begreppsdefinitioner och förklaringar samt ekvivalenter på ett eller flera språk. — Termen **ordlista** används också om en samling ordförklaringar i anslutning till en textutgåva. — Termen förekommer (genom påverkan av den engelska direkta motsvarigheten *word list*) marginellt också i betydelsen lemmalista.

3.3.6 Kulturella skillnader

Vissa förklaringar är mer encyklopediskt inriktade än andra. Ibland har det ansetts vara av intresse för läsarna att få veta något om skillnader i ordbokskultur mellan olika länder och språkområden. Ett typiskt sådant fall utgör ordböckerna över främmande ord och över s.k. svåra ord [20].

ordbok över främmande ord [20]

— [...] Ordböcker över främmande ord skiljer sig när det gäller urvalet ganska mycket åt i olika länder, och de spelar också olika stor roll inom olika språkområden. Inom det tyska språkområdet och i Skandinavien är uppfattningarna om vad som är främmande ord ganska likartade; inget land kan uppvisa en så rikhaltig och varierad flora av *Fremdwörterbücher* som Tyskland. Inom de romanska språken räknas exempelvis de lånord som bygger på grekiskt-latinskt språkgods som *mots savants* och inte som främmande ord i vår mening; ordböcker över främmande ord i romanska språk är i stället ofta anglicism-ordböcker. Inom det anglosaxiska språkområdet är ordbokstypen sällsynt och ställs helt i skuggan av den för detta område typiska *dictionary of hard words*. [...]

3.3.7 Exempel på ordbokstyper och individuella ordböcker

Till de mer jordnära pedagogiska inslagen i förklaringarna hör att det till en ordbokstyp av mer allmänt slag ges exempel på mer speciella ordbokstyper, som utgör exponenter för den allmänna typen [21]. Man kan också gå ett steg längre i konkretion och hänvisa till individuella ordböcker [22], och ibland får även NLO själv tjänstgöra som exempel [23].

kumulativ ordbok [21]

ordbok som enbart räknar upp de utvalda lexikaliska enheterna utan att kommentera dem individuellt

— Kumulativa ordböcker uppträder huvudsakligen bland vissa typer av specialordböcker: uttrycksparadigmatiska (t.ex. anagramordböcker och rimordböcker), innehålls-paradigmatiska (t.ex. vissa synonymordböcker) och syntagmatiska (t.ex. satsordböcker och vissa citatordböcker). [...]

sakordbok [22]

ordbok vars syfte är att i första hand ge information om världen utanför språket

— [...] Typiska exempel på sakordböcker är encyklopedier, t.ex. den svenska *Nationalencyklopedin*, och uppslagsverk över enstaka fackområden, t.ex. det svenska *Sohlmans musiklexikon*.

allordbok [23]

ordbok vars syfte är att ge information både om språket och om världen utanför språket
— *Nordisk lexikografisk ordbok* är på flera sätt ett gott exempel på en allordbok. I den enskilda artikeln ger NLO dels termdefinitioner, dels upplysningar om den del av ämnesområdet där termen hör hemma. Den systematiska inledningen till ordboken tjänar också detta dubbla syfte genom att den dels sätter de olika termerna i en systematisk relation till varandra, dels ger en allmän översikt över hela ämnesområdet.

3.4 Synonymer

Efter förklaringen ges eventuell synonym, föregången av ett likhetstecken [24]. Likhets-tecknet hänvisar alltid från en definierad term till en odefinierad, vilket samtidigt innebär

att det hänvisar från en rekommenderad term till en icke-rekommenderad. Synonymen ger däremot upphov till en hänvisningsartikel [25]. Den dubbla pilen hänvisar alltid från en odefinierad term till en definierad synonym, vilket samtidigt innebär att den hänvisar från en icke-rekommenderad term till en rekommenderad.

receptionsordbok [24]

ordbok avsedd för förståelse av text

[...]

= avkodningsordbok

avkodningsordbok [25]

=> receptionsordbok

3.5 Antonymer

Antonymer blir det inte särskilt ofta anledning att ange, men det förekommer ibland. Av [26–27] framgår att det inte brukar vara fråga om antonymer i särskilt sträng mening utan snarare ”pedagogiska” antonymer.

onomasiologisk ordbok [26]

ordbok vars makrostruktur är baserad på lexikaliska enheters egenskaper på innehålls-
sidan

[...]

<—> semasiologisk ordbok

semasiologisk ordbok [27]

ordbok vars makrostruktur är baserad på lexikaliska enheters egenskaper på uttrycks-
sidan

[...]

<—> onomasiologisk ordbok

3.6 Hänvisningar till besläktade begrepp

Hänvisningar till besläktade begrepp utgör en viktig del av begreppssystematiken. Överordnade begrepp skall det i princip inte behöva hänvisas till, eftersom de som regel anförs som genus proximum i definitionen. Hänvisningar till sidoordnade begrepp kan däremot ibland vara motiverade [28–30].

idiomordbok [28]

syntagmatisk ordbok som för varje lemma anger i vilka fasta och till betydelsen
ogenomskinliga ordförbindelser lemmat kan förekomma

[...]

—> kollokationsordbok, valensordbok

kollokationsordbok [29]

syntagmatisk ordbok som för varje lemma anger vilka ord som tenderar att förekomma
tillsammans med lemmat i icke-idiomatiska ordförbindelser

[...]

—> idiomordbok, valensordbok

- valensordbok** [30]
 syntagmatisk ordbok som för varje lemma anger vilka typer av komplement som är syntaktiskt möjliga, i vissa fall även vilka semantiska restriktioner som gäller för dessa komplement
 [...]

—> kollokationsordbok, idiomordbok

Den här typen av hänvisning tillgrips också i sådana fall där alla begrepp inom ett visst begreppssystem förklaras på ett och samma ställe [31–37].

- familjenamnsordbok** [31]
 —> namnordbok
- förnamnsordbok** [32]
 —> namnordbok
- namnordbok** [33]
 ordbok i vilken lemmalistan utgörs av egennamn och i vilken föremålet för beskrivningen är namnen själva, inte deras referenter
 — Namnordböcker kan indelas i två huvudgrupper: **personnamnsordböcker** och **ortnamnsordböcker** (ibland kallade **topografiska ordböcker**; en bättre term vore **toponymordböcker**). [...] Personnamnsordböckerna kan i sin tur indelas i **förnamnsordböcker** och **familjenamnsordböcker**.
- ortnamnsordbok** [34]
 —> namnordbok
- personnamnsordbok** [35]
 —> namnordbok
- topografisk ordbok** [36]
 —> namnordbok
- toponymordbok** [37]
 —> namnordbok

3.7 Hänvisningar till den systematiska inledningen

Hänvisningarna till paragrafer i den systematiska inledningen behöver knappast någon särskild kommentar. Av [38] framgår exempelvis att den som vill ha en sammanhängande översikt över specialordböckerna bör gå till paragraferna 29–33.

- specialordbok** [38]
 ordbok som har en särskild inriktning när det gäller behandlade språkvarieteter, texter, lemmatyper, informationstyper, presentationssätt eller avsedda användarkategorier
 [...]

§ 9, 29–33

3.8 Litteraturreferenser

Litteraturreferenserna, som avslutar de flesta av artiklarna, avser ibland nordisk metalexikografisk litteratur, men i de allra flesta fall hänvisas till trebandsverket *Wörterbücher* [39].

översättningsordbok

[39]

ordbok avsedd för överflyttning av text från ett eller flera språk till ett annat eller flera andra språk

[...]

Litt.: Hausmann, F.J. & Werner, R.O.: Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher. I: *HSK 5:2729–2769* (särskilt avsnitt 9) — Kromann, H.-P., Riiber, Th. & Rosbach, P.: Principles of bilingual lexicography. I: *HSK 5:2711–2728* (särskilt avsnitt 3–4) — Kühn, P.: Typologie der Wörterbücher nach Benutzungsmöglichkeiten. I: *HSK 5:111–127*

4 Projektets fortsättning

Så snart den nu pågående fasen avslutats, vidtar slutredigeringen. Då skall alla bidrag samordnas och alla danska, finska, isländska och svenska artiklar översättas till norskt bokmål. Den systematiska inledningen, som författades innan definitionsskrivandet hade kommit igång på allvar, måste samordnas med artiklarna, och eftersom den i sin preliminära version är avfattad på svenska, måste den översättas till bokmål.

Ett moment som delvis kan pågå parallellt med slutredigeringen är en genomgång av de engelska, tyska och franska ekvivalenterna, vartill krävs insatser från metalexikografer som var och en har ett av de tre språken som modersmål.

Efter slutredigeringen återstår generering av ekvivalentregistren och diverse kontroll-läsningar. Tryckningen beräknas äga rum under andra halvåret 1996 och utgivningen är planerad till december samma år.

Litteratur

- Bergenholtz, Henning/Bo Svensén 1994: Systematisk inledning till *Nordisk lexikografisk ordbok (NLO)*. I: *LexicoNordica 1*, 149–185.
- Fjeld, Ruth Vatvedt 1994: Ekvivalensproblematik i Nordisk lexikografisk ordbok. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993*. Århus: Nordisk Forening for Leksikografi, 78–86.
- Mikkelsen, Hans Kristian 1994: Lemmaselektion til Nordisk Leksikografisk Ordbog. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993*. Århus: Nordisk Forening for Leksikografi, 193–203.
- Svensén, Bo 1994a: Lexikografins terminologi: systematiska aspekter. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993*. Århus: Nordisk Forening for Leksikografi, 263–278.
- Svensén, Bo 1994b: *Nordisk lexikografisk ordbok* — terminologisk systematik. I: *LexicoNordica 1*, 229–238.

Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie ...
(1:1989, 2:1990, 3:1991), red. Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta. (HSK 5.1–3.) Berlin/New York: de Gruyter.

Lars Törnqvist

Engelsk-svensk namnordbok

English-Swedish Name Dictionary contains 3600 English proper names which are written in a different way in Swedish. The dictionary includes names of countries, rivers and cities, kings and saints, people and places of the Bible and the Greek mythology, as well as titles of books, comics and plays. The main sources are monolingual encyclopaedias.

1 Bakgrund

Engelsk-svensk namnordbok är en tvåspråkig ordbok över egennamn som har olika skriftformer i engelska och svenska. Den har utarbetats på fritid av artikelförfattaren, som till vardags är handläggare på Tekniska nomenklaturcentralen (TNC). Namnordboken kommer att bli färdig för utgivning under hösten 1995 och beräknas omfatta cirka 3600 uppslagsord. Den är i första hand avsedd att användas aktivt vid textproduktion av svenskar som översätter från engelska till svenska eller som refererar engelsk text på svenska. Naturligtvis kan den dessutom användas även som passiv ordbok.

Det finns sedan tidigare flera typer av namnordböcker. Ofta har de som syfte att beskriva etymologi eller uttal, och de är i sådana fall vanligen enspråkiga. Namnordböcker för översättning omfattar oftast många språk, exempelvis Adrian Rooms sexspråkiga namnordbok (*Dictionary of Translated Names and Titles*) och EG-kommissionens ordböcker som täcker alla EU:s officiella språk. Renodlade tvåspråkiga namnordböcker är sällsynta; däremot finns som regel en del egennamn med i allmänna tvåspråkiga ordböcker, antingen som lemman i ordbokens huvuddel eller i en lista som bihang till ordboken. Ingen av de tidigare namnordböckerna har tjänat som direkt förebild till *Engelsk-svensk namnordbok*, utan den är utformad med utgångspunkt från vilka funktioner den skall fylla.

2 Val av uppslagsord

Ordboken omfattar i huvudsak namn av följande kategorier:

- geografiska namn, t.ex. namn på länder, hav, floder, berg och större städer
- namn på historiska personer, t.ex. kungar, helgon och ryska politiker
- person- och ortnamn ur antik mytologi
- person- och ortnamn ur Bibeln
- namn på internationella organisationer

- namn på personer och platser i skönlitteratur och tecknade serier
- namn på böcker, skådespel, filmer och andra litterära och konstnärliga verk

För att ordlisteprojektet skulle bli hanterligt sattes målet för ordlistans omfång redan från början till högst 4000 uppslagsord. Antalet namn som har olika skriftformer på engelska och svenska är ju mycket större än så, och detta innebar att ett urval måste göras. Arbetet koncentrerades i första hand på de mest allmänna delarna av namnförrådet — namn på större geografiska objekt och namn på sådana verkliga och fiktiva personer som ofta omnämns i allmänna texter. Mindre vikt lades vid organisationsnamnen, eftersom de i allmänhet finns tillgängliga i olika officiella dokument. Urvalet av litterära och konstnärliga verk måste också av utrymmesskäl bli ytterst ofullständigt. Vilka enskilda namn som har tagits med i ordboken har i stor utsträckning styrts av källorna. Valet av källor har därför varit mycket betydelsefullt för ordbokens slutliga innehåll.

3 Uppläggning

Ordboken är som sagt i första hand avsedd för aktiv användning av svenskar som översätter från engelska till svenska. Artiklarna är därför alfabetiskt sorterade med den engelska namnformen som uppslagsord. Varje artikel i ordboken innehåller ett engelskt namn och minst ett motsvarande svenskt namn. Ordlistan är alfabetiskt sorterad med den engelska namnformen som uppslagsord. Olika engelska stavningsvarianter (t.ex. *Faeroes* och *Faroes*) står på sina respektive platser i den alfabetiska ordningen utan inbördes hänvisningar.

Eftersom ordlistan skall användas i bara en riktning har urvalet av namnformer gjorts efter olika principer för de båda språken. Urvalet av engelska namn är rätt vildvuxet — det innehåller bl.a. många föråldrade namnformer — medan de svenska namnen måste vara helt korrekta enligt nutida språkkuppfattning.

De flesta artiklarna innehåller någon form av förklaring för att översättaren skall kunna avgöra om det svenska namn som anges i ordboken verkligen kan användas om den referent som nämns i källtexten. Ordboken avser inte att ge encyklopedisk information, och förklaringarna har därför gjorts så kortfattade som möjligt.

En sida i ordboken visas i figur 1.

4 Arbetsgång

Projektet har genomförts på fritid av en person under en period av två år. Den tid som har varit tillgänglig för utarbetandet av ordlistan har följaktligen varit ganska begränsad, vilket har ställt stora krav på uppläggningsarbetet och valet av källor. Det var bland annat nödvändigt att välja källor med relativt litet omfång och stor informationstäthet för att arbetet skulle kunna slutföras inom rimlig tid.

I huvudsak har ordlistan utarbetats genom parallell läsning av olika enspråkiga uppslagsverk. Olika engelskspråkiga encyklopedier har bläddrats igenom från pärm till pärm, och artiklar med många egennamn har jämförts med motsvarande artiklar i ett eller flera svenska uppslagsverk. Förutom de namn som utgör uppslagsord i källorna har många namn

noterats i själva artikeltexterna. Ett exempel visas i figur 2, där motsvarande namnformer har markerats i två källtexter.

D

- Dacia** historiskt rike i Europa Dakien
Daedalus person i grekisk mytologi Daidalos
Daghestan land i Europa Dagestan
Dagwood Bumstead seriefigur Dagobert Krikelin
Daisy Duck Disney-figur Kajsa Anka
Dalecarlia landskap i Sverige Dalarna
Dalmatia landskap på Balkanhalvön Dalmatien
Damascus stad i Syrien Damaskus
Damian helgon Damianus
Damian, Peter italiensk biskop, helgon Petrus Damiani
Damocles hovman i Syrakusa Damokles
Dances with Wolves film av Kevin Costner Dansar med vargar
Danelaw, [the] del av England under vikingatiden Danelagen
Danes' Island ö vid Spetsbergen Danskön
Danube, [the] flod i Europa Donau
Daphne person i grekisk mytologi Dafne
Daphnis and Chloe roman av Longos Dafnis och Chloe
Dardanelles, [the] sund mellan Europa och Asien Dardanellerna
Dargomizhsky, Alexander rysk tonsättare Aleksandr Dargomyzskij
Darius Codomannus kung av Persien Dareios Kodomannos
Darius Nothos kung av Persien Dareios Nothos
Darius Ochus kung av Persien Dareios Ochos
Darius the Median gammaltestamentlig kung Darejaves av Medien
Darkness at Noon roman av Arthur Koestler Natt klockan tolv på dagen
Dashkova, Ekaterina Romanovna rysk furstinna Jekaterina Romanovna Dasjkova
Dateline: Danger tecknad serie Toppreportrarna; Våghalsarna
Dathan man i Gamla testamentet Datan
Dauphiny historisk provins i Frankrike Dauphiné
Davis Strait sund mellan Canada och Grönland Davis sund
Dead Sea, [the] insjö i främre Asien Döda havet
Dead Souls roman av Nikolaj Gogol Döda själar
Deák, Francis ungersk politiker Ferenc Deák
Death of a Salesman drama av Arthur Miller En handelsresandes död
Deborah gammaltestamentlig profetissa Debora
Decameron, [the] novellcykel av Boccaccio Decamerone
Decapolis historiskt stadsförbund i Syrien Dekapolis
Decline and Fall of the Roman Empire, The History of the historieverk av Edward Gibbon Det romerska rikets nedgång och fall
Dee Jen-djeh kinesisk domare Di Jen-djeh
Deerslayer, The roman av James Fenimore Cooper Hjortdödaren
Delilah kvinna i Gamla testamentet Delila
Delphi stad i Grekland Delfi
Delphinus stjärnbild Delfinen
Demetrius kung Demetrios
Demetrius rysk tronföljare Dmitrij
Demetrius Phalerus grekisk politiker Demetrios från Faleron
Demetrius Poliorcetes kung av Makedonien Demetrios Poliorketes
Democratic Party, [the] politiskt parti i USA Demokratiska partiet

Figur 1. En sida i Engelsk-svensk namnordbok.

Charles II the Bad (1332–87), king of Navarre from 1349, made various short-lived attempts to expand Navarrese power in both France and Spain. Related to the French royal house by birth and marriage, he took advantage of France's difficulties in the Hundred Years' War with England, forcing the French king John II to grant him extensive lands in Normandy (1354). He lost most of these to John's successor, Charles V, in 1366, but he was awarded part of the western Basque provinces by Pedro of Castile in return for an alliance in the same year. These lands in turn were lost in 1372, and, by the end of his reign, Charles was subservient to Castile. He was known as "the Bad" because of his reputation for political treachery.

Karl II (fr. *Charles II*), kallad "**Karl den onde**" (*le Mauvais*), 1332–87, kung av Navarra från 1349. Som greve av Évreux hade K. stora områden i Normandie. I allians med England sökte han vinna inflytande i Frankrike men förlorade 1378 de flesta av sina normandiska områden. K. stred även utan framgång mot Kastilien. Han byggde ut Navarras förvaltning och gynnade näringslivet.

Figur 2. *Texter ur två källdokument med motsvarande namn markerade. Den övre texten är hämtad från Encyclopædia Britannica, den undre från Nationalencyklopedin.*

Som källor för de engelska namnformerna valdes i första hand enbands uppslagsverk, eftersom urvalet av fakta i dessa är lagom noggrant för det aktuella ändamålet. Bland de viktigare engelska källorna kan nämnas *The Cambridge Encyclopedia*, *Chambers Biographical Dictionary*, *Chambers Dictionary of World History*, *Pears Cyclopaedia* och *The Times Atlas of European History*. De svenska namnformerna har om möjligt hämtats ur moderna, relativt omfattande uppslagsverk, främst *Nationalencyklopedin* och *Focus*.

Som komplement till uppslagsböckerna har även andra texter av varierande ålder och karaktär använts, bl.a. skönlitteratur, tidskrifter och skivomslag. Vissa flerspråkiga källor har också utnyttjats, bl.a. den svenska standarden för landsnamn, SS-ISO 3166, och den tidigare nämnda *Dictionary of Translated Names and Titles* av Adrian Room. Denna ordbok ger inga svenska namnformer, men den innehåller många originaltitlar på tyska och franska böcker och filmer med motsvarande engelska titlar. Med hjälp av en svensk uppslagsbok som anger originaltitel och svensk titel har det sedan gått att para ihop den engelska titeln med den svenska.

De noterade namnen registrerades i en ordbehandlingsfil med koder för de olika fälten i varje artikel. I stort sett användes samma fältindelning som i TNC:s normala termformat, vilket gjorde det möjligt att använda i stort sett samma rutiner som TNC använder för produktion av terminologiska ordlistor. Figur 3 visar några registrerade poster.

5 Utvecklingsmöjligheter

Den föreliggande ordlistan kan användas som utgångsmaterial till nya ordboksprojekt. Det som ligger närmast till hands är en motsvarande svensk-engelsk namnordbok. En sådan bör ha ett något annat urval av uppslagsord; exempelvis bör man ta med namn på svenska organisationer, boktitlar, litterära personer m.m. som sällan förekommer i engelska originaltexter. Många sådana namn har redan noterats i samband med utarbetandet av den engelsk-svenska ordlistan, men de har lagts undan i en särskild fil för eventuellt framtida behov.

Det är också tänkbart att utnyttja det insamlade namnförrådet som underlag till motsvarande ordböcker för andra språkkombinationer, t.ex. från engelska till norska eller danska eller från tyska eller franska till svenska.

+G	lakenasser
*enTE	Lake Nasser
*UR	CamEnc
*svFK	uppdämd sjö i Egypten
*svTE	Nassersjön
*UR	NE
+G	lakemanzala
*enTE	Lake Manzala
*UR	CamEnc
*svFK	lagun i Egypten
*svTE	Manzilasjön
*UR	NE
+H	walburga
*enTE	Walpurga
*enSY	Walpurgis
*UR	CamEnc
*svFK	helgon
*svTE	Valborg
*UR	Focus
+H	walburga
*enTE	Walburg[a]
*UR	ChamBiog
*svFK	helgon
*svTE	Valborg
*UR	Focus

Figur 3. Registrerade poster. Den översta raden innehåller selekterings- och klassifikationskod samt sorteringsnyckel. På följande rader finns engelskt namn, källa för det engelska namnet, kort förklaring, svenskt namn samt källa för det svenska namnet.

Litteratur

The Cambridge Encyclopedia 1990. Cambridge: Cambridge University Press.

Chambers Biographical Dictionary 1974. Edinburgh: Chambers.

Chambers Dictionary of World History 1993. Edinburgh: Chambers.

Encyclopædia Britannica: Micropædia 1983–. 15th edition. Chicago: Encyclopædia Britannica.

Focus uppslagsbok 1977–. Fjärde, reviderade upplagan. Stockholm: Focus Uppslagsböcker.

Nationalencyklopedin 1989–. Höganäs: Bokförlaget Bra Böcker.

Pears Cyclopaedia 1970. Seventy-ninth edition. Pelham Books.

Room, Adrian 1986: *Dictionary of Translated Names and Titles*. London: Routledge & Kegan Paul.

SS-ISO 3166: *Landsbeteckningar* 1990. Stockholm: SIS – Standardiseringskommissionen i Sverige.

The Times Atlas of European History 1994. London: BCA.

Problem kring ordutvalet i Norsk Ordbok

Norsk Ordbok (The Norwegian Dictionary) is a multi-volume dictionary of the Norwegian dialects and the *Nynorsk* written language. Three volumes have appeared until now, reaching into the letter *g*. In the paper, problems concerning lemma selection are discussed. Three main types of selection criteria are identified, namely structural, pragmatic and normative criteria. The main structural criterion is morphological status (simple words vs. derivatives and compounds). The main pragmatic criteria are frequency, area of usage and age. The main normative criterion in this context derives from the purist tradition of *Nynorsk*, where words of Danish and (Low) German origin have been regarded as problematic. The different types of problematic words are discussed in turn, and some more specific selection criteria are tentatively suggested.

Norsk Ordbok er ei vitenskapleg ordbok over det nynorske skriftspråket og dei norske dialektane. Den har vore i arbeid sidan 1930, og band 3 av i alt tolv planlagde band kom ut vinteren 1994/95. Vi er komne eit stykke ut i bokstaven *g*, og omfanget peikar alt no i retning av seksten band heller enn tolv. Vi må derfor gjennomføre ei kraftig innstramming i redigeringa, og spørsmålet om ordutvalet er eitt av dei viktige punkta å drøfte i den samanhengen.

Ordboka baserer seg delvis på eit setelmateriale med innsende opplysningar frå dialektane, delvis på eldre ordboksverk og lokale ordsamlingar, delvis på ekserpert skjønnlitteratur og sakprosa (jf. Ore 1995). Hovudmengda av materialet vårt stammar frå tida frå ca. 1850, da Ivar Aasen gav ut sine grunnleggjande grammatiske og leksikografiske arbeid, og fram til i dag. (For ein nærare presentasjon av Norsk Ordbok, sjå Hageberg 1992.)

