

# Nordiske Studier i Leksikografi



# Nordiske Studier i Leksikografi

9

Rapport fra konference om  
leksikografi i Norden  
Akureyri 22.–26. maj 2007

Redigeret af

Ásta Svavarsdóttir

Guðrún Kvaran

Gunnlaugur Ingólfsson

Jón Hilmar Jónsson

Skrifter udgivet af Nordisk Forening for Leksikografi  
Skrift nr. 10

I samarbejde med Språkrådet i Norge og  
Árni Magnússon instituttet for islandske studier

Reykjavík 2008

Nordiske studier i leksikografi, nr. 9

© Nordisk forening for leksikografi og forfatterne, 2008

Redaktion: Ásta Svavarsdóttir, Guðrún Kvaran, Gunnlaugur Ingólfsson og Jón Hilmar Jónsson

Bogtilrettelæggelse og sats: Bessi Aðalsteinsson

Sat med Andron Corpus

Omslag: Bessi Aðalsteinsson og redaktionen

Tryk: Leturprent

ISBN 978-9979-654-05-6

ISSN 0803-9313

Støtte fra:

Nordplus Sprog

Clara Lachmanns Fond

Letterstedska föreningen

Minningarsjóður Margrétar Björgólfsdóttur

MS

JPV-útgáfa

Glitnir

Edda – útgáfa

Akureyrarbær

Dalvíkurbyggð

Sparisjóður Svarfdæla

Rapporten kan bestilles hos:

Språkrådet

Postboks 8107 Dep

NO-0032 Oslo

e-post: [post@sprakradet.no](mailto:post@sprakradet.no)

## Indhold

Forord . . . . .	ix
<i>Torben Arboe</i> Fraseologiske enheders indplacering i Jysk Ordbog . . . . .	1
<i>Ari Páll Kristinsson</i> Den nye islandske rettskrivningsordboka . . . . .	19
<i>Kristin Bakken og Oddrun Grønvik</i> Materialsortering på digital plattform Eit steg mot høgare dataintegritet i den vitskaplege leksikografien . . . . .	31
<i>Anna Braasch</i> Verber med fælles syntaktiske konstruktioner og semantisk slægtskab . . . . .	43
<i>Lars Brink</i> Leksikalske huller . . . . .	61
<i>Torben Christiansen</i> Fra redaktørens værksted Om færdigredigeringen af stor Tysk-Dansk Ordbog . . . . .	69
<i>Silvie Cinková</i> “Movement towards Structure” Basic Verbs in Learners’ Lexicons . . . . .	77
<i>Mette Gismerøy Ekker og Christian-Emil Ore</i> Norrøn metaordbok – en rapport fra en øvelse . . . . .	89
<i>Eli Johanne Ellingsve</i> Digitale vegar til Norsk Ordbok . . . . .	105
<i>Ken Farø og Erla Hallsteinsdóttir</i> Hvad er en videnskabelig flersproget idiomordbog? . . . . .	113
<i>Ruth Vatvedt Fjeld</i> Plan for leksikalsk dokumentasjon av moderne bokmål . . . . .	131
<i>Alexandra Granström</i> Ordböcker och informationsbeteende En gränsöverskridande studie i ordboksanvändning . . . . .	143

<i>Oddrun Grønvik</i> Fast redigerings skjema for verb? Semantikk versus grammatikalisering . . .	155
<i>Asgerd Gudiksen og Henrik Hovmark</i> Båndoptagelser som kilde til Ømålsordbogen . . . . .	173
<i>Guðrún Kvaran og Heimir Steinarsson</i> En ny tysk-islandsk ordbog. Metode og problemer . . . . .	183
<i>Halldóra Jónsdóttir og Þórdís Úlfarsdóttir</i> ISLEX: Islandsk-skandinavisk webordbog. . . . .	197
<i>Inger Schoonderbeek Hansen</i> Jysk Ordbog – Rapport fra en digital ordbog . . . . .	209
<i>Helgi Haraldsson</i> Om noen praktiske spørsmål i tospråklig leksikografi . . . . .	219
<i>Henrik Hovmark</i> Konventionaliserede forbindelser med danskeretningsadverbier Leksikografisk repræsentation og funktion . . . . .	229
<i>Ann-Kristin Hult</i> Från ord till handling En studie i ordboks användning på nätet . . . . .	243
<i>Anita Ågerup Jervelund og Jørgen Schack</i> På strejftog indenfor Retskrivningsordbogens territorie Retskrivningsordbogen 2011 . . . . .	259
<i>Jón Hilmar Jónsson</i> Lemmatisering og tilgangsstruktur Refleksjoner på bakgrunn av Íslenskt orðanet . . . . .	269
<i>Kjeld Kristensen</i> Svensk-Dansk Ordbog – tradition og fornyelse . . . . .	283
<i>Patrick Leroyer</i> Maultasche og Novillada? En teori for turistordbøger . . . . .	291
<i>Nina Martola</i> Angivelse av verbkonstruktioner i svenska ordböcker . . . . .	309
<i>Henrik Nikula</i> Valensinformation i ordböcker ur användarens synvinkel . . . . .	327

<i>Kristina Nikula</i> <i>Pensionär – aldrig i livet</i> De äldre i ordböckerna . . . . .	337
<i>Bolette Sandford Pedersen, Sanni Nimb og Lars Trap-Jensen</i> DanNet: udvikling og anvendelse af det danske wordnet . . . . .	353
<i>Dagfinn Rødningen og Knut E. Karlsen</i> Leksikografi i pedagogisk perspektiv Ein presentasjon av Norsk Ordbok for skoleelevar og studentar . . . . .	371
<i>Henrik Købler Simonsen</i> Virksomhedsleksikografien viser tænder Leksikografiske løsninger i København Zoo og Fagerberg A/S . . . . .	383
<i>Martin Skjekkeland</i> Noreg og Nederland – eit tett samband i to hundre år . . . . .	399
<i>Simon Skovgaard Boeck</i> Fire tematiske ordbøger fra renæssancens Danmark . . . . .	409
<i>Emma Sköldberg</i> Från vrångstruppen till fotabjället Om presentationen av unika konstituenten i Svensk ordbokutgiven av Svenska Akademien . . . . .	421
<i>Emma Sköldberg og Maria Toporowska Gronostaj</i> Modell för beskrivning av kollokationer i ett medicinskt lexikon (MedLex) . . . . .	433
<i>Lars Trap-Jensen</i> Tilgangs- og henvisningsstruktur i digitale ordbøger Overvejelser baseret på ordnet.dk . . . . .	447
<i>Lars S. Vikør</i> Behandling av fraseologi i Norsk Ordbok . . . . .	465
<i>Dagfinn Worren</i> “Bjørn Haldorsens islandske Lexicon” og Ivar Aasens ordbøker . . . . .	477
Stikkordregister . . . . .	487





## Forord

Den niende *Konference om leksikografi i Norden* blev afholdt i Akureyri i dagene 22.-26. maj 2007. Arrangørerne var Nordisk forening for leksikografi, Árni Magnússon instituttet for islandske studier og Språkrådet i Norge. Den lokale organisationskomité bestod af Ásta Svavarsdóttir, Guðrún Kvaran, Halldóra Jónsdóttir og Jón Hilmar Jónsson. Som ved de tidligere konferencer bistod Rikke Hauge fra Språkrådet i Norge med praktisk støtte ved arrangementet.

Der var registreret 107 deltagere på konferencen, og der blev holdt 40 foredrag. Af disse er 37 foredrag omarbejdet til de artikler som udgives i denne konferencerapport. Som sædvanlig var temaet for konferencen ikke afgrænset til et bestemt område, men man kan trygt hævde at programmet og artiklerne i rapporten vidner om bredden og mangfoldigheden i nordisk leksikografi.

Under konferencen, som fandt sted i Hotel KEAs lokaler, blev der arrangeret en udstilling af ordbøger i trykt og elektronisk form og andre leksikografiske skrifter fra nordiske forlag og institutioner.

Det sociale program bestod bl.a. af en udflugt til Dalvík og det omliggende område på vestsiden af Eyjafjörður og en reception af bystyret i Akureyri kommune.

Redigeringen af konferencerapporten fandt sted på Árni Magnússon instituttet for islandske studier under redaktion af Ásta Svavarsdóttir, Guðrún Kvaran, Gunnlaugur Ingólfsson og Jón Hilmar Jónsson. Bessi Aðalsteinsson har haft ansvaret for layout. Bortset fra én artikel, som er skrevet på engelsk, med et sammendrag på svensk, er samtlige artikler skrevet på et af de nordiske sprog dansk, norsk bokmål, nynorsk eller svensk, og alle artikler har et engelsk resumé. Bagest i rapporten er der et stikordsregister over centrale faglige ord og udtryk med sidehenvisninger til relevante artikler. Hver artikel bliver efterfulgt af oplysninger om forfatteren (arbejdsplads, e-mail-adresse m.m.).

Konferencen har modtaget økonomisk støtte fra mange instanser, både til arrangementet og til udgivelsen af konferencerapporten. Nordiske bidrag kommer fra Nordplus Sprog, Clara Lachmanns Fond og Letterstedtska föreningen. Islandske bidragsydere er: Margrét Björgólfssdóttirs mindefond, mælkeproducenten MS, forlagene Edda-útgáfa og JPV-útgáfa, Glitnir Bank, sparekassen Sparisjóður Svarfdæla og kommunerne Akureyri og Dalvíkurbyggð.

Redaktørene takker alle som har bidraget til udgivelsen af denne rapport og retter en varm tak til forfatterne for godt samarbejde under redigeringsfasen.

Reykjavík, december 2008

Ásta Svavarsdóttir    Guðrún Kvaran    Gunnlaugur Ingólfsson    Jón Hilmar Jónsson



TORBEN ARBOE

## Fraseologiske enheders indplacering i Jysk Ordbog

The topic is presented within a framework of relevant overview schemata in Lüger 1999 and on the basis of especially the latest published parts of the Jutlandic Dictionary ([www.jyskordbog.dk](http://www.jyskordbog.dk)), from HA- to HU-. Comments are made as to the terminological discussion of e.g. “phraseologisms”, “phrasemes”: they may be used as terms not only for units below sentence-level, but also for idioms with sentence-structure (cf. chapters 2–3). Particular attention is drawn to “one-word-phrasemes” (juxtapositions, chap. 2.3), and to “phraseme-lemmas”, i.e. lemmas created just to account for a special phraseme (chap. 4). Concludingly, the microstructures of very large articles (nouns and verbs) are inspected as to the precise placing of the large amount of set phrases (chap. 5).

### 1. Indledning

I *Jysk Ordbog*, [www.jyskordbog.dk](http://www.jyskordbog.dk), skelner vi for det fraseologiske stofs vedkommende grundlæggende mellem FASTE FORBINDELSER og TALEMÅDER. I princippet medtages alle forbindelser, der ud fra materialet må betegnes som faste, for at ordenes grammatiske brug kan blive belyst så fuldstændigt som muligt. Også mange af de forbindelser, der kendes fra rigsmål, nævnes, men behandles ikke indgående. For talemåderne er vi ikke forpligtet til at medtage alle, da vi ved mange ord har en mængde talemåder, hvoraf mange er velkendte fra rigsmål; her præsenterer vi først og fremmest de, der er jyske eller især jyske. Vi har en forpligtelse til at angive udbredelsesområdet, i det mindste omtrentligt, for både de faste forbindelser og de talemåder, vi bringer. Vi har tidligere skelnet mellem ordsprog og talemåder, men det viste sig for besværligt, usikkert og tidkrævende i praksis, så vi placerer nu ordsprog blandt talemåderne.

Men inden jeg går nærmere ind på indplaceringen af faste forbindelser og talemåder i ordbogens mikrostruktur, i de enkelte artikler, vil jeg som ramme om emnet gennemgå en oversigt over de forskellige typer af fraseologiske enheder. Oversigten stammer fra Lüger 1999 og bygger på Lügers disputats et par år tidligere; den omfatter ganske vist flere fraseologiske størrelser, end vi finder i *Jysk Ordbog*, men giver det bedste overblik, jeg hidtil har set, over fraseologiske enheders “placering” i forhold til hinanden. Det samlede oversigtsskema ses som figur 1; det ses, at i det grundlæggende “koordinatsystem” er FLERLEDDETHED afsat som dimension ud ad x-aksen, FASTHED tilsvarende som dimension ud ad y-aksen. Af tryktekniske grunde vises her ikke totalskemaet med betegnelser for alle de fraseologiske enheder, men kun de mest overordnede størrelser; de øvrige enheder og termer vises på fig. 2–4.

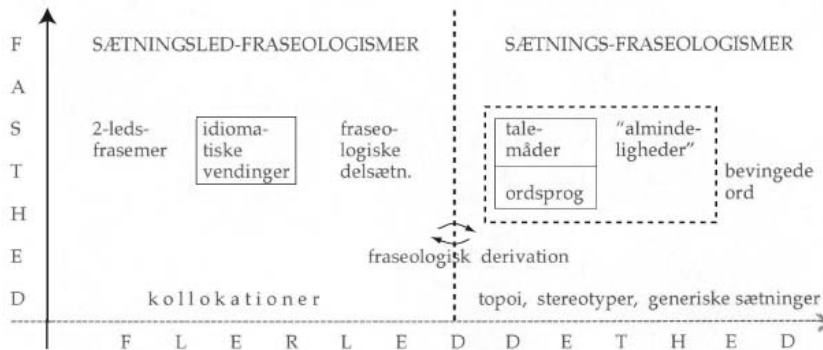


Fig. 1: Fraseologiske enheder: kerneområder og enkelte randområder. Oversigt ud fra Lüger (1999) p. 49ff., 129ff.

Betegnelserne SÆTNINGSLEDS-FRASEOLOGISMER og SÆTNINGS-FRASEOLOGISMER afspejler Lügers “satzgliedwertige Phraseologismen” hhv. “satzwertige Phraseologismen”. Men jeg har ikke dermed lagt mig fast på betegnelsen fraseologisme frem for frasem; jeg bruger de to ord som (næsten) synonyme.<sup>1</sup> Lügers betegnelser “satzgliedwertige” og “satzwertige” ville i direkte oversættelse hedde “sætningsledsværdige” og “sætningsværdige”; men denne teoretiske præcisering med “-værdig” er næppe nødvendig i ordbogssammenhæng. Hvis en fraseologisme i et ekscerpt fra en kilde faktisk optræder som et sætningsled, må den uden videre kunne betegnes som blot sætningsledsfraseologisme eller sætningsledsfrasem. Tilsvarende kan hele sætninger, der optræder som fraseologismer, betegnes som sætningsfraseologisme eller sætningsfrasem. Her ligger i øvrigt et terminologisk stridspunkt: kan faste udtryk på sætningsniveau i det hele taget kaldes fraseologismer hhv. frasemer, eller bør disse termer forbeholdes faste udtryk *under* sætningsniveau, altså sætningsled? Efter min mening må også faste udtryk på sætningsniveau betegnes som fraseologismer (ellers ville jeg ikke tage Lügers skemaer som udgangspunkt for min beskrivelse); dermed hører talemåder, ordsprog mv. ind under kategorien fraseologismer. De kan ses som FRASEOLOGISMER I VIDERE BETYDNING; tilsvarende kan faste udtryk under sætningsniveau betragtes som FRASEOLOGISMER I SNÆVRERE BETYDNING.<sup>2</sup>

1 De fraseologiske enheder kan også blot kaldes FRASER (jf. Brink 2006: 43), da der ikke er tale om et systematisk skel frase – frasem på linje med distinktionen morf - morfem, fon - fonem og lign. Hos Falster Jakobsen (2005a,b) bruges udelukkende termen “frasem”; Almind og Bergenholtz (2007) benytter “frasem” og “fraseologisme” som ensbetydende (f.eks. p.25). I den germanske forskningsretning er “Phraseologismus” den mest anvendte term, men Burger et al. (2007: 3,12) plæderer for, at man gør “Phrasem” hhv. engelsk “phrase” til den knæsatte betegnelse for fraseologisme i *snævrere betydning* (jf. note 3) — med “set phrase” som engelsk ækivalent.