Ser vi på det som finst av offisielle retningslinjer for redaksjonen når det gjeld ordutval, så finn vi lite konkret. Ein må da sjå på den innleiinga den første hovudredaktøren, Alf Hellevik, skreiv til det første heftet av ordboka (Hellevik 1994 [1950]:10–11). Det går der fram at det meste er overlate til redaktørane skjønne. To faste prinsipp finst det likevel:

1. Det samla ordtilfanget i dei grunnleggjande nynorskordbøkene (Aasen 1873, Ross 1971 [1895]) skal vere med i *Norsk Ordbok* (med mogleg unntak av "ei eller anna sjølvsgd samansetning").
2. Særnamn, medrekna stadnamn, blir ikkje tekne med (men derimot eit utval av innbyggjarnamn og dei vanlegaste etterledda i slike namn).

Alt anna er altså skjønsspørsmål. Sjølvsgt er dei aller fleste konkrete tilfella opplagde, men ei rekkje problemområde står igjen.

1 Kriterium

Vi kan dele kriteria for ordutvalet inn i tre typar, nemleg *strukturelle*, *pragmatiske* og *normative* kriterium.

Det *strukturelle* kriteriet går på forholdet mellom grunnord på den eine sida og avleiingar og samansetningar på den andre. Sidan norsk liksom dei nærskyldte språka er svært produktivt når det gjeld samansetningar, og til dels òg når det gjeld avleiingar, er det uråd å dekkje alt som finst av slikt. Sjølv om ein tek mål av seg til å ha ei tilnærma fullstendig dekning av grunnorda i eit språk — med dei avgrensingane dei følgjande kriteria gir grunnlag for — så må ein sortere ut mange samansetningar og avleiingar sjølv om ein ikkje har andre kriterium å avvise dei etter.

Dei *pragmatiske* kriteria er av to slag. Det eine går på *frekvens*, det andre på *bruksområde*. Korleis skal ein behandle ord som er svært sjeldsynte, dvs. som vi har svært få belegg på — heilt ned til berre eitt? Og kva med ord som i og for seg er vanlegare i bruk, men berre på svært avgrensa bruksområde, særleg faglege og vitskaplege? Dessutan kan *alderskriteriet* reknast som eit pragmatisk kriterium: Nye ord stiller i ein kategori for seg i høve til eldre og etablerte ord, som på si side står i motsetning til arkaismar: ord som var i bruk før, men er sjeldsynte eller borte frå det levande språket no.

Dei *normative* kriteria kan vi òg dele i to. Det eine er *etymologisk* basert, og har ei særskild tyngd i eit språk med puristiske tradisjonar, slik nynorsken har. Det går på det at ord som har unorsk opphav, vanskelegare blir aksepterte enn ord som er nedarva frå norrønt. Desse "unorske" orda kan igjen delast i fleire grupper: dansk-tyske ord, nyare fremmedord frå andre språk, og særskilt i vår tid: anglisismar. Også her kjem alder inn som eit tilleggskriterium: Dei dansk-tyske lånorda er gjerne gamle i språket, ein del av dei er typiske arkaismar, mens anglisismane like klart er nyord. — Det andre normative kriteriet er det *estetiske*: "Stygge" ord (bannord, ord for seksualitet, avføring osv.) er ofte problematiske å ta stilling til for ein leksikograf.

Eg skal da drøfte nærare ein del problem som knyter seg til desse kriteria. Dei fleste eksempla eg nemner har eg teke frå den delen av ordboka eg sjølv har arbeidd med, strekninga *hav* — *havrand* (førebels berre førstegongsredigert).

2 Samansetningar

Om samansetningar står det i Hellevik (1994:10) at ein har freista å halde ute "den store mengd der tydinga er heilt grei når ein kjenner dei sjølvstendige orda samansetningane er laga av. Under artikkelar med fyrstelekken til oppslagsord er det ofte teki med nokre døme på slike lettskjønelege samansetningar, utan at det på noko vis har vori meininga å få med alle."

Desse kriteria kan vere meir problematiske i praksis enn i teorien. Tydingstilhøvet mellom ledda i ei samansetning kan verke opplagt for språkbrukarane, men over tid kan kunnskapen om denne samanhengen gå tapt. Det er no ein gong slik at dette tilhøvet er konvensjonelt, ikkje "naturleg". Eit *overfolka* land har for mye folk, eit *overfiska* hav har for lite fisk. For å ta nokre døme frå samansetningane med *hav*: *Havbrasme* er ein brasme (ein slags fisk) som lever i havet. *Havkatt* er derimot ingen katt som lever i havet, men namnet på ein bruskfisk som også blir kalla m.a. *havmus*. *Havklår* kunne ha betydd 'klar som havet', men betyr hos Aasen 'klar over Havet; om Luften'. *Havkiste* er ei kiste til å lagre brød i, som ein tek med seg til havs under fiske (belegg frå Hitra). Det kunne ha betydd ei skipskiste generelt, eller for den del brukast biletleg om havet som gravplass, men det har vi ikkje notert belegg på. Slikt kan ein ikkje vite utan vidare. Ved det siste ordet eg har

redigert til no, *havrand*, tenkte eg på førehand at det kunne vere eit synonym til *havkant*, nemleg kystline. Men havranda viste seg å liggje på motsett side av havet; det er ei av mange folkelege nemningar på horisonten til havs.

I mange tilfelle har vi dessutan fleire ulike tydingar av same samansetninga. I samband med *hav* er det såleis eit stort innslag av ord som ikkje har nokon basis i verkelegheita slik vi opplever ho, men som hører til i tradisjonell folketro — samtidig som dei same orda kan brukast med ei meir daglegdags tyding. *Havbonde* kan bety bonde ved kysten, eller, i folketroa, eit overnaturlig vesen som levde i havet og dreiv gard der. Særskilt tydingsrikt er ordet *havhest*, som kan brukast om tre dyreartar: 1) ein fugl, stormfugl; 2) eit pattedyr, kvalross — rett nok truleg forelda i dag; 3) ein fisk, som i dag vanlegvis heiter sjøhest. I tillegg kan det vere 4) eit havuhyre på skap som ein hest, og så kan vi ha ulike overførte tydingar: 5) skip, 6) bølgedrag på havet — og 7) ein dyktig kar på havet. Det einaste det ikkje kan bety, er ein hest som lever i havet, og nettopp fordi det av biologiske grunnar manglar primærttyding, er det ledig for så mange sekundærttydingar.

Det er viktig å leggje vekt på om samansetninga er leksikalisert med ei eiga tyding, eller om ho er ei tilfeldig danning som berre fungerer innanfor ein bestemt kontekst. Eit døme på det siste er *havkollega*. Det er avisa *Gula Tidend* som i ein artikkel skriv om den mellomalderlege trua at det i havet fanst ein tru kopi av menneskesamfunnet med kyrkje og alt anna som hører til, og at det ein gong vart fanga ein havbiskop i Austersjøen. ”Då to romersk-katolske biskopar etterpå fylgde havkollegaen sin attende til sjøen og fridomen, gjorde havbispnen krossteikn og vart borte i djupet”, står det i artikkelen. Det var ikkje noko vanskeleg val å utelate både *havbis(ko)p* og *havkollega*.

Men det kan altså vere risikabelt å utelukke samansetningar ut frå at redaktøren og dei fleste andre norsktalande i samtida oppfattar ho som sjølvsagt, i ei ordbok som langt på veg er diakron. Trass i dette er det uunngåeleg at slike avvegingar blir gjorde, og det i ganske stort omfang. Generelt gjeld det at når ein kuttar ut samansetningar ein finn belegg på, bør ein helst ha støtte i nokre av dei kriteria eg no skal sjå på, nemleg dei pragmatiske.

3 Frekvens — Hapax legomenon

Frekvens er svært viktig. Ei samansetning med mange belegg vil vanlegvis vere leksikalisert og ha mange bruksmåtar, også overførte, og slike må vi sjølvsagt ha med. Men svært mange ord, og det gjeld i særleg grad samansetningar, har få belegg, og ber sterkt preg av å vere tilfeldige lagingar. Dei fleste som det kan vere tvil om, er slike der det berre er notert einstaka belegg frå ein eller to spreidde dialektar, eller frå skjønnlitterære kjelder, svært ofte poesi, eller det kan vere nyord tekne frå aviser — som også kan vise seg å vere døgnfluger.

Hapax legomenon er eit velkjent begrep for leksikografar. Det kan vere ekte hapax legomena, altså ord som er laga av ein enkeltperson og brukt berre éin gong eller i ein bestemt samanheng, t.d. eit dikt — eller einmannslagingar av meir tilfeldig type. Men det kan òg vere ord vi berre har eitt belegg på, samtidig som det kan vere eit meir vanleg brukt ord — t.d. ord frå moderne mediespråk, som unekteleg er noko tilfeldig ekserpert og registrert. Endeleg kan det vere eit sjeldsynt belegg frå ein dialekt.

Vi er vel ikkje heilt avklart når det gjeld handlinga av slike tilfelle. Sjølv har eg laga meg visse kriterium som eg forsøksvis legg fram her (dei er enno ikkje drøfta i redaksjonen).

Normalt vil eg gå svært langt i å ta opp ord som lever eller har levd i dialektane, jamvel om dei er så lokale at det berre finst eitt eller eit par belegg. Sjølv sagt må eg da vurdere sjølve belegget, for å sjå om opplysninga kan reknast som påliteleg. Mange av dei orda Ross (1971) har med, er av ein slik type, og dei skal vi som nemnt alltid ta med. Det er naturleg for meg å ta inn flest moglege slike ord for å gi eit mest mogleg fullstendig bilete av ords-katten i målføra.

Når det gjeld skriftlege belegg, vil eg vere restriktiv overfor tre typar kjelder. Det er for det første skjønnlitteratur, særleg lyrikk. Diktarar har ein rett og ei evne til å skape ord som ikkje andre utan vidare har del i, veit vi, og ofte kan slike ord meir vere dikterte av rytme og stemning i diktet enn av trongen til å gi ord til ei ny tyding. Ikkje minst i nynorsk lyrikktradisjon er ordlaging svært oppvurdert, særleg i noko eldre tid enn i dag. Slike "diktarord" kan nok ha ei "tyding" i konvensjonell meining slik vanlege ord i språket har, men ofte får dei inga eigentleg tyding som ein kan referere til utanfor diktets eiga ramme. Dei må altså tolkast litterært før dei eventuelt kan definerast leksikografisk, og etter mi meining kan slik litterær tolking ikkje høre til ein leksikografs oppgåver. Som døme nemner eg *havnatt*, som i og for seg er ei grei nok samansetning, og som i arkivet vårt har sju belegg, men alle frå skjønnlitteraturen: tre sitat kvar av lyrikarane Olav Aukrust og Anders Hovden, og eitt av romanforfattaren Olav Duun. Sitata viser tydeleg at *havnatt* ikkje berre betyr 'natt over havet', men har eigne konnotasjonar innanfor diktets ramme knytte til vondskap, trugande fare, nattsvart sinnsstemning, storm og uvêr på havet om natta o.l. Og desse konnotasjonane er meir primære i dikta enn denotasjonen av den i og for seg tilfeldige (ikkje leksikaliserte) danninga *havnatt*. Eg har valt å ikkje ta med ordet, sidan berre utførlege sitat frå dikta der dei står med litterær analyse attåt kan belyse det. Denne haldninga tek eg generelt til ord som berre finst hos éin diktar, kanskje berre i eitt dikt. Unntak er dei største klassikarane i nynorsk diktning: Ivar Aasen, Aasmund Olavsson Vinje, Arne Garborg og nokre få andre.

Spesielt restriktiv er eg ved omsetjingar, der omsetjaren har laga nye ord for å gi att ein fremmendspråkleg original. I slike tilfelle må ordet ha komme i bruk i språket og såleis vise att i andre kjelder før eg vil ta dei inn. Dette gjeld òg i omsetjingar frå norrønt, der særleg eldre omsetjarar hadde ein tendens til å ta inn norrøne ord også av ideologiske grunnar. Også slike ord må dokumentarast frå generelle nynorske kjelder før det kan komme på tale å ta dei med. Døme på eit slikt problemord er *havguve*, som berre finst i den nynorske omsetjinga av *Konungs skuggsjá* (Kongsspegelen) som namn på eit havuhyre som på norrønt heiter *hafgufa*. Det er rimeleg å sjå bort frå det som eit ord i nynorsk.

Ei tredje tvilsam kjelde, dvs. kjelde som ikkje blir brukt utan støtte av andre allmennspråklege kjelder, er ordbøker frå fremmendspråk. Dei har same ulempa som omsetjingar: Ordboksredaktøren kan ha kjent seg fri til å lage ord eller ta inn "daude" ord for å gi att eit ord eller ein nyanse på det fremmende språket. Eit litt problematisk eksempel er *havkabel*, som berre finst belagt i Schjøtt (1909) som omsetjing av *undervandsledning*. Denne samansetninga er grei og tilforlateleg, og det er vel truleg at ordet er brukt, men sidan samansetninga vel må reknast som temmeleg sjølv sagt samtidig som vi ikkje har belegg på konkret bruk av ordet (som kunne antyde leksikalisering eller terminologisering), har eg valt å ikkje ta det med.

4 Fremmendord

Dette problemet er i grunnen to problem, for fremmendord-begrepet kan oppfattast på to måtar. Vi har eit historisk fremmendordbegrep, der fremmendord er alle ord som har komme inn i språket utanfrå utan å tilpasse seg språksystemet strukturelt (ein kan da skilje dei frå lånord, som har fremmdt opphav, men har tilpassa seg det norske systemet). Det andre fremmendordbegrepet er synkront: Det inneber at fremmendord er ord av fremmdt opphav som er lite kjende i allmennspråket, men har sitt særlege bruksområde innanfor bestemte fagspråk eller stillag.

I Hellevik (1994:10) er det sagt klart i frå at "hovudarbeidet må leggjast på det heimlege ordtilfanget". Grunnen til det er at fremmendorda er sams for bokmål og nynorsk, og grundig behandla i *Norsk Riksmålsordbok*. Men så blir det i neste setning sagt at ein likevel tek med eit avgrensa utval av dei mest vanlege fremmendorda, fordi ein elles ville få eit skeivt bilete av målet. Ein særleg grunn til å ta dei med er at defineringa av dei kan vere tenleg i arbeidet med å finne norske avløyसारord.

Után at det er presisert, er det lett å sjå at det er det diakrone fremmendord-begrepet som ligg til grunn for desse refleksjonane. Her kjem det ideologiske momentet inn som vi snart skal gå nøyare inn på i samband med dei dansk-tyske orda: Ordboka skal ha eit normativt aspekt, ikkje berre vere deskriptiv. Denne posisjonen må reknast som svært problematisk for eit vitskapleg ordboksverk. Problematisk er òg tilvisinga til *Norsk Riksmålsordbok*, fordi den impliserer at *Norsk Ordbok* ikkje skal ha som utgangspunkt at nynorsk er ei autonom språkform som skal beskrivast på eigne premisser, men eit slags "særpråk" i forhold til bokmålet. Skilnader i frekvens og bruk av fremmendord mellom bokmål og nynorsk vil ikkje komme fram i ordboksverket om ein skal gå ut frå denne tankegangen som Hellevik her målber (og implisitt vedgår Hellevik da også dette i det som følgjer i det same avsnittet).

Vi må forstå Helleviks formuleringar som eit forsøk på kompromiss i ein ideologisk konflikt blant nynorskbrukarane — derfor vagheita. Men uansett er det ikkje lett å finne fram til ein rett grenseoppgang her — fordi den enorme veksten i teknolektar i dei siste tiåra knapt kan avspeglast i eit verk som *Norsk Ordbok*. I dei faktiske avvegningane vi må gjere her, vil det vere mest naturleg å ta utgangspunkt i det synkrone fremmendordbegrepet — altså sjå etter korleis ordet fungerer i språket heller enn kor det kjem frå. Ut frå tilfanget sitt er det greitt at *Norsk Ordbok* skal beskrive allmennspråket først og fremst, og ord som berre høyrer til i spesielle fagterminologiar må vi vere restriktive med. Men alt er ikkje sagt med det, for grensa mellom allmennspråk og fagspråk kan vere vag. Fagbøker er såleis med i grunnlagsmaterialet for *Norsk Ordbok*, men da gjerne av eldre dato, og helst slikt som er skrivi for "den opplyste ålmenta": skolebøker, populærvitskap, artiklar i tidsskriftet *Syn og Segn*, Johannes Lids standardverk *Norsk flora*, og også somme universitetslære-bøker. Skjønnet må altså først og fremst komme inn i vurderinga av kor utbreidd ordet er og kva plass det har i språket; fagord som inngår i allmennutdanninga vil t.d. ha plass i ordboka, men ikkje slike som er uaktuelle for ikkje-spesialistar. I lista over forkortingar i brukarretteiinga til *Norsk Ordbok* (Grønvik o.a. 1994:66–70) står det òg nemnt kva for fagfelt som har levert ord og tydingar til ordboka, og der finn vi dei fleste fag som det finst utdanningar for i Noreg, men ikkje felt som edb, oljeteknologi, havbruk. Det seier seg sjølv at ord frå slike område vil måtte få plass i ordboka (og det har òg skjedd i det nyaste bandet), men det seier seg like mye sjølv at den rivande utviklinga på desse felta ikkje kan avspeglast i eit verk som

dette. Det sentrale kriteriet vil framleis måtte bli: Kva status har ordet i allmennspråket på det aktuelle tidspunktet?

5 Nyord

Dei fleste av dei nye orda som kjem inn i språket, høyrer eigentleg til dei kategoriane vi nettopp har behandla: Dei er innlånte frå andre språk, dei er bundne til bestemte språkbrukskontekstar (særleg i eigenskap av fagord), og/eller dei er samansetningar eller avleiingar. Dei kan da vurderast frå ulike synsvinklar (den diakrone, den pragmatiske og den morfologiske), og avvegningane må skje ut ifrå dei prinsippa vi har skissert ovanfor. Den viktigaste synsvinkelen er om det aktuelle ordet kan seiast å ha fått fast fot i det allmenne bruksspråket, og her kan ordboksredaktøren aldri vere sikker på at han greier å vurdere rett, sidan først ettertida får den endeleg fasiten i handa her. Dessutan er begrepet "det allmenne bruksspråket" ekstra problematisk nettopp her, for det skal ikkje nektast at materialet vårt er ekstra tynt når det gjeld moderne publikasjonar. Dei områda der nynorsken står sterkast i skrift, er skjønnlitteratur, delar av pressa, særleg lokalpressa, og lærebøker og annan faglitteratur. Derimot står det heller dårleg til i populærkulturen, reklame, teknisk fagspråk osv. der mye av ordskapinga i vår tid går føre seg. For min eigen del har eg funne ein del ord som truleg er nye i samband med framvoksteren av havbruksnæringa på Vestlandet, reflektert i lokale aviser som *Fiskaren*, *Gula Tidend* og *Strilen*, og også i den landsdekkjande vekeavisa *Dag og Tid*. Belegga er frå sytti- og åttiåra, og det trengst ofte ei fagleg innsikt for å sjå om orda er blitt permanente og har fått eit allmennspråkleg preg i alle fall i Kyst-Noreg.

Ein annan og problematisk type nyord er alle anglisismane. Her kjem det etymologisk-normative aspektet inn med full styrke, og det er vel rett å seie at vi i *Norsk Ordbok* er og vil vere svært restriktive med å ta inn slike ord. Eit godt pragmatisk argument for det er at det er vanskeleg eller umogleg å vite kva slags anglisismar som vil setje seg fast i språket på lang sikt, og kva for nokre som berre er døgnfluger. Ein viss konstans i språket er i det heile eit krav vi må stille til eit ord før vi tek det inn, og det avgrensar sjølvsagt nyordsinnslaget sterkt.

6 Dansk-tyske lånord

Nynorsk språk er bygd på ein puristisk tradisjon, som først og fremst rettar seg mot lånord frå dansk og tysk. Det gjeld særleg ord med visse innlånte affiks, som er blitt frekvente og til dels produktive i norsk, særleg *-heit*.

Mange av desse orda er svært vanlege i norsk talemål, også i dei dialektane som ligg nærast nynorsk, dei blir truleg nokså mye brukte i nynorsk standardtale (som er friare i ordbruken enn skriftspråket), og ein god del har fått hevd også i skrift, andre blir brukte meir sporadisk. Ikkje desto mindre har dei ein status som "ikkje-godkjende", dvs. at dei ikkje blir tekne inn i ordbøker og ordlistar som er autoriserte av Norsk språkråd til bruk i skolen, og heller ikkje i den normgjevande eittbandsordboka *Nynorskordboka* (Hovdenak o.a. 1986), som også er utgitt av Norsk språkråd i samarbeid med Avdeling for leksikografi på Universitetet i Oslo. I desse ordbøkene er altså avvegingsproblemet løyst for ordboksredaktørane sin del: Språkrådet bestemmer kva slags ord av denne typen som kan takast inn eller ikkje.

Med *Norsk Ordbok* stiller det seg annleis. Mens *Nynorskordboka*, for ikkje å snakke om skoleordlistene, er prinsipielt normative, er *Norsk Ordbok* prinsipielt deskriptiv. Språkrådet eller andre normgjevande instansar har ingenting å seie her; det er redaksjonen som suverent bestemmer kva ord som skal takast inn og korleis dei skal behandlast.

Forholdet mellom deskriptivitet og normativitet i ei ordbok er svært innfløkt og vagt. Inga ordbok kan beskrive heile språkbruken; det utvalet som blir gjort, utgjer i seg sjølv ei normering. Dessutan fungerer ordboka i eit samfunn som er tufta på normer, normene styrer språkbruken, så ordboka må beskrive normer og dermed befeste og spreie dei ved å spreie kunnskapar om dei. På den andre sida kan ikkje ei normativ ordbok unngå å beskrive; eit ord må vere i bruk før det kan beskrivast som del av ei norm. (Dette ideelle kravet er likevel ikkje alltid like godt følgt i praksis.)

For *Norsk Ordboks* vedkommande er situasjonen den at vi først og fremst har setelaterialet og tekstkorpuset vårt å gå ut ifrå. Truleg ligg det der i seg sjølv ei feilkjelde: Dei som har sendt inn belegg frå dialektane har nok vore meir interesserte i "genuine" og "gamle" ord i dialekten enn i nyare lånord. Også skriftmålskjeldene vil typisk ha låg frekvens av slike lånord, fordi mye av det er av eldre dato, og nynorsknorma (dvs. den nynorsknorma som er internalisert hos kvar enkelt skribent) er ein ganske effektiv sensurinstans. Utvalet av tekster som blir tekne med i materialet kan òg ha sitt å seie.

Kva slags kriterium skal ein så bruke ved utvalet av slike ord?

Dei allmenne kriteria må sjølvsagt takast i bruk her òg, dvs. frekvens, stilverdi, konstans i språket, status i allmennspråket. Men den særeigne karakteren til *Norsk Ordbok* reiser to prinsipielt vanskelege, men viktige problem: 1) Kva med ord som er vanlege i dialektane, men lite brukte i skrift? 2) Kva med ord som blir brukte i skrift, men som ideologisk og normativt blir betrakta som "ikkje-nynorske" eller "bokmålske"?

Hellevik (1994:11) opnar for ei liberal haldning. Han tek først avstand frå akademi-prinsippet, altså det normative prinsippet, men slår klart fast at ordboka skal vere deskriptiv. Så seier han:

Vi har ikkje sett det som vår oppgåve å døma ut frå eit subjektivt syn på kva ord som bør reknast for godtakande i nynorsk skriftmål, men har freista å byggja på ei varsam omdøming av det tilfanget som røyneleg ligg føre. Så får då seinare mindre ordbøker som er meir esla til skulebruk, gjera eit høveleg utval på grunnlag av det tilfanget som her er lagt fram.

Med andre ord: Normeringa av ordtilfanget blir overlata til skoleordlistene.

Dette er nok den lina som det er mest naturleg å stå på. Det vil seie at ein bør gå ut frå materialet vi byggjer på, og så gjere eit utval bygd på dei same kriteria som vi nyttar for ordtilfanget elles, m.a.o. ikkje behandle denne typen ord som ei særskild gruppe. Det har like stor interesse både vitskapleg og for normeringa å få framstilt den røynelege utbreiinga og bruken av slike ord som for alle andre ord.

Men problemet forskuvar seg: Kva haldning skal ein ta til tekster (skjønn- eller faglitterære) som har mye av slike ord? Korleis skal representativiteten til det korpuset vi byggjer på, sikrast best råd?

Det er klart at dei tekstene som blir lagde inn (ekserperte eller skanna) ikkje blir valde ut frå noko mekanisk prinsipp, men etter skjønn, og hovudprinsippet er å sikre ein viss representativitet, kronologisk og sjangermessig. Denne representativiteten bør omfatte også

sjølve språkbruken, frå det tradisjonalistiske til det tilnæringsvennlege og dei moderne, urbaniserte nynorskvariantane. Ein slik representativitet vil truleg også sikre eit allsidig ordtilfang, og dermed ein betre beskrivelse av det nynorske kultur målet i full breidd. Det må også omfatte ord som lever i nynorsken, men som normerarane har eit spent forhold til.

Ser vi på ordutvalet i *Norsk Ordbok* til no, finn vi at dette omsynet i stor grad er ivareteke. Ei rekkje ord med prefikset *be-* er såleis komne inn — eg har talt 156, av dei somme temmeleg sjeldsynte som *bedara*, *bedyra*, *bekoma*, *berå*, *beskikka*. Rett nok manglar andre *be-*ord som i dag er ganske vanlege, som *bearbeida*, *beredskap*, *berekna* og *beskriva*, men ein kontroll i setelarkivet viser at det ikkje eller nesten ikkje finst belegg på desse. Til ein viss grad er nok dette eit uttrykk for tilfældigheiter, men det kan òg ha noko å seie at redigeringa av *be-*skjedde på femti- og tidleg på sekstialet og at desse orda på den tida var meir ”markerte” som bokmålsord enn dei er i dag. Det finst faktisk ein setel på *beredskap*; den er blyantskriven og lyder slik: ”*beredskap* m dessverre nytta mykje av målmen i Stortinget, i staden for reiugskap, vernebuing, klarhald”. Trass i eit slikt døme må vi generelt sett konstatere at så langt det gjeld *be-*, har redaksjonen vore trufast mot materialet og ikkje drive normativ eller språkideologisk ”sensur”.

I *Norsk Ordbok* 1–2 (*a-flusken*) har eg funne desse *heit*-orda:

alminnelegheit	dårligheit	elskverdigheit
argfastheit	egenheit	enkelheit
audmjukheit	eindrygheit	etterlatugheit
avguheit	einfaldigheit	evigheit
beinheit 'beine'	einfaldugheit	falskheit
blidheit	einheit	farlegheit
blygheit	einigheit	fastheit
botferdigheit	einsamheit	fegenheit
djupheit	einsheit	feigheit
drektigheit	einsidigheit	ferdigheit
dristigheit	einslegheit	festlegheit
drygheit	einspønenheit	finheit
dug(e)legheit	einsvørslegheit	fjørenheit
dumheit	ekkelheit	flauheit
dyrverdigheit	ektheit	flinkheit
dåneklenheit	elendigheit	flottheit
		/ 48 ord

Frå Oddvar Nes (1984:88)

Det same kan ein seie om det svært kontroversielle suffikset *-heit*. Oddvar Nes (1984) har granska dei to første banda i verket, og funne 48 ord på *-heit* (sjå ramma). Frå band 3 kan ein leggje til døme som *forlegenheit*, *frekkheit*, *friheit*, *friskheit*, *frodigheit*, *fromheit*, *fruktarheit*, *fulheit*, *fullheit*, *galenheit*, *galningsheit*, *gapingsheit*, *geistlegheit*. Det som verkar litt pussig, er at svært mange av desse orda (jf. lista til Oddvar Nes) er ukurante på bokmål og i mange tilfelle truleg heller ”sære” også i talemålet. Dette kan nok vere eit resultat av tendensen til å vurdere det mest ”særmerkte” i dialektane som mest interessant; mange heimelsmenn kan ha hatt sperrer mot å sende inn ord dei rekna som ”bokmålsord”.

Også her er *Norsk Ordbok* trufast mot materialet sitt. Det normative aspektet som trass i alt også er der, kjem fram i bruken av merkinga *lbr*, ev. *lbr i skr*, det vil ifølgje forkortingslista i brukarrettleiinga (Grønvik o.a. 1994:68) seie 'lite bruka, og vanleg rekna for lite brukande (i skrift)'. Dei fleste eller alle *-heit*-orda har slik merking, som i og for seg er deskriptiv, men likevel forkledd normativ: Ved å beskrive ei normativ haldning, verkar ordboka i seg sjølv normativt, sjølv sagt. Den alternative og reint deskriptive merkinga er *sj*, som betyr 'sjeldan'. — Men ein må seie at normativiteten i *Norsk Ordbok* er så svak som det er mogleg for ei ordbok å bli: Det minimum av normativitet som ligg i å beskrive normstyrt språkbruk, kan no ein gong inga ordbok unngå. Men det er av same grunn uunngåeleg at ei ordbok som spenner over fleire generasjonar reflekterer endringar ikkje berre i språkbruken, men òg i normsynet. Det er god grunn til å tru at ord som *nyheit*, *staheit* og *ærlegheit* kjem til å få plass i *Norsk Ordbok* utan noka "lbr"-merking.

Eg skal berre bruke få ord på den reint estetiske normativiteten. Eg har sett på nokre få "stygge" ord (*dass*, *drit(a)*, *fanden*, *fitte*), og slike ord blir medtekne og behandla på vanleg vis (med bruksmerking) så sant dei er etablerte nok i folkemålet til å ha komme inn i setelmaterialet. Her har redaktørane altså teke ei reint deskriptiv haldning.

7 Konklusjon

Som eg sa i innleiinga, må det skjerast ned på redigeringa av *Norsk Ordbok* om prosjektet skal komme vel i hamn. Ei slik nedskjering må omfatte andre aspekt enn berre ordutvalet, men om vi avgrensar oss til det her, så vil det etter mi meining vere naturleg å leggje hovudvekta på dei to første utvalskriteria, det strukturelle og det pragmatiske. I alle fall er det viktig at kriteria for ordutvalet, same korleis vi til slutt vel å formulere dei, blir uttrykte så klart og eintydig at ikkje berre noverande og framtidige redaktørar, men også brukarane av ordboka kan kjenne til dei og forhalde seg til dei — slik at det blir mest mogleg samsvar mellom det ordboka lovar, og det ho gir.

Litteratur

Aasen, Ivar 1873: *Norsk Ordbog med dansk Forklaring*. Christiania: P. T. Mallings Boghandel. [Fleire fotografiske opptrykk.]

Grønvik, Oddrun o. a. (red.) 1994: *Norsk Ordbok. Brukarrettleiing*. Oslo: Det Norske Samlaget.

Hageberg, Arnbjørg 1992: Norsk Ordbok — i 1950 eit nybrottsverk, i 1991 ein anakronisme? I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi*. Oslo: Nordisk forening for leksikografi, 225–243.

Hellevik, Alf 1994 [1950]: Om oppbygginga av ordboka. I: Grønvik o. a. 1994, 9–17. [Opphavig i *Norsk Ordbok*, hefte 1, under "Innleiing".]

Hovdenak, Marit o. a. (red.) 1986: *Nynorskordboka*. Oslo: Det Norske Samlaget.

Nes, Oddvar 1984: *Om nokre affiks i nynorsk — mest om -heit*. Bergen: Nordisk institutt.

Norsk Ordbok. 1966–. Oslo: Det Norske Samlaget.

Norsk Riksmålsordbok. 1937–95. Oslo: Aschehoug/Kunnskapsforlaget.

Ore, Christian-Emil 1995: Korpus og seddelarkiv — fredelig sameksistens mellom det beste og det gode? [I denne konferanserapporten, 331–337.]

Ross, Hans 1971 [1895]: *Norsk Ordbog*. Oslo: Universitetsforlaget.

Schjøtt, Steinar 1909: *Dansk-norsk Ordbog*. Kristiania: H. Aschehoug & Co.

Et arkiv over forandringer i moderne norsk ordforråd og dets betydning for norsk leksikografi

Since 1968 a project for registrating lexical changes in Norwegian has been going on at the Section of Lexicography, University of Oslo. The number of excerpts is now about half a million, covering new words, new meanings of words, new phraseology and various characteristic lexical features in modern Norwegian. The greatest part of the material is computerized, and each excerpt is provided with a code which gives information about word class, gender and various lexical and syntactic features. The topic of this paper is furthermore what the material has been used for, particularly its importance for dictionary making in Norway the last 20 years, and finally how useful material of this kind is in dictionary making.

1 Innledning

Vi leksikografer i Norge blir litt misunnelige på kollegaene våre i andre nordiske land når vi hører om deres innholdsrike elektroniske tekstkorpus med millioner på millioner av ord som de kan boltre seg i når de lager sine ordbøker. Språkdata i Gøteborg har tekstkorpus med flere titalls millioner ord. Dansk Ordbog kan øse fra et på over 40 millioner. Et hovedmål for leksikografien i Norge må bli å komme etter på dette området, og vi kan nå konstatere at det er ting på gang også i vårt land. Men det er langt fram, og dette foredraget skal handle om noe av det vi har hatt å hjelpe oss med til nå.

Ved Avdeling for leksikografi ved Universitetet i Oslo (UiO) har en fra 1968 registrert forandringer i ordforrådet i moderne norsk allmennspråk med særlig vekt på nyord. Tallet på ekserpter er nå rundt en halv million. Jeg skal si litt om hvordan arkivet over det innsamlede materialet er bygd opp og hva det er blitt brukt til, særlig i hvilken grad det er blitt benyttet under utarbeidelsen av norske ordbøker. Til slutt vil jeg gi en vurdering av i hvilken grad et slikt ekserpert materiale egner seg som grunnlag for ordbøker.

Det var daværende ordboksredaktør Dag Gundersen ved Det norske litterære ordboksverk (som seinere ble en del av Norsk leksikografisk institutt (NLI) ved UiO, og nå er en del av Avdeling for leksikografi) som satte i gang prosjektet, som til daglig har gått under navnet Nyordsprosjektet. Prosjektet fikk støtte fra NAVF de første årene før NLI ble opprettet.

Innsamlingen av materiale er organisert på den måten at folk som er interessert, får et gratisabonnement på en avis eller et tidsskrift mot at de streker under nye ord, nye bruksmåter av ord, og ellers slikt de synes viker av fra vanlig språkbruk eller finner språklig interessant på annen måte. Tallet på frivillige ekserptorer har variert fra vel 30 til snaut 20. Opplegget ligner etter det jeg forstår, på det som Dansk Ordbog har med sine "sporhunde". I tillegg til disse frivillige hjelperne leser og streker de tilsatte ved avdelingen så mye avis- og tidsskriftstoff som de har tid til ved siden av andre gjøremål. I en kort veiledning for ekserptorene, står det at hensikten med innsamlingen er "å bidra til beskrivelsen av det norske

språk. Senere kan materialet brukes til praktiske formål som ordbøker eller til å fastslå hva som bør regnes for rett eller galt". Det er derfor slik at dette materialet inneholder mye interessant språkstoff utover akkurat nyord.

Mens materialet de første to-tre årene ble skrevet ut på sedler, begynte en i 1971 de første forsøkene på å skrive inn og lagre det elektronisk. Samtidig ble det utarbeidet et system av bokstavkoder, slik at det innsamlede materialet kan deles inn i emner, og en får muligheter til ikke-alfabetiske søkeinnganger og kryss-søk i det.

Den første plassen i koden er forbeholdt markering av ekserptets ordklasse og når det gjelder substantiv, samtidig kjønn på ekserptet. Den andre viser om det ekserperte ordet er sammensatt eller ikke. Er det usammensatt, markeres det med en *x*. Hvis det er sammensatt, blir ordklasse og eventuelt kjønn på førsteleddet angitt. Videre kan en så føye til så mange tilleggskoder som en måtte ha bruk for. Hver av dem består av to bokstaver, og angir språklige særtrekk ved det aktuelle ekserptet:

asylant, *cxsi* = usammensatt felleskjønnsord dannet med suffiks
brevgiro, *cn* = sammensatt felleskjønnsord med et nøytrumsord som forledd
eldreomsorgstjeneste, *ccsfks* = sammensatt felleskjønnsord, sammensatt felleskjønnsord som forledd og komposisjonsfuge *s*.
hyttebygge, *vcrg* = retrograd verbdannning med et felleskjønnsord som forledd
ostehøvelprinsipp, *ncsfbb* = sammensatt nøytrumsord, sammensatt felleskjønnsord som forledd, brukt billedlig
versting, *cxsc* = usammensatt felleskjønnsord lånt fra svensk

De tilleggskodene som blir brukt, kan deles inn i 12 grupper:

1. Koder for sammensatte ord (f.eks. for bindestreksord, komposisjonsfuge, sammensatt forledd o.fl.)
2. Koder for affiks, avledninger og analogidannelser
3. Koder for lån fra andre språk
4. Koder for ordklasseovergang (også overgang fra affiks til selvstendig ord)
5. Koder for allusjoner, bildebruk, stilfigurer
6. Koder for lydlige trekk
7. Koder for påfallende bøyingsformer, bøyning, ortografi og ordbruk
8. Koder for forkortinger og forkortede uttrykksmåter
9. Koder for fraser og andre ordsammenstillinger
10. Koder for passiv og refleksiv form av verb
11. Koder for syntaktiske trekk
12. Andre koder (f.eks. koder som markerer overgang fra abstrakt til konkret betydning, ekserpt som står i anførselstegn, slang o.fl.)

Kodesystemet kan utvides etter hvert som behovet måtte melde seg, og det utgjør et ganske finmasket nett som vi anser som et godt hjelpemiddel til å kartlegge viktige utviklingstrekk i moderne norsk.

Med god hjelp av Dokumentasjonsprosjektet er nå også seddelmaterialet lagt inn og lagret elektronisk, og en er i ferd med å legge alt inn i en database med f.eks. muligheter for forskjellige kryss-søk. Foreløpig er dette bare en mindre prøvebase, men den skal utvikles etter hvert, og tanken er at den med tiden skal bli tilgjengelig på Internett.

2 Hva materialet er blitt brukt til

I veiledningen som ble skrevet for ekserptorene for 27 år siden, ble det sagt at materialet bl.a. skulle brukes til å fastslå hva som bør regnes for rett eller galt. Den funksjonen har det også hatt. Når Norsk språkråd skal normere et nyord, eller vurderer å forandre skrivemåten av et ord, kontakter de svært ofte vår avdeling for å høre hvilke skrivemåter vi har registrert i våre arkiver før de tar en avgjørelse. Nyordsmaterialet utgjør m.a.o. en viktig del av det empiriske grunnlaget for språknormeringsvedtak.

Mange av dem som tar vår emnekrets i leksikografi, henter naturlig nok stoffet til sin semesteroppgave fra vårt materiale. Som andre eksempler på hva materialet blir brukt til, kan jeg nevne en hovedfagsoppgave om komposisjonsfuge i sammensatte ord, og en omfangsrik avhandling om verb i passiv fulgt av perfektum partisipp, som kom i fjor. En svært stor del av materialet for moderne norsk i denne avhandlingen kunne hentes fra våre arkiv, ved å søke på koden *pp* (påfallende passiv).

Når nå nyordsmaterialet blir gjort lettere tilgjengelig for søking og annen utnyttelse elektronisk, håper vi at det skal bli bedre utnyttet til forskningsformål, både av hovedfagsstudenter og andre.

Når det gjelder ordbøker, har det knapt kommet en seriøs enspråklig ordbok i Norge de siste tjue årene som ikke i større eller mindre grad har gjort seg nytte av nyordsmaterialet, om vi ser bort fra reine fagordbøker. Jeg skal nevne de viktigste.

Nyord i norsk 1945–1975 (1982) var det norske bidraget i det nordiske samarbeidet med å registrere og vurdere tilveksten av nye ord i de nordiske språkene. Også når det gjaldt arbeidet med den danske og svenske nyordboka var nyordsmaterialet til hjelp når en skulle finne norske ekvivalenter til de danske og svenske nyordene.

Videre kan vi nevne den allmennspråklige, ettbinds *Riksmålsordboken* (1977) utgitt av Det Norske Akademi for Sprog og Litteratur. I forordet takker utgiverne NLI for adgang til instituttets arkiver. Denne ordboka kom i ny og atskillig større utgave i 1993 med tittelen *Norsk Illustrert Ordbok*.

Bokmålsordboka og *Nynorskordboka* var samarbeidsprosjekter mellom Norsk språkråd og Avdeling for leksikografi. Bøkene ble utarbeidet ved Avdeling for leksikografi, og en hentet selvfølgelig oppslagsordene bl.a. fra avdelingens store nyordssamlinger. Førsteutgavene kom i 1986 og de reviderte utgavene i 1993.

I mellomtida, i 1992, hadde også *Bokmålsordlista* kommet, en rettskrivningsordliste som bygde på *Bokmålsordboka*, men som det også var naturlig å supplere med nyord som hadde kommet til etter 1986, fra nyordssamlingene. I disse dager er den tilsvarende *Nynorskordlista* på det nærmeste ferdig.

I mange år har det vært arbeidet med supplementsbind til den fire bind store *Norsk Riksmålsordbok* (ferdig i 1957). Resultatet er blitt to supplementsbind som skal lanseres høsten 1995. Redaktøren har hatt sin arbeidsplass ved avdelingen, og vi blir igjen takket i forordet for den store hjelp og nytte han har hatt av avdelingens samlinger.

Et annet ordboksprosjekt, *Norsk anglisismeordbok*, har fått utskrifter av våre lånord fra engelsk og norske avløserord (kodet henholdsvis *ag* og *na*), som redaktørene har brukt særlig som utfyllende eksempelstoff.

Til spesialordbøker som *Bevingede ord* og Kunnskapsforlagets *Fremmedordbok* har også redaktøren hentet stoff fra nyordsmaterialet til nye utgaver.

3 Representativitet

En har forsøkt å gjøre utvalget av aviser og tidsskrifter representativt ved å spre det på både riksdekkende aviser og på lokale aviser. Alle landsdelene, Østlandet, Sørlandet, Vestlandet, Trøndelag og Nord-Norge er representert. På samme måte har en prøvd å få til et variert utvalg av allmenne tidsskrifter: *Naturen*, *Tidsskrift for samfunnsforskning*, *Filosofisk Tidsskrift*, *Tidsskrift for teologi og kirke*, *Geografisk tidsskrift* o.fl. Men når en er avhengig av frivillige og ulønnede medarbeidere, kan en ikke stå for hardt på kravene om den ideelle representativitet.

I veiledningen til ekserptorene er det angitt noen retningslinjer for innsamlingen. Som nye ord regner en ord som ikke står i ordbøker som *Tanums store rettskrivningsordbok* eller andre nyere norske ordbøker. Nå kan en vanskelig forlange at ubetalte medarbeidere skal bruke mye tid på å slå opp i ordbøker for å sjekke om et ord er ordboksført; det er ikke sikkert de har disse ordbøkene en gang. Da er det naturlig at de streker under ord som de oppfatter som nye, og at de tilsatte ved Avdeling for leksikografi siler det materialet som kommer inn.

På samme måte vil nye betydninger eller nye bruksmåter av ord si mye for ekserptor. Og det vil til enhver tid være det som ekserptoren oppfatter som avvikende syntaktiske konstruksjoner, feil eller vakkende preposisjonsbruk han/hun vil streke under.

Jeg får legge til at det overfor medarbeiderne blir presisert at de ikke bare skal feste seg ved det som er "rart". De blir bedt om å ta med også f.eks. vanlige preposisjonsuttrykk og vanlige ord, særlig når de er brukt på en særlig opplysende og helst kortfattet måte, slik at de egner seg som ordbokssitat.

Men utvalget er basert på skjønn. Det som er avvikende for én, trenger ikke være det for en annen. Evner og kunnskap og hvor mye arbeid den enkelte er villig til å legge ned i materialinnsamlingen, varierer selvsagt mye. Det er derfor ikke vanskelig å finne svakheter metodisk sett ved en slik måte å drive materialinnsamling på.

Våren 1995 var det oppslag i media i Norge om en kvinneordbok som er under arbeid, og i den forbindelse ble det fokusert på at norske ordbøker mangler mange vanlige ord for mer typisk kvinnelige aktiviteter som håndarbeid, matlaging, kvinnelige klesmoter, sminke o.l. Det er pekt på at det ofte er menn som lager ordbøker, og det er menn som står for materialinnsamlingen.

Det første er ikke helt riktig, f.eks. har både *Bokmålsordboka* og *Nynorskordboka* også hatt kvinnelige redaktører. Men jeg har ikke problemer med å innrømme at det kan være skjevheter ved materialinnsamlingen. De fleste ekserptorene er menn, og det er rimelig å

tenke seg at mange av dem heller leser f.eks. teknisk stoff eller sportssidene i avisene og streker ord derifra enn fra mote- og matspalter. Dette er kritikk som vi må ta alvorlig og prøve å rette på i framtiden.

Men en slik materialsamling er ikke verdiløs fordi det kan reises metodiske innvendinger mot innsamlingen. Alt i alt tror jeg vi har greidd ganske godt å fange opp nye ord, betydninger, uttrykk og andre endringer som har gjort seg gjeldende i norsk allmennspråk de siste 25 åra. Men vi vil ikke få svar på hvor fullstendig eller eventuelt hvor mangelfullt materialet er uten å lage ett så fullstendig korpus som mulig for den samme perioden og sammenligne. Det blir det neppe aktuelt med det første. Problemet med representativitet har en forresten når en skal bygge opp store, elektroniske tekstkorpus også, men jeg innbiller meg at det tross alt er lettere å løse da enn når en er avhengig av frivillige ekserptorer.

4 Hvordan egner et slikt ekserpert materiale seg som grunnlag for ordbøker?

Jeg har sett på noen filer av det materialet som ble skrevet inn i 1994, til sammen nærmere 1000 ekserpter. Av disse valgte jeg vilkårlig ut 300 ved å ta de 20 øverste ordene på hver side fra en 15-siders utskrift av en liste over de ekserperte ordene.

Av de 300 ekserptene er det etter min vurdering ca. 85 (knappt 28%) som er interessante i ordbokssammenheng. Da tenker jeg på en vanlig ettbinds allmennspråklig ordbok. 63 (vel 21%) av dem vil bli vurdert som eventuelle nye oppslagsord, 19 (vel 6%) som nye betydninger av eksisterende oppslagsord, eventuelt som brukseksempel som viser betydningsnyanser, og 4 (vel 1%) har interesse, eller kan få interesse for normering. (Utvalget er for lite til at en kan legge stor vekt på disse prosentene.) Ikke alle ordene, betydningene eller betydningsnyansene er nye, men de er ikke med i 1993-utgaven av *Bokmålsordboka*. Heller ikke alle vil uten videre gå inn i ordboka, men de kan settes på venteliste mens en ser om de blir noe mer enn døgnfluer.

Her får vi nøye oss med noen eksempler fra de forskjellige gruppene.

Nye oppslagsord:

asylant (ordet er blitt svært frekvent i norsk etter at flyktninger, særlig fra det tidligere Jugoslavia, begynte å søke tilflukt i norske kirker for å unngå å bli sendt tilbake av norske myndigheter)

brevgiro (alle nordmenn med bankkonto har fått tilbud om brevgiro)

børstraktor (spøkefull betegnelse på de store privatbilene med firehjulstrekk som er blitt så vanlige)

frislipp (økonomisk frislipp: f.eks. om norske bankers hemningsløse utlånsvirksomhet i 80-årene)

kinaisering (= flytting av arbeidsplasser til lavkostland)

krematør (norske ordbøker har ikke ført opp noe nomen agentis til *kremere* og *kremasjon*)

lobbying (*lobbisme*, *lobbyvirksomhet* er ordboksført før)

pubertil (norske ordbøker har ikke noe adj. til pubertet, så det fyller et tomrom i leksikon. *Svenska Akademiens Ordlista* har *pubertal*, og vi har også belegg

på *pubertistisk* i nyordsmaterialet, så her trenger redaktøren å undersøke litt nærmere hvilken form ordet bør få i norsk)
tagge, tagger (spraye merker på fasader, trikker o.l.; person som tagger)

Nye betydninger og betydningsnyanser:

A4 (= ordinær, vanlig; i sms. *A4-barnehage*)
flyktning (i sms. som *avgiftsflyktning*, *skatteflyktning* = person som flytter til et annet land for å unngå avgifter og skatter)
på golvet (*de på salgsgolvet* = de vanlige selgerne; tidlige bare om folk som arbeider på fabrikkgolvet, til forskjell fra ledelse og funksjonærer)
harehjerte (= hjerteflimmer)
house party (= ungdomsfest i nedlagt fabrikklokale)
nullstille (overført, om person: *være nullstilt*)
spikre (= fastsette: *programmet er spikret*)
økse (= redusere, skjære ned: *økse budsjettene*)

Former som har interesse for normering:

føyk (pret. av *fyke*; i offisiell rettskrivning bare *fauk* el. *føk*)
spott (eng. = kort innslag eller opptreden; i offisiell rettskrivning *spot*)
et spise (= mat, kost; i offisiell rettskrivning *en spise*)
et eksistensial (bare ordboksført i fl. *eksistensialer*, og bare i *Norsk fremmedordbok*)

Kaster vi et blikk også på resten av de ca. 1000 ekserptene fra 1994, finner vi bl.a. noen forledd og etterledd som er særlig produktive.

I dette lille materialet fins det f.eks. 23 sammensetninger med *miljø-* som forledd: *-ansvar*, *-bølge*, *-charter*, *-evaluering*, *-forbedring*, *-hazop*, *-hensyn*, *-innpakning*, *-konsekvens*, *-krav*, *-ledelse*, *-merke*, *-merking*, *-parameter*, *-profil*, *-proteksjonisme*, *-reklame*, *-revisjonskompetanse*, *-riktig*, *-standard*, *-strategi*, *-ulempe*, *-vennlig*.