2 Men også her findes terminologisk forskellig praksis. Skellet mellem *snævrere betydning* og *videre betydning* blev vist indført af Pilz (1981: 21ff.), men mere præcist fastlagt

Som det fremgår, er de i skemaet nævnte sætningsledsfraseologismer oftest af typen KOLLOKATION, der her kort kan defineres som “fast ordfølge”.<sup>3</sup> Ved sætningsfraseologismer benytter man sig af videre kategorier, især kendt fra retorikken. TOPOI (plur. af græsk TOPOS, ‘sted’) er sætninger, hvor almen viden udtrykkes; STEREOTYPER er sætninger, der udtrykker gængse vurderinger, fordomme mv., og GENERISKE SÆTNINGER giver generelle udsagn uden inddragelse af eventuelle forbehold. Også disse termer kan diskuteres, og deres præcise indhold er i litteraturen blevet defineret meget forskelligt, afhængigt af formålet med brugen af dem. Men de kan i hvert fald tjene som orienteringspunkter her.<sup>4</sup>

## 2. Fraseologiske enheder på sætningsledsniveau

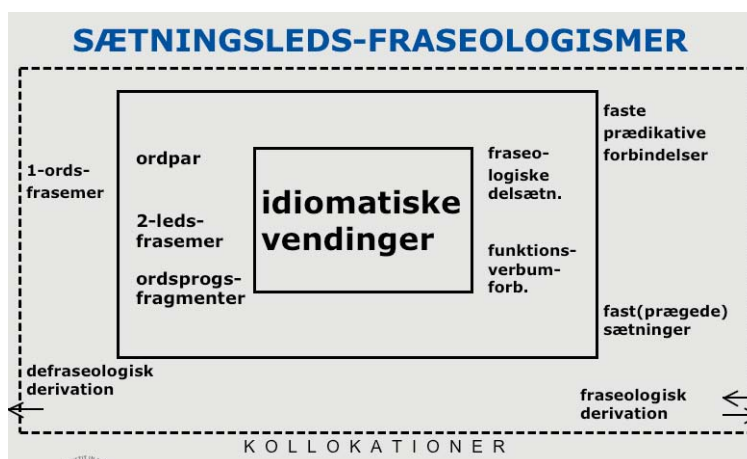


Fig. 2. Sætningsleds-fraseologismer.

af Burger et al. (1982: 1), hvor “fraseologisme” blev defineret som en flerordsenhed eller forbindelse, der 1) ikke kan forklares fuldt ud via forbindelsens “syntaktiske og semantiske regulariteter”, og som 2) bruges på lignende måde som et leksem. Hvis både 1) og 2) er opfyldt, er der tale om en “fraseologisme i snævrere betydning”, hvorimod en “fraseologisme i videre betydning” foreligger, hvis kun 2) er opfyldt (jf. Burger 2003: 32). I den videre betydning kræves altså ikke nogen egentlig uklarhed i flerordsenheden, men blot, at forbindelsen hos brugeren aktiveres ligesom et leksem. Det må betyde, at ordsprog og talemåder på sætningsniveau kan medregnes som “fraseologismer i videre betydning”. Hos Lüger diskuteres denne del af definitionen yderligere (1999: 38ff). En kort historisk opsummering og en redegørelse for de nu mest udbredte betydninger af de to begreber foreligger hos Farø (2005: 14f.) og Burger (2005: 24ff).

3 Kollokation er som term blevet defineret på forskellig vis, afhængig af forskernes tilgang til fraseologien, jf. opsummeringer hos Svensén (2004: 206ff.) og Brasch (2005: 98ff.) I det følgende bruges termen som overbegreb for idiomatiske vendinger og de øvrige enheder inden for den stiplede ramme i fig. 2.

4 Nærmere omtale hos Lüger: “verbal fixierte Topoi” (1999: 98ff.), verbale stereotyper (1999: 59ff.) og “Generizität” (1999: 91ff.) En udførlig behandling af generiske sætningers roller i tekster findes i Falster Jakobsen 2005a.

### 2.1 Idiomatiske vendinger

I fig. 2 er de IDIOMATISKE VENDINGER sat i centrum, altså vendinger eller faste forbindelser, der har en overført betydning. Lüger nævner (1999: 48) eksempler, der på dansk svarer til *have en høj stjerne hos nogen* ('bei jemandem einen Stein im Brett haben'), *være tredje hjul til en gig* ('das fünfte Rad am Wagen sein'), *granske hjerte og nyrer* ('jemandem auf Herz und Nieren prüfen').<sup>5</sup> I den del af *Jysk Ordbogs* redaktion, jeg har hentet eksempler fra, nemlig HA–HU, optræder de nævnte idiommer med ordene *hjul* og *hjerne* ikke. De idiomatiske forbindelser, der er medtaget i *Jysk Ordbog*, kan for *hjul*'s vedkommende ses i bilag 1, f.eks. *sætte de høje hjul 'for*,<sup>6</sup> dvs. sætte dem forrest, hvilket betegner at gøre tingene på en forkert eller overlegen måde, idet de høje hjul var baghjulene på de gammeldags vogne.

### 2.2 Øvrige kollokationer

Blandt de næste fraseologismer i skemaet vil jeg først omtale de to betegnelser til højre i den stiplede ramme, dernæst de tre til venstre i rammen. Som eksempel på FRASEOLOGISKE DELSÆTNINGER nævner Lüger idiommet *jemandem zeigen, was eine Harke ist*, der på dansk kan betyde 'få krammet på', 'lære på hårdhændet måde'. Det er jo et i hvert fald delvist idiomatisk udtryk; dette illustrerer udmærket, at der i disse skemaer må regnes med uskarpe grænser og glidende overgange, idet udtrykket også kan henregnes til de idiomatiske vendinger.

Forbindelser med FUNKTIONSVERBER (på tysk "Funktionsverbgefüge"), f.eks. *træffe beslutning om*, tilhører især skriftsproget, oprindeligt vel især kancellistilen, og vi har relativt få af dem i dialekterne, der jo er talesprog.

Med ORDPAR, også kaldet TVILLINGFORMLER, sigtes til forbindelser som *fiks og færdig*, eller med eksempler fra *Jysk Ordbog* afsnit HA–HU: *herlig og pæn* = 'ganske rigtigt', *over hals og hoved, i handel og vandel, hverken hoved eller hale, hus og hjem*, hvor de sidste fire er fælles med rigsmålet. De tre sidstnævnte er anført i bilag 1; det ses, at *i handel og vandel* og *hverken hoved eller hale* i de respektive artikler er placeret efter en betydningsplanke sammen med andre faste forbindelser, mens *hus og hjem* er placeret som selvstændig underbetydning.

2-LEDS-FRASEMER betegner forbindelser som *roter Faden* = 'den røde tråd', *freier Markt* = 'det frie marked', altså på dansk for så vidt 3-leds-frasemer, hvis man tæller den bestemte artikel med. Hertil hører forbindelser som *røde hunde* (sygdommen) og *gale hunde* = et kortspil (jf. bilag 1: *hund 5.1*). I bilag 2 gives et udvalg af 2-leds-frasemerne fra afsnit HA–HU, især ud fra personnavne og stednavne. F.eks. ses, at *Hamborg lås* eller *hamborger lås* er navn på et slynget symmetrisk mønster; *blanke Hans* er en betegnelse for Vesterhavet eller Vadehavet; *Holmblads katekismus* eller *Holmblads salmebog* betegner et spil kort. Den grammatiske struktur er altså: proprium (evt. i genitiv) + substantiv, eller: adjek-

5 Om problematikken vedr. egnede idiom-ækvivalenter i oversættelse fra et sprog til et andet jf. f.eks. Hallsteinsdottir (2006: 108ff).

6 Her og i det følgende markerer trykstregen ' hovedtryk.

tiv (evt. afledt af *proprium*) + substantiv. Det ses i slutningen af artiklen *Hamborg*, at 2-leddede frasemer også kan skifte status og blive til sammensatte ord: *hamborg(er)-rakker*, hvilket også ses i sammensætningsledsartiklen *2.hamborger* med henvisning til bl.a. artiklen *hamborger-garn*. Det samme ses under adj. *hollandsk*, der i sideformen *hollands* indgår i sammensætninger, fx *hollands-fad* som alternativ til frasemet *hollandsk fad* om et stentøjsfad.

Med ORDSPROGSFRAGMENTER hentydes til, at dele af ordsprog kan optræde som kollokationer eller idiomer, f.eks. *grave en grav for andre*; dette vil hos de fleste frembringe associationen: *men falder selv deri* som komplettering til fuldstændigt ordsprog. Her er naturligvis knyttet en forbindelse til skemaets anden del, sætningsfraseologismerne.

### 2.3 Øvrige sætningsledsfraseologismer

Det kan diskuteres, om det er nødvendigt at regne med 1-ORDS-FRASEMER: en fraseologisme / et frasem / en frase er pr. definition en flerordsenhed og består da af mindst to ord, kan man hævde. Men på grænsen mellem ord og frasem ligger ifølge Lüger sammensatte ord som *sortbørs*, der stammer fra fraseologismen *den sorte børs*, *tykhudethed* ('have tyk hud'), *halsbrækkende* ('så farlig, at man risikerer at brække halsen'). Disse sammensætninger er afledt af fraseologismer mv. og har derunder (næsten) mistet deres status som sådanne; deraf betegnelsen DEFRASEOLOGISK DERIVATION i oversigten (fig. 2). Ord af disse typer placeres i *Jysk Ordbog* og andre ordbøger blot som sammensætninger.

Men en anden ordtype kan være berettiget til betegnelsen 1-ords-frasem, nemlig de såkaldte JUXTAPOSITIONER, altså ordsammenstillinger (til forskel fra almindelige sammensætninger), f.eks. plantenavnet *forglemmigej*. Her er der tale om en hel sætning (imperativ + dativobjekt + adverbium), der i rigsdansk ortografi trænges sammen til ét ord og dermed ikke på overfladen kan skelnes fra en sammensætning. I *Jysk Ordbog* markeres, af hensyn til brugerne, grænsen mellem 1. og 2. led i sammensætninger i opslagsformerne (lemmaerne) med et hævet punktum (kaldet en "sammensætnings-prik"), mens leddene i en juxta-position for tydelighedens skyld angives med bindestreger, her altså *forglem-mig-ej* – eller rettelig: *for'glem-mig-ej*, idet vi fremhæver placeringen af hovedtrykket via en trykstreg, når dette ikke ligger på første stavelse. Fra ordbogsafsnittet HA–HU kan tilsvarende nævnes f.eks. substantiverne *hjælp-i-'nød* og *hjælp-i-'trang*; disse er sammenstillinger af substantiv + præpositionsforbindelse (eller sætningsfragmenter af verbet *hjælpe* og en præpositionsforbindelse: *hjælpe-i-'nød* etc.), jf. bilag 3, hvor et udvalg af afsnittets juxta-positioner vises. Det ses, at begge nævnte lemmaer kan betegne en løs medhjælper, der kan tilkaldes, når der er behov for det, og at *hjælp-i-'trang* desuden kan have den specifikke betydning 'jordemoder'. – En anden type juxta-position ses i lemmaet *hjørne-og-konge*: en sammenstilling af substantiv + konjunktion + substantiv, der er fast sammenknyttet lydligt og har specifik betydning: 'slag i keglespil'. Noget tilsvarende gælder for *hundrede-*

og-ét som navn på et kortspil. Af andre ordsammenstillingstyper i bilag 3 kan nævnes omdannede indlån: *herom-den-svans* fra tysk, *hov-en-drov* fra hollandsk.<sup>7</sup> Man kunne vælge at undlade bindestregerne i de nævnte opslagsformer og dermed skjule, at der er tale om fraser; men vi finder det vigtigt at signalere dette allerede i opslagsformerne, så her ses 1-ords-fraser som en relevant kategori. Alternativt kunne man anføre hvert frasem som 3 adskilte ord; men derved ville enheds-/lemma-præget mistes.

De FASTE PRÆDIKATIVE FORBINDELSER i højre side af skemaet i fig. 2 udgøres hos Lüger (1999: 50) af udtryk som *jemandem sticht der Hafer*, svarende til dansk *være højt oppe* eller *være i højt humør*. Hvilke kriterier man i praksis skal bruge for at adskille disse fra de "fraseologiske delsætninger" i rammen indenfor, fremgår ikke tydeligt (jf. omtalen i kap. 2.2 ovenfor). Betegnelsen FAST(PRÆGED)E SÆTNINGER er overtaget fra Lüggers "festgeprægte Sätze" og kan måske oversættes med "fast sætning" eller "fast udtryk", men bliver derved næsten for generelt og i hvert fald vanskeligt at adskille fra de netop omtalte termer "fast prædikativ forbindelse" og "fraseologisk delsætning". Som eksempel nævner Lüger (1999: 50) idiomet *da liegt der Hase im Pfeffer*, der på dansk har ækvivalenten: *der ligger hunden begravet* el. lign. Men det synes jo ikke at være en sætningsledsfraseologisme, derimod en sætningsfraseologisme. Lüger omtaler det da også som et overgangstrin til sætningsfraseologismen, men forklarer ikke nærmere, hvorfor han bibeholder det under sætningsledsfraseologismen.

### 3. fraseologiske enheder på sætningsniveau

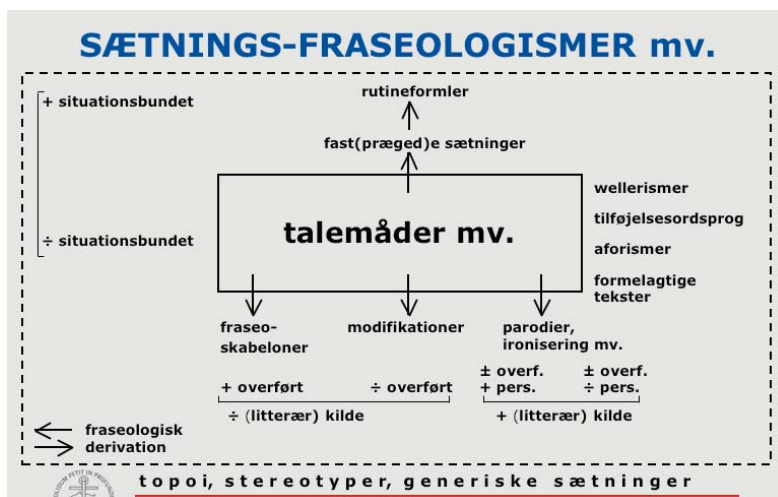


Fig. 3. Sætningsfraseologismer 1.