Nå mener jeg ikke at alle disse sammensetningene skal inn i ordboka, men ord som *miljøbølge*, *-hensyn*, *-innpakning*, *-krav*, *-merking*, *-riktig*, *-ulempe* vil nok bli vurdert i forbindelse med neste utgave. Av disse 23 sammensetningene er det bare *miljøvennlig* som står i *Bokmålsordboka* 1993.

Etterleddet *-kåt* (= sugen på, som har lyst til) er også produktivt: *bade-*, *demokrati-*, *innsparings-*, *markeds-*, *påske-* (*påskekåt aprilsol*), *sol-* (*solkåte nordmenn*), *utlåns-*.

I *Bokmålsordboka* 1993 er det bare *PR-kåt* som er med, og det er neppe noen av disse sammensetningene her som vil få plass i ordboka. Men vi bør få med denne utvidede betydningen under oppslagsordet *kåt*.

Ordet *park* har også utvidet sitt betydningsområde de siste årene (= område for en spesiell aktivitet), og vi finner den i sammensetninger som: *kunnskaps-*, *miljø-*, *nærings-* og *travpark*. Dette må nedfelles i ordboka både ved at en tar med noen sammensetninger av denne typen, og at en fører opp denne utvidede betydningen under *park*.

Når en arbeider med dette materialet, ser det både rikt og mangfoldig ut. Men problemet er at du ser hva du har, men ikke det du eventuelt mangler. Det er viktig å være klar over

de begrensningene et slikt ekserpert materiale har, og ikke gå ut fra at det forteller hele sannheten om utviklingen i ordforrådet.

Et eksempel: I dette materialet jeg har omtalt her, finner vi tre ord som betegner musikkformer innenfor vår tids ungdomskultur: *ambient*, *etno* og *hardcore*. Det fins mange flere slike musikkbetegnelser, noen har vi nok i nyordsmaterialet, andre ikke. For å oppfylle kravet om konsistens i ordboka, må en i tillegg finne fram til faglitteratur eller kontakte eksperter, f.eks. folk som dekker det som skjer på denne fronten, i avisene, slik at en får med i alle fall de viktigste betegnelse innenfor området.

5 Hva slags materiale er det ordboksredaktøren legger til side?

En god del av ekserptene er tatt med på grunn av syntaktiske særtrekk, og de har i mindre grad direkte interesse for ordboksredaktøren (jf. kodegruppe 11).

Svært mange av sammensetningene er tilfeldige sammensetninger laget for behovet i øyeblikket, og vil ikke bli tatt med i vår ordbok, f.eks. *bilvrakdyng*, *bondesøppel*, *brukt-boksalg*, *bunadsaktivist*, *budsjettkaos*, *forsikringsløs* (*forsikringsløse biler*), *en frida* (etter en populær norsk film om ei jente som heter Frida), *neonhumør*, *oberstbesøk*, *OL-skapt* (*OL-skapte arbeidsplasser*), *sannhetskommisjon*, *valgkampopplagt*.

Noen av ekserptene er fagord som blir for spesielle. Eks.: *brannflekkyske*, *porselenssyke* (betegnelser for krepsesykdommer).

Noe stoff fra lokalavisene er såpass lokalpreget (dialektord o.l.) at det ikke har interesse for en allmennspråklig standardordbok, f.eks. *bøvelsbygder* (nedsettende betegnelse på folk fra en spesiell bygd på Vestlandet), *noppemjøl* (lokal betegnelse på byggmjøl).

Noen av ekserptene er allerede ordboksført, f.eks. *etterbruk*, *blind alarm*, *bohemer* (i utvidet bruk: *bohemer i politikken*), *brikkene faller på plass*. Men de kan jo tjene som en bekreftelse på at redaktøren gjorde rett da han/hun tok inn ordet, betydningen eller brukseksemplet i ordboka.

Men dette materialet som ikke umiddelbart er så relevant i ordbokssammenheng, har selvfølgelig interesse for språkforskere i andre sammenhenger.

6 Konklusjon

Nyordsmaterialets bidrag til ordbøkene består i en komplettering når det gjelder nye ord, nye betydninger og betydningsnyanser, ny fraseologi og gode eksempler på hvordan ord brukes. Når en arbeider med et slikt ekserpert materiale, må en være klar over de begrensningene det har. Selv om det inneholder mye informasjon om utviklingen av ordforrådet i norsk, er det ikke fullstendig. Og jeg stiller meg stadig spørsmålet: Hva har vi gått glipp av sammenlignet med våre nordiske korpuskolleger?

Litteratur

Berulfsen, Bjarne/Dag Gundersen (red.) 1986: *Fremmedordbok*. Oslo: Kunnskapsforlaget.

Guttu, Tor 1993: *Norsk illustrert ordbok*. Oslo: Kunnskapsforlaget.

Kristensen, Kjeld 1994: Den Danske Ordbogs tekstkorpus og spORDhunde. I: Anna Garde/Pia Jarvad (red.): *Nordiske studier i leksikografi* II. København: Nordisk forening for leksikografi, 138–142.

Landrø, M. I./B. Wangenstein 1993: *Bokmålsordboka*. Oslo: Universitetsforlaget.

Anvendelse af korpora i opbygningen af ordbøger til maskinoversættelsessystemet METAL

This paper presents some of the results that the Danish METAL group has achieved as to the production of lexicons for the MT system METAL, and especially our recommendations regarding the use of various documentation categories.

Existing MT lexicons from the German-English language pair, lists of terms for inhouse usage, as well as domain specific corpora form the basis of the lexicons. From experience we find that lexicons should mainly consist of words and terms extracted from technical corpora.

However, technical documentation often consists of words that should not be entered into lexicons; this might be due to terminological inconsistency, the formation of ad hoc compounds, etc. I close this paper by dealing with some of the lexical problems we had to cope with when translating a specific technical text.

1 Indledning

I denne artikel vil jeg gøre rede for, hvordan vi ved Handelshøjskole Syd, DK har opbygget ordbøger til maskinoversættelsessystemet METAL, og hvordan vi har bearbejdet de kilder, der har dannet grundlaget for vore maskinoversættelsesordbøger (herefter anvendes forkortelserne MO-system og MO-ordbog). Vi har forsøgt at tillempe terminologiske arbejdsmetoder og principper i ordbogsopbygningen til et morfo-syntaktisk MO-system. Nøgleordene i artiklen er således fagsprog, terminologi og maskinoversættelse.

2 Ordbøger i METAL-systemet

I den danske METAL-gruppe har vi udarbejdet en prototype for et tysk-dansk modul til MO-systemet METAL; det blev færdigt i 1994. Mine erfaringer bygger hovedsageligt på ordbogsarbejdet, hvor jeg bl.a. har arbejdet med en bilteknisk ordbog.

Ordbogsopbygningen i METAL, der er opbygget efter transfermodellen, er skitseret i følgende figur:

<i>Tysk monoordbog</i>	<i>Tysk — Dansk transferordbog</i>	<i>Dansk monoordbog</i>
Auto NST	Auto NST ⇒ bil NST	bil NST

Som det fremgår, har METAL to monoordbøger for vores sprogpår, en tysk og en dansk, og desuden én transferordbog, her tysk-dansk. Substantivet *Auto* vil således være indlæst i den tyske monoordbog, ligningen *Auto* ⇒ *bil* i transferordbogen og *bil* i den danske monoordbog.

Monoordbøgerne indeholder, som det ses i eksemplet *kobling*, primært oplysninger om et ords bøjninger:

Ensproget ordbog

TRÆK	VÆRDI	FORKLARING
(„kobling“	NST	indgangsord, ordklasse
ALO	„kobling“	stamme
CL	(P-ER S-EN)	bøjninger
GD	(C)	grammatisk køn
LINK	S	fugebogstav for komposita
SX	(N)	naturligt køn
TYN	(CNC)	semantisk træk
...)		

Transferordbøgerne indeholder ordets egentlige betydning i form af ordligninger, der som et minimum omfatter oplysninger om grundform og ordklasse for både udgangs- og målsprog samt fagområde. Det ses i ordbogsartiklen for *Kupplung*, der ligger i den biltekniske ordbog (ME-CAR):

Tosproget ordbog

„Kupplung“ NST ⇒ „kobling“ NST
 Pref S.0.0.00 Tag (ME-CAR)
 <<HHS Lw 29-Apr-91>>

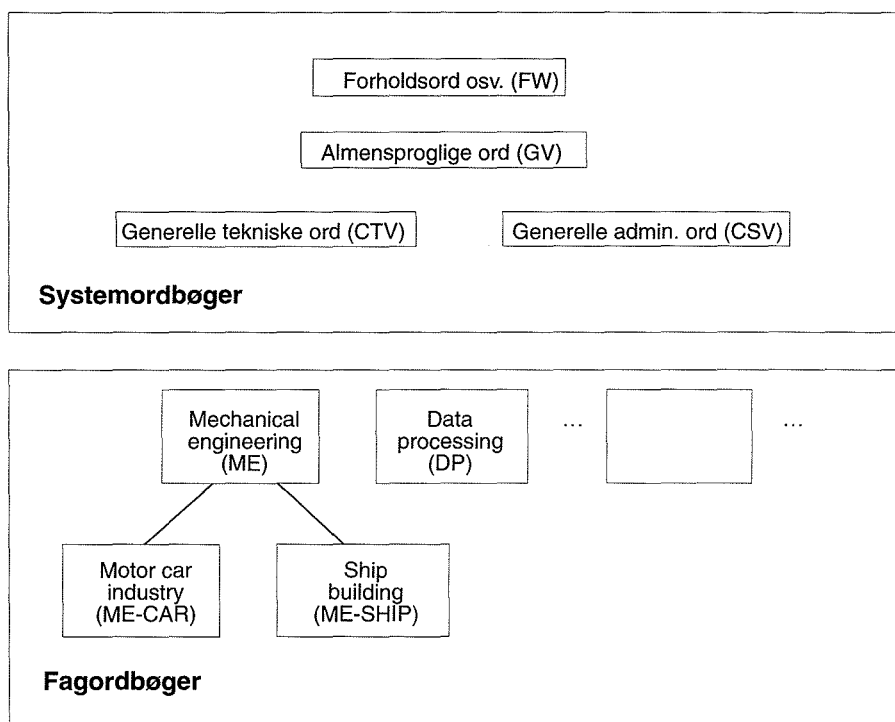
Transferordbøgerne i METAL er hierarkisk opbygget og består overordnet af en gruppe af systemordbøger, bl.a. en almensproglig ordbog og en generel teknisk ordbog, samt en gruppe af fagspecifikke ordbøger. Ved at inddele transferordbøgerne i fagspecifikke ordbøger opstår der en systematisk makrostruktur, der kan være mere eller mindre detaljeret.

3 Makrostruktur som hjælp til disambiguering

Hvis en term har flere betydninger, kan en human oversætter vha. definitioner og forklaringer finde frem til den rigtige oversættelse.

Den mulighed har maskinen ikke! Derfor må man i ordbøgerne anføre oplysninger, der er så entydige, at maskinen kan foretage det rigtige valg.

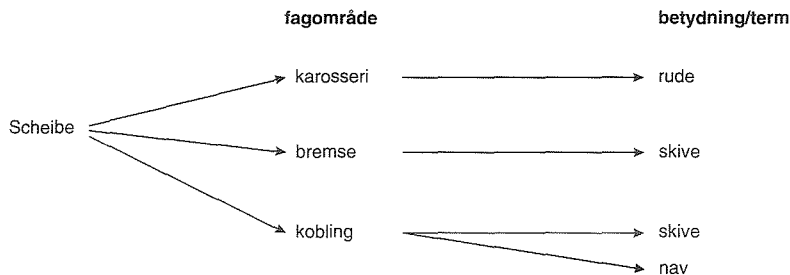
Angivelse af fagområde i transferartiklerne er én måde at hjælpe maskinen til at disambiguere korrekt. Det vil med andre ord sige, at mulighederne for disambiguering også afhænger af den valgte makrostruktur.



Transferordbøger i METAL

I den hierarkiske struktur i transferordbøgerne er **systemordbøgerne** med almene og almentekniske ord placeret øverst. Nedenunder findes **fagordbøger**; de kan være underinddelt, f.eks. er den maskintekniske ordbog inddelt i en ordbog for biler og én for skibe. Inden en tekst sendes til oversættelse, angives den mest specifikke af de ordbøger, systemet skal slå op i; for de biltekniske tekster er det naturligvis den biltekniske ordbog. Dér starter systemet sin søgning. Hvis den søgte term ikke findes i den biltekniske ordbog, bevæger systemet sig op på næste niveau, dvs. i den overordnede maskintekniske ordbog (ME), og så fremdeles.

Som den biltekniske ordbog ser ud i dag, vil *Scheibe* blive oversat med alternativerne *rude/skive/nav*:



Men som det ses af ovenstående figur, vil problemet til dels kunne løses ved at underinddele den biltekniske ordbog i flere mindre ordbøger med hver sit delområde, her karosseri,

bremse og kobling. Inden for delområdet kobling vil der dog stadigvæk være alternativerne *skive* og *nav*.

Kilderne og de fagområder, der behandles i dem, vil være bestemmende for, hvordan den endelige systematik i transferordbøgerne kommer til at se ud. Men grunden til, at vi opbygger ordbøgerne med en systematisk makrostruktur er den søgestrategi, METAL bygger på. Ud fra en terminologisk betragtning, kan man sige, at det er første skridt på vejen mod begrebsystemer.

4 Kilder anvendt i opbygningen af METAL-ordbøgerne

I dette kapitel skal jeg kort skitsere, hvordan vi har brugt vores kildemateriale. Vi havde nok en ide om, hvordan den ideelle vej så ud, men bl.a. fordi vi er en del af et internationalt projekt, har vi undertiden været nødt til at gå anderledes til værks, end det, der efter vores mening, var mest hensigtsmæssigt.

4.1 Systemordbøgerne

Til opbygningen af systemordbøgerne for det tysk-danske modul tog vi først udgangspunkt i de eksisterende METAL-systemordbøger fra det tysk-engelske sprogpar. — Det viste sig dog, at ordbøgerne efter vores opfattelse var for store og emnemæssigt for brede.

Vore erfaringer i dag peger i retning af, at man kan tilvejebringe systemordbøger mere målrettet ved at opbygge dem på baggrund af ordenes eller fagordenes frekvens i fagsproglige tekstkorpora end ved at overtage eksisterende systemordbøger fra andre sprogpar. **Korpusbaserede systemordbøger** vil komme til at indeholde ord, der optræder på samme måde i mange forskellige typer tekniske fagtekster, og som altid oversættes ens. Da vores mål for vores ordbøger er kun at optage de ord, der er relevante for maskinoversættelse af tekniske tekster, vil vi for det dansk-tyske modul afprøve en metode, hvor systemordbøgerne så godt som udelukkende er baseret på tekstkorpora. Vi håber på den måde at opnå en bedre oversættelses kvalitet og bedre hastigheder.

Vi har således fokuseret på indholdet i system- eller basisordbøgerne, og ikke som de fleste på deres størrelse.

4.2 Fagspecifikke ordbøger

Udviklingen af de fagspecifikke ordbøger har overvejende været korpusbaseret. I modsætning til systemordbøgerne har det lige fra starten været vores strategi, at det vigtigste grundlag for de fagspecifikke ordbøger var de **fagsproglige korpora**, vi har arbejdet med i forbindelse med systemudviklingen. Vi har dermed skullet tage højde for alle de termer og termstrukturer, der fandtes i vore korpora.

Det lyder ideelt og dermed burde lemmaselektionen eller termselektionen, som jeg vil kalde det, være klart afgrænset, men fagtekster indeholder ofte ord, som man ikke vil og heller ikke bør medtage i ordbøgerne. Det vil jeg komme tilbage til.

4.2.1 Kildemateriale til den biltekniske ordbog

For det tysk-danske modul har jeg primært arbejdet med terminologien inden for det biltekniske område.

Den biltekniske ordbog er den største af de fagspecifikke ordbøger, vi har opbygget. Den er tillige den af vores fagspecifikke ordbøger, der er baseret på det mest alsidige kildemateriale både mht. kildekategorier og kvalitet.

Grundlaget for den biltekniske ordbog har primært været tysk undervisningsmateriale for mekanikere og en reparationsvejledning samt sekundært en virksomhedsintern ordliste.

Undervisningsmaterialet er et parallelt korpus, der foreligger på tysk som udgangssprog og i dansk oversættelse. Da fagfolk har efterredigeret den danske oversættelse meget grundigt, har vi været i en ideel situation, hvor kildematerialet på begge sprog er formuleret præcist på såvel sætnings- som termniveau. Sproget i disse tekster opfylder kravene til fagsproglige tekster om entydighed, terminologisk konsistens og sprogøkonomi.

Kvaliteten i reparationsvejledningen, der kun forelå på tysk, er ikke på højde med undervisningsmaterialet. En sidste og meget vigtig kilde i oversættelsesarbejdet har været fagfolkene selv.

Fra reparationsvejledningen har vi oversat kapitler inden for forskellige delområder, som også er behandlet i undervisningsmaterialet, f.eks. motor, køler, udstødning, bremses. Ved at behandle flere mindre områder fra begge korpora, kunne vi opnå en god terminologisk afdækning af et antal klart afgrænsede delområder og samtidig en vis bredde i ordbogen.

5 Oversættelsesopgave fra BMW

Vi har for nylig lavet en oversættelsesopgave for BMW. På baggrund af den skal jeg nu prøve at eksemplificere det praktiske ordbogsarbejde i forbindelse med maskinoversættelse. Mit første indtryk var, at teksten fra BMW med sine korte, overskuelige sætninger ville være velegnet til maskinel oversættelse.

Et af emnerne på denne konference er, hvordan de enkelte kilder påvirker lemmaselektionen. Det skulle vise sig, at denne tekst er af en kvalitet, der ikke bør overføres til ordbøgerne, og langt fra den kvalitet, der specielt var kendetegnende for undervisningsmaterialet.

Nogle af de forhold, der har vanskeliggjort termselektionen er:

- at teksten er skrevet i **versaler**
- at den indeholder mange **inkonsekvente forkortelser**
- at teksten er præget af andre **inkonsekvente stavemåder**
- at teksten er præget af **synonymi**
- at der er anvendt mange **engelske låneord**
- at teksten er skrevet i **telegramstil**

Jeg skal i det følgende gennemgå de nævnte forhold nærmere, og se på, hvordan vi så vidt muligt har undgået at overføre disse „unoder“ til ordbøgerne. En stor del af dette arbejde består i at forbedre kildeteksten ved manuel eller automatisk preeditering.

Først er det dog vigtigt at slå fast, at nogle termer er bedre egnede til at kommunikere en teksts indhold videre end andre. I valget af termer følger vi de regler, man i terminologilæren har opstillet for termdannelse.

Til de vigtigste **termkrav** hører, at termer skal være motiverede, at de skal være afledningsproduktive og følge det almindelige sprogs syntaktiske regler.

Det har vist sig, at der kan opnås en bedre kvalitet i oversættelserne, hvis kravene følges. Nogle krav vil med maskinoversættelse endda kunne overholdes og opfyldes bedre og mere konsekvent, end humane oversættere har gjort det.

Det er især ved vores valg af de danske ækvivalenter disse termkrav har betydning, eftersom den tyske terminologi på forhånd er givet i teksten. Men i det omfang det er muligt, inddrages termkravene også for den tyske dels vedkommende.

5.1 Versaler

Teksten er skrevet i versaler. Vi ønsker naturligvis ikke at optage termerne i ordbøgerne i denne form. METAL kan ikke uden videre håndtere hele tekster, der er skrevet i versaler. Men systemet behandler enkeltord i versaler som konstanter, der ikke skal oversættes. Det vil oftest være firma- eller produktnavne, der netop skal overtages direkte i oversættelsen.

Vi konverterede først dokumentet til tekstbehandlingsprogrammet Word. I Word er der mulighed for automatisk at ændre versaler til små bogstaver. Da de fleste ord var substantiver, udnyttede vi en facilitet, der ændrer versalerne til små bogstaver på nær det første bogstav, som i overensstemmelse med den tyske retskrivning for substantiver skrives med stort.

5.2 Inkonsekvente forkortelser

Forfatteren har af edb-tekniske pladshensyn anvendt flere inkonsekvente forkortelser. Eksempler på disse forkortelser er:

- (1) *Bremsflüssigkeitsbeh.* for 'Bremsflüssigkeitsbehälter'
- (2) *Zylinder Sch.Bügel* for 'Zylinderschutzbügel'
- (3) *Z.-Vert.* for 'Zündverteiler'
- (4) *Geri.* for 'gerissen' i konteksten *Naht Aufgeg., Geri.*

Som det ses, er der stor forskel i gennemskeligheden af disse forkortelser. Ofte hjalp det mig, at der var tale om en terminologi, jeg før havde beskæftiget mig med. Eksempel (1) *Bremsflüssigkeitsbehälter* hører til de mere gennemskelige. Eksemplerne (2)–(3) kunne jeg finde i deres fulde form andre steder i teksten ved at søge på dele af forkortelserne og sammenligne deres kontekst, andre kunne jeg gætte mig til ud fra konteksten. Eksempel (4) *Geri* med stort G voldte en del besvær, men også her hjalp konteksten *Naht Aufgeg., Geri* mig til sidst. Det store G var en rest fra versalerne i det oprindelige dokument. Da *gerissen* forekom over 300 gange skrevet ud, var det ikke umiddelbart indlysende, at der var tale om en forkortelse for *gerissen*.

5.3 Inkonsekvente stavemåder

I teksten er der flere inkonsekvente stavemåder i forbindelse med binde- eller fugestreg i komposita, og der er eksempler på anvendelse af flere mulige ortografiske varianter.

Mange tyske **komposita** ses skrevet både med og uden bindestreg. For oversættelses-systemet opfattes disse som to forskellige termer, ligesom det også vil være tilfældet ved forskellige fugebogstaver. Da vi ikke ønsker at have redundans i vore ordbøger, har vi besluttet os for en strategi, hvor vi principielt kun indlæser formerne uden bindestreg.

- (5) a *Tankentlüftungs-Ventil* (4)
- b *Tankentlüftungsventil* (1)
- (6) a *Kupplungs-Seite* (1)
- b *Kupplungsseite* (1)
- c *Kuppl.-Seite* (1)

Tankentlüftung i eksempel (5)a og (5)b optræder fire gange med bindestreg og en gang uden. Men det er mit generelle indtryk, at anvendelsen af bindestreger er ret tilfældig.

Termen *Kupplungsseite* i eksempel (6) forekommer både med og uden bindestreg samt i forkortet form; de er alle hver især realiseret en gang. Denne term er blot et af mange eksempler på, at flere af de nævnte forhold kan optræde side om side.

- (7) a *Blasbildung* (11)
- b *Blasenbildung* (40)

I eksempel (7) *Blasbildung/Blasenbildung* er termerne dannet med og uden fugebogstaver. *Blasbildung* optræder i alt 11 gange i teksten, mens *Blasenbildung* optræder 40 gange; noget taler altså for at vælge denne form. Frekvensoplysningerne fremgår af nogle for-analyselister, som METAL genererer, og som er udgangspunkt for leksikonarbejdet.

5.3.1 Metal Pattern Matcher (MPM)

Til METAL-systemet hører forskellige hjælpeværktøjer. Et af disse hedder Metal Pattern Matcher; det er et **mønstergenkendelses- og batch-hjælpeprogram**, der automatisk kan redigere en tekst enten før eller efter oversættelse.

Dette hjælpeværktøj kan i kildefilen erstatte de stavemåder, som vi ikke ønsker at optage i ordbøgerne, med den rigtige stavemåde, dvs. den stavemåde vi ønsker at optage i ordbøgerne.

Det anføres i en fil, at *Blasbildung* — fra eksempel (7)a — hver gang det forekommer, skal læses som (7)b *Blasenbildung*, der er den term, der er indlæst i ordbøgerne.

5.4 Synonymer

Teksten er som nævnt præget af synonymer.

- (8) a *Anhängermodul*
- b *Anhängemodul*
- (9) a *Antriebseinheit*
- b *Antriebs-Einheit*
- c *Antriebsteil*

Nogle af disse synonymer har vi valgt at behandle ved hjælp af Metal Pattern Matcher. Det er synonymer af typen som vist i eksempel (8)a *Anhängermodul* og (8)b *Anhängemodul*. Der er tale om en slags ortografiske varianter.

Synonymer som i eksempel (9) *Antriebseinheit/Antriebsteil* optages som selvstændige artikler i vores ordbøger, da de som i eksemplet kan have forskelligt køn og dermed forskellig bøjning af artikler og adjektiver mv. Det kan der ikke tages højde for i Metal Pattern Matcher.

5.5 Engelske låneord

I BMW-teksten er der anvendt en del engelske låneord som *Airbag*, *Kickdown*, *Memory*. Min umiddelbare holdning var, at jeg måtte undgå at optage disse anglicismer i ordbøgerne. Det skyldtes bl.a., at de engelske låneord også skulle indlægges i den tyske monoordbog, som den tyske METAL-gruppe har ansvaret for. Det var efter min opfattelse den tyske gruppe, der skulle afgøre, om den kritikløst ville overtage de engelske låneord.