<sup>7</sup> I bilag 3 ses flere eksempler på opslagsformer af typen verbum + præpositionsforbindelse, fx subst. *hoppe-i-græs* 'græshoppe'; jf. også eksemplerne under *havre-sød 5* i slutningen af bilaget: subst. *finpe-* hhv. *gynge-(i)-havresød* mv. som fuglenavn.



Her er talemåder mv. den vigtigste kategori (se nærmere fig. 4 og kap. 3.3), mens de øvrige enheder måske bør betegnes som fraseologisme-relaterede, ikke som egentlige fraseologismer på grund af deres ofte større omfang (flere sammenknyttede sætninger) og mindre faste ordstruktur. Men først et par ord om skemaets opbygning.

### 3.1 Ekstra dimensioner i oversigten

Det er forsøgt at indsætte 3 ekstra skalaer eller dimensioner i diagrammet, markeret med skarpe klammer (1 lodret, 2 vandrette). Den vigtigste skala i *Jysk Ordbogs* sammenhæng er +/- OVERFØRT, der gengiver Lügers "bildhaft" og giver mulighed for at skelne f.eks. mellem talemåder, der kan rubriceres som +overført, og erfaringssætninger, der hører til -overført på skalaen. Den (eller rettere: de) næste vandrette skalaer vedrører identifikation af kilden; Lügers term er her (jf. 1999: 129) "belegbar", som betegner, at en (litterær) kilde kan angives. Kan et bestemt værk eller forfatter angives som ophav for en given sætningsfraseologisme, betegnes den som "belegbar", i skemaet markeret som +(LITTERÆR) KILDE; men drejer det sig om en talemåde som *Når enden er god, er alting godt*, betegnes den "-belegbar" hos Lüger og -(LITTERÆR) KILDE i fig. 3, idet der næppe eller kun vanskeligt kan spores nogen ophavsmand til talemåden.<sup>8</sup>

Også i diagrammets lodrette dimension kan der skelnes, nemlig mellem SITUATIONSBUNDNE og IKKE-SITUATIONSBUNDNE fraseologismer mv. Her betegnes RUTINEFORMLER som situationsbundne, f.eks. *hav det godt* ('Mach's gut!'), *held og lykke* ('Hals- und Beinbruch!', hvor man på dansk kun i visse situationer kan bruge *knæk og bræk* som ækvivalent). Fra *Jysk Ordbogs* afsnit HA-HU kan nævnes *nej hold* = 'nej, stop, nu går det for vidt' (jf. 2.holde 5, i bilag 5). Som eksempler på FAST(PRÆGED)E SÆTNINGER nævner Lüger her dels *das geht auf keine Kuhhaut* (som vi ikke har en oplagt ækvivalent til på dansk; nærmest kommer nok 'nu går det for vidt'), dels *So ist das Leben* ('sådan er livet'). Begge er vel bundne til bestemte situationstyper; men den første kan også ses som en talemåde, den anden som en truisme (jf. kap. 3.3). Det fremgår ikke klart, hvordan Lüger egentlig vil skelne skarpt i praksis, jf. også bemærkningerne til de fast(præged)e sætninger i kap. 2.3; der må som tidligere nævnt regnes med flydende overgange i skemaet. De øvrige størrelser i skemaet er ikke nødvendigvis situationsbundne, jf. det følgende.

### 3.2 Enheder i oversigten

Indholdet i centralenheden TALEMÅDER MV. venter af praktiske grunde til næste afsnit, og enhederne "rutineformler" og "fast(præged)e sætninger" er netop omtalt ovenfor, så her blot nogle få ord om de øvrige termer. Med WELLERISMER

<sup>8</sup> Ordsproget bag det i kap. 2.2 nævnte fragment ('den, der graver ...') hører egentlig til i kategorien "+(litterær) kilde", idet det kan spores tilbage til Ordsprogenes bog i Bibelens gamle testamente og dermed er "belegbar" i Lügers betydning af ordet.

betegnes talemåder, der indeholder udtryk som “*sagde NN, da ...*” – f.eks.: *gudske lov for den gode varme, sagde kældingen, da brændte huset.*<sup>9</sup> Vi har adskillige eksempler på denne type i *Jysk Ordbog*, især med formlerne “*sagde drengen / manden*” hhv. “*sagde fanden*” (f.eks.: *noget er bedre end intet, sagde han fanden, da åd han kærnemælk med en høfork*). Med TILFØJSESORDSPROG menes udtryk som: *det er ingen skam at være fattig – bare man har penge* (jf. Lüger 1999: 41). Denne type næsten selvophævende talemåde findes også ind imellem i *Jysk Ordbog*, hvor den oftest er placeret blot som citat. Sætningens ordstruktur er ikke nøje fastlagt, så der er ikke tale om egentlige fraseologismer, men om fraseologisme-relaterede enheder. Hvorvidt vi overhovedet anfører AFORISMER fra dialektforfattere m.fl., er vanskeligt at efterspore; i givet fald er de blot nævnt som eksempel i betydningsdelen. Som eksempler på FORMELAGTIGE TEKSTER nævnes velkomstord, taksigelser og berigtigelser (Lüger 1999: 42); hvis sådanne optræder i vore kilder, vil vi næppe citere dem, da de formentlig vil være stærkt rigsmålsprægede.

I linjen under ‘talemåder mv.’ behøver MODIFIKATIONER og PARODIER MV. ikke nærmere omtale, blot kan som eksempel nævnes, at i tilføjelsesordsprogene ovenfor kan tilføjelsen naturligvis ses som en modifikation. FRASEOSKABELONER, dvs. talemåder eller udtryk, hvor syntaksen er overtaget fra et fast udtryk, f.eks. *arbejde er (nu engang) arbejde, hun vil og vil alligevel ikke* (svarende til formler som “X er X”, “X og (samtidig) ikke-X”) (Lüger 1999: 42f.), kan nok forekomme i *Jysk Ordbog*, men nævnes i så fald blot blandt citaterne.

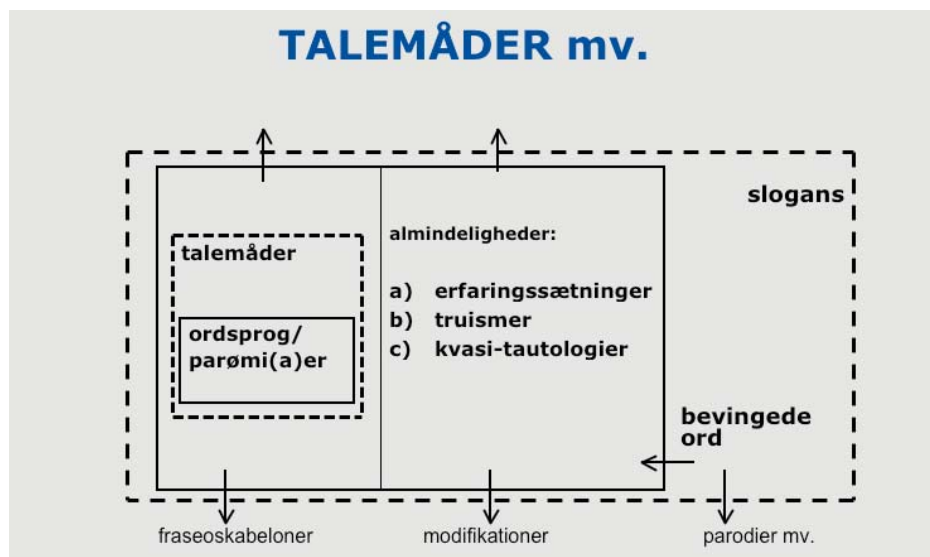


Fig. 4. Sætningsfraseologismer 2 (jf. kap. 3.3).

9 Termen er dannet ud fra en fiktiv person, Mr. Weller i Charles Dickens: *The Pickwick Papers*, der yndede at bruge denne type talemåde; jf. Kjær og Holbek (1969: 18, danske eksempler p. 249ff).

### 3.3 Talemåder, ordsprog mv. (jf. fig. 4).

I *Jysk Ordbog* betegner TALEMÅDE en fast, gængs vending, evt. med særlig (overført) betydning – i overensstemmelse med ordets betydning 2.4 i *Ordbog over det Danske Sprog* og fjernt fra den betydning, der antydes i *Nudansk Ordbog* med eksemplet “tomme talemåder”.<sup>10</sup> Som nævnt i indledningen betragter vi ORDSPROG som en gruppe inden for talemåderne og særmærker dem ikke i den redaktionelle tekst. I skemaet er termen TALEMÅDER indsat af mig: Lüger nævner ikke “Redensarten” i oversigten (1999: 129), hvor det dog klart må være underforstået. Desuden har jeg i ordsprogsrammen tilføjet betegnelsen PARØMI(A)ER, idet man ofte møder termene “Parömie” og “Parömiologie” i tysk faglitteratur om ordsprog og forskning heri; den tilsvarende “danske” term er “parømia”.<sup>11</sup>

Ved siden af talemåderne, der har en overført betydning, er nævnt ALMINDELIGHEDER (som oversættelse af Lügers “Gemeinplätze”) uden overført betydning, jf. omtalen under skalaerne i 3.1. ovenfor. Herunder hører ERFARINGSSÆTNINGER som *al begyndelse er svar*, TRUISMER som *man lever kun én gang* og KVASI-TAUTOLOGIER som *for meget er for meget* (“Was zuviel ist, ist zuviel”; alle jf. Lüger 1999: 132). Disse typer har vi adskillige eksempler på i *Jysk Ordbog*. Ved siden af talemåderne og “almindelighederne” ses kategorien BEVINGEDE ORD, som næppe (eller i hvert fald meget lille omfang) er repræsenteret i *Jysk Ordbog*, da den vel kræver skriftlig udbredelse (via rigsmålet), og dialekterne især er talesprog. Dialektforfattere som Blicher og Aakjær har måske eksempler på det; men i så fald vil vi kun citere det efter en nøje vurdering af, om det også indgår som genuin dialekt. Kategorien SLOGANS optræder ikke i *Jysk Ordbog*; de hører en senere tid og den trykte presse til.<sup>12</sup>

Desværre tillader tid og plads ikke nærmere omtale af talemåder mv.; men der kan henvises til Toftgaard Andersen (2001) for videre diskussion og afgrænsning af emnet i leksikografisk sammenhæng.<sup>13</sup> Som nævnt i indledningen er vi her inde i området FRASEOLOGI I VIDERE BETYDNING; jeg vender nu tilbage til aspekter af FRASEOLOGI I SNÆVRERE BETYDNING.

## 4. Særlige sætningsledsfraseologismer

Som nævnt i afsnit 2.2 fremhæver vi sætningsledsfraseologismer med understregning hhv. kursivering i den redaktionelle tekst. Men vi opretter faktisk også

<sup>10</sup> Jf. karakteristikken af “talemåde” som både den mest udbredte og samtidig den mest problematiske af alle danske fraseologiske termer (Farø 2007: 949f.).

<sup>11</sup> Ifølge *Dansk Fremmedordbog*, 2. udg. 1999, af Karl Hårbøl, Jørgen Schack og Henning Spang-Hanssen; men ordet er vist ikke etableret som egentlig term endnu.

<sup>12</sup> De to sidstnævnte kategorier nævnes blot for fuldstændighedens skyld; i Lügers dagblads- og tidsskriftmateriale fra 1990’erne er de velafhjemlede.

<sup>13</sup> Jf. kap. 3: Fraseologibegrebet, p. 333ff.; kap. 4: Idiomatiske vendingers karakteristika, p. 363ff.

mange artikler med det ene formål at redegøre for kollokationer eller idiomatiske udtryk, idet et givet ord kun forekommer i en sådan forbindelse. Disse artikler indeholder i princippet kun opslagsformen (lemmaet) og den redaktionelle angivelse “kun i forbindelsen ...” (hvor frasemet nævnes); de kan derfor kort betegnes FRASEM-LEMMAER. Bilag 4 viser et udvalg af de ca. 50 eksempler, der findes på denne kategori alene i afsnittet HA–HU. F.eks. kendes *helbreds-stund* kun i forbindelsen: *ikke el. aldrig have en helbredsstund*, med betydningen ‘ikke el. aldrig have en stund med godt helbred’, dvs. ‘være dårlig, utilpas hele tiden’. Ved det parallelle udtryk *hellends dag* ser vi et eksempel på, at det kan være vanskeligt at afgøre, om der er tale om en sammensætning, en juxtaposition eller blot en forbindelse af 2 ord (jf. diskussionen af “1-ords-frasemer” i kap. 2.3 ovenfor).

Flere af ordene og forbindelserne her er tæt afhjæmlede inden for velafgrænsede områder, hvorfor vi har en forpligtelse til at tegne såkaldte spaltekort over deres udbredelse. Det gælder f.eks. adj. *himmel-betændt*, der kun kendes i forbindelse med subst. *rende-fog* ‘stærk snefygning’, med udbredelse i det nordligste Jylland, jf. bilag 4 (med kortet placeret nederst ved bilag 2).

Blandt de mere specielle frasem-lemmaer er verberne *2.hjerne* og *hovede*. Der er tale om afledninger til substantiverne *1.hjerne* og *hoved*, der fremhæver forskellige aspekter af hjernens hhv. hovedets funktion: *hjerne på* ‘stole på, bande på’ (hvor kilden selv fremhæver, at der er tale om et egnsudtryk), over for: *det hoveder ham ikke* ‘det passer ham ikke’.

I nogle tilfælde er det ikke helt så kategorisk kun i en given forbindelse, men især dér, man træffer et bestemt opslagsord. Det gælder f.eks. ved *ha'sard*, hvor der ved betydning 1 præciseres: især i forb. *stå sin egen ha'sard* osv. ‘selv tage risikoen’, jf. bilag 4 nederst.

## 5. Frasemrækkefølge i mikrostrukturen

I de enkelte artikler i *Jysk Ordbog* anfører vi som nævnt faste forbindelser enten med understregning eller i kursiv, ofte til sidst i artiklerne efter citater til understøtning af den givne definition. Skal en eller få talemåder nævnes, sker det også helst til sidst i artiklen, men således, at den eller de talemåder, der direkte vedrører definitionen, sættes foran de faste forbindelser. Disse kan evt. placeres efter planke, og vi vil her til sidst se på ordbogens praksis mht. rækkefølgen i de artikler, hvor der behandles mange faste forbindelser. I bilag 1, som der tidligere er refereret til, har jeg opstillet de faste forbindelser, der nævnes i de største substantiv-artikler i afsnit HA–HU. Det ses, at der ved adskillige betydningsangivelser skal redegøres for temmelig mange faste forbindelser, f.eks. ved *1.hoved 1.1* og *hund 1.1.3*. Et overordnet princip er, at forbindelserne (altså frasemerne) forsøges anført i alfabetisk orden, se f.eks. *1.hammel 4*, hvor der bl.a. anføres 5 forbindelser med *gå*, derpå forbindelser med *holde* og *hænge* etc. frem til de 2 sidste med *være*. Dette princip kan naturligvis kollideres med indholdsmæssige grupperinger,

f.eks. i *1.hjul*, der først har forbindelser med *rende*, *sætte* og *være*, derpå med *have*, *trække* og *holde*, etc. Her er analyser som Lügers af idiomatiske vendinger mv. (jf. kap. 2) afgørende for, hvilken rækkefølge fraserne præsenteres i. Ved *1.hus* er det alfabetiske princip gennemført i betydning 10, præpositionsforbindelserne, og det betones, at det netop sker på tværs af betydningsinddelingen.

Alle de nævnte forbindelser er understreget i den redaktionelle tekst (understregning dog udeladt i gengivelsen her). Det ses, at der i pkt. 10.1, 10.3 m.fl. er tale om *gamle kasusformer* efter bestemte præpositioner: *af/fra/i/til/ved huse*; disse har vi en forpligtelse til at bringe. Desuden ses i pkt. 10.3 m.fl., at der kan foretages en underopdeling ved hjælp af små bogstaver, her a og b, og at disse kan anføres fed skrift: det er gjort i enkelte artikler, hvor teksten til disse underpunkter er meget lang; normalt skrives blot med ordinær, f.eks. i *1.hoved 2: (have/få) til hoveds: a), b)* – i øvrigt også her ved angivelse af gamle kasusformer.

Ved de “store” verber (dvs. verber med mange betydninger, underbetydninger mv.) anfører vi tryktabsforbindelser i et særligt afsnit til sidst i artiklen, jf. bilag 5, hvor de 3 “største” verber i afsnit HA–HU er opstillet: *4.have*, *2.holde* og *hugge*; tryktabsforbindelserne ses under hovedpunkterne henholdsvis D, F og C. Ved *4.have* kunne vi have placeret alle disse forbindelser som betydning 25.1 – 25.10, og tilsvarende for de to andre verber. Men vi har af hensyn til overblik og læservenlighed valgt at give hver forbindelse eget betydningsnummer, under *4.have* altså fra betydning 25 *have 'for* til betydning 34 *have 'ved*. I disse afsnit anføres fraserne alfabetisk ud fra efterfølgende trykstærk(t) præposition el. adverbium, f.eks. fra *holde 'ad* til *holde 'ved*; nogle af disse har adskillige underbetydninger, helt op til 7 under *holde 'af*. De markeres lige som ved substantiverne ovenfor med små bogstaver a) - g) og sættes for overblikkets skyld med fed skrift i artiklen. Betydningsforskelle, der registreres på dette niveau, er f.eks. (i betydning 24), at *holde 'af* kan betyde så vel *holde 'ud* som *holde 'op*.

Brugen af de særlige afsnit med tryktabsforbindelser kan også illustreres via verbet *hugge* (i bilag 5). F.eks. ses det, at betydning 1.2 angår brug af ordet i forbindelse med *høst*, og at tryktabsforbindelserne vedr. denne betydning især henvises til den tværgående opregning i afdeling C, betydningsnumrene 11–19. Således henvises *hugge 'af* til betydning 11b: her angives – når man slår efter på [www.jyskordbog.dk](http://www.jyskordbog.dk) – betydningen ‘blive færdig med at meje’, mens samme frase i 11c angives som havende betydningen ‘fare af sted’, hvilket er henvist fra betydning 10: ‘ile, fare af sted’. Tilsvarende henvises *hugge 'ud* fra betydning 1.2 til betydning 19b, hvor der angives en særlig måde at meje stråene på, mens der for samme frase i 19a angives betydningen ‘tynde ud blandt træer’, henvist fra betydning 1.1.

Grundlaget for disse overvejelser og valg er naturligvis analyser af de grammatiske strukturer i de faste forbindelser (sætningsledsfraseologismer, -fraser), der medtages i artiklerne, og her har Lügers analyser og klassifikationer været nyttige som pejlepunkter og inspiration for videre overvejelser, jf. kapitel

1–4. Disse analyser kan måske raffineres til at give den helt korrekte lingvistiske beskrivelse af hver enkelt fraseologisme<sup>14</sup> og derigennem udskille et antal syntaktiske el. semantiske strukturer, som obligatorisk skal indgå i ordbogen, jf. eksemplerne i Vikørs artikel i nærværende skrift (2008: 465ff.). Men generelle, lingvistiske konklusioner om, hvilke strukturtyper der er fremherskende mv., er egentlig kun til begrænset hjælp i arbejdet som leksikografer og ordbogsredaktører: vi er forpligtede til at redegøre for *alle* de faste forbindelser, som er tilstrækkeligt sikkert afhængende i ordbogsmaterialet, uanset de pågældende frasemers grammatiske struktur.

## 6. Afslutning

Det har været spændende at anvende Lügers (pragma)lingvistiske fraseologismemodel på et leksikografisk materiale i form af det fraseologiske stof, der skal overvejes før eventuel optagelse i *Jysk Ordbog*. Modellen viste sig rigeligt at kunne rumme *Jysk Ordbogs* fraseologismetyper; den og de tilhørende analyser skærpede opmærksomheden på skellet mellem sætningsleds- hhv. sætningsfraseologismer (kap. 2–3), hvilket bevirkede fremhævelse af dels ordbogens 1-ords-frasemer (kap. 2.3), dels de kun (eller især) fraseologisk motiverede ordbogsartikler, de såkaldte frasem-lemmaer (kap. 4). Ligeledes har modellen og dens analysepraksis kunnet bidrage til overvejelserne omkring den optimale interne strukturering af de fraseologiske afsnit i ordbogens ”store” artikler (kap. 5).

## Litteratur

- Almind, Richard & Henning Bergenholtz 2007: Fysisk genstand med ben og bagdel: funktionel fraseologi. – i: Widell, Peter og Henrik Jørgensen (udg.): *Det bedre argument*. Århus: Wessel og Huitfeldt, 13–34.
- Braasch, Anna 2005: Kollokationer som fraseologisk kategori set fra forskellige synsvinkler. – i: *Hermes* 35, 97–117.
- Brink, Lars 2006: Den fraseologiske terminologi. I: *Nordiske Studier i Leksikografi*. 8. København: Gyldendal, 39–51.
- Burger, Harald 2003: *Phraseologie*. 2. Aufl. Berlin: Erich Schmidt.
- Burger, Harald 2005: 30 Jahre germanistischer Phraseologieforschung. I: *Hermes* 35, 17–43.
- Burger, Harald, Annelies Buhofer und Ambros Sialm 1982: *Handbuch der Phraseologie*. Berlin / New York: Walter de Gruyter.
- Burger, Harald, Dmitrij Dobrovolskij, Peter Kühn & Neal S. Norrick (Hrsg.) 2007: *Phraseologie / Phraseology*. 1–2. (= HSK 28.1–2). Berlin / New York: Walter de Gruyter.

14 Men her er tidsaspektet helt afgørende, jf. bemærkningen: “en fraseolog kan arbejde med et eller få frasemer i flere år, hvorimod en leksikograf skal beskrive dusinvis hver eneste dag for at færdiggøre en ordbog” (Almind og Bergenholtz 2007: 14).

- Falster Jakobsen, Lisbeth 2005a: Det er sundt med daglig motion. Om fraser og generiske ytringer i tekster. I: Cramer, Jens, Mette Kunøe og Ole Tøgeby (udg.): *Teorien om alt*. Århus: Wessel og Huitfeldt, 135–147.
- Falster Jakobsen, Lisbeth 2005b: Fraser som konstituent i løbende tekst. I: *Hermes* 35, 135–151.
- Farø, Ken 2005: Dänisch-europäische Phraseologieforschung. I: *Hermes* 35, 11–16.
- Farø, Ken 2007: Danish phraseography. I: Burger et al. (Hrsg.), bd. 2, 949–958.
- Hallsteinsdóttir, Erla 2006: Phraseographie. I: *Hermes* 36, 91–128.
- Hermes. = *Hermes. Journal of Linguistics / Journal of Language and Communication Studies*. 1ff. 1988ff. Århus: Handelshøjskolen.
- Kjær, Iver og Bengt Holbek 1969: *Ordsprog i Danmark*. København: Jørgen Paludan.
- Lüger, Heinz-Helmuth 1999: *Satzwertige Phraseologismen. Eine pragmalinguistische Untersuchung*. Wien: Edition Praesens.
- Pilz, Klaus Dieter 1981: *Phraseologie: Redensartenforschung*. Stuttgart: J.B. Metzler.
- Svensén, Bo 2004: *Handbok i lexikografi*. Stockholm: Norstedts Akademiske Förlag.
- Toftgaard Andersen, Stig 2001: *Talemåder*. 2. udg. København: Gyldendal.
- Vikør, Lars S. 2008: Behandling af fraseologi i Norsk Ordbok. I: dette bind, 465–476.

Torben Arboe  
videnskabelig medarbejder, ordbogsredaktør, f. 1948  
Peter Skautrup Centret for Jysk Dialektforskning, Nordisk Institut  
Aarhus Universitet  
Jens Chr. Skous Vej 4  
Nobelparken, bygn. 1483  
DK-8000 Århus C  
jysta@hum.au.dk

## BILAG 1: Frasemers placering i mikrostrukturen ved store substantiver

<sup>1</sup>hammel

1) || (et par) hamler, et par h.; den korte h., den lange h., lige/ens h. > 4)  
 4) ... – fise/sidde på hamlen. a) b) c) – gå på <sup>1</sup>ens/<sup>1</sup>lige h. – gå <sup>1</sup>godt i h. med – gå med/på den lange h. el. den lange ende af hammelen – gå med/på den korte/stakkede h. el. den korte ende af hammelen – gå med /få /tage hammelen – holde h. med – hænge på hamlerne – slå h. – trække i hammel(en) – trække / være på <sup>1</sup>én/<sup>1</sup>ens/<sup>1</sup>lige h. – være rent / > *ryt* ud af hammelen

<sup>2</sup>handel

1.1) ... || i h. og vandel – komme til handels – ud(e) på handelen/æ handel – levende h. > *levende*

5) ... || få en anden h. på – ikke hverdags h. – det har ingen h. – der er ingen h. ved én

6) ... til daglig h. – til jævn h. > 12.1)

12.1) jævn h. ... 12.2) til ren h.

<sup>1</sup>hjul

1) || rende/løbe som et vindt/skævt h. – sætte/vende (/køre med) de høje h. <sup>1</sup>for – være på det store h. – have sine ben i hjulene – trække/sætte (nogen) i hjulet – holde hjulet i furen – sætte en kæp i hjulet – som kæppe(r) i h. > *kæp*

6) navne på lege: > *dyble* h. – rende/køre/trille h. (i nav) – sløve (et) h. – slå h. – trille/løbe (med) h.

<sup>1</sup>hoved

1.1) || hverken h. eller hale – over hals og h. > *kytte* h., slå > *ast* over hovedet – (falde, styrte, stå etc.) på hovedet – stå på <sup>1</sup>hovedet – (vende/stille etc. ngt) på hovedet – (give/få) nogle på / ved / om hovedet – sige ngt til hovedet (af én)

2) || få ngt i hovedet – have i <sup>1</sup>hovedet – vil igennem med hovedet – (have/få) til <sup>1</sup>hoveds:

a) b) – være til <sup>1</sup>hoveds

hund

1.1.2) ... røde hunde

1.1.3) i faste forb. med verb.: have en hund i rebet – have en lykke / være så glad som en enøjet hund – have (etc.) hunden i fårestien – holde hundene fra døren(e) – holde sin(e) hund(e) inde/hjemme – ikke have hunds ret – komme i de vilde hunde – leve som hund i helmis > *helmis* 2) – værre end at malke en hund

5.1) || gale hunde (kortspil) ... 5.3) hund efter hare > *hare'tavl* ... 5.4) tage/rende hund efter hare, tage hund for hare (leg) ... 5.5) pudse hund efter hare (dans)

<sup>1</sup>hus

1) || det bitte/lille hus

2.2) husen(e) = hele gården || husens folk – husens skik – til husens brug

4) ... hus og hjem || tage hus i helvede (... i Paris, i Pellworm, i Brasilien)

6) || holde<sup>1</sup>hus (for nogen) – holde<sup>1</sup>hus (med noget)

7) || holde <sup>1</sup>hus (= lave ballade mv) – der ligger et hus (= d.s.)

10) faste forb. af præp. + nøgen form el. gl. genitiv-dativ (på tværs af betydn.inddelingen ovf.)

10.1) af huse ... 10.2) fra hus ... 10.3) fra huse. a) b) ... 10.4) i hus. – a) || tage i <sup>1</sup>hus; b) ... 10.5) i huse. a) b) ... 10.6) imellem huse ... 10.7) til huse. a) b) || sidde, være til <sup>1</sup>huse – være (velstand, smalhals, småt etc.) til huse; c) stå, bo, sidde, være (etc.) til <sup>1</sup>huse ... 10.8) ved huse



## BILAG 2. 2-leds-frasemer.

**Hamborg** stedn. ...

|| **Hamborg/hamborger** lås, nøgle, port, portlås, portlåge (el. lign.) = navn på slyngt symmetrisk mønster, anvendt i magisk sammenhæng; (senere) behændighedsøvelse blandt skolebørn [spor. i Nørrejylland; jf. @mare:kors] – varianter af mønstret er vist i JKamp.DFm.323 og SprK.XI.132. || **hamborg(er)-rakker** (kortspil) se @rakker x.

<sup>2</sup>**hamborger-** sms.led. – udtales som <sup>1</sup>hamborger el. med svækket 2.led (-bår).

1) = fra **Hamborg**, som i **Hamborg** etc., fx *-hamborger-alen*, *-hamborger-skilling*.

2) (overført) = af **prima kvalitet**, fx *-hamborger-garn*, *-hamborger-mellemmad*.

<sup>2</sup>**Hans** mn. ...

= **rigsm.** ... || **blanke Hans** = Vesterhavet, Vadehavet [vist < nedertysk el. frisisk, fastholdt i litterært sprog, jf. DSt.1979.129ff.; spor. i VSønJy] ... *Her har vi ingen Jordskælv og ildsprudende Bjerger ... ingen Slanger og vilde Dyr, ingen Røvere og Mordere". Den "Blanke Hans", som hver Dag kunde komme og sætte hans Bo under Vand, tænkte han ikke paa.* SJyMSkr.1934f.96.

**Holland** stedn.

– i navn på leg: **hoppe i 'Holland** = hoppe i paradys (hinkeleg). "*Hoppe i Mand*" (jf. @mand x) *hed "Hop' i Holland"*. \*Ålborg. \*BarnAalborg.150. ...

**hollandsk** adj. – med sideformen **hollands** (også opfattet som 1.sms.led). ...

= **rigsm.** (fra Holland; af hollandsk type mv.). – ud over nedennævnte forb. også fx *hollandsk vejrmølle*; *hollandske kakler*, *klinker*. || faste forb.: **hollandsk fad**, **tallerken** = stentøjsfad, -tallerken [spor. i Nørrejylland] ... **hollandsk hede** = klokkelyng; jf.

JLange.ODP.I.544 [jf. @'hede 2; Fjølde] – **hollandsk kløver** = gul kløver, humlekløver; jf. JLange.ODP.II.727. \*Fjølde. – **hollandsk vægt** = mål for tærsket korns renhed, tørhed [spor. afhjemlet; mht. nærmere definition mv. se LandbO.II.339f.] ...

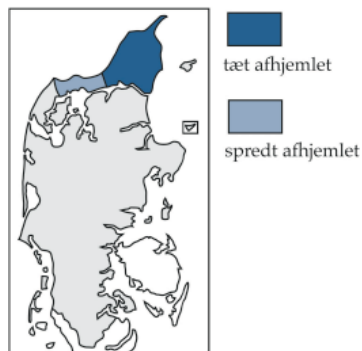
**Holmblad** propr. – kun afhjemlet i genitiv: *hålmblas* °Vroue, °Voldby; *hålmblajs* MØJy.