På det tidspunkt stødte jeg på en artikel om engelske låneord i tyske bildokumenter af Sverre Vesterhus (1991). I sin undersøgelse kommer Vesterhus frem til, at de tyske bilmærker alle ligger under den gennemsnitlige frekvens for engelske låneord på 0,85 pr. side, dog med en undtagelse, nemlig BMW.

Gennemsnitlig frekvens	0,85
BMW	1,07
Ford	0,77
Opel	0,74
Audi	0,67
Porsche	0,41
Volkswagen	0,41
Mercedes-Benz	0,39

(Vesterhus 1991)

For de tyske bilmærker tegnede BMW sig med 1,07 for den højeste frekvens for engelske låneord, mens Mercedes-Benz lå lavest med 0,39. — På baggrund af disse tal er jeg efter min mening nødt til at acceptere de engelske låneord og optage i hvert fald nogle af dem i ordbøgerne. Det er interessant, at de engelske låneord ikke kun optræder i én teksttype, nemlig reklamematerialet, men også i BMWs mere interne fagdokumentation, hvorfra låneordene sandsynligvis endda stammer.

Mange engelske låneord findes endnu ikke i tyske ordbøger, og for flere mangler der endnu et bøjningsmønster, fordi lånene er relativt nye. I maskinoversættelse skaber det problemer mht. de grammatiske oplysninger, som skal anføres i den tyske monoordbog aht. sætningsanalysen.

Og i BMW-teksten er der anvendt en telegramstil, som bevirker, at det oftest ikke er muligt at udlede køn eller bøjningsformer.

Nogle af de engelske låneord har jeg optaget i ordbøgerne, mens jeg har valgt at behandle andre med Metal Pattern Matcher. Jeg har i det følgende grupperet nogle af låneordene efter, hvordan jeg har behandlet dem.

Da *Recall Campaign* i eksempel (10) også forekommer som oversættelseslånet *Rückrufaktion* har jeg i Metal Pattern Matcher erstattet *Recall Campaign* med *Rückrufaktion*.

Recall Campaign bliver således slet ikke optaget i ordbøgerne, men registreret i en MPM-fil.

(10) *Recall Campaign* → *Rückrufaktion*

Metal Pattern Matcher indeholder også en anden del, hvor man i en fil kan anføre ord, som systemet skal markere som konstanter. På den måde ved systemet, at det ikke skal forsøge at oversætte ordene. Det vil som nævnt meget ofte være produktnavne. I eksemplerne (11)–(12) har den fagmand, jeg har arbejdet sammen med, opgivet, at *Park Distance Control* og *Interlock* skal overtages til dansk.

(11) *Park Distance Control* → *Park Distance Control*

(12) *Interlock* → *interlock*

Da jeg ikke umiddelbart har kunnet finde de nødvendige grammatiske oplysninger, har jeg valgt at behandle disse låneord i Metal Pattern Matcher, indtil jeg har fundet de manglende grammatiske oplysninger.

(13) *Fluidbloc* → *hydroleje*

(14) *Silentblock* → *gummibøsning*

Eksemplerne (13) *Fluidbloc* og (14) *Silentblock* har voldt mig de største problemer. Da disse låneord ikke overtages direkte til dansk, men har danske ækvivalenter, er jeg tvunget til at optage dem i ordbøgerne. *Fluidbloc* oversættes til *hydroleje* og *Silentblock* til *gummibøsning*. Jeg har valgt at kode dem med det køn og bøjningsmønster, jeg anså for mest sandsynligt, og har så i et kommentarfelt i ordbogsartiklerne anført, at kodningen senere skal efterredigeres. På grund af telemgramstilen er artikler ofte udeladt i teksten, derfor vil en forkert ordbogsoplysning om artikelbrug ikke komme til at genere sætningsanalysen i denne oversættelse.

5.6 Telegramstil

I BMW-teksten forekommer der en form for reduceret artikelbrug, der skaber en slags **pseudo-komposita** eller pseudo-termer, som vi ikke ønsker at optage i ordbøgerne. Nogle eksempler er *Frontverkleidung-Ziergitter* og *Korrektur-Buchung*:

(15) a *Frontverkleidung-Ziergitter*

b *Frontverkleidung des Ziergitters*

c *Frontverkleidung am Ziergitter*

(16) a *Korrektur-Buchung*

b *Buchung der Korrektur*

De fleste eksempler er reducerede genitivkonstruktioner som *Frontverkleidung-Ziergitter*, der i sin komplette form er *Frontverkleidung des Ziergitters* eller *Frontverkleidung am Ziergitter*. Genitiv-artiklen eller præpositionen 'an + artikel' erstattes af en bindestreg. Derved bliver der dannet nogle konstruktioner, der minder om komposita og derved får status som selvstændige termer uden at være det.

Eftersom der ikke er tale om regulære termer, bør man så vidt muligt undgå at optage dem i ordbøgerne. I det omfang det er muligt, vil jeg altså lade dem oversætte kompositionelt, hvor systemet oversætter *Frontverkleidung* for sig og *Ziergitter* for sig og derefter sammenføjer dem til et kompositum på målsproget.

Men er der tale om udtryk, hvor strukturen er forskellig på udgangs- og målsprog, er man tvunget til at optage dem i ordbøgerne for at kunne foretage de nødvendige strukturelle ændringer.

6 Konklusion

Maskinen har ingen viden om verden og derfor ikke samme muligheder for at vurdere en ordbogsartikel som en person. Derfor er det uhyre vigtigt, at de data, der indlæses i ordbogsartiklerne i et MO-system, er korrekte og præcise.

Desuden er vi i MO-ordbøgerne nødt til at tage højde for hver eneste af de ord og termer, der findes i en tekst, der skal maskinoversættes. Opbygningen af vores systemordbøger vil fremover ganske vist være korpusbaseret, men som i traditionelt leksikografarbejde kan vi så at sige springe et udtryk over. Men i forbindelse med konkrete maskinelle oversættelsesopgaver er der ingen vej udenom. Der er man nødt til leksikalsk at bearbejde korpus ned til mindste detalje. Vores erfaringer med fagtekster er, at de ofte lader noget tilbage at ønske. BMW-teksten er et af de værre eksempler, men ikke enkeltstående.

Endelig har vores arbejde med MO-oversættelse vist, at det i meget høj grad er ordbøgerne, der bestemmer, hvor god en maskinoversættelse bliver.

Mine eksempler fra BMW-teksten viser, at det er vigtigt at behandle teksten og korpora — i det hele taget — systematisk og lægge strategier for, hvordan forskellige problemer skal håndteres. — Det nævnte foranalyseprogram er én hjælp til at behandle tekster systematisk, men er langt fra tilstrækkeligt, f.eks. giver det ikke nogen oplysninger om semantiske relationer, hvilket er vigtigt for at kunne opbygge en konsistent terminologi i det mindste for målsprogets vedkommende og for at kunne placere termerne i de rigtige fagspecifikke ordbøger, hvilket har betydning for makrostrukturen.

Den systematiske gennemgang af BMW-teksten har vist, at det var nødvendigt med ganske meget pre-editering, hvis ordbøgerne bagefter skulle have en kvalitet, der var til at leve med. For at kunne oversætte teksten er den oprindelige tekst blevet en hel del forbedret, enten manuelt eller automatisk vha. Metal Pattern Matcher. Det vil med andre ord sige, at ordbogsarbejdet har været med til at forbedre kvaliteten i kildeteksten. Når der er tale om maskinoversættelse, kan man derfor også vende problematikken om og spørge: Hvordan påvirker ordbogsarbejdet kildematerialet? Men man kan og skal selvfølgelig ikke omskrive hele teksten; det er klart! Hvis det er nødvendigt, skal en tekst ikke maskinoversættes.

Bibliografi

- Vesterhus, Sverre A. 1991: Anglicisms in German car documents. I: *Language International* 3.3, 10–15.
- Weilgaard Christensen, Lotte 1992: Ordbøger i METAL. I: Henrik Holmboe (ed.): *Skriften På Skærmen*; Datalingvistisk Forening, Årsmøde nr. 2, Handelshøjskole Syd, 13.–14. maj 1992. Århus: Afdeling for Datalingvistik, Handelshøjskolen i Århus, 31–41.
- Weilgaard Christensen, Lotte 1993: Ordbøger i METAL — Anvendelse af terminologiske principper til opbygning af ordbøger i METAL. I: Bertha Toft (ed.): *Terminologi, edb & videnssteknik*, 2. nordisk symposium 1992, Varde. Kolding, 202–221.

- Weilgaard Christensen, Lotte 1994: Termvalg i fagtekster og maskinoversættelsesordbøger. I: *Jubilæumsskrift for HHS*. Kolding, 135–149.
- Weilgaard Christensen, Lotte/Ellen Christoffersen 1993: Oversættelse med METAL. I: *Sprog & Erhverv, Erhvervsprogligt Forbund 7–8/93*, 13–15.
- Weilgaard Christensen, Lotte/Ellen Christoffersen/Gert Engel 1991: *Aufbau und Wartung der METAL-Wörterbücher* (internt papir).
- Weilgaard Christensen, Lotte/Ellen Christoffersen/ Margrethe M. Møller 1991: Ordbøger i METAL. I: *Handelshøjskole Syd Forskning, Temanummer 3 af HHS-Information om erhvervsproglig forskning ved HHS*. Kolding, 6–7
- Weilgaard Christensen, Lotte/Ellen Christoffersen/ Margrethe M. Møller 1994: Metoder og strategier i et maskinoversættelsesprojekt. I: *Festskrift for Gert Engel*. Kolding, 233–247.

Redigeringsystemet RiSt **— udviklet og anvendt i et færøsk ordbogsmiljø**

The objective of this paper is to present a computer program for dictionary editing. RiSt, as the program is called, is used in some dictionary projects in the Faroe Islands. The program enables the lexicographer to write and edit dictionary entries in a structural manner applying elements from a predefined set of categories and use familiar editing tools known from text processing programs. The entry elements are shown in category-dependent type styles, and RiSt automatically inserts punctuation marks between the entry elements. Optional input control can be placed on selected categories according to tables with preset or ongoing defined values. The user can pick out a selection of dictionary entries for printing and export (in a tag-format) according to search conditions including for instance word occurrence and user name. RiSt supports PostScript-printers, and the user can adjust some lay-out parameters to get various printing setups. The RiSt system has been developed as a joint project involving institutes at Fróðskaparsetur Føroya (University of the Faroe Islands) and the company Z-teld.

1 Indledning

I det følgende vil der blive givet en præsentation af et ordbogsredigeringsystem, som anvendes i forbindelse med udarbejdelsen af den — i øvrigt første — færøske modersmålsordbog. Redigeringsystemet går under navnet RiSt, som er afledt af ordet *ritstjórn*, det færøske ord for redigering. Systemet har også været anvendt ved udarbejdelse af en ny udgave af dansk-færøsk ordbog, der forventes at udkomme ved årsskiftet 1995/96. Desuden har et dansk forlag købt et kopi af RiSt.

2 Edb-systemer til ordbogsredigering

Den kraftige udvikling inden for edb-teknologi de seneste 10–15 år har medført en voksende anvendelse af edb-systemer til opbevaring og redigering af tekst. Det fremgår af en oversigt over ordbøger i Danmark fra 1987 (DANLEX-gruppen 1987a), at der anvendes edb i langt de fleste ordbogsprojekter, hvoraf der i en række projekter anvendes specialsystemer, medens man i andre projekter anvendte en tilpasset udgave af et standardprogram til tekstbehandling eller databaseadministration.

PC'ernes fremmarch siden midten af firserne har givetvis forstærket anvendelsen af standardprogrammer i ordbogsmiljøer, men samtidig har brugere savnet faciliteter, som har betydning for en ordbogsredaktør, men som ikke forefindes i et standardprogram til tekstbehandling eller databaseadministration.

2.1 Tekstbehandlings- og databaseprogrammernes begrænsninger

Tekstbehandlingssystemer indeholder en række faciliteter på tegn- og blokniveau (indskyde, slette, flytte, skriftsnit m.v.), som gør dem velegnede til at skrive og redigere løbende tekst, så som breve og dokumenter. Ønsker man en automatisk strukturering af teksten, bl.a. i form af hierarkisk opbyggede elementer og alfabetisering af nøgleord, er der sædvanligvis ikke meget at hente i de gængs tekstbehandlingssystemer. Nogen hjælp kan man få ved selv at udvikle makroer, hvorved der kan udføres en række basisoperationer i et enkelt makrokald. Denne mulighed har givetvis været anvendt af mange ordbogsredaktører, specielt i tilfælde, hvor der har været tale om relativt simple ordbogsstrukturer.

Behovet for automatisk alfabetisering og fast artikelstruktur bliver bedre imødekommet i et databasesystem end i et tekstbehandlingssystem. Af den grund har der været en række forsøg med at anvende databasesystemer til ordbogsredigering. Eksisterende redskaber til forespørgsler i databasen og rapportering giver redaktøren muligheder for bl.a. at foretage udtræk af artikler og automatisk styring af skriftsnit, hvilket er svært at udføre i et sædvanligt tekstbehandlingssystem. På den anden side: i tekstbehandlingssystemer er redigeringsfaciliteterne på tegn- og blokniveau som regel mere alsidige end i databasesystemerne. *Maksimal postlængde og fast antal/rækkefølge af kategorier* begrænser på afgørende vis anvendelsen af standard databasesystemer til ordbogsredigering.

2.2 Egentlige redigeringsystemer

De tidligste egentlige redigeringsystemer blev udviklet på enkelte større institutter og forlag. Sådanne systemer kørtes typisk fra et centralt anlæg i organisationen; de var „lukkede“, hardwareleverandørafhængige og blev ikke anvendt på andre institutter/forlag. Standardiseringen af styresystemer, der skete med fremkomsten af specielt DOS og UNIX, har — foruden et stort udbud af standardprogrammer til tekstbehandling m.v. — fået enkelte softwarefirmaer til at udvikle specialsystemer til ordbogsredigering.

Et af de første egentlige redigeringsystemer, som har været anvendt i en række ordbogsprojekter, er det DOS-baserede COMPULEXIS, der har været på markedet siden midten af firserne. Systemet består af to hovedelementer: en indtastnings- og redigeringsdel CX-Basic, der bygger på WordPerfect 5.1, og CX-Master, hvor alfabetisering og systematiseringen foregår, og hvorfra forlaget kan generere satsfiler til den trykte ordbog (Geertsen 1993).

Et af de nyere redigeringsystemer er GestorLEX, der er baseret på anbefalingerne fra den såkaldte DANLEX-gruppe (DANLEX-gruppen 1987b). Systemet, som kører under OS/2, anvender strukturbeskrivelsesproget SGML, der gør det muligt at beskrive vilkårligt komplicerede ordbogsstrukturer. GestorLex indeholder et syntakskontrolprogram, der kontrollerer om markeringer i teksten er i overensstemmelse med strukturbeskrivelsen (Norling-Christensen 1991).

3 Redigeringsystemet RiSt

Da man i 1989 skulle påbegynde redigering af en færøsk modersmålsordbog, besluttede man selv at forestå udviklingen af et redigeringsystem. Dels fandtes der gode programmerings-

redskaber på PC-plattformen, og dels skønnede man, at de systemer, der var tilgængelige på markedet, enten var for „lukkede“ og leverandørfafhængige eller for „kostbare“ (både i indkøb og i krav til hardware) i forhold til, hvad man selv ville være i stand til at udvikle.

Ved udviklingen af RiSt lagde man bl.a. vægt på at inkludere faciliteter, som brugeren kender fra tekstbehandlingssystemer, og at basere sig på en simpel datastruktur, hvor man undgår begrænsningerne fra standard databasesystemer med maksimal postlængde og fast antal/rækkefølge af kategorier. Der blev også tilstræbt at have så vidt mulig indskrivningskontrol og automatisering mht. skriftsnit og skilletegn.

RiSt tager sigte på brugere, der anvender systemet dagligt. Ved udfærdigelsen af brugergrænsefladen har man således lagt vægt på, at frekvente funktioner er hurtige og kræver få tastetryk. De fleste funktioner udføres ved anvendelse af funktionstaster eller kombinerede tastetryk, men kan også aktiveres gennem et udbygget menusystem, der desuden omfatter en række mere sjældent anvendte faciliteter (til bl.a. søgning i ordbogen og vedligeholdelse af hjælpe- og indekstabeller).

Set i en mere datateknisk vinkel er redigeringsystemet RiSt et dataprogram, der kører på PC'ere under styresystemet DOS. Programmet kan også køre på PC-net med flere samtidige brugere, og der kan skiftes mellem flere ordbøger. Systemet kan køre på ældre 80286-baserede PC'ere med 1 Mb hukommelse, men der anbefales hurtige 386/486 maskiner med mindst 2 Mb hukommelse.

3.1 Datastruktur

Data for en ordbog i RiSt kan i store træk opdeles i fire grupper:

- A tabel over *opslagsord*
- B tabel med *artikelindhold*
- C tabeller med *skilletegn* og *kategoriattributter* (skriftsnit m.m.)
- D tabeller til *indskrivningskontrol*

Tabellen med opslagsord (A) indeholder desuden nogle administrative oplysninger (redaktør, tidsstempling og status) vedrørende den enkelte artikel. I B-tabellen gemmes en artikels indhold som en sammenhængende tegnsekvens i den rækkefølge, hvorved indholdet fremtræder i selve ordbogen. Sekvensen er opdelt i artikelelementer med markeringer af elementernes oplysningstype (kategori). Artiklernes rækkefølge i B-tabellen bestemmes af den rækkefølge, hvori artiklerne indskrives. Et indeks knyttet til A-tabellen sammen med en „pegepind“ fra opslagsord til artiklens placering i B-tabellen giver hurtig adgang til artikelindholdet.

C-gruppen omfatter skilletegn, der indsættes automatisk mellem to artikelelementer, samt oplysninger om kategoriernes attributter, der bestemmer elementernes fremtræden på skærm og papir. Disse data er med til at fastlægge ordbogens mikrostruktur og vil typisk være defineret inden indskrivning og redigering af artikler påbegyndes. Dette vil også være tilfældet for de tabeller i D-gruppen, der indeholder et givet sæt „tilladte“ værdier for en bestemt kategori (fx ordklasse). Andre tabeller i D-gruppen vil evt. skulle vedligeholdes løbende i redigeringsforløbet.

Udover de fire nævnte grupper data genererer systemet nogle indekstabeller, der gør det muligt for brugeren at foretage søgninger efter indhold i alle artikelkategorier. Der er særlige

indekstabeller for første og sidste del af de ord, der forekommer, således at søgninger efter fx ord* og *bog giver hurtige svar (se senere).

3.2 Kategorier, hierarki og skilletegn

Udover *opslagsord* og *homografnummer*, som er obligatoriske kategorier i enhver RiSt-ordbog, definerer brugeren efter behov et sæt kategorier eller oplysningstyper, der fastlægger den pågældende ordbogs „detaljeringsgrad“. Ordklasse, bøjning, anvendelse, definition og eksempel vil være blandt de typiske kategorier i en almen ensproglig ordbog.

Til hver kategori fastlægges der et *navn* (1–2 bogstaver) med angivelse af *type* (fri tekst: fx i definition; tabeltekst: fx i ordklasse; ordenskategori: fx i subnummer) og *skrift* (snit, størrelse, form, linjeskift, op-/indrykning), der bestemmer elementernes fremtræden i redigeringskorrektoren og andre udskrifter.

En artikels hierarkiske struktur kan styres ved at anvende særlige *niveau-kategorier*. Ved at indlede et underafsnit med et element fra en sådan kategori fremkommer en *automatisk nummerering*. Der kan indføres flere niveauer ved at definere et tilsvarende antal niveau-kategorier, fx således at nummerering i 2. niveau nulstilles efter indføjelser af et element fra 1. niveau-kategori, eksempelvis: . . . A(5) . . . B(1) . . . B(2) . . . A(6) . . . B(1) . . . hvor A og B er elementer fra hhv. 1. og 2. niveaukategori, og hvor tallene i parentes angiver den automatiske nummerering. I nummereringen kan der vælges mellem arabertal, romertal eller bogstaver.

Mellem to elementers indhold indsættes der i artikelvinduet (se senere) og korrektoren *skilletegn automatisk*. Skilletegnene, bestående typisk af 2–3 tegn blandt tegnene , ; : () [] inklusive blanktegn, fastlægges af de kategorier, som de to sammenstødende elementer tilhører, og de gemmes i en tabel, der kan vedligeholdes fra programmet. I tilfælde af, at fx parentes skal omslutte en kategoris elementer, vil „(“ og „)“ være inkluderet i skilletegnene foran hhv. efter elementet. Når de automatiske skilletegn er blevet fastlagt, behøver redaktøren normalt ikke at tænke på semikolon, parenteser og andre skilletegn. I skilletegnstabellen er der mulighed for at angive et eller flere *alternative skilletegn* for et kategori-par. Indsættelse af et alternativt skilletegn angives i elementvinduet (se senere).

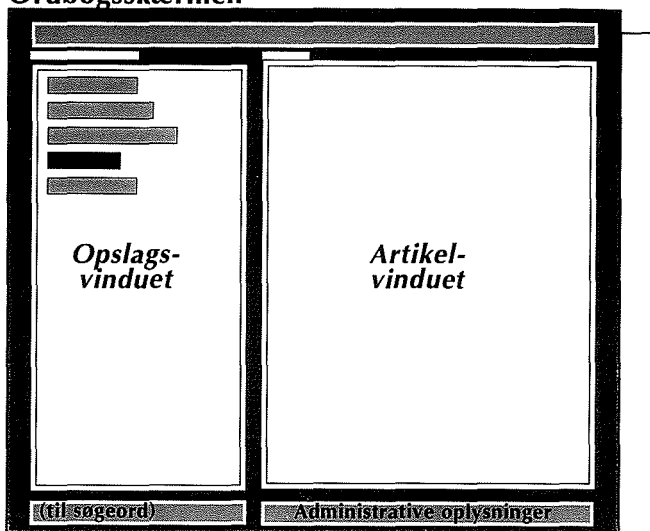
Der gives yderligere mulighed for at ophæve tabelopslaget for skilletegn ved at anvende en bestemt *annullér-kategori*, der har tomme skilletegn med alle andre kategorier. Således kan man i selve annullér-elementet indsætte de tegn, man måtte ønske til at adskille to kategori-elementer i stedet for de automatiske skilletegn. Der er desuden forud defineret en *fodnote-kategori*, af hvilken der i hver artikel findes ét element, hvori redaktøren kan skrive bemærkninger angående artikelen, fx „skal spørge NN om betydning 3“ og „figur foreslås“.

3.3 Alfabetisering og indskrivningskontrol

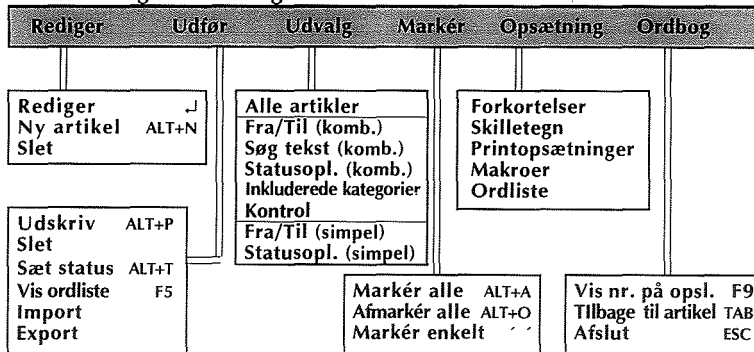
I hver RiSt-artikel findes netop ét opslagsord, og for ordbogen som helhed skal opslagsordene, inklusive evt. homografnummer, være unikke. Et forsøg på at oprette en ny artikel med et tidligere anvendt opslagord vil blive afvist af systemet. Opslagsordenes alfabetiske rækkefølge fastlægger artiklernes rækkefølge i ordbogen. Af hensyn til variationer i de vesteuropæiske sprogs alfabeter — både med hensyn til bogstaver og placering — gives der mulighed for at definere de tilladte karakterers alfabetiseringsrækkefølge, hvilket foregår i en af systemets opsætningsfiler.

I forbindelse med visse kategorier, fx for stil- og fagkoder, kan der pålægges en *kontrol af inddata*, således at programmet i forbindelse med redigering gør opmærksom på, hvis en anvendt kode/forkortelse ikke findes i den teksttabel, som er tilknyttet den pågældende kategori. Der kan vælges mellem en *streng* indskrivningskontrol, hvor inddata skal forefindes i teksttabellen, og en *lempelig* kontrol, hvor redaktøren i redigeringsfasen kan supplere teksttabellen med en ny kode eller forkortelse.