[spor. afhjemlet; spøgende]

– kun i forb. **Holmblads katekismus** hlv. **salmebog** = sæt spillekort. *ska vi ta æn jænlø væs atør hålmblas* <sup>1</sup>sal'm<sub>1</sub>bo'q = skal vi tage et enkelt vers efter Holmblads salmebog (dvs. tage et lille slag kort). °Vroue.

## Kort fra bilag 4:

himmelbetændt rendefog



himmelbetændt

## BILAG 3. 1-ords-fraser / juxtapositioner

1) Som opslagsformer / lemmaer:

**hejt-om-dejt** adv. [muligvis < *hejtre*] = **hulter til bulter**.

**herom-den-svans** subst. – *hær'om-an-'svan's* [formentlig < tysk *herum den Schwanz*; jf. Dialektstudier.V.315f.] = **navn på leg**.

**herom-hjalde** subst. [< udbruddet *herom, hjalde!*, jf. @'hjalde 2] = **skuffelse, "afbrænder"**.

**hitte-på-som** adj. – også (med tryksvagt sidsteled) **hitte-'på-<sub>s</sub>om** [< ?*hitte 1*; kun anførte kilder] = **opfindsom**.

**hiv-i-land-kone** subst. [< @*hive 3*; jf. Ord&Sag.1981.5ff.] (spøgende) = **jordemoder**. \*Him-V.

|| også kaldt **hiv-i-land-kælling** (\*Hards-V), **hiv-i-land-mutter** (\*Him-V).

**hjerter-lunge-urt** subst. [sammenstillingen af *hjerter* og *lunge* skyldes antagelig medikamentets virkning på hjerter- og lungeregionen] = **droge af (planten) hjertensfryd**. \*Vends (JLange.ODP.II.53). \*Hards.

**hjørt-her** interj. ... = **første replik i (og derfor navn på) gætteleg for børn** (alm. anvendt i juletiden og med pebermødder som indsats).

**hjælp-i-nød** subst. [afhjemlet fra to kilder i TyrstingH (MØJy-S) og NimH (SØJy-N); jf. @*hjælp-i-'trang 1*] = **løs medhjælper**.

**hjælp-i-'trang** subst. 1) = **løs medhjælper** [spredt i Nordjy, spor. i Midtjy; jf.

@*hjælp-i-'nød*] || (også) = **dårlig medhjælper** (man må tage til takke med) [spor. i Nordjy]

2) = **jordemoder** [spor. i Nordjy] || hertil: **hjælp-i-'trang-mand** = **jordemoderens mand**.

\*Him.

**hjørne-og-konge** subst. – *'hjønå,koŋ* [< @'hjørne 5, @*konge x*] = **slag i keglespil**.

\*°Gosmer.

**hoppe-i-græs** subst. ... [< *græs'hoppe*, jf. Dialektstudier.IV.123; afhjemlet i to kilder fra

Rømø og to fra Sønderjy-S] = **græshoppe**.

**hoppe-i-kurv** subst. – *håpi'kåra* °Bov. 1) = **ung, livlig** (evt. letsindig) **kvinde** [spor. i

Sønderjy] 2) = **barn, der hopper på en voksens skød**. \*Sønderjy.

**hoppe-i-to** subst. [<@*to x*] *hop-i-'to'w*, **spøgende betegnelse for en halt person**.

\*AEsp.VO.IV.355. \*HundborgH (Thy).

**hov-en-drov** adv. [formentlig < hollandsk *hou en trouw* (= fuldstændig trofast); Læsø;

fiskersprog] *'hōw'ən'drōw*, **i lige retning, tværs over altting**. °AEsp.Læsø.

**hundrede-og-ét** subst. – el. **hundrede-og-én** (Vestjy, sideform i Sønderjy) [spredt i Midt-

Syd- og Sønderjy] = **navn på ældre kortspil** (hvor man vinder, når man har fået 101

points), jf. PDavid.GS.43f. ...

**huslig-lykke** subst. [jf. @*hus'lykke*] = **navn på forskellige potteplanter**. a) *huw'sli lø'k* ...

*Linaria cymbalaria* (egl. = vedbend-torskemund, jf. JLange.ODP.I.866f., vel misforståelse

for planten husfred, jf. @*hus'lykke*). \* Skyum.Mors.I.116. \*°Vroue. – b) *Rosenrød*

*Surkløver* (kaldes) *Huslig Lykke*. \*Sønderjy (JLange.ODP.II.169).

2) I underbetydning:

**havre-sæd** subst. ... 5) = **hvid vipstjært** (stedvis også om blå el. grå vipstjært); navnet

hentyder til, at når vipstjærten kommer, er det tid at så havre (jf. @*byg'sæd 3*); i MØJy og

på Djurs ofte med foranstillet verbum el. verbalforbindelse som går på fuglens vippende

hale (jf. Brøndegård.Fauna.II.340): *fimpe-(i)-havresæd*, *gynge/gynke-(i)-havresæd*,

*gynge-i-havresæd* [NØJy-S, Midtjy-Ø (± Anholt og SSamsø), MVJy-SØ, Sydjy (±SV)]

## BILAG 4. Frasem-lemmer – markeret med "kun i forb."

<sup>2</sup>**halse** verb. [af uvis opr.; Hards-V, desuden ndf. anførte optegnelse fra Han; se kort] – kun i forb. halse græs/hø 'af = slå græs, idet leen føres fra det stående græs og udefter (modsat *odde* 'ind, se @odde x). ...

<sup>2</sup>**hang** adj. – *han'* [< 'hang] – kun i forb. være hang efter = have voldsom lyst til. ...

<sup>2</sup>**havlig** adj. ... [jf. vestnordisk *haglig* (= tilpas, behagelig); formen med *å* skyldes vel sammenblanding med *-hovlig*; SydJy (÷NØ og Fanø), SønJy-N (÷ Mandø); spor. i MØJy-S]

– kun i forb. hverken havlig eller -lavlig = ikke tilpas, utilfredsstillende.. ...

**helbredsstund** subst. [< *stund* x; spor. i Vends og Hards; jf. @hellends] – kun i forb.

ikke/aldrig have en helbredsstund. *han høj et en 'hælbre, stå'η = han følte sig aldrig rask og stærk*. Hards-V.

**helkastet** adj. ... [2.sms.led egl. ptc. > *kaste* x; spor. i Hards-NV] kun i forb. helkastet tørv (og som sms. helkastet tørv) = gravet tørv, skæretørv (i modsætning til *-ælte tørv* x).

**hellends** adj. – *hælens* °Darum. – kun i forb. hellends dag el. (som sms.): hellends dag. *han hå'r ålør æn hælens daw'* = han har aldrig en dag med godt helbred. \*°Darum. \*SlaugsH (SVJy).

**helmis-tid** subst. [< *helmis* I; spor. i Nørrejy] – kun i forb. ved helmistid el. (oftest, i plur.) ved helmistider = i tiden omkring allehelgensdag, 1. november.

**hel-skid** subst. – 1.sms.led: *hè'-* °Hellum; *hæl'-* °Erslev; *hæl'-* °Vroue [< 'hel; spor. i Nord- og Midtjy] – kun i forb. slå sin helskid = dø. ... – (overført:) *no slo'r han hans h. = nu giver han tabt*. Vestjy (F.).

**himmel-betændt** adj. [Vends, spredt i Han og Thy-NØ, spor. i Him; se kort //s. 4//]

– kun i forb. himmelbetændt -rendefog = snestorm med voldsom fygning. *hves de va e hemmelbetæntj ræjffog, så skul e nok rammel te, at så skul madammen hentjes* = hvis det var et himmelbetændt rendefog, så skulle det nok passe, at så skulle madammen (dvs. jordemoderen) hentes (til en fødsel). Vends.

<sup>2</sup>**hjerne** verb. – \**hjærn* Hards [< @'hjerne] – kun i forb. hjerne på. *de ka do li' hjærn o'* (= det kan du lige hjerne på), *dvs. stole på, måske rigtigere: bande på – er en gammel sæj* (= vending, udtryk, jf. @sige x) fra *Sunds*. \*Hards (HPHansen.Opt.) || (hertil, i ed: *Den Onde hjærne mig*. \*Djurs (HPHansen.FB.81).

**hovede** verb. – *hoj/'hoj* (K 1.9). – bøjning: -er -et -et (K 6.2, K 6.1) [< 'hoved 2; spor. i SønJy; jf. @hove I] – kun i forb. som det hoveder ham (ikke) = det passer ham (ikke), er (ikke) efter hans hoved. ...

**hunde-bønner** subst. plur. [spor. i AndstH (SØJy)] – kun i forb. bede/læse hundebønner = bande.

**hunde-korn** subst. [kun anførte kilder fra SydJy] – kun i forb. male hundekorn = mukke, brokke sig....

## fra gl. bilag: især i forb.

**ha'sard** subst. – kun i sideformen **a'sart** ...

I) = risiko; ansvar; især i forb. stå sin egen hasard, stå hasarden, på egen hasard [jf. ØMO.; Him-S, Ommers, Fjends, MØJy, SVJy-V og SønJy-NV; spredt i Djurs; spor. i øvrige Nørrejy; se kort] ... *han stoer Ersarten* (= han står hasarden), *dvs. det tager han paa sin Kappel, staar han inde for*. Mols. 'gjø som 'do vet, 'a stæ'ə mi 'ejən a'sart = *gør som du vil, jeg staar min egen Risiko – følger mit eget Hoved*. °Gosmer. ... || stå i hasard for = stå inde for, garantere. ... \*AarbSkive.1966.29.

## BILAG 5: Frasemers placering i mikrostrukturen ved store verber

<sup>1</sup>have

1) have<sup>1</sup>fat, h. <sup>1</sup>i, h. <sup>1</sup>ved; ... 2) || h. det = "være den"; ... 4) || h. (fart) på > efterflg. subst.; ... 6) || h. vejr > *vejr* || h. omgang med, h. hastværk o.l.; ... 7) || h. brug for, nytte af > subst.; ...12) || h. det i ... = h. ondt i; || h. det med = h. for vane; ...13) h. det = være stillet; ...16) det har sig = forholder sig || (hertil mul.) det har (det) på sig; ... 17) det har = der er; ... 18) || h. godt af > <sup>2</sup>god 13.2; ... 24) || h. ledelse = væmmes, o.l. (< *led* adj.)

D) tryktabsforb. m. efterflg. præp

25) h. <sup>1</sup>for. a) ... b) > 35) || ... 26) h. <sup>1</sup>fra > 35) || ... 27) h. <sup>1</sup>i (ngn/ngt) > 1), 35) || ... 28) h. <sup>1</sup>mod (ngn/ngt) > citat i 24); ... 29) h. <sup>1</sup>med (sig) = bringe ...30) h. <sup>1</sup>med at gøre ... 31) h. <sup>1</sup>over = blive bearbejdet ...32) h. <sup>1</sup>på ... 33) h. <sup>1</sup>til = påstå ...34) h. <sup>1</sup>ved (ngn/ngt) > 1

E) som hjælpeverb.

35) = få, udvirke ... || med underforstået obj., hoved-vb. mv.; h. <sup>1</sup>for/<sup>1</sup>fra (om forspand), h. <sup>1</sup>i (om andel)

<sup>2</sup>holde

1) || h. barnet (dåb) ... || h. sig ... 2) – h. <sup>1</sup>op for sol og måne (leg) ... 3) || h. mælken; h. mund; h. for = forhindre; h. sig: a, b) ... 5) ... nej hold! ... 6) || godt/vel holdt ... 7) || h. tæt a) b) || h. sig (foderstand mv.) ... 8) || h. sig = stå sig ... 11) || h. mon = h. takt; h. hug/slag ...; h. hammel med ... 14) h. snak med ngn ... 15) – h. <sup>1</sup>af sig, <sup>1</sup>fra sig = styre til venstre/højre ... 16) || h. <sup>1</sup>lade-/<sup>1</sup>tømmevogn ... 17) || h. <sup>1</sup>halvt = investere fælles > <sup>1</sup>halv 1.5 b... 18) || h. hus > *hus* ... 19) || h. for nar – h. <sup>1</sup>af > 24); ... 21) h. sig (over) = klage

F) tryktabsforb. m. efterflg. præp.

22) holde <sup>1</sup>ad = blive ved ... 23) h. (ngn) <sup>1</sup>ad sig = indsmigre sig hos ... 24) h. <sup>1</sup>af. a) b) c) d) e) f) g) ... 25) h. <sup>1</sup>an. a) b) || ... 26) h. <sup>1</sup>bi > <sup>2</sup>bi ... 27) h. <sup>1</sup>for. a) b) c) ... 28) h. <sup>1</sup>frem. a) b) ... 29) h. <sup>1</sup>hen ... 30) h. <sup>1</sup>her ... 31) h. <sup>1</sup>hid ... 32) h. <sup>1</sup>igen ... 33) h. <sup>1</sup>i ... 34) h. <sup>1</sup>inde. a) b) ... 35) h. <sup>1</sup>med ... 36) h. <sup>1</sup>op. a) b) c) d) e) f) ... 37) h. sig <sup>1</sup>op. a) b) ... 38) h. <sup>1</sup>på. a) b) ... 39) h. <sup>1</sup>sammen. a) b) ... 40) h. <sup>1</sup>til. a) b) c) d) ... 41) h. sig <sup>1</sup>til ... 42) h. <sup>1</sup>ud. a) b) c) ... 43) h. <sup>1</sup>ved. a) b) c)

hugge

1.1) || h. <sup>1</sup>af > 11a); h. <sup>1</sup>for > 12a; h. <sup>1</sup>ud > 19a

1.2) h. <sup>1</sup>om, <sup>1</sup>ned (høste) || h. <sup>1</sup>af > 11b; h. <sup>1</sup>for > 12b; h. for<sup>1</sup>bi > 13; h. <sup>1</sup>ind > 14; h. <sup>1</sup>om > 16b; h. <sup>1</sup>op > 17a; h. <sup>1</sup>sammen > 18; h. <sup>1</sup>ud > 19b-c

1.3) huggede/hugne tørv

2.2) || h. <sup>1</sup>ud (isfiskeri) || hverken til at h. el. stikke i

3.1) h. <sup>1</sup>op > 17b || h. (et) hus ... 3.2) h. <sup>1</sup>for > 12c ... 4.1) h.-t sukker/melis ... 4.2)

h.-n skulde ... 5) || h. til blok (vippe) ... 7) || h. <sup>1</sup>fråde > *fråde* ... 8) || h. i det, a) b)

10) || h. og tærse/strække – h. <sup>1</sup>af > 11c; h. <sup>1</sup>ud > 19d

C) tryktabsforb. m. efterflg. adv. (på tværs af betydn.inddeling ovf.)

11) h. <sup>1</sup>af. a) b) c) ... 12) h. <sup>1</sup>for. a) b) c) ... 13) h. for<sup>1</sup>bi ... 14) h. <sup>1</sup>ind ... 15) h. <sup>1</sup>ned.

a) > 1.2), b) c) d) ... 16) h. <sup>1</sup>om. a) > 1.2), b) c) d) ... 17) h. <sup>1</sup>op. a) b) ... 18) h.