Ordbogsskærmen



Menuoversigten i ordbogsskærmen



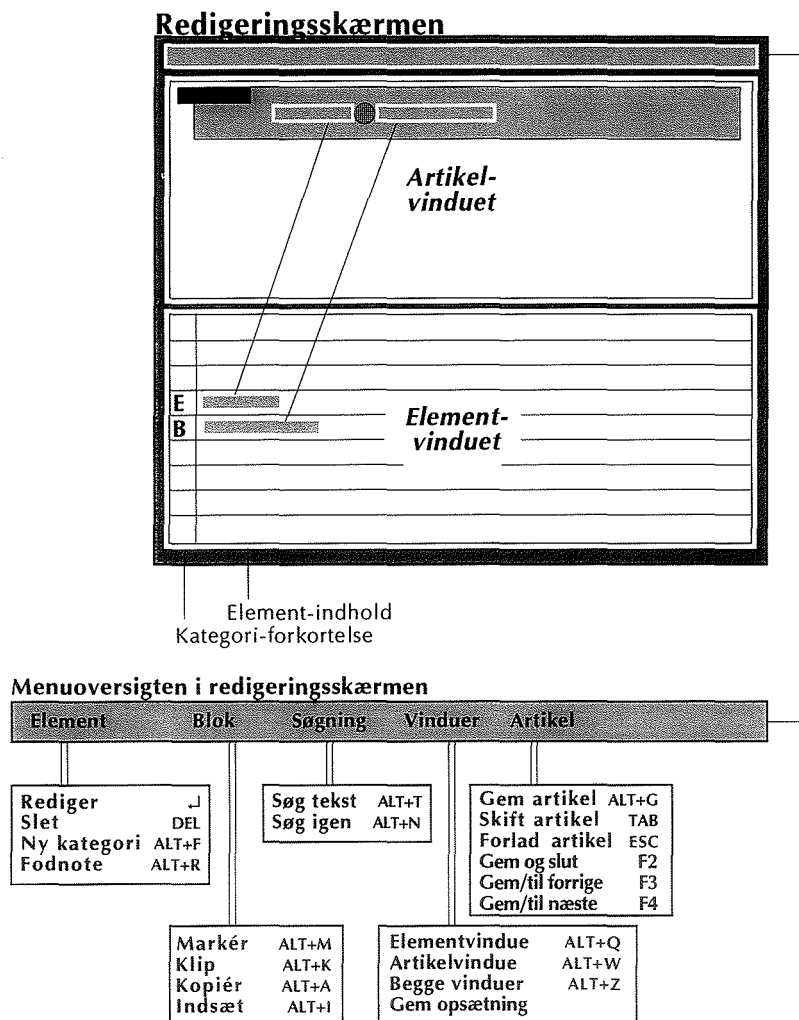
Figur 1: Ordbogsskærmens indretning og valgmulighederne i ordbogsskærmens menuoversigt.

For at opnå størst mulig fleksibilitet er redaktøren frit stillet mht. rækkefølge af kategorier i en ordbogsartikel. Dette øger dog samtidig risikoen for, at redaktøren kommer til at anvende ikke tilladte kategorisekvenser. Med en simpel ordbogsstruktur vil man kunne opnå en „svag“ kontrol af kategorirækkefølgen v.hj.a. skilletegnene, idet ved sammenfald

af kategorier, som ikke er inkluderet i skilletegnstabelen, vil systemet bruge „#_“ som skilletegn. Optræder „#_“ som skilletegn i artikelvinduet, tyder det på, at der enten er fejl i kategorikombinationen, eller også skal skilletegnstabelen ajourføres.

3.4 Brugergrænseflade

Programmet starter op med en *systemskaerm*, hvor man kan vælge sig ind på den ønskede ordbog. Som bruger har man adgang til RiSt i to tilstande eller på to niveauer, enten via *ordbogsskaermen* (Figur 1) eller via *redigeringskaermen* (Figur 2). Afhængig af hvilke funktioner, der ønskes udført, vil brugeren skifte mellem disse to skærme.



Figur 2: Redigeringskaermens indretning og valgmulighederne i redigeringskaermens menuoversigt.

Ordbogsskærmen er todelt: et *opslagsvindue* til venstre og et *artikelvindue* til højre. I opslagsvinduet, som indeholder indtastede opslagsord i alfabetisk rækkefølge, kan der bladres v.h.j.a. tasterne ↑/↓/PgUp/PgDn eller ved at skrive det søgte ord eller de første bogstaver i ordet på linjen under opslagsvinduet og trykke på Enter-tasten.

I ordbogsskærmen udføres de operationer, der vedrører flere artikler, fx i forbindelse med udskrivning eller eksport af flere artikler. I ordbogsskærmen kan der indskrives nye opslagsord eller allerede eksisterende opslagsord kan ændres/slettes. Desuden kan man i ordbogsskærmen — gennem menuoversigten øverst på skærmen — udføre forskellige *vedligeholdelsesfunktioner* (bl.a. vedr. skilletegn, forkortelser og printeropsætning) foruden indeksering af ordbogsdatabase. Artikelvinduet til højre i ordbogsskærmen viser artikelindholdet under det opslagsord, som markøren opholder sig ved. Indholdet optræder i et *tilnærmet trykformat* med skilletegn og i de farver, som er tildelt kategorierne. Nederst til højre i artikelvinduet står de administrative oplysninger vedr. artiklen, bl.a. brugerkoder, status, dato for oprettelse af artikel og sidste dato for redigering.

Når man står ved et opslagsord i ordbogsskærmens opslagsvindue og trykker på Enter-tasten, kommer man til redigeringskærmen, hvor redigering af den enkelte ordbogsartikel foregår. Redigeringskærmen er delt i to: øverst et artikelvindue (som viser den pågældende artikels indhold) og nederst *elementvinduet*, hvor selve redigeringen af de enkelte artikelelementer foregår, og hvor hvert element står venstrejusteret på hver sin linje på skærmen. Skilletegn ses ikke i elementvinduet, men kommer frem i artikelvinduet, der løbende bliver ajourført i overensstemmelse med ændringer i elementvinduet. I elementvinduet kan man anvende *makroer*, som frembringer ofte forekommende tegnkombinationer, til bl.a. tegn til „indramning“ af tekst, som afviger fra kategoriskriften. Makrodefinitionerne gemmes i tabeller, som kan vedliges af brugeren.

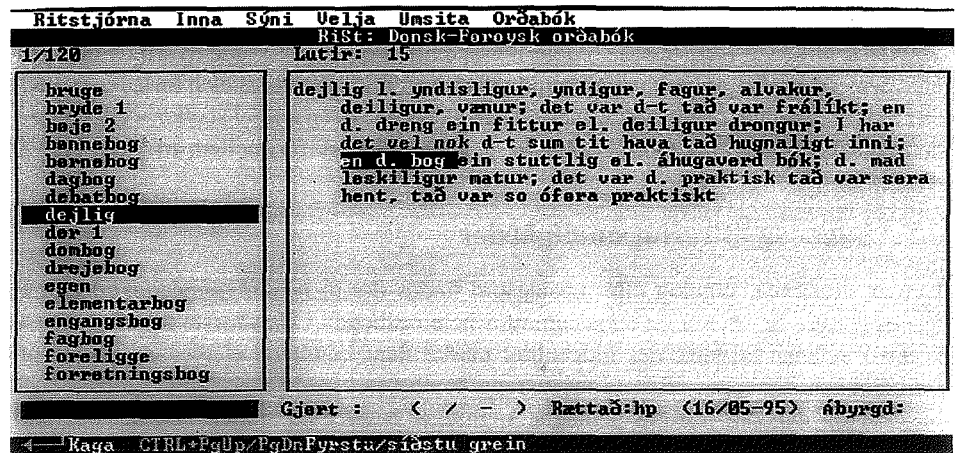
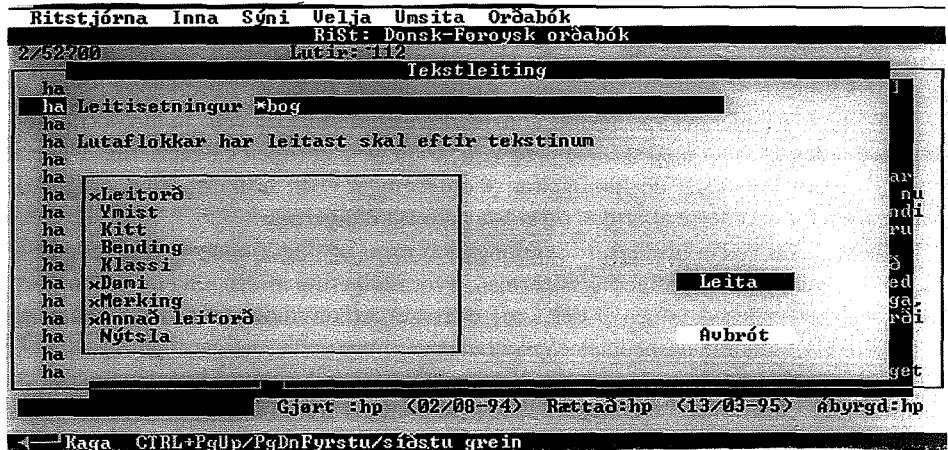
Ved at anvende TAB-tasten i redigeringskærmen kan man forlade en artikel midlertidig, men den er dog stadig åben (i modsætning til tryk på ESC, hvor artiklen forlades og lukkes). Derved føres man til ordbogsskærmen og kan vælge en anden artikel at efterse eller rette i. På den måde har brugeren *to artikler åbne på én gang* og kan skifte imellem artiklerne ved TAB.

3.5 Søge- og filtreringsmuligheder

Udover alfabetisk søgning efter opslagsord findes der to *tekstsøgningsfunktioner* i RiSt: tekstsøgning i én artikel og tekstsøgning i hele ordbogen. Den første funktion kan være praktisk i lange artikler, der indeholder måske flere hundrede elementer. I den anden funktion kan der søges efter en tekststreng kombineret med jokertegn (*) forrest og/eller bagerst, og der kan udføres gentagne søgninger baseret på simple boolske operationer. Denne søgemulighed kan være til stor hjælp, når man fx ønsker at finde de artikler, hvori der forefindes sammensatte ord med en fast orddel. I et *valgvindue* vil brugeren kunne afgrænse søgningen til udvalgte kategorier (Figur 3, øverst).

En tekstsøgning i hele ordbogen resulterer i et *udvalg af artikler*, og i opslagsvinduet i ordbogsskærmen vil man derefter kun se opslagsordene for de udvalgte artikler; man har foretaget en filtrering af artiklerne (Figur 3, nederst). Denne facilitet er nyttig, når samme operation skal udføres på et artikeludvalg, fx når en række artikler skal udskrives/eksporteres, efterses eller måske slettes. Inden en operation aktiveres på et artikeludvalg, eller en del deraf, skal de pågældende artikler markeres.

Man kan også udføre en filtrering i henhold til artiklernes administrative oplysninger, fx kan en redaktør afgrænse navigeringen i opslagsvinduet til de artikler, han selv har oprettet og/eller sidst har rettet i i et givet tidsinterval. Statuskoden kan også danne grundlag for en filtrering. Ved undersøgelse af forskellige kategoriers forekomst kan der laves en filtrering, således at netop de artikler medtages — eller medtages ikke — som indeholder en eller flere bestemte kategorier.



Figur 3: Øverst: Valgvinduet ved tekstsøgning med søgestrengen *bog begrænset til udvalgte kategorier bl.a. leitorð (opslagsord) og dømi (eksempler). Nederst: Resultatet af tekstsøgningen ovenover. Ordbogsskærmen viser en del af de ialt 120 opslagsord, hvis artikler indeholder søgestrengen. I artikelvinduet vises artikelen *dejlig*, hvor elementet *en d. bog* (der tilhører eksempel-kategorien) er fremhævet.

3.6 Udskrifter, export og import

Med RiSt kan brugeren få udskrifter af udvalgte artikler i et rigtigt ordbogsformat, såvidt det er muligt med den PostScript-laserprinter, der er koblet på systemet. I programmet kan brugeren selv indstille en række *layoutparametre*, bl.a. antal spalter, linjehøjde, marginbredde, samt placering af sidenummer og udskriftdato. Dog kan man ikke opnå højre justeret tekst og orddeling.

Til den samme ordbog kan brugeren definere og gemme flere *printopsætninger* til forskellige formål (korrektur, speciallister m.v.). Til hver printopsætning er der et valgvindue, hvor brugeren kan specificere en række karakteristika vedrørende opsætningen. Udover ovennævnte layoutparametre kan man vælge skrift (snit, størrelse, form), samt bestemme hvilke kategorier skal udskrives. Eksempler på forskellige printopsætninger: *korrektur*-udskrift, alle kategorier, inkl. fodnote, skriftstørrelse 11–12 punkt og linjehøjde 18 punkt, *ordbogs*-udskrift med tre spalter, uden fodnoter og mindre skriftstørrelse og linjehøjde, samt liste med *opslagsord* i fire spalter, kun med opslagsord og ordklasse (Figur 4).

Der er mulighed for at eksportere ordbogsdata i et RiSt-format (ASCII), der indeholder alle artikelinformationer inkl. artikelens administrative oplysninger. Programmet kan også importere data i dette format, så de bliver flettet med en eksisterende RiSt-ordbog. Der er således mulighed for, at en del af en ordbog, der er redigeret på en ekstern PC'er, vil kunne inkorporeres i ordbogens hoveddatabase.

Ordbogsdata i RiSt, inkl. skilletegn og oplysninger om skriftsnit, kan desuden eksporteres i et *SGML-lignende opmærknings-format*, der kan importeres af de kendte DTP-programmer (Corel Ventura, PageMaker, Quark-Xpress o.a.), hvor den endelige opsætning af ordbogsteksten med henblik på udgivelse kan finde sted.

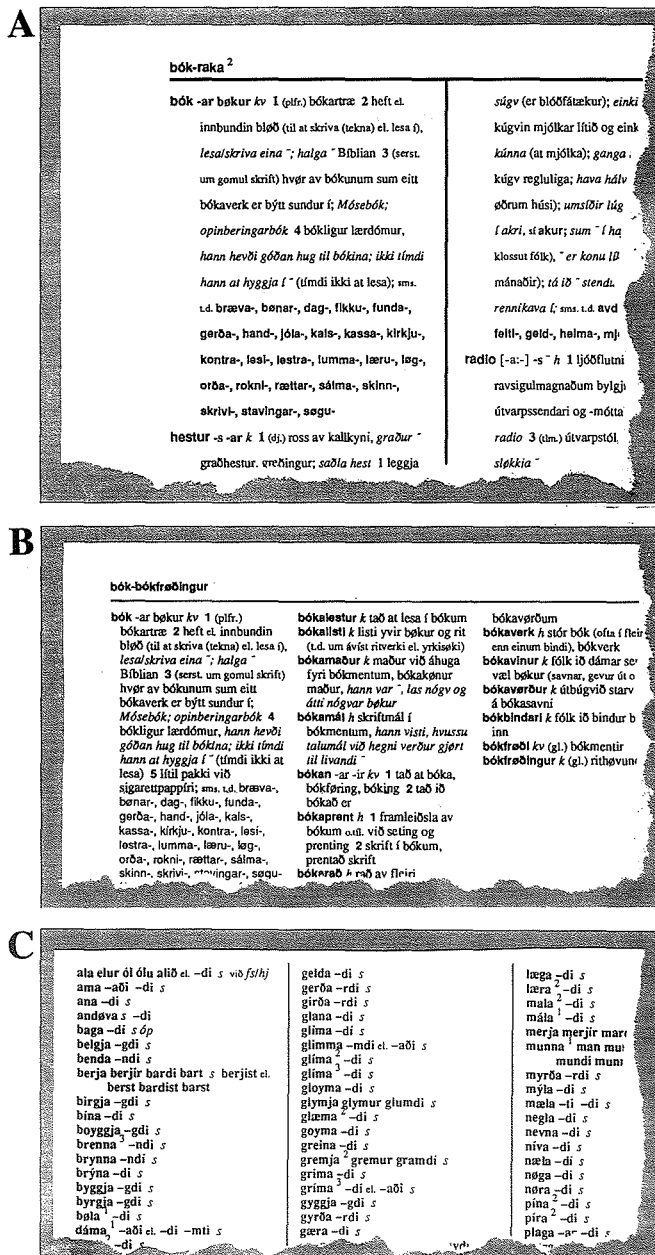
4 Afsluttende bemærkninger

RiSt er primært udviklet som et redskab for redaktørerne af *móðurmálsorðabókin* (den færøske modersmålsordbog), men er samtidig så generelt — i hvert fald på en række basale punkter — at systemet vil kunne anvendes i andre ordbogsprojekter. Her tænkes specielt på det *simple kategorikoncept*, der giver brugeren stor fleksibilitet mht. artikelstruktur og attributstyring. Til trods for at kategorierne i RiSt ikke kan tilskrives en niveau-attribut, så gør de specielle niveau-kategorier (med automatisk nummerering) det muligt at operere med et begrænset „pseudo“-hierarki. I andre redigeringsystemer er der muligheder for at definere mere komplicerede taxonomier.

Móðurmálsorðabókin bliver redigeret på baggrund af en omfattende seddelsamling samt andet skriftligt materiale. Vi har ikke i konstruktionen af systemet tilstræbt en integration af redigering og evt. edb-baseret kildemateriale i form af fx ordregistre og tekstsamlinger på edb-medium. RiSt er — i sin nuværende form — først og fremmest et redigeringsystem. I RiSt har brugeren alligevel mulighed for at gøre opslag i en forud valgt ekstern ordliste, imens der arbejdes i redigeringskærmen.

Til redigering af ordbøger af „normal“ størrelse, d.v.s. 50–100 000 opslagsord og 500–1000 sider, kommer man ikke til at støde på nogle systemmæssige begrænsninger i RiSt, men overvejes systemet til specielle redigeringsopgaver, bør man være opmærksom på bl.a. følgende maksimalværdier for en enkelt ordbog: 350 000 opslagsord, 32 kategorier og

4000 elementer pr. artikel. Angående *móðurmálsorðabókin* er her nogle nøgledata: 65 000 opslagsord, 24 kategorier, 600–700 elementer i de længste artikler og 27 Mb på disken, hvoraf godt halvdelen er indekstabeller.



Figur 4: Eksempler på udskrifter fra RiSt med forskellige printopsætninger: A: korrektur-udskrift, B: ordbogs-udskrift og C: liste med opslagsord.

Der er ikke blevet tilføjet nogle væsentlige nye faciliteter i RiSt de sidste par år. Løbende er systemet blevet „tunet“, så det nu kører strømlinet (uden betydelige afbrydelser) i et PC-lokalnet med 15 arbejdsstationer, deriblandt de fem redaktører på *móðurmálsorðabókin*. Man må forudse — udviklingen med grafiske brugergrænseflader taget i betragtning — at der inden for 1–2 år vil være behov for en opgradering. En evt. Windows-version af RiSt vil — i forhold til den eksisterende DOS-version — medføre en nemmere håndtering af karaktersæt og udskrifter samt bedre mulighed for integration mellem ordbogsredigering og udnyttelse af edb-baseret kildemateriale.

Litteratur

- DANLEX-gruppen 1987a: *Ordbøger i Danmark — en oversigt*. København.
- DANLEX-gruppen 1987b: *Descriptive Tools for Electronic Processing of Dictionary Data*. Studies in Computational Lexicography. Lexicographica Series Maior 20. Tübingen: Max Niemeyer Verlag.
- Norling-Christensen, O. 1991: Struktureret redigering af ordbøger. I: R. V. Fjeld (red.): *Nordiske Studier i Leksikografi* I. Oslo: Nordisk forening for leksikografi, 447–454.
- Geertsen, A. 1993: Ordbogsredaktion med COMPULEXIS. I: Anna Garde/Pia Jarvad (red.): *Nordiske Studier i Leksikografi* II. København: Nordisk forening for leksikografi, 87–95.

Oversikt over forfatterne

Henrik Andersson Den Danske Ordbog, Njalsgade 80, DK-2300 København S

Jens Axelsen Gyldendal, Klareboderne 3, DK-1001 København K

Kristin Bakken Avdeling for navnegransking, INL, Universitetet i Oslo, Postboks 1011 Blindern, N-0315 Oslo

Henning Bergenholtz Handelshøjskolen i Århus, Fuglesangs Allé 4, DK-8210 Århus V

Frede Boje Institut for tysk, Handelshøjskolen i København, Dalgas Have 15, DK-2000 Frederiksberg

Anna Braasch Center for Sprogteknologi, Københavns universitet, Njalsgade 80, DK-2300 København S

Anne Duekilde Supplement til Ordbog over Det Danske Sprog, Njalsgade 80, DK-2300 København S

Greta Duvå Handelshøjskolen i Århus, Fuglesangs Allé 4, DK-8210 Århus V

Gunhild Dyrberg Handelshøjskolen i København, Dalgas Have 15, DK-2000 Frederiksberg

Eiríkur Rögnvaldsson Háskóla Íslands, Árnagarði v/Suðurgötu, IS-101 Reykjavík

Ruth Feil Handelshøjskolen i Århus, Fuglesangs Allé 4, DK-8210 Århus V

Ruth Vatvedt Fjeld Avdeling for nordisk språk og litteratur, INL, Universitetet i Oslo, Postboks 1013 Blindern, N-0315 Oslo

Martin Gellerstam Institutionen för svenska språket, Göteborgs universitet, S-412 98 Göteborg

Oddrun Grønvik Avdeling for leksikografi, Universitetet i Oslo, Postboks 1001 Blindern, N-0315 Oslo

Guðrún Ingólfsdóttir Smyrilsvegur 28, IS-107 Reykjavík

Guðrún Kvaran Orðabók Háskólans, Neshaga 16, IS-107 Reykjavík

Anna Helga Hannesdóttir Institutionen för svenska språket, Göteborgs universitet, S-412 98 Göteborg

Zakaris S. Hansen Føroyamálsdeildin, Fróðskaparsetur Føroya, V.U. Hammershaimbs gøtu 16, FR-100 Tórshavn

Helgi Haraldsson Institutt for østeuropeiske og orientalske studier, Universitetet i Oslo, Postboks 1030 Blindern, N-0135 Oslo

Berit Heer Henriksen Den Danske Ordbog, Njalsgade 80, DK-2300 København S

Marit Hovdenak Norsk språkråd/Avdeling for leksikografi, INL, Universitetet i Oslo, Postboks 1001 Blindern, N-0315 Oslo.

Jógvan í Lon Jacobsen Føroyamálsdeildin, Fróðskaparsetur Føroya, V.U. Hammershaimbs gøtu 16, FR-100 Tórshavn

Tove Jacobsen Romansk institutt, Universitetet i Bergen, Sydneplass 9, N-5007 Bergen

Pia Jarvad Dansk Sprognævn, Njalsgade 80, DK-2300 København S

Anfinnur Johansen Føroyamálsdeildin, Fróðskaparsetur Føroya, V.U. Hammershaimbs gøtu 16, FR-100 Tórshavn

- Jón Hilmar Jónsson* Orðabók Háskólans, Neshaga 16, IS-107 Reykjavík
- Kristín Bjarnadóttir* Orðabók Háskólans, Neshaga 16, IS-107 Reykjavík
- Anna-Lise Laursen* Handelshøjskolen i Århus, Fuglesangs Allé 4, DK-8210 Århus V
- Carl-Erik Lundbladh* Svenska Akademiens ordbok, Tornavägen 9 B, S-223 63 Lund
- Sven-Göran Malmgren* Institutionen för svenska språket, Göteborgs universitet, S-412 98 Göteborg
- Nina Martola* Forskningscentralen för de inhemska språken, Sörnäs strandväg 25, FIN-00500 Helsingfors
- Magne Myhren* Avdeling for leksikografi, INL, Universitetet i Oslo, Postboks 1001 Blindern, N-0315 Oslo
- Henrik Nikula* Institutionen för tyska språket och litteraturen, Vasa universitet, Puolalaparken 1 b B 12, FIN-20100 Åbo
- Kerstin Norén* Institutionen för svenska språket, Göteborgs universitet, S-412 98 Göteborg
- Christian-Emil Ore* Dokumentasjonsprosjektet, Universitetet i Oslo, Postboks 1123 Blindern, N-0317 Oslo
- Bo Ralph* Institutionen för svenska språket, Göteborgs universitet, S-412 98 Göteborg
- Inger Hesslin Rider* Norstedts Förlag, Box 2052, S-103 12 Stockholm
- Leena Savolainen* Forskningscentralen för de inhemska språken, Box 57, FIN-00401 Helsingfors
- Rik Schutz Van Dale* Lexicografie, Postbus 19232, NL-3500 De Utrecht
- Pål Steigan* Cappelen forlag, Postboks 350 Sentrum, N-0101 Oslo
- Bo Svensén* Svenska Akademien, Box 2118, S-103 13 Stockholm
- Joan Tournay* Handelshøjskolen i København, Dalgas Have 15, DK-2000 Frederiksberg
- Lars Törnqvist* Älvkarleövägen 13, S-115 43 Stockholm
- Lars S. Vikør* Avdeling for leksikografi, INL, Universitetet i Oslo, Postboks 1001 Blindern, N-0315 Oslo
- Boye Wangenstein* Avdeling for leksikografi, INL, Universitetet i Oslo, Postboks 1001 Blindern, N-0315 Oslo
- Lotte Weilgaard Christensen* Handelshøjskole Syd, Engstien 1, DK-6000 Kolding
- Petur Zachariassen* Teldudeildin, Fróðskaparsetur Føroya, Nóatún, FR-100 Tórshavn

Refererte ordbøker

- A New English Dictionary on Historical Principles ⇒ The Oxford English Dictionary
- Aasen 1850 (A¹) ⇒ Ordbog over det norske folkesprog
- Aasen 1873 (A²) ⇒ Norsk Ordbog med dansk Forklaring
- Agricola 1962 ⇒ Wörter und Wendungen ...
- Allén et al. 1970–1980 ⇒ Nusvensk frekvensordbok ...
- Anglo/Amerikansk-Dansk, Dansk-Anglo/Amerikansk specialordbog inden for revision, regnskabsvæsen mv. (1990), 94
- AUSLAN Dictionary. A Dictionary of the Sign Language of the Australian Deaf Community (1989), 359
- Ásgeir Blöndal Magnússon 1989 ⇒ Íslensk orðsifjabók
- BBI ⇒ The BBI Combinatory Dictionary of English ...
- Becker-Christensen/Widell 1991 ⇒ Nudansk Ordbog (elektronisk udgave)
- Benson/Benson/Ilson 1986 ⇒ The BBI Combinatory Dictionary of English ...
- Bergstrøm-Nielsen/Lange/Verner Larsen 1991 ⇒ Dansk-tysk ordbog (Munksgaard 1991)
- Berkov 1994 ⇒ Russisk-norsk ordbog
- Berulfsen/Gundersen 1986 ⇒ Fremmedordbok
- Bidrag til en Ordbog over jyske Almuesmål (1886–1914), 83, 297
- Björn Halldórsson 1992 ⇒ Orðabók. Íslensk – latnesk – dönsk
- Blinkenberg/Høybye 1991 ⇒ Dansk-fransk Ordbog
- Blöndal = Sigfús Blöndal ⇒ Islandsk-dansk ordbog
- Bokmålsordboka. Definisjons- og rettskrivningsordbok (1993), 151–153, 155–156, 375–376, 417–418
- Bokmålsordlista (1992), 205
- Bork 1990 ⇒ Dansk-tysk ordbog
- BOB ⇒ Bokmålsordboka
- Brien 1992 ⇒ Dictionary of British Sign Language ...
- Bruun 1977 ⇒ Dansk Sprogbrug
- Brüel m.fl. 1993 ⇒ Fremmedordbok
- Brynildsen 1927 ⇒ Norsk-Engelsk Ordbok
- B&H ⇒ Dansk-fransk Ordbog (1991)
- Cambridge International Dictionary of English (1995), 355
- Cannelin/Cannelin/Hirvesala/Hedlund 1976 ⇒ Finsk-svensk ordbok
- CAPLEX. Cappelens ettbinds leksikon (1991), 380
- Cappelens musikkleksikon (1978–80), 380
- Chambers Biographical Dictionary (1974), 400
- Chambers Dictionary of World History (1993), 400
- CIDE ⇒ Cambridge International Dictionary of English
- CIED 1989 ⇒ Concise Icelandic-English Dictionary

- Collins Concise Dictionary Plus (1989), 353
- Collins COBUILD English Language Dictionary (1987/1995), 159, 354
- Collins COBUILD Student's Dictionary (1989), 162
- Collins COBUILD Student's Dictionary. Bridge Bilingual: Portuguese (1995), 160
- Collins Plus ⇒ Collins Concise Dictionary Plus
- Concise Icelandic-English Dictionary (1989), 188
- COBUILD ⇒ Collins COBUILD English Language Dictionary (1995)
- Dahl/Hammer/Dahl 1907 ⇒ Dansk Ordbog for Folket
- Dahlerup 1919–1956 ⇒ Ordbog over det danske sprog
- Damms store leksikon (1985–90), 378
- Dansk ordbog (1833), 309
- Dansk Ordbog for Folket (1907), 38
- Dansk sprogbrug. En stil- og konstruktionsordbog (1977), 140, 142
- Dansk-engelsk ordbog (1984), 24–25
- Dansk-engelsk ordbog (V&B 1990; elektronisk udgave 1992), 2, 19, 23, 142–143
- Dansk-fransk ordbog (1991), 2, 229
- Dansk-norsk Ordbog (Schjøtt 1909), 406
- Dansk-tysk ordbog (Bork 1990), 57
- Dansk-tysk ordbog (Munksgaard 1991; elektronisk udgave 1993), 56, 142
- Dansk-tysk ordbog. Undervisning (Hjorth 1993), 57
- Den Danske Ordbog, 1–2, 4–5, 87, 195, 198, 413
- Deutsches Wörterbuch (Grimm/Grimm 1854), 175
- Deutsches Wörterbuch (Wahrig 1986/91), 57–58
- Dictionarium Færoense. Færøsk-dansk-latinsk ordbog (1966), 234, 238
- Dictionary of British Sign Language / English (1992), 359
- Dictionary of Translated Names and Titles (1986), 397, 400
- DIO 1957 ⇒ Dönsk-íslensk orðabók
- DIO 1992 ⇒ Dönsk-íslensk orðabók
- Duden Deutsches Universalwörterbuch (1989), 57
- Duden GWB ⇒ Duden. Das große Wörterbuch der deutschen Sprache ...
- Duden. Das große Wörterbuch der deutschen Sprache in sechs Bänden (1978), 38
- DUW ⇒ Duden Deutsches Universalwörterbuch
- Dönsk-íslensk orðabók (1992), 187
- Dönsk-íslensk orðabók (1957), 187
- EIO 1984 ⇒ Ensk-íslensk orðabók ...
- Encyclopædia Britannica: Micropædia (1983ff.), 400
- Engelsk-dansk dansk-engelsk ordbog ⇒ Gads store ordbøger ...
- Engelsk-dansk Ordbog (1977), 356
- Ensk-føroysk orðabók (1992), 240

- Ensk-íslensk orðabók með alfræðilegu ívafi (1984), 187–188
- Erhvervs- og økonomiordbog Dansk-Spansk (1989), 95
- Evensen 1905–09 ⇒ Føroysk orðabók
- Falk/Torp 1910–11 ⇒ Norwegisch-dänisches etymologisches Wörterbuch
- Feilberg 1886–1914 ⇒ Bidrag til en Ordbog over jyske Almuesmål
- Finsk basordbok ⇒ Suomen kielen perussanakirja
- Finsk-svensk ordbok (1976), 283
- Focus uppslagsbok (1977ff.), 400
- Fransk-norsk ordbok ⇒ Kunnskapsforlagets fransk-norske ordbok
- Fremmedordbok (1986 – norsk) 416, 418
- Fremmedordbog (1993 – dansk), 224
- Freysteinn Gunnarsson 1957 ⇒ Dönsk-íslensk orðabók
- Fritzner 1883/86–96; 1954; 1972 ⇒ Ordbog over det gamle norske sprog
- Fryd 1990 ⇒ Anglo/Amerikansk-Dansk, Dansk-Anglo/Amerikansk specialordbog ...
- Føroysk orðabók (1905–09), 236
- Føroysk-donsk orðabók (1927–28), 238
- Føroysk-donsk orðabók (1961; 2. útgáva), 238
- Føroysk-donsk orðabók (1974; Eykabind), 238
- Gad ⇒ Gads store ordbøger engelsk-dansk dansk-engelsk
- Gads store ordbøger engelsk-dansk dansk-engelsk (1993), 24
- Galindo 1989 ⇒ Erhvervs- og økonomiordbog Dansk-Spansk
- Gammeldansk Ordbog
- Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi (1992), 45
- Glossarium suiogothicum ... (1769), 340
- Glossarium Sveo-Gothicum eller Swensk Orda-book ... (1712), 340
- Grimm, Jacob und Wilhelm 1854 ⇒ Deutsches Wörterbuch
- Grolier Multimedia Encyclopedia (1993), 40
- Grundt 1991/1994 ⇒ Stor norsk-fransk ordbok
- Gundersen 1974 ⇒ Norsk synonymordbok
- Guttu 1993 ⇒ Norsk Illustrert Ordbok
- Gyldendals Røde 1984 ⇒ Dansk-engelsk ordbog
- Handwörterbuch der deutschen Gegenwartssprache (1984), 38
- Henriksen 1991 ⇒ Juridisk ordbog spansk – dansk
- Hjorth 1993 ⇒ Dansk-tysk ordbog
- Holm/Aðalsteinn Davíðsson 1982 ⇒ Svensk-isländsk ordbok
- Hovdenak o. a. 1986 ⇒ Nynorskordboka
- Hróbjartur Einarsson 1987 ⇒ Norsk-isländsk ordbok
- Hustad 1979/1984 ⇒ Stor Norsk Tysk Ordbok

- IBM-morfologien (elektronisk ordliste), 205, 207–208
- Ihre 1769 ⇒ Glossarium suiogothicum ...
- Islandsk-dansk ordbog (1920–24), 344
- IS 1985 ⇒ Íslensk samheitaorðabók
- Íslensk orðabók handa skólum og almenningi (1983), 11, 13–15, 344
- Íslensk orðsifjabók (1989), 125, 178, 257
- Íslensk orðtíðnibók (1991), 127, 256
- Íslensk samheitaorðabók (1985), 185
- Johnston 1989 ⇒ AUSLAN Dictionary ...
- Jón Árnason 1994 ⇒ Nucleus latinitatis
- Jón Hilmar Jónsson 1994 ⇒ Orðastaður
- Juridisk ordbog spansk – dansk (1991), 113
- Jysk Ordbog
- Jörgen Pind [ed.]/Friðrik Magnússon/Stefán Briem 1991 ⇒ Íslensk orðtíðnibók
- Kalkar 1881–1918 ⇒ Ordbog til det ældre danske Sprog
- Kaufmann/Bergenholtz 1992 ⇒ Genteknologisk Ordbog
- Kjærulff-Nielsen 1977 ⇒ Engelsk-dansk Ordbog
- Kunnskapsforlagets fransk-norske ordbok (1986), 219
- Landrø/Wangensteen 1993 ⇒ Bokmålsordboka
- Landsbeteckningar ⇒ SS-ISO 3166
- Le Nouveau Petit Robert (1993), 108, 219
- Lewandowski 1976 ⇒ Linguistisches Wörterbuch
- Lexicon Færoense (1877–88), 235, 238
- Linguistisches Wörterbuch (1976)
- Longman 1992 ⇒ Longman Dictionary of language teaching ...
- Longman Dictionary of Contemporary English (1978), 339
- Longman Dictionary of language teaching and applied linguistics (1992), 38
- Molbech 1833 ⇒ Dansk ordbog
- Munksgaard ⇒ Dansk-tysk ordbog (1991/1993)
- Murray et al. 1884–1928 ⇒ The Oxford English Dictionary
- Nationalencyklopedin (1989ff.), 400
- Nationalencyklopedins ordbok, 275–277, 280–281
- NDO ⇒ Nudansk Ordbog (elektronisk udgave 1991)
- NEO ⇒ Nationalencyklopedins ordbok
- NIO 1987 ⇒ Norsk-islandsk ordbok
- Norsk Illustrert Ordbok (1993), 415
- Norsk landbruksordbok (1979), 298
- Norsk ordbok (1950ff.), 307–308, 403, 407–411
- Norsk ordbok (Schjøtt 1914), 298, 301

- Norsk Ordbog (Ross 1895/1971), 301, 305, 307, 403, 406
Norsk Ordbog med dansk Forklaring (Aasen 1873), 295–298, 308, 403
Norsk riksmålsordbok (1937–57), 298, 300–301, 380, 407
Norsk synonymordbok (1974), 185–186
Norsk-Engelsk Ordbok (1927), 24
Norsk-islandsk ordbok (1987), 186, 188
Norsk-tysk ⇒ Stor Norsk Tysk Ordbok
Norwegisch-dänisches etymologisches Wörterbuch (1910–11)
NO ⇒ Norsk ordbok (1950ff.)
NOB ⇒ Nynorskordboka
NRO ⇒ Norsk riksmålsordbok
NS 1974 ⇒ Norsk synonymordbok
Nucleus latinitatis (1994), 180
Nudansk Ordbog (1992; elektronisk udgave 1991), 38, 41, 139–140, 142, 227, 229–230
Nusvensk frekvensordbok baserad på tidningstext (1970–1980), 348
Nusvensk ordbok (1981), 317
Nynorsk etymologisk ordbok (1919)
Nynorskordboka (1986/1993), 205–206, 208, 298, 408
Nyord i svenskan från 40-tal till 80-tal (1986), 277
ODS ⇒ Ordbog over det danske Sprog
ODS-S ⇒ Ordbog over det danske Sprog. Supplement
OED ⇒ The Oxford English Dictionary
Ordbog over det danske Sprog (1918–56), 2, 20, 23, 83, 141–142
Ordbog over det danske Sprog. Supplement (1992ff.), 86–91
Ordbog over det gamle norske sprog (Fritzner), 32, 125, 126, 130, 170, 172, 308
Ordbog over det norske folkesprog (Aasen 1850), 295–296, 298
Ordbog til det ældre danske Sprog (Kalkar 1881–1918), 83, 85
Ordbok öfver svenska medeltidsspråket (Söderwall 1884–1918), 277
Ordbok över svenska språket, utgiven av Svenska Akademien (1898ff.)
Orðabók. Íslensk – latnesk – dönsk (Björn Halldórsson 1992), 181
Orðaskyn. Mín fyrsta orðabók (1994), 240
Orðastaður. Orðabók um íslenska málnotkun (1994), 126, 134, 245–247, 249, 254, 256
Pears Cyclopaedia (1970), 400
Perussanakirja ⇒ Suomen kielen perussanakirja
Petit Robert ⇒ Le Nouveau Petit Robert
Politikens Nudansk Ordbog ⇒ Nudansk Ordbog (1992)
Retskrivningsordbogen (1986), 2, 24
Rey-Debove/Rey 1993 ⇒ Le Nouveau Petit Robert
Richards/Platt/Platt 1992 ⇒ Longman Dictionary of language teaching ...

- Rommetveit 1979 ⇒ Norsk landbruksordbok
 Ross 1971 [1895] ⇒ Norsk Ordbog
 RO ⇒ Retskrivningsordbogen
 Russisk-norsk ordbog (1994), 185, 188, 190, 193
 SAOB ⇒ Ordbok över svenska språket ...
 SAOL ⇒ Svenska Akademiens ordlista
 Schjøtt 1909 ⇒ Dansk-norsk Ordbog
 Schjøtt 1914 ⇒ Norsk ordbok
 Sigfús Blöndal 1920–24 ⇒ Islandsk-dansk ordbog
 SIO 1982 ⇒ Svensk-isländsk ordbok
 Slovar' sinonimov russkogo jazyka (1968), 185
 Slovar' sinonimov russkogo jazyka (1970), 186, 189
 SO ⇒ Svensk ordbok
 SOB ⇒ Svensk ordbok
 Spegel 1712 ⇒ Glossarium Sveo-Gothicum ...
 SS-ISO 3166. Landsbeteckningar (1990), 400
 Statsnavne og nationalitetsord (1994)
 Stor norsk-fransk ordbok (1991/1994), 213
 Stor Norsk Tysk Ordbok, Studentutgave (1979/1984), 55, 57
 Suomen kielen perussanakirja (1990–94), 283–286, 289–290
 Svabo 1966 ⇒ Dictionarium Færoense ...
 Svavar Sigmundsson 1985 ⇒ Íslensk samheitaorðabók
 Svensk ordbok (1986)
 Svensk-isländsk ordbok (1982)
 Svenska Akademiens ordbok ⇒ Ordbok över svenska språket ...
 Svenska Akademiens ordlista över svenska språket (1986)
 Svenska ord — med uttal och förklaringar (1992)
 Sverrir Hólmarrson/Sanders/Tucker 1989 ⇒ Concise Icelandic-English Dictionary
 Söderwall 1884–1918 ⇒ Ordbok öfver svenska medeltidsspråket
 Sören Sörenson 1984 ⇒ Ensk-íslensk orðabók ...
 The BBI Combinatory Dictionary of English. A Guide to Word Combinations (1986), 246
 The Cambridge Encyclopedia (1990)
 The Oxford English Dictionary (1884–1928; 1989), 83, 88, 272, 376
 The Pictorial Dictionary of Sign Language ⇒ Viittomakielen kuvasanakirja
 The Thai Sign Language Dictionary (1990), 359
 Torp 1919 ⇒ Nynorsk etymologisk ordbok
 Verben in Feldern, Valenzwörterbuch zur Syntax und Semantik deutscher Verben (1986)
 Viittomakielen kuvasanakirja (1973)
 Vinterberg/Bodelsen 1990/1992 ⇒ Dansk-engelsk ordbog

VIF ⇒ Verben in Feldern ...

V&B ⇒ Dansk-engelsk ordbog (1990/1992)

Wahrig ⇒ Deutsches Wörterbuch

WDG ⇒ Wörterbuch der deutschen Gegenwartssprache

Webster 1978 ⇒ Webster's New Twentieth Century Dictionary of the English Language

Webster's New Twentieth Century Dictionary of the English Language (1978), 188

Wrigley 1990 ⇒ The Thai Sign Language Dictionary

Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch (1962), 246

Wörterbuch der deutschen Gegenwartssprache (1970–77)

Östergren 1981 ⇒ Nusvensk ordbok

Stikkordregister

A

- absolutte uttrykk, 154
- abstrakt betydelse, 325
- abstrakt verb, 325
- ad hoc-afledninger, 3, 4
- ad hoc-sammensætninger, 3, 4
- adjektiv
 - ikke-restriktivt, 149
 - participielt, 269–270
 - restriktivt, 149
- adjektivisk particip, 268, 270
- affikser, 221
 - bundne, 223
 - frie, 223
 - skabs-, 223, 226
- afledninger, 223
 - ad hoc-, 3, 4
 - skabs-, 221, 223, 226, 230
- akronym, 113, 116
- aktionsart, 53, 54, 292, 327
 - durativ, 54
 - inchoativ, 54
- aktionsartsbetydelse, 327
- aktiv ordbog, 52, 63
- aktiv ordbok, 389
- albogsleksikografi, 40, 41
- alfabetisering, 209, 436
- alfabetisk makrostruktur, 44
- alfabetisk ordliste, 94, 99
- allmänordböcker, 342, 388
- allmänspråklig ordbok, 16, 353, 388
- allmänspråkligt lexikon, 16
- allordbok, 391
- almen ordbog, 19
- almensprog, 43
- almensproglig leksikografi, 45
- almensproglige ordbøger, 7, 42, 96
- alordbøger, 40, 41
- analogiordbøger, 247
- analogisk informasjon, 247
- analogoversættelse, 119
- andraspråksinlærning, 159
- angivelser
 - uttals-, 280, 354
- anglicismer, 21, 428
- anglismar, 408
- antonymi, 3
- anvendt leksikologi, 40
- anvendt lingvistik, 39
- anvendt sprogteknologi, 75
- användaranvisningar, 356
- användarvänlig, 351
- arkaisme, 20, 404
- artikelelementer, 435
- artikelindhold, 435, 439
- artikelkategorier, 435
- artikelkema, 114
- artikelstruktur, 107, 110, 114
- attribut, 53
- autenticitet, 314
- autentisk belæg, 176
- autentiska exempel, 311, 313
- automatisk nummerering, 436
- automatisk ordudvælgelse, 3
- automatisk selektion, 3
- automatisk snuing, 213
- automatisk sortering, 98, 99
- automatisk udvælgelse, 3
- avkodningsordbok, 392
- avledda sammansætninger, 269
- avledning, 27, 153, 265, 292
 - kort-, 267
 - prefix-, 266
 - processuell, 292
 - resultativ, 292
 - substantiv-, 285
 - suffix-, 266
- avledningsbase, 266, 267
- avledningssuffix, 272
- avleiingar, 206, 404, 408
- avløserord, 416
- avløysarord, 407
- avstavning, 278, 355
- avstavningsmarkeringar, 279

B

basic meaning, 168
 basisordbøger, 71
 begreb, 44, 46
 begrebsinhold, 46
 begrebsmæssige relationer, 94
 begrebsordbøger, 45
 begrebssammenhæng, 44, 45
 begrebstilordninger, 44, 45
 begreppssystematik, 390
 begrepsfelt, 247
 begreps sammensmelting, 30, 31
 belegg, 336, 405, 406
 belæg, 311
 belæg

- autentisk, 176

 belægsamlinger, 95
 benämningar, 13

- prefix-, 14
- sammensatta, 13
- simplex-, 13
- öronmärkenas, 13

 benævnelse, 46
 beskrivelsesmodel, 79
 betydelse

- abstrakt, 325
- aktionsarts-, 327
- konkret, 325
- kärn-, 15
- rums-, 326, 327

 betydelseanalyse, 283
 betydelsebeskrivning, 160
 betydning

- del-, 120
- ikke-komposisjonell, 30
- kausativ, 54
- komposisjonell, 29, 30
- kompositionel, 53
- konkret, 195
- konnotativ, 253
- leksikalsk, 53
- overført, 195

 betydningsdifferensiering, 155
 betydningsdifferentiering, 196
 betydningsinddeling, 84

betydningsnummer, 154
 betydningsnyanse, 154, 417, 418
 betydningsordbog, 1
 bridge dictionaries, 159
 brolexikon, 159
 brugergruppe, 42
 brugergrænseflade, 438
 brugervejledning, 108, 109
 brugskode, 239
 brugssituation, 42
 bundne affikser, 223
 bundne rødder, 222
 bøyning, 153

C

CD-ROM, 375
 citation form, 128
 COBUILD-definitioner, 160
 collocation, 133
 compounding, 129, 258
 compounds, 255, 256

- lexicalized, 257
- multiple, 262
- productive, 258

 concordance, 123, 124, 167

- lemmatized, 123

 connotation, 168, 169
 constructions à verbe support, 54

D

danismer, 239, 242
 DANLEX-taksonomi, 72, 74, 76, 77, 79
 database

- faksimile-, 335
- lexikalisk, 17

 datamatisk leksikografi, 71
 datamatstøttet leksikografi, 75
 datasamling

- leksikalsk, 71, 72, 79, 80
- leksikografisk, 80

 datastruktur, 435
 datorbaserad leksikografi, 342
 definitionsordbok, 160, 275, 342
 deiksis, 151
 deiktisk, 152

delbetydninger, 120
delækvivalent, 118
demotivering, 31
denotatum, 29
derivations, 260
deskriptiv, 44, 284
deskriptiv ordbok, 385
deskriptivitet, 409
dialektord, 21
dictionary
 bridge, 159
 interlingual, 389
 sign language, 359

dictionary for foreign learners, 386
dictionary for native speakers, 389
disambiguering, 422
durativ aktionsart, 54
durativa verb, 327
durativitet, 139

E

edb-simulering, 6
edb-støj, 7
eftertext, 356, 384
egennamn, 344, 397
eksempler, 120
 illustrerende, 253
ekserpering, 372
ekserpt, 335, 413, 414, 417, 418
ekserptor, 413, 415–417
ekserptseddel, 331, 336
ekvivalent, 187, 287
 oversettelses-, 213, 217
elektronisk korpus, 58
elektronisk ordbog, 69
elektronisk ordbok, 377
elektronisk ordsamling, 207
elektronisk snuing, 213
elektronisk tekst, 334
elektronisk tekstkorpus, 195, 196, 198,
 413, 417
elektroniske faksimiler, 334
elektronisk leksikon, 342, 348, 376
elliptiske former, 115

empirisk basis, 94
encyklopedisk ordbok, 377
encyklopedisk informasjon, 11, 16
encyklopædi, 40–42, 46
encyklopædiske ordbøger, 40, 41
encyklopædologi, 41
engelske låneord, 428
enkeltordstermer, 98
enspråkig ordbok, 387
eponymordbok, 385
etnonym, 193
etnosemantisk fält, 11
etterledd, 418
ettordslemma, 153
etymologier, 277
eufemisme, 5
excerpere, 98
excerpering, 176, 177, 233, 236, 237, 240
excerpting, 256
excerption, 128, 130–132
exempel, 311
 autentiska, 311, 313
 grammatisk, 316
 konstruerade, 311, 313
 multifunktionella, 315
 ordboks-, 314
 tautologisk, 315
 äkta, 311

F

fagekspert, 43, 44
fagleksikografi, 43, 44, 46, 93
fagord, 24, 241, 308
fagordbog, 43–45, 93, 98, 423, 432
fagspecifikke ordbøger, 422, 424, 430
fagsprog, 42, 43, 60, 421
fagsproglig ordbog, 107
fagsproglige korpora, 424
fagsprogsordbøger, 96
fagspråk, 407
fagsystematik, 44, 45
fagterm, 43, 63
fagterminologi, 308
faksimiledatabase, 335
fast sammansättning, 324