<sup>1</sup>sammen ... 19) h. <sup>1</sup>ud. a) b) c) d)

ARI PÁLL KRISTINSSON

## Den nye islandske rettskrivningsordboka

*Stafsetningarorðabókin* (The [Icelandic] Spelling Dictionary), edited by Dóra Hafsteinsdóttir, 2006, can be called the official Icelandic spelling dictionary. It contains ca 65,000 lemmas with ca 73,000 words. In addition, there is a chapter, 61 pp, with comprehensive and detailed rules on Icelandic orthography and punctuation, with various examples. The dictionary shows recommended spelling, and the recommended conjugations and inflections, not only of the traditional vocabulary but of a number of recent loans as well. The present article describes the background to this publication and goes on to discuss the general principles which were followed, as well as some matters of opinion which were dealt with in the editing process, e.g. which inflections were to be recommended, and the amount of loan words adopted in the book.

### 1 Innledning

Denne artikkelen handler om en ny islandsk rettskrivningsordbok: *Stafsetningarorðabókin* 2006, heretter *S*.<sup>1</sup> Boken er av betydelig omfang, dvs. den inneholder ca 65.000 lemmer på 672 sider, og dessuten finnes det omfattende skriveregler på 61 sider.

Det blir referert om bakgrunnen til denne utgave mht. Íslensk málnefnd (Islandsk språkråd) oppgaver ifølge islandsk lov inntil 1. september 2006. Jeg diskuterer kortfattet de generelle premissene for en rettskrivningsordbok av denne slags mht. normering av skriftspråket. Så beskriver jeg de retningslinjene som Islandsk språkråd ønsket at redaksjonen tok hensyn til, og der kan man godt merke språkpolitikken som hersker i islandsk språksamfunn. Enkelte problemstillinger ved redigeringen blir diskutert, bl.a. bokens veiledninger om korrekt bøyning og det spørsmålet om en del ord fra det daglige talespråket skulle tas opp som lemmer i rettskrivningsordboken. Man måtte bl.a. lage nye egne arbeidsregler om skrivemåten av fremmedord i islandsk.

### 2 Bakgrunn – gammel og ny lov

Den organisatoriske bakgrunnen til *S* er den tidligere loven om Islandsk språkråd hvor det het bl.a. at språkrådet skulle gi ut skrifter for informering og veiledning om islandsk språk, bl.a. en rettskrivningsordbok. Det var nettopp med henvisning til denne paragrafen at språkrådet hadde gitt ut *Réttritunarorðabók*

<sup>1</sup> Takk til Ruth V. Fjeld og Lars S. Vikør for kommentarer, på NFL-konferansen i 2007, til innlegget som artikkelen bygger på, og takk til redaksjonen for kommentarer til artikkelen.

*handa grunnskólum* (Rettskrivningsordbok for grunnskolen) i 1989, heretter *R*. Den boken var ganske lille, 146 sider, 14.700 ord. Det var helt klart allerede kring dette arbeidet på 80-tallet at språkrådet ønsket senere å gi ut en mye større rettskrivningsordbok, i retning av de ordbøkene som fantes for f.eks. dansk, norsk og svensk. Når man startet arbeidet med den større boken, *S*, etter en forberedningsfase i 1996–2001, fikk Islandske språksekretariat, språkrådets kontor, dette som en praktisk oppgave, på vegne av språkrådet. Dóra Hafsteinsdóttir ble ansatt som redaktør.

Etter en lovendring i 2006 har Islandske språkråd fått en endret rolle og rettskrivningsordbøker er ikke lenger blant språkrådets oppgaver. Ordbøkene er blitt levert til et nytt institutt, Stofnun Árna Magnússonar í íslenskum fræðum (Árni Magnússon-instituttet for islandske studier). Ifølge loven skal det nye instituttet gi språklige råd og veiledninger på teoretisk grunnlag, og rådgivningens mål skal være å “styrke og bevare” det islandske språket (jf. *Lög um Stofnun Árna Magnússonar í íslenskum fræðum* 2006). Denne forpliktelsen, som er i tråd med den tradisjonelle islandske språkpolitikken, har betydning i denne sammenheng. Instituttet har allerede planer om en ny utgave av *S*.

### 3 Ordbokstype – noen sidestykker til *Stafsetningarorðabókin*

Her dreier det seg selvsagt om en rettskrivningsordbok, dvs. en “ordbok der formålet er å angi ordenes rettskrivning ifølge en viss norm” (*NLO*). I *NLO* heter det videre at det i rettskrivningsordbøker, foruten skrivemåten, normalt også angis “ordklasse, genus, bøyning, uttale av visse ord (særlig fremmedord) osv.” *S* er antakelig en nokså typisk rettskrivningsordbok ifølge disse kriteriene.

Ordforrådet i *S*, ca 65.000 lemmaer (med i alt ca 73.000 ord), er all vanlig islandske ordforråd, pluss de fleste islandske personnavn, veldig mange stednavn (både på islandske steder og i andre deler av verden, bl.a. alle statsnavn og respektive nasjonalitetsord), m.m. Det finnes i boken en masse vitenskapsord, f.eks. innenfor kjemi og fysikk osv., også navn på dyr og planter, bergarter m.m.

Målet med *S* var at den kunne fungere som den “offisielle islandske rettskrivningsordboken”. Dette innebærer ikke bare at boken viser den anbefalte skrivemåten, men også: Hvis et ord finnes i ordboken antyder det at ordet ikke strider mot normen om godt skriftspråk, ifølge Islandske språkråd. Man gir også veiledninger om “korrekt” bøyning i boken. I de tilfeller hvor det finnes i faktisk islandske språkbruk noen andre bøyingsformer, som ikke blir representert i boken, så betyr deres fravær antakelig at de faller utenfor språkmyndighetenes forestillinger om normen for optimalt skriftspråk.

Det finnes ingen regler som sier hvem som bør skrive i tråd med rettskrivningsordbokens anbefalinger når det gjelder valg av ord, deres bøyninger og språkbruken ellers. Allikevel viser boken en norm som en språkmyndighet publiserer og derfor kan man kalle den for “den offisielle rettskrivningsordboken”.

De fleste vurderer det vel slik at det er naturlig å forvente at iallfall skolene, og myndighetene ellers, følger de anbefalte formene som finnes i boken. Fra et normteoretisk synspunkt kan man si at publiseringen av *S* på en måte er et eksperiment eller et forsøk på å se om og hvordan en slik omfattende norm for skriftlig språkbruk får gjennomslagskraft i islandsk språksamfunn. Jeg vurderer det slik på nåværende tidspunkt at boken vil ha innflytelse i skolene og på mange måter i det offentlige, og, som følge, at den senere vil ha indirekte innflytelse i den private sektoren.

*S* bygger mye på *R* (1989) hvilket er naturlig ettersom det dreier seg om samme organisatoriske og språkpolitiske premisser, jf. 2 ovenfor. Dessuten hadde man selvsagt også noen utenlandske rettskrivningsordbøker som et slags forebilde.

Det viktigste av dem har uten tvil vært Dansk Sprognævns rettskrivningsordbøker, og dette er nærmere beskrevet i ordbokens forord. Den danske *Retskrivningsordbogen*, 3. udg. 2001, hadde ca 66.000 oppslagsord, dvs. den var av omtrent samme størrelse som *S*. Den organisatoriske rammen kring den var også et mønster for det islandske arbeidet, ettersom den blir redigert av Dansk Sprognævns personale. I etteråret 2001 fikk Dóra Hafsteinsdóttir, redaktør i Islandsk språksekretariat (Islandsk språkråds kontor), to måneders forskningsorlov for å dra til København og bl.a. sitte i Dansk Sprognævns lokaler og treffe medarbeiderne der med spesiell tanke på å bli kjent med arbeidsrutinene kring deres rettskrivningsordbok, de redaksjonelle prinsippene m.m. Det er derfor kanskje ingen tilfällighet at *S* omtrent er av samme omfang som den danske utgaven, og at det i den islandske boken også finnes skrive regler som man henviser ofte til fra ordbokens lemnaer. (Jf. Anita Ågerup Jervelunds og Jørgen Schacks artikkel i denne rapporten om arbeidet omkring de danske rettskrivningsordbøker.)

Det finnes også bl.a. noen norske sidestykker til *S* som det var viktig å se på, f.eks. disse tre bøkene som Boye Wangenstein redigerte/reviderte: *Bokmålsordlista* (1992), *Tanums store rettskrivningsordbok* (1996) og *Bokmålsordboka* (2005).

Av ikke-nordiske bøker som det ble sett på ved forberedelsen av *S* bør særskilt nevnes *Duden. Die deutsche Rechtschreibung* (2000) som har ca 120.000 lemnaer. Der finnes det skrivemåte, skrive regler, bøyning og mange definisjoner.

Premissene for en offisiell rettskrivningsordbok for islandsk er naturligvis av samme slags som for rettskrivningsordbøkene for de andre språkene: de respektive myndighetene har påtatt seg det ansvar at det finnes en norm for rettskrivning som kan brukes i skolen og i offentlige dokument.

#### 4 Retningslinjene og deres gjennomføring

I dette kapitlet kommer jeg inn på hvilke prinsipp som ble lagt til grunn som utgangspunkt i *S* (jf. (1) under 4.1), og hvordan det lyktes å følge dem (jf. avsnitt 4.2–4.6 nedenfor).

#### 4.1 Redaksjonens prinsipp

Islandsk språkråd vedtok følgende prinsipp som redaksjonen skulle bygge på.

##### (1) Prinsippene som redaksjonen bygget på

1. Ettersom det dreier seg om en rettskrivningsordbok behøves det ikke ta med forskjellige fremmedord som kanskje bare er delvis tilpasset, m.m., og som tilhører det uformelle talespråket.
2. Det bør legges vekt på veiledning om korrekt bøyning.
3. Boken skal presentere på en klar måte detaljerte skrivereregler som bygger på utdannings- og kulturdepartementets sirkulærer om rettskrivning og tegnsetning, og er i tråd med de arbeidsregler som Baldur Jónsson har laget, for Islandsk språkråd, om islandsk rettskrivning.
4. Man skal, så vidt det er mulig, anbefale én skrivemåte i de tilfeller der det kan finnes valgfrihet.
5. Man skal ha som prinsipp å vise de formene som kan anses som de mest opprinnelige i islandsk, eller kan anses som den beste språkbruken i den forstand at den bør anbefales overfor andre former.

Pkt. 4 og 5 ovenfor bør forklares litt nærmere.

Prinsipp 4 betød at man i ordboken skulle prioritere i noen tilfeller der det finnes valgfrihet i de gjeldende skriveregelene fra utdannings- og kulturdepartementet, og også i noen tilfeller der man ikke tidligere hadde hatt støtte av noen offisielle regler i det hele tatt. Dette beskrives nærmere i avsnitt 4.6 nedenfor.

Når man snakker om “former” i prinsipp 5 så gjelder det både skrivemåte og bøyning.

#### 4.2 Analyse av noen lemmaer

For å kunne analysere litt hvordan prinsippene i (1) i 4.1 ble satt ut i livet i *S* er det mest hensiktsmessig å se, i (2) nedenfor, på noen typiske lemmaer, fra kapitlet *A* i boken. Dette gjør jeg også for å gi leserne et bedre innblikk i bokens faktiske innhold og i lemmastrukturen.

##### (2) Noen typiske lemmaer i *S*

**annáll** *-inn* annáls; annálar þessa er getið í annál; annálaritari; annálsverður

**Anný** Annýjar

**antígskur** antígsk; antígskt STIGB. -ari, -astur

**Antígva** -ð Antígva Antígva og Barbúda; Antígvamaður

**Anton** Anton, Antoni, Antons

**anóða** *-n* anóðu; anóður, EF.FT. anóðna

**apa** apaði, apað *hún apaði allt eftir honum*



**apalegur** -leg; -legt STIGB. -ri, -astur – **apalega**  
**aragrúi** -*nm* -grúa þar var aragrúi af fólki/bókum

I lemmaene i (2) har man altså veiledninger om skrivemåte, bøyning, ordforbindelser, sammensetningsmetoder, m.m. Eksemplene blir nå forklart litt, ikke minst for å peke på noen (problematiske) trekk som det måtte tas beslutninger om ved redigeringen.

**annáll** ‘annal’: Dette er et substantiv i hankjønn og de opplysningene får brukeren gjennom *-inn* hvilket er endelsen i bestemt form entall; rettskrivningsreglene sier at hankjønnsord har *nm* i nominativ i bestemt form mens hunkjønnsordene har *n*. Dette viser i tillegg den korrekte skrivemåten i bestemt form, hvilket er av stor nytte ettersom det er et svært vanlig problem i islandsk rettskrivning om den bestemte artikkelen med *n* bør skrives med *-n(-)* el. *-nm*. I trykksvak stilling i islandsk uttale hører man nemlig ikke forskjellen på *n* og *nm*; dette må innlæres i rettskrivningsundervisningen. Formene *annáls* og *annálar* viser respektive genitiv entall og nominativ flertall. Eksemplet *þessa er getið í annál* er et konstruert eksempel for å vise at bøyningsformen *annál*, dativ entall, er uten et *-i*. I sterk bøyning av substantiv i hankjønn finnes det nemlig i dativ entall enten endelsen *-i* eller ingen endelse, *-Ø*; også finnes det valgfrihet i en del ord. Derfor er bøyningsformen i dativ entall av sterke hankjønnsord blant de opplysningene som man forventet at kunne være av nytte for brukeren. Eksemplene *annálaritari* og *annálsverður* blir gitt for å vise at man lager sammensetninger av dette ordet enten med genitiv flertall (*annála-*) eller genitiv entall (*annáls-*), og eksemplene antyder i tillegg at det første passer for sammensatte substantiv og det siste for sammensatte adjektiv. Fravær av sammensetning av ordstammen, *annál-*, antyder at man ikke lager sammensetninger av dette ordet på den måte i det hele tatt.

**Anný**: Dette er et kvinnenavn, man angir genitiv entall (*Annýjar*), hvilket også er nyttig i dette tilfellet spesielt for å vise at det bør være et *j* etter *y-* og foran *-ar* ifølge islandske skrivereregler. Fravær av spesielle akkusativ- og dativformer i lemmaet antyder at man der regner med samme form som i nominativ.

**antígskur** ‘antiguansk’: Adjektivet angis i hankjønn entall nominativ, deretter kommer formene i hunkjønn og intetkjønn. “STIGB.” er en forkortelse for *stígbreyting* ‘gradbøyning’; endelsene *-ari* og *-astur* viser at det heter *antígskari* i komparativ og *antígskastur* i superlativ form (hankjønn). – Det var et spørsmål som redaktøren overveide frem og tilbake under hele redaksjonsarbeidet om det var hensiktsmessig å angi gradbøyning av nasjonalitetsadjektivene. Sikkert er det mange som reagerer på dette og mener at ingen har bruk for å si ‘mer antiguansk’ el. ‘mest antiguansk’, eller at det i så fall var et meget periferisk problem. Men redaktøren begrunnet sin beslutning bl.a. på at det finnes fraser i islandsk som *hún er kínverskari en allt sem kínverskt er* ‘hun er mer kinesisk en alt (annet) som er kinesisk; hun er typisk kinesisk’.