- felt, 206
 begreps-, 247
 kollokations-, 114, 120
 lemma-, 114
 ord-, 2
 semantisk, 150
 ækvivalent-, 114, 117
 filtrering, 439
 finit verbum, 138
 fleksiv, 222
 flerfagsordbog, 45
 flerordsforbindelse, 202
 flerordstermer, 98, 99
 fogmorfem, 272
 folkdefinition, 349
 folketymologi, 349
 folktaxonomi, 349
 forbindelse
 flerords-, 202
 fri, 51
 forkortelser, 24
 forledd, 418
 fortetting, 185, 193
 frekvens, 153
 frekvenssuppgifter, 348
 fremmedord, 21
 fremmedsprogsækvivalenter, 4
 fremmendord, 404, 407
 frie affikser, 223
 frie forbindelser, 51
 frie rødder, 222
 fuld ækvivalens, 119
 fuldverbum, 137, 138, 140
 fullstendighet, 149
 function word, 131
 funktionsnomen, 52
 funktionsverbum, 51, 52, 137, 138, 139,
 141
 funktionsverbumbeskrivelse, 51, 52,
 137
 fyllnadsbestämningar, 278
 fyllnadsled, 278, 317
 fårmärkenbeteckningar, 16
 fält
 etnosemantiskt, 11
 förkortningar, 355
 förkortningsordbok, 386
 förtext, 356, 384
- G**
- gangbarhet, 149
 genbrug, 70
 genbrug af data, 72
 generell referanse, 29
 gennemskuelighed, 22
 genuint formål, 254
 gjenbruk, 210
 gloss, 359
 glosse, 253
 glossering, 117, 254
 glossografi, 38
 grammatikkontrol, 69
 grammatisk subjekt, 53, 54
 grammatiskt eksempel, 316
 grunnord, 206
 gränsrelaterade verb, 327
 gränsrelatering, 327, 328
- H**
- handlingsverber, 144
 hapax legomenon, 177, 180, 405
 head sign, 360
 henvisning
 implicit, 196
 henvisningsstruktur, 110
 Herübersetzungswörterbuch, 389
 Hinübersetzungswörterbuch, 389
 historisk ordbog, 175, 176
 Hjulstad-format, 206
 hjælpeverber, 137
 homografer, 206
 homografseparering, 209
 homonym, 124, 125, 259
 homonymi, 45, 155
 homonymy, 361
 hovedlemma, 195, 247
 hovedværk, 83, 87, 89, 90
 humanordbøger, 70, 71, 78
 husterminologi, 302
 huvudord, 265, 267

huvudstickord, 354
huvuduppslagsord, 267, 270, 271
hyperonymer, 276
hyperonymirelationer, 281
hyponymer, 276, 281
 ko-, 276, 281
hyponymirelationer, 281
håndbøger, 46

I

icke-förutsägbara sammansättningar, 290
icke-gränsrelaterade verb, 327
idiolet, 20
idiom, 196, 200, 249
idiomaticitet, 139
idiomatiska uttryck, 283
idiomatiska uttryk, 53
idiomordbok, 392
ikke-komposisjonell betydning, 30
ikke-lemmatisering, 202
ikke-restriktivt adjektiv, 149
ikkje-alfabetiske teikn, 209
illustrerende eksempler, 253
implicit henvisning, 196
implicit information, 57
implisitt informasjon, 250
inchoativ aktionsart, 54
indeks, 96
 polysemi-, 118
indgangslemma, 45
indskrivningskontrol, 436
informant, 43, 44, 237
informasjon
 analogisk, 247
 implisitt, 250
 leksikografisk, 245
 paradigmatisk, 247
informasjonstyper, 250
information
 encyklopedisk, 11, 16
 implicit, 57
 lexikalisk, 342, 347
 uttals-, 280
informationskategorier, 275
inkoativitet, 139

inlärningslexikon, 161
inlärningsordböcker, 311
interlingual dictionary, 389
introspektion, 20, 98

K

kanonisk form, 347, 349
karitativa bildningar, 285
kategori
 artikel-, 435
 informations-, 275
 niveau-, 436
 ord-, 11
kausativ betydning, 54
kausativitet, 53, 139
kilderepræsentation, 5
kohyponymer, 276, 281
kollokasjon, 249
kollokater, 102
kollokation, 111, 120, 160, 195, 196
 lexikalisk, 160
kollokationsfelt, 114, 120
kollokationsordbok, 392
kollokatorer, 252
kollokvialismer, 21
kolumntitlar, 356
kombinatorisk ordbok, 246
kommunikationsniveau, 58
kommunikativ funktion, 314
kommunikativ kontekst, 314
kommunikativ relevans, 314
komplement, 278
komposisjonell betydning, 29, 30
komposita, 427
komposisjonell betydning, 53
konkordans, 331
 KWIC-, 331
konkordansliste, 198, 199, 202
konkordansprogram, 59
konkret betydelse, 325
konkret betydning, 195
konkret verb, 325
konnotativ betydning, 253
konsistens, 149
konstruerade exempel, 311, 313

- konstruktion
 funktionsverbums-, 51, 52, 137
 nominaliseringsverbums-, 54
 passiv-, 55
 støtteverbums-, 54
 konstruksjonstype, 52
 konstruksjonssuppgifter, 277
 kontekst
 kommunikativ, 314
 stark, 315, 318
 kontroll
 grammatik-, 69
 indskrivnings-, 436
 stave-, 69, 237
 stil-, 69
 konvensjonalisering, 29, 31
 konvensjonalisert sammensetning, 30
 konversationsmaximer, 314
 kopulaverbum, 138
 korpus, 331
 elektronisk, 58
 fagsproglig, 424
 maskinlæsbart, 111
 tekst-, 96, 97, 99, 237, 331, 413, 417
 korpusbaserte systemordbøker, 424
 korpusforekomst, 4
 korpusleksikografi, 365
 korrekthet, 149
 kortavledning, 267
 kryptorødder, 223
 kryptosammensætninger, 221, 223, 229
 kumulativ ordbok, 388, 391
 KWIC-konkordans, 331
 källspråk, 284
 kärnbetydelse, 15
- L**
- lejlighedsdannelser, 226
 leksem, 222
 leksikalisere, 382, 405
 leksikaliserte sammensætninger, 230
 leksikalisering, 27, 153, 246
 leksikaliseringsgrad, 22, 30, 32
 leksikaliseringsprosess, 27, 28
 leksikaliseringsskala, 31, 32
 leksikalsk betydning, 53
 leksikalske data, 72
 leksikalske datasamlinger, 71, 72, 79, 80
 leksikograf, 43
 sprog-, 40
 leksikografi, 37–39, 41–44, 46, 375
 albogs-, 40, 41
 almensproglig, 45
 datamatisk, 71
 datamatstøttet, 75
 fag-, 43, 44, 46, 93
 korpus-, 365
 meta-, 39, 40, 42
 multimedia-, 375
 praktisk, 39, 40
 sag-, 41
 sprog-, 40–42
 teoretisk, 39
 leksikografisk beskrivelse, 77
 leksikografisk informasjon, 245
 leksikografisk opslagsværk, 40, 41, 46,
 73
 leksikografiske datasamlinger, 80
 leksikologi, 37, 40
 anvendt, 40
 leksikon, 40, 41, 43, 46
 elektronisk, 376
 lemma, 115
 ettords-, 153
 hoved-, 195, 247
 indgangs-, 45
 monosemt, 118
 polysemt, 118
 sub-, 195, 196, 201, 247
 lemmabestand, 94, 107, 246, 254
 lemmafelt, 114
 lemmakandidater, 3, 6, 8
 lemmalista, 384
 lemmaliste, 1, 5, 245
 nestalfabetisk, 245
 systematisk, 94
 lemmaluckor, 285
 lemmaseleksjon, 245, 249
 lemmaseleksjon, 3, 4, 8, 19, 25, 93, 98,
 99, 110, 111, 238, 283, 424, 425

lemmatisera, 289
lemmatisere, 196, 200, 201, 241
lemmatisering, 153, 195, 196, 201, 202,
277, 284, 285, 347
 ikke-, 202
lemmatization, 124
lemmatized concordance, 123
lemmaoppsättning, 283, 284
lexeme, 127
lexicalization, 255, 258
lexicalized compounds, 257
lexikalisering, 267, 285, 321, 325, 326,
328, 347, 349
lexikaliseringsgrad, 324, 328
lexikaliseringsprocess, 349
lexikalisk information, 342, 347
lexikalisk kollokation, 160
lexikalisk kompetens, 343, 346
lexikaliska databaser, 17
lexikaliska regler, 313
lexikaliskt delsystem, 11, 15, 16
lexikaliskt system, 11
lexikograf, 339
lexikografi, 339
 datorbaserad, 342
 praktisk, 340
lexikografisk teori, 339
lexikografisk tradition, 344
lexikologi, 317
lexikon
 allmänspråkligt, 16
 bro-, 159
 elektronisk, 342
 inlärnings-, 161
lexikonmodell, 342–344, 346, 347, 349
lista, 387
loan words, 259
logisk subjekt, 53
låneord, 21, 119, 238, 241, 428
 engelske, 428
låneoversættelse, 119
lös sammansättning, 324

M

macrostructure, 255
makrostruktur, 46, 93, 108–110, 245, 284,
384
 alfabetisk, 44
 striktalfabetisk, 245
 systematisk, 44, 45, 93, 424
markeringar
 avstavnings-, 279
 uttals-, 280
maskinbrugbar ordbog, 70, 71, 77
maskinel oversættelse, 69, 75
maskinel sprogbehandling, 71, 75
maskinlæsbare ordbøger, 70
maskinlæsbart korpus, 111
maskinlæsbart tekstkorpus, 111
maskinoversættelse, 52, 64, 421, 424
maskinoversættelsesordbøger, 421
maskinoversættelsessystem, 79
maskinoversættelsessystemet METAL,
421
maskinstøttet oversættelse, 69, 75
materialinnsamling, 416
maximer, 314
mediostruktur, 110, 111
megastruktur, 384
mellanspråklig ordbok, 389
meta-, 39
metaleksikografi, 39, 40, 42
metonymi, 116
mikrostruktur, 46, 108, 109, 246, 284,
384, 435
modeord, 226
momentana verb, 327
monoordbøger, 422
monosemere, 55
monosemt lemma, 118
morfem
 fog-, 272
multifunktionella exempel, 315
multimedialeksikografi, 375
multiple compound, 262
multispektion, 98, 99
mutuel statistik, 197, 198, 202

målgruppsanpassad, 351
målspråk, 284

N

namnordbok, 393, 397
nationalitetsord, 344, 345
nemne, 296
neologisme, 7, 20, 21
neologisms, 259
nestalfabetisk lemmaliste, 245
nesting, 354
niveau-kategorier, 436
nomina actionis, 279
nominaliseringsverbumbeskrivelse, 54
nominalstil, 60
normativ, 284
normativ ordbok, 388, 390, 409
normativitet, 409, 411
normere, 415
normering, 409, 417, 418
 språk-, 415
nummerering
 automatisk, 436
nydannelser, 239, 241
nyord, 404, 405, 408, 413, 415
nyordbok, 415
nyordsproduksjon, 229
nyordssamling, 415
næsten-synonymer, 3

O

omsetjingar, 406
omtekst, 94
onomasiologisk, 44
onomasiologisk ordbok, 246, 392
oplysninger
 real-, 42
 sproglike, 42
oplysningstyper, 2, 66, 74, 77, 436
opplagsverdi, 246
oppføring, 366
opplagsenheter, 72
ord, 42
ord-for-ord-metode, 113
ord for ord-utvalg, 3

ordbygging, 285
ordbok, 41, 43, 46
 aktiv, 52, 63
 al-, 40, 41
 almen, 19
 almensproglig, 7, 42, 96
 basis-, 71
 begreps-, 45
 betydnings-, 1
 elektronisk, 69
 encyklopædisk, 40, 41
 fag-, 43-45, 93, 98, 423
 fagspesifikk, 422, 424, 430
 fagsproglig, 107
 fagsprogs-, 96
 flerfags-, 45
 historisk, 175, 176
 human-, 70, 71, 78
 maskinbrukbar, 70, 71, 77
 maskinlesbar, 70
 maskinoversettelses-, 421
 mono-, 422
 oversettelses-, 107
 papir-, 43, 45
 papirbasert, 66
 passiv, 63
 periode-, 86, 91
 samtid-, 91
 spesial-, 107
 system-, 423
 transfer-, 422
 valens-, 46
ordbogsrepresentasjon, 5
ordbogsstrukturer, 40, 46, 109
ordbogsverk, 83
ordbok
 aktiv, 389
 allmän, 342, 388
 allmänspråklig, 16, 353, 388
 al-, 391
 analogi-, 247
 avkodnings-, 392
 definisjons-, 160, 275, 342
 deskriptiv, 385
 elektronisk, 377

encyclopedisk, 377
 enspråkig, 387
 eponym-, 385
 förkortnings-, 386
 idiom-, 392
 inlärnings-, 311
 kollokations-, 392
 kombinatorisk, 246
 kumulativ, 388, 391
 mellanspråklig, 389
 namn-, 393, 397
 normativ, 388, 390, 409
 ny-, 415
 onomasiologisk, 246, 392
 paradigmatiske, 387
 passiv, 389
 perceptions-, 343
 produktions-, 246
 produktions-, 284, 343, 389
 receptions-, 384, 385, 389, 392
 rese-, 353
 resepsjons-, 246
 rettskrivings-, 205
 sak-, 391
 semasiologisk, 245, 246, 392
 special-, 342, 390, 393
 stor-, 283
 synonym-, 246
 syntagmatiske, 387
 tospråklig, 185
 valens-, 393
 verbvalens-, 317
 vetenskaplig, 388
 översättnings-, 389, 394
 ordbok för främmandespråksinläring,
 386
 ordbok för modersmålsbrukare, 389
 ordbok över främmande ord, 391
 ordboksexempel, 314
 ordfamiljer, 285
 ordfelter, 2
 ordgransking, 308
 ordkategori, 11
 ordledsanalyser, 279
 ordlista, 390

ordliste, 108, 179, 180, 205
 alfabetisk, 94, 99
 rettskrivings-, 205
 ordsamling, 177, 178
 elektronisk, 207
 ordudvælgelse, 1, 3
 automatisk, 3
 ordutval, 410
 overført betydning, 195
 overhyppighed, 199
 oversettelsesekvivalent, 213, 217
 oversættelse, 42
 analog-, 119
 låne-, 119
 maskinel, 69, 75
 maskin-, 52, 64, 421, 424
 maskinstøttet, 69, 75
 oversættelsesforslag, 119
 oversættelsesordbog, 107

P

papirbaserede ordbøger, 66
 papiordbøger, 43, 45
 paradigmatiske informasjon, 247
 paradigmatiske ordbok, 387
 particip, 269, 285
 adjektivisk, 268, 269
 perfekt-, 268
 presens-, 268
 verbal-, 22
 participiellt adjektiv, 268–270
 partiel ækvivalens, 119
 partikelverb, 196, 270, 292, 321, 326
 partikkelverb, 247
 passiv ordbog, 63
 passiv ordbok, 389
 passiv översättningsordbok, 389
 passivkonstruktion, 55
 perceptionsordbok, 343
 perfektparticip, 268
 periodeordbog, 86, 91
 personnavne, 24
 polysem, 55, 154
 polysemi, 44, 45, 154, 155
 polysemiindeks, 118

polysemt lemma, 118
 polysemy, 362
 post, 206
 praktisk leksikografi, 39, 40
 praktisk lexikografi, 340
 preeditering, 425, 430
 prefix, 266
 prefixavledning, 266
 prefixbenämning, 14
 presensparticip, 268
 presentasjonsform, 376
 processuella avledningar, 292
 procesverber, 144
 productive compounds, 258
 productivity, 258
 produksjonsordbok, 246
 produktion, 42
 produktionsordbok, 284, 343, 389
 proprier, 24
 prosodisk sammensetning, 29, 33
 prospektiv, 314
 prototypisk, 315
 præfiks, 221
 præskriptiv, 44
 pseudo-komposita, 429
 purisme, 238, 241
 puristisk sprogpolitikk, 238

R

rammestruktur, 109
 realoplysninger, 42
 reception, 42
 receptionsordbok, 384, 385, 389, 392
 redigeringsystem, 433
 referanse, 29
 generell, 29
 spesifikk, 29
 referensologi, 41
 referent, 11
 regelanvändningar, 314
 regelformulering, 313
 register
 stikords-, 96
 systematisk, 98
 reinholdsterminologi, 307

relationer

begrebsmæssige, 94
 hyperonymi-, 281
 hyponymi-, 281
 representativitet, 409, 416
 reseordbok, 353
 resepsjonsordbok, 246
 restriktivt adjektiv, 149
 resultativa avledningar, 292
 retrogradbildninger, 270
 retrospektiv, 314
 rettskrivingsordbok, 205
 rettskrivingsordliste, 205
 rod, 221
 bunden, 222
 fri, 222
 krypto-, 223
 rumsbetydelse, 326, 327
 running words, 127

S

sag, 41, 42
 sagleksikografi, 41
 sakordbok, 391
 samansetningar, 206, 404, 405, 408
 samansetningsfuge, 206
 sambeskrivning, 86
 samforekomst, 195, 198, 199
 samleartikkel, 185, 190
 samleillustrasjoner, 380
 samling
 belæg-, 95
 data-, 71, 72, 79, 80
 nyords-, 415
 ord-, 177, 178, 207
 seddel-, 233, 236, 237
 sitat-, 336
 term-, 71
 sammansatt benämning, 13
 sammansättning, 265, 266, 290, 324
 avledd, 269
 fast, 324
 icke-förutsägbar, 290
 lös, 324
 varierende, 324

- sammansättningstyp, 324
sammenfletning, 87, 89, 90
sammensetning, 27, 28, 418
 konvensjonalisert, 30
 prosodisk, 29, 33
sammenskriving, 32, 33
sammenskriving av sammensetninger, 33
sammensætning, 223
 ad hoc-, 3, 4
 krypto-, 221, 223, 229
 leksikalisert, 230
samsortering, 209
samtdsordbog, 91
seddelarkiv, 331
seddelsamling, 233, 236, 237
seleksjon
 automatisk, 3
 lemma-, 3, 4, 8, 19, 25, 93, 98, 99,
 110, 111, 238, 283, 424, 425
 term-, 424, 425
seleksjonskriterier, 6
semantic class, 171, 173, 174
semantic classification, 167, 168, 171,
 174
semantic field, 174
semantisk felt, 150
semantisk feltteori, 150
semantisk korrelat, 31
semasiologisk ordbok, 245, 246, 392
SGML, 207, 434
sign language dictionaries, 359
simpleks, 31
simpleksord, 3
simpleksverbum, 138, 139, 141
simpleksbenämning, 13
simplexverb, 328
sitatsamling, 336
sitatteknikk, 149
skabsaffikser, 223, 226
skabsafledninger, 221, 223, 226, 230
slang, 24
slangord, 24
slangudtryk, 20
snuing
 automatisk, 213
 elektronisk, 213
sortering
 automatisk, 98, 99
 sam-, 209
specialordbog, 107
 systematisk, 98
specialordbok, 342, 390, 393
specialspråk, 16
spesifikk referanse, 29
spontanlaging, 29, 31
sprog, 41
sprogleksikograf, 40
sprogleksikografi, 40–42
sproglige opplysninger, 42
sprogligt tegn, 44, 46
sprogteknologi, 71, 75
 anvendt, 75
sprogteknologiske produkter, 69
sprogteknologiske værktøjer, 73
språkinlärning, 348
språknormering, 415
språkteknologi, 213
stamme, 222
standardisering, 71
STANLEX-taksonomi, 76, 77
stark kontekst, 315, 318
stavekontrol, 69, 237
stickord, 355
 huvad-, 354
 under-, 354
stikordsregister, 96
stilkontrol, 69
stillag, 65
stilvärde, 292
storordbok, 283
Streckverbum, 51
striktalfabetisk makrostruktur, 245
struktur
 artikel-, 107, 110, 114
 data-, 435
 henvisnings-, 110
 makro-, 46, 93, 108–110, 245, 284,
 384
 medio-, 110, 111
 mega-, 384

mikro-, 46, 108, 109, 246, 284, 384,
 435
 ordbogs-, 40, 46, 109
 ramme-, 109
 tilgangs-, 46
 strukturingshierarki, 109
 støtteverbum, 51
 støtteverbumbeskrivelse, 54
 subjekt
 grammatisk, 53, 54
 logisk, 53
 sublemma, 195, 196, 201, 247
 sublemmatisere, 153
 substantivavledning, 285
 substantivering, 55
 substantivfølger, 32, 33
 suffiks, 221
 suffix, 266
 avlednings-, 272
 suffixavledning, 266
 supplement, 83, 86, 87, 89, 90, 236, 356
 synkron, 44
 synonym ækvivalent, 118
 synonymer, 185, 186
 næsten-, 3
 synonymi, 3, 99
 synonymordbok, 246
 synonymrekke, 185, 186, 188
 syntagmatisk ordbok, 387
 systematik
 begrepps-, 390
 fag-, 44, 45
 term-, 99
 syntaktisk mønster, 195
 syntetiske uttrykk, 154
 systematisk lemmaliste, 94
 systematisk makrostruktur, 44, 45, 93,
 424
 systematisk register, 98
 systematisk specialordbok, 98
 systemordbøker, 423
 korpusbaserede, 424
 særskrivning, 33

T

tabuord, 24
 taggerlemmatiser, 367
 taksonomi, 79
 DANLEX-, 72, 74, 76, 77, 79
 STANLEX-, 76, 77
 talemåte, 196
 talemålsgransking, 308
 tautologisk, 318
 tautologisk eksempel, 315
 tekstkorpus, 96, 97, 99, 237, 331, 413,
 417, 424
 elektronisk, 195, 196, 198, 413, 417
 maskinlæsbart, 111
 tekstproduksjon, 44
 tekstsøgning, 439
 teoretisk leksikografi, 39
 term, 44, 46, 111, 296
 enkeltords-, 98
 fag-, 43, 63
 flerords-, 98, 99
 termbaser, 69
 terminografi, 42–44, 46, 71
 terminolog, 43, 44
 terminologi, 42–44, 46, 421
 fag-, 308
 hus-, 302
 reinhalds-, 307
 terminologibaser, 78
 termkrav, 426
 termsamlinger, 71
 termseleksjon, 424, 425
 termsystem, 16
 termsystematik, 99
 textualitet, 313
 tilgangsstruktur, 46
 tillægsord, 278
 tilstandsverber, 144
 toolkit, 367
 tospråklige ordbøker, 185
 transferordbøker, 422
 transitivitet, 160
 tunggrepp, 356
 typografi, 356

U

- udtryk
 - idiomatiska, 53
 - slang-, 20
- udvælgelse
 - automatisk, 3
 - ord for ord-, 3
- understickord, 354
- uppgifter
 - frekvens-, 348
 - konstruktions-, 277
 - valens-, 292
- uppslagsord, 344, 347, 349, 356
 - huvud-, 267, 270, 271
- uttalsangivelser, 280, 354
- uttalsinformation, 280
- uttalsmarkeringar, 280
- uttryck
 - idiomatiska, 283
- uttrykk
 - absolutte, 154
 - syntetiske, 154

V

- valens, 195, 277
- valensforskning, 317
- valensled, 325
- valensmønster, 195, 198
- valensordbog, 46
- valensordbok, 393
- valensroller, 326, 328
- valensuppgifter, 292
- varierande sammansättning, 324
- vedligeholdelsesfunktioner, 439
- verb
 - abstrakt, 325
 - durativa, 327
 - gränsrelaterade, 327
 - icke-gränsrelaterade, 327
 - konkret, 325
 - momentana, 327
 - partikel-, 270, 292, 321, 326
 - partikkel-, 247
 - simplex-, 328
- verbalabstrakt, 267

- verbalparticipier, 22
- verbalsubstantiv, 138, 144, 267, 278, 279
- verbum
 - finit, 138
 - fuld-, 137, 138, 140
 - funktions-, 51, 52, 137, 138, 139, 141
 - handlings-, 144
 - hjælpe-, 137
 - kopula-, 138
 - partikel-, 196
 - proces-, 144
 - simpleks-, 138, 139, 141
 - Streck-, 51
 - støtte-, 51
 - tilstands-, 144
- verbvalensordbok, 317
- verifisering, 366
- vetenskaplig ordbok, 388
- vulgarismer, 20

W

- word formation, 258, 259
- word frequency, 126
- Wörterbuch
 - Herübersetzungs-, 389
 - Hinübersetzungs-, 389

Ä

- äkt exempel, 311
- ämnnesområde, 355

Æ

- ækvivalens
 - fuld, 119
 - partiel, 119
- ækvivalent, 118
 - del-, 118
 - fremmedsprogs-, 4
 - synonym, 118
- ækvivalentfelt, 114, 117

Ö

- öronmärken, 11, 12, 14, 16
- öronmärkenas benämningar, 13

öronmärkenas form, 13
öronmärkeskombinationer, 13
översättningsordbok, 389, 394