**Antígva** ‘Antigua’: Dette proprium er en del av statsnavnet *Antígva* og

*Barbúda* som angis altså også. Genitiv entall er *Antígva*, hvilket viser at ordet tilhører samme bøyingsklasse som ord av typen *auga*. Med *-ð* viser man at ordet bør brukes i intetkjønn. Dette forklares slik at *-ð* er endelsen i bestemt form av substantiv i intetkjønn som ender på vokal (jf. om *-inn* ved *annáll* ovenfor). Slik heter det f.eks. *auga* ‘øye’ i ubestemt form men *augað* ‘øyet’ i bestemt form. Det er ikke vanlig språkbruk å tilføye bestemt endelse til statsnavn men for å være konsekvent med hvordan man angir opplysninger om kjønn av substantiv i ordboken, valgte man altså å bruke endelsene for bestemt form (*-ð* osv.), også i de tilfeller der det ikke er vanlig å tilføye selve endelsen til det respektive ordet. Man kan ha behov for å henvise til grammatisk kjønn i statsnavn i islandske setninger, f.eks.: *Antígva er fallet* ‘Antigua er vakkert (intetkjønn)’ (det bør altså ikke hete *fallegur/falleg* ‘vakker (hankjønn/hunkjønn)’). Mot slutten gir man nasjonalitetsordet *Antígvamaður* ‘antiguaner’.

**Anton:** Mannsnavn; det angis akkusativ, dativ og genitiv form av navnet.

**anóða** ‘anode’: Med *-n* viser man at dette er et substantiv i hunkjønn, dvs. endelsen i bestemt form er *-n* (jf. om *annáll* ovenfor). Så viser man genitiv entall *anóðu*, og nominativ flertall *anóður*. Siden har man “EF.FT.” hvilket er forkortningen for *eignarfall fleirtölu* ‘genitiv flertall’. Den formen skrives ut: *anóðna*. Dette sistnevnte blir diskutert nærmere i avsnitt 4.4.

**apa** ‘imiterere’: Et svakt verb: man viser formen i preteritum (3.p. entall, ind.) og i partisipp. Så kommer det et konstruert eksempel på en frase for å vise bruken av preposisjonen *eftir* + dativ med verbet *apa*.

**apalegur** ‘tåpelig’: Adjektivet angis i hankjønn, og det rekker å vise formene i hunkjønn og intetkjønn med etterleddet *-leg* / *-legt*. Gradbøyningen følger deretter, og man legger merke til at den er annerledes i komparativ enn den er i ordet *antígskur* ovenfor. Til slutt har vi her adverbet **apalega**, i fet stil, hvilket bør tolkes slik at dette ordet egentlig har status som et eget oppslagsord.

**aragrúi** ‘overhendig mengde’: Dette er et substantiv i hankjønn og det har bestemt form på *-nn* (jf. om *annáll* ovenfor). Siste leddet *-grúa* viser formen i oblike kasus. Så gir man et konstruert eksempel for å vise bruken av preposisjonen *af* + dativ med ordet *aragrúi*.

### 4.3 Fremmedord – diskusjon om prinsipp 1

Språkrådet presiserte i dets oppfordringer til redaktøren, at det her dreier seg om en rettskrivningsordbok og derfor skulle man ikke nødvendigvis ta med forskjellige utenlandske ord som hovedsakelig tilhører uformelt talespråk snarere enn formelt skriftspråk (jf. (1)1 i 4.1 ovenfor).

Ved redigeringen av boken var det et av de store spørsmålene å bestemme i hvilken grad fremmedord av den slags skulle tas med. Spørsmålet dreide seg selvsagt ikke om alminnelig godkjente låneord som *banani* eller *píanó* men snarere om fremmedord som ikke hadde sterk fotfeste i skriftspråket, og som faktisk i mange tilfeller også har ekvivalenter av mer “genuint” islandsk ordmateriale. Det

dreier seg altså om ord av typen *meil*, *situasjón*, *djóka* ‘e-post, situasjon, skøye’. Disse ord er vanlige i talespråket men de tilhører ikke de mere formelle registrene i språkbruken og, som følge av det, besluttet man at de ikke skulle med i denne boken.

Det er et kjennetegn ved islandsk språkbruk at man i mange tilfeller har fremmedord/låneord i talespråk som brukes mye mindre i skriftspråk. Stort sett besluttet man altså å unngå å ta med i *S* de ord som ikke er godt tilpasset til islandsk språk og som først og fremst bare finnes i uformelt talespråk. Det finnes svært mange andre låneord i islandsk og de er godt representert i boken, som det også faktisk framgår av eksemplene i (2) i 4.2 ovenfor. Men som et eksempel på at man i *S* har et mere avslappet forhold til låneord enn i mange tidligere islandske ordbøker, kan nevnes at det her finnes fremmedord som *týpískur* ‘typisk’, selv om dette ordet faktisk angis som ”uformelt” i *Íslensk orðabók* (Islandsk ordbok) 2002. Ordet finnes ikke i *R*.

Det var to motsatte argument i diskusjonen om andel fremmedord som skulle tas med. På den ene siden var man oppmerksom på at når man publiserer et fremmedord i boken så betyr det at ordet har vært godkjent av “normeringsinstansen”. På den annen side så anerkjente man det faktum at folk skriver innenfor forskjellige registre og at de derfor har bruk for veiledninger ved den skrivingen, og man trenger veiledning ikke minst akkurat om hvordan fremmedord skal tilpasses slik at de blir en naturalisert del av islandsk ortografi og bøyning. Dette synspunktet var antakelig det viktigste når redaktøren besluttet seg for å ta ordet *týpískur* med i boken.

#### 4.4 Korrekt bøyning – diskusjon om prinsipp 2 og 5

Språkrådet presiserte at boken skulle legge stor vekt på veiledning om korrekt bøyning (jf. (1)2 i 4.1). Eksemplene, som ble gitt og diskutert i forbindelse med (2) i 4.2 ovenfor, viser hvordan man implementerte dette. Men det er særlig grunn til å si litt mere om veiledningen om bøyningens endelsen i genitiv flertall i en klasse svake substantiv i hunkjønn (jf. ordet *anóða* i 4.2 ovenfor). Grunnen til at man angir genitiv flertall i svake hunkjønnordsord er den at i islandsk språkbruk finnes det enten *-a* el. *-na* i genitiv flertall i mange ord av denne typen, og man besluttet å normere dette i boken. Endelsen *-na* er den opprinneligste i denne bøyningssklassen, men det finnes unntak som skyldes bestemte fonotaktiske begrensninger i en del ord, f.eks. *kría*, *lilja*, *buna* m.m. Med hensyn til prinsippet om at man skal vise (og dermed anbefale) den opprinneligste formen i tilfeller som dette (jf. (1)5 i 4.1 ovenfor), velger man altså i *S* å vise bare formene med *-na*-endelsen, hvor det er historisk legitimt og fonotaktisk akseptabelt. Derfor angis det bare formen *anóðna* osv. i de tilfellene. Denne beslutningen har allerede fått kritikk fra noen av brukerne av *S* – selv om dette ble gjort i tråd med tradisjonen fra språkrådets første rettskrivningsordbok, *R* (1989). I forord til *R* sier redaktøren at selv om den svake bøyningen til vanlig er så enkel at det ikke er nødvendig å

vise den, så velger man i *R* allikevel ofte å påminne om endelsen *-na* i genitiv flertall ettersom det kan forventes at mange ikke husker den eller at det råder uklarhet om den (1989:7). Som det framgår av (3) så følger *S* altså tradisjonen fra dens forgjenger:

(3) Anbefalte endelser i genitiv flertall i svake hunkjønnsord fra side 107 i *S*, sammenlignet med tilsvarende ord i *R*

<i>R</i> (1989)	<i>S</i> (2006)
<b>skriða</b> ef. ft. skriðna	EF.FT. skriðna
<b>skrína</b>	EF.FT. skrína
<b>skrúfa</b> ef. ft. skrúfna	EF.FT. skrúfna
<b>skrýtla</b> ef. ft. skrýtlna	EF.FT. skrýtlna
<b>skræða</b> ef. ft. skræðna	EF.FT. skræðna
<b>skræfa</b> ef. ft. skræfna	EF.FT. skræfna
<b>skupla</b> ef. ft. skuplna	EF.FT. skuplna
<b>skurfa</b>	EF.FT. skurfna
<b>skutla</b> ef. ft. skutlna	EF.FT. skutlna

Et annet spørsmål, som også har med prinsippet i (1)2 å gjøre, men allikevel ikke var så problematisk som genitiv flertall endelsen i svake hunkjønnsord, gjelder at det finnes uklarhet i islandsk språkbruk om endelsen i dativ entall i de sterke substantiv i hankjønn som tilhører historisk *a*- og *i*-stammer i norrønt. Spørsmålet dreier seg om de skal ha bøyningendelsen *-i* eller ingen endelse, jf. eksemplet *annáll* i (2) i 4.2 ovenfor. Man besluttet seg for å konstruere fraser i *S* for å gi veiledninger for brukeren om dette fenomenet, ettersom erfaringen fra språkrådgivningen i Islandsk språksekretariat hadde vist at dette spørsmålet var noe som mange språkbrukere ofte lurer på.

#### 4.5 Detaljerte skriveregler – diskusjon om prinsipp 3

Et av språkrådets prinsipp sa at *S* skal presentere på en klar måte detaljerte skriveregler som bygger på utdannings- og kulturdepartementets sirkulærer om rettskrivning og tegnsetning, jf. (1)3 ovenfor. Et problem i forbindelse med dette er at det var nødvendig i en del tilfeller å ta en prinsipiell beslutning om skrivemåten der det ikke fantes tydelige regler fra departementet. Dette dreide seg faktisk hovedsakelig om problemer som liten eller stor bokstav og ett eller to ord. Erfaringen viser at dette stort sett er det mest problematiske spørsmålet for den skrivende allmennheten. I (4) har vi eksempler på ord som det måtte tas stilling til i denne sammenhengen:

(4) *alexandertækni* ‘aleksanderteknikk’, *Alpafjöll* ‘Alpene’, *alpahúfa* ‘alpelue’

Ordet *alexandertækni* bygger på propriet *Alexander* men man vurderte det slik at ordet ikke lenger har noen semantisk forbindelse med det og derfor skrives det *alexandertækni* med liten forbokstav. Stednavnet *Alpafjöll* bør ha stor forbokstav, som proprium, mens *alpahúfa* til gjengjeld, som også bygger på propriet *Alpar*, ikke anses ha tilknytning til “fjellene” og det skrives derfor med liten forbokstav.

#### 4.6 Prioriterte skrivemåter - diskusjon om prinsipp 4

I de tilfeller hvor utdannings- og kulturdepartementets sirkulærer sier at man kan velge mellom to skrivevarianter, så skulle redaksjonen ifølge språkrådets prinsipp (jf. (1)4 i 4.1) så ofte som mulig prøve å prioritere på en eller annen måte, istf. bare å vise begge variantene uten noen anmerkninger.

Dette gjelder f.eks. skrivemåten av adverbiale ledd av følgende type: *alls konar, einhvers konar, eins konar, þess konar* ‘al slags, noen slags, en slags, den slags’ osv.

Grunnprinsippet i de ministerielle skrivereglene (fra 1974) sier at ordforbindelser, som blir brukt i adverbial betydning, skal skrives ifølge deres opprinnelse, dvs. man skal skrive hvert enkelt “ord” for seg (*Aughýsing um íslenska stafsetningu*, paragraf 37, pkt. 5). I tråd med dette skrives det f.eks. *þess háttar* ‘den slags’ mens “*þessháttar*” ikke blir tillatt. Men reglene fra departementet har allikevel valgfrihet, som unntak, i noen adverbiale fraser, bl.a. *alls konar, einhvers konar, eins konar, þess konar* osv. Det kan enten skrives slik, eller *allskonar, einhverskonar, einskonar, þesskonar*, ifølge reglene. Men denne valgfriheten kan anses som uheldig, slik at man i *S*, i tråd med prinsippet i (1)4 (jf. 4.1 ovenfor), valgte den metode å ta i selve ordboken bare lemmaer med særskriving:

(5)

**alls konar** § 76  
**einhvers konar** § 76  
**eins konar** § 76  
**þess konar** § 76

Lemmaene forekommer på deres respektive alfabetiske plass, mens variantene *allskonar, einhverskonar, einskonar, þesskonar* osv. ikke tas med der i det hele tatt. Symbolet § 76 i lemmaene i (5) ovenfor henviser til paragraf nr. 76 i skrivereglene bakerst i boken. Der finner man det som departementets regler sier om adverbiale ledd, altså at ordforbindelser med *konar* eller *kyns* riktignok kan skrives som en enhet, men man tilføyer allikevel at man anbefaler skrivemåten med to ord. Dette beskrives videre med følgende eksempler:

(6)

alls kyns	allskyns
eins konar	einskonar

(*S* 2006, skrivereglene, s. 707)

Med hensyn til prinsippet i (1)4, så tror jeg faktisk at redaksjonen kanskje kunne ha gjort litt mere for å imøtekomme dette ovennevnte ønsket om å prioritere. Det finnes f.eks. valgfrie skrivemåter i islandsk som *ítarlegur* / *yítarlegur*; *ítarlega* / *yítarlega*. I *S* har man følgende lemmer:

(7)

**ítarlegur** -leg; -legt STIGB. -ri, -astur – **ítarlega**  
**yítarlegur** -leg; -legt STIGB. -ri, -astur – **yítarlega**

Her anbefeler man altså ikke den ene skrivemåten overfor den andre, og man har heller ikke henvisninger fra den ene til den andre, slik at brukeren godt kan være seg uvitende om at det overhodet finnes dette valg mellom *ítarlegur* og *yítarlegur*, og mellom *ítarlega* og *yítarlega*. En prioritering, iallfall en henvisning, hadde vært til nytte synes jeg nå, og mere i tråd med språkrådets ønskemål.

## 5 Skrivemåte av fremmedord

Det finnes veldig mange ord av utenlandsk opprinnelse i islandsk særlig innenfor områder som kjemi, musikk, krydderi og noen andre. Men her har man i mange tilfeller ofte manglet normert skrivemåte. Jeg mener at *S* ikke minst er viktig som normerende arbeid når det gjelder disse typer ord. Noen eksempler vises i (8) nedenfor:

(8) *apasji* 'apasje', *aragónit* 'aragonit', *arameíska* 'arameisk', *timjan* 'timian', *talibani* 'talibaner', *panafrikanismi* 'panafrikanisme'

Man kan si at mangelen av fremmedord i godkjent norm for islandsk skriftspråk (iallfall inntil 2006) bl.a. hadde hatt som en følge at man i mange tilfeller ikke hadde klare prinsipper om skrivemåten av fremmedord i en islandsk tekst. Derfor var det blant utfordringene ved forberedelsen av *S* å lage egne nye arbeidsregler om skrivemåten av fremmedord i islandsk. Altså bortsett fra spørsmålet om i hvilken grad man valgte å ta med i boken låneord/fremmedord (jf. diskusjon i avsnitt 4.3), så måtte man allikevel ta stilling til normeringen av skrivemåten i de låneord som man valgte å ta med. I mange tilfeller var det faktisk blitt veldig aktuelt å fastlegge skrivemåten. Det var f.eks. blitt svært vanlig i språkrådgivningen i Islandsk språksekretariat med spørsmål av typen: Hvordan skal jeg skrive *timjan* eller *talibani* eller *panafrikanismi* i en islandsk tekst?

## 6 Sluttord

I våre moderne skriftspråks- og kunnskapssamfunn viser det seg at en stadig større del av befolkningen har stadig mer behov for stadig mer detaljerte

veiledninger innenfor stadig flere områder av språkbruken. I denne sammenheng spiller rettskrivningsordbøker en viktig rolle. Som jeg har vært inne på ovenfor, så søker disse bøkene ikke bare å normere ortografien. Normeringsinstansene i de enkelte land anbefaler i preskriptive ordbøker av denne slags også ordvalg, ordforbindelser og bøyningsformer, og det gjør de i tråd med normer som ligger i retning av noe vi kan kalle for “godt skriftspråk” ifølge den herskende språkpolitikken i de respektive språksamfunn.

Vi vil få se i de kommende årene hvilken kritikk *S* får fra leksikografer og fra de alminnelige brukerne. Forhåpentlig vil noen undersøke bruken av *S*, og muligens blir det også noen som prøver å vurdere om den påvirker det islandske skriftspråket og, i så fall, på hvilken måte. Fra et normteoretisk perspektiv er *S* bl.a. av interesse når det gjelder spørsmål som vid vs. snever språknorm, og forholdet mellom preskriptiv og deskriptiv leksikografi.

### Litteraturliste

- Auglýsing um íslenska stafsetningu* [Sirkulære om islandsk rettskrivning]. Reykjavík: Ut-dannings- og kulturdepartementet 1974.
- Bokmålsordlista*. Rettskrivningsordliste for skole, hjem og kontor. Gjennomgått og godkjent av Norsk språkråd. Red. Boye Wangenstein. Oslo: Universitetsforlaget 1992.
- Bokmålsordboka*. Definisjons- og rettskrivningsordbok. Utarbeidet av Universitetet i Oslo, Institutt for lingvistiske og nordiske studier og Språkrådet. Red. Boye Wangenstein. 3. utgave. Oslo: Kunnskapsforlaget 2005.
- Duden. Die deutsche Rechtschreibung*. Das umfassende Standardwerk auf der Grundlage der neuen amtlichen Regeln. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2000.
- Íslensk orðabók*. Red. Mördur Árnason. Reykjavík: Edda 2002.
- Lög um Stofnun Árna Magnússonar í íslenskum fræðum* [Lov om Árni Magnússon-instituttet for islandske studier]. 2006. <<http://www.althingi.is/lagas/nuna/2006040.html>> 04.03.2008.
- NLO = Nordisk leksikografisk ordbok*. Red. Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson, Bo Svensén. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 4 i samarbeid med Nordisk språksekretariat. Oslo: Universitetsforlaget 1997.
- Retskrivningsordbogen*. 3. utgave. København: Dansk Sprognævn 2001.
- R = Rétritunarorðabók handa grunnskólum*. Red. Baldur Jónsson. Reykjavík: Íslensk málnefnd og Námsgagnastofnun 1989.
- S = Stafsetningarorðabókin*. Red. Dóra Hafsteinsdóttir. Reykjavík: Íslensk málnefnd og JPV ÚTGÁFA 2006.
- Tanums store rettskrivningsordbok. Bokmål*. Revidert av Boye Wangenstein. Gjennomgått av Norsk språkråd. Oslo: Kunnskapsforlaget 1996.

30 ARI PÁLL KRISTINSSON

Ari Páll Kristinsson  
Forskningsdosent, f. 1960  
Stofnun Árna Magnússonar í íslenskum fræðum – Háskóli Íslands  
Neshaga 16  
IS-107 Reykjavík  
aripk@hi.is



## Materialsortering på digital plattform

### Eit steg mot høgare dataintegritet i den vitskaplege leksikografien

For Norsk Ordbok 2014 linguistic documentation comes in the forms of scanned paper slips, electronic slips and corpus quotations, organised through the index system called the Meta Dictionary. Each entry in Norsk Ordbok is directly linked to the corresponding materials in the Meta Dictionary. However, with up to several thousand instances to sort into a complex entry structure, editors needed a sorting tool for classifying instances and storing the sorting itself. This would both improve editorial stringency, allow for variation in approach method and save time. This article, entitled *Classifying instances of linguistic documentation on a digital platform – one step further towards data integrity within academic lexicography*, describes the tool developed for Norsk Ordbok to meet this need, and sums up user experiences through the first two years as positive but still with space for improvement.

#### 1. Innleiing

I 2002 blei redaksjonen ved Norsk Ordbok omorganisert. Omlegginga innebar nyordning både når det galdt administrasjon og struktur, finansiering, faglege løysingar og ikkje minst digitalisering og utvikling av nye verktøy. Éi viktig erfaring vi har gjort desse fem første åra, er at dei faglege omleggingane og alle tiltaka som har med digitalisering å gjere heng uløyseleg saman. Vi vil her sjå nærare på det verktøyet som vi i redaksjonen av Norsk Ordbok kallar *Sorteraren*. Dette er kanskje det beste dømet på korleis utviklinga av tekniske løysingar hos oss både heng tett saman med faglege prioriteringar og får faglege konsekvensar.

Eit krav redaksjonen hadde på seg i 2002 var å dra nytte av at kortarkivet vårt alt var digitalisert. Under eit slikt krav låg sjølvstøtt primært ønsket om effektivisering. Dei første åra vart det då utvikla ein redigeringsapplikasjon som hadde direkte kontakt med det digitaliserte arkivet via den digitale indeksen for språksamlingane, kalla Metaordboka (for nærare omtale sjå Ore 2000). Ordutval og dimensjonering blir no gjord delvis automatisk med utgangspunkt i det digitaliserte arkivet, og underlagsmaterialet er tilgjengeleg på skjerm. I teorien treng altså ikkje lenger redaktørane forhalde seg til skuffar med arkivmateriale. Men eitt problem stod att: korleis sortere digitale arkivkort? Mange av redaktørane

såg seg framleis nøydde til å sortere underlagsmaterialet på skrivebordet i små bunkar med arkivkort. Dette problemet blei akutt for den delen av redaksjonen som arbeider ved universitetet i Trondheim, for det hadde ikkje tilgjenge til det fysiske arkivet, og var såleis nøydde til å bruke dei digitaliserte arkivkorta, som dei berre fekk sjå eitt for eitt på skjermen.

Mot denne bakgrunnen gjekk vi i dialog med utviklarane ved Eining for digital dokumentasjon. Vi ville ha ein digital sorterar som løysing på eit praktisk problem. Sorteraren skulle vere eit tenleg verktøy i materialanalysen, og helst skulle ein spare tid i denne arbeidskrevjande fasen av det redaksjonelle arbeidet. Vi ønskte også å hente kvalitative gevinstar, dvs. å sikre at kvar og ein gjorde så godt empirisk grunnarbeid som mogleg. Metodisk såg vi her ei moglegheit for at Norsk Ordbok i større grad kunne bli empirisk etterprøvbar, med andre ord at dataintegriteten kunne bli høgare med ein digital sorterar på plass. Til sist var det svært viktig for oss at verktøyet vi etablerte ikkje innebar ei einsretting av arbeidsformene. I forkant hadde vi nemleg kartlagt korleis redaktørane faktisk arbeider med materialet. Vi hadde sett at dei arbeider forbausande ulikt. Og her vil vi understreke at dette dreier seg om heilt grunnleggjande tilnæringsmåtar som heng saman med personlegdomen, og at det er umogleg å seie at den eine tilnæringsmåten er "betre" enn den andre. Dette kjem vi tilbake til, i denne omgangen held det med å understreke at vi ønskte å ta vare på denne metodiske diversiteten. Vi ville at det nye sorterarverktøyet skulle bli eit fleksibelt verktøy som kunne tilpassast ulike, individuelle arbeidsformer. Det måtte for all del ikkje bli ei metodisk tvangstrøye. Dette var då ønsket vårt: Vi ville ha eit fleksibelt verktøy som opnar for metodisk diversitet, som sparar tid, hevar kvaliteten og aukar dataintegriteten.

## 2. Løysing: Eit funksjonsrikt system

### 2.1 Plassering i NO-systemet

IKT-løysinga for Norsk Ordbok 2014 er eit system av databasar og IT-verkty til ulike føremål. Norsk Ordbok blir redigert i ein applikasjon som matar data rett inn i ein database. Det er lagringsstaden for ordboksartikkelen, som skal vere resultatet av ei mogleg, men ikkje ei obligatorisk sortering. Tilfanget kjem frå Metaordboka og blir hekta rett på artikkelhovudet, og på denne måten er det direkte samband mellom Metaordbok og ordboksdatabase.

Sorteraren er eit tilbod, eit slags siderom på dette sambandet, der ein kan ordne og annotere tilfanget. Sorteraren ligg derfor utanfor redigeringsapplikasjonen og ordboksdatabase. Sorteraren er tilgjengeleg berre frå applikasjonen, via den einskilde artikkelen.

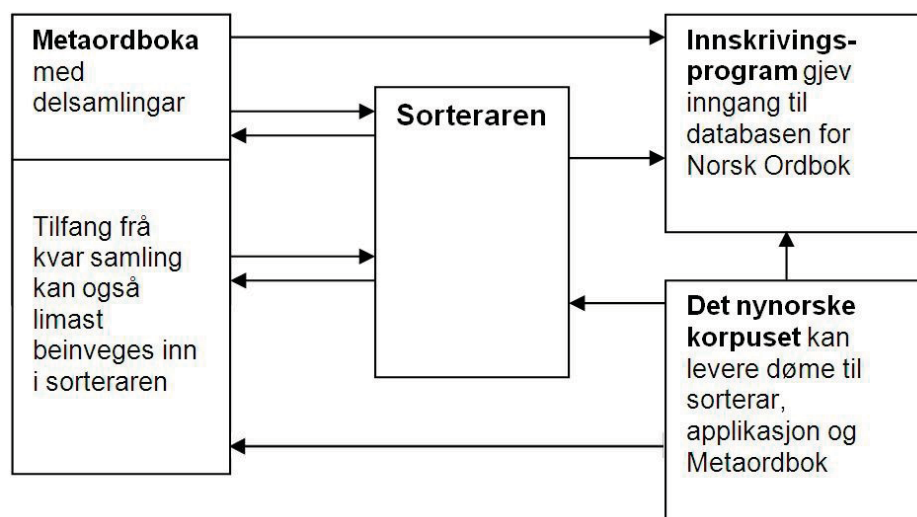
## 2.2 Utforming

Først eit ord om ein tanke vi forkasta. Det hadde vore mogleg å lage eit simultert "skrivebord" på skjermen, der ein brukte musa til å dra belegg over til rett "bunke". Dette ville vere visuelt attkjenneleg. Men setelbunkane har aldri vore noka ideell løysing. Dei har den veikskapen at ein berre ser øvste setel, og er avhengig i postit-lappar på bunkane for å minnst sorteringskriterium. Det var også eit poeng å unngå musbruk og halde seg mest mogleg til tastaturet.

Vi valde i staden ei løysing som tek vare på den mentale sida ved sorteringsprosessen, nemleg sjølv kategoriseringa. Poenget er då å kunne (1) merkje eller annotere kvar setel, (2) sjølv velje sorteringskriterium, (3) kunne merkje opp setlane etter så mange kriterium som ein vil, (4) kunne sortere framlengs og baklengs etter eitt eller fleire kriterium samstundes, (5) lagre sorteringa som grunnlag for framtidig arbeid.

Sorteraren ser ut som eit rekneark, der venstre kolonne listar opp alle belegg linje for linje. Kvar belegg frå Metaordboka er representert med eit symbol. Når ein klikkar på symbolet kjem setelen opp til venstre på skjermen. Ein kan leggje til kolonnar (for sorteringskriterium) og gje dei tittel etter eige skjøn i høve til kvart ord.

Ein stor føremon ved denne arbeidsmåten er at når tilfanget er ferdig annotert og sortert ser ein straks kva ein har i kvar "bunke", og ein ser heilskapen i sitt eige sorteringssystem. Sorteraren gjer dermed tilfanget langt meir oversiktleg enn ein nokon gong før har hatt det.



Figur 1. NO 2014: Kopling av tilfang til og frå sorteraren. Pilene syner kva veg tilfang kan koplast.

### 2.3 Innhald, deling og flytting

Når ein opprettar ein sorterar, kan ein hente inn alle belegg som finst i Metaordboka under det aktuelle oppslaget. Men ein kan også hente inn tilfang seinare, frå korpuset eller nytt tilfang som kjem inn i samlingane under Metaordboka. Dette er nyttig både under redigering og manusrevisjon.

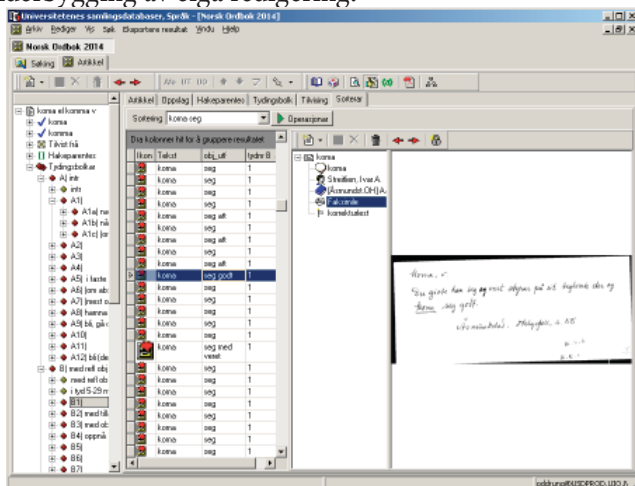
Det kan vere mange grunnar til at ein sorterar bør delast i fleire, frå homogراف-separering til å halde styr på til dømes tydingsinndeling i store materialmengder. Dette er lett å få til gjennom den innebygde funksjonsmenyen. En kan også flytte ein sorterar til ein annan artikkel, eller slette ein mislykka sorterar.

Frå sorteraren kan ein kopiere belegg direkte over på alle nodar i artikkeltreet. Dette er særleg kjekt når ein skal grunnge ei uventa form eller ei sjeldsynt tyding frå eit ellers stort artikkeltilfang.

### 2.4 Skrivning og innsyn

Berre redaktøren kan skrive direkte i sorteraren. Det er såleis hans eller hennar private notatblokk. Men Norsk Ordbok er eit ope prosjekt der alt heng saman med alt. Andre redaktørar kan ha bruk for å sjå korleis ein kollega arbeider med eit visst ord; systemet for kvalitetskontroll av manus kan ha bruk for å sjekke både heilskap i materialbruk og ein detalj her og der. Derfor kan alle redaktørar sjå på alle sorterarane attom kvar artikkel, om dei vil.

Når Norsk Ordbok ein dag blir ei elektronisk ordbok, vil det vere mogleg å gjere sorterarane tilgjengelege for innsyn om ein så vil. Det som i alle fall blir tilgjengeleg, er den dokumentasjonen som er knytt direkte til nodar i artikkeltreet. Førstels avgjer kvar redaktør kor mykje dokumentasjon han eller ho vil leggje fram som underbygging av eiga redigering.



Figur 2. NO 2014: Sorterar til verbet *koma* som dekkjer sambandet "koma seg" (Tyding nr B1). Til venstre ser ein artikkeltreet, med tyding B1 markert, til høgre ser ein det belegget som er markert i sorteraren.

### 3. Resultat: erfaringar og evalueringar

Intensjonen ein har med ein applikasjon blir ikkje alltid realisert av brukarane slik ein hadde tenkt. Etter knappe to år med sorteraren på plass var det då svært interessant for oss å kartleggje bruken av han. 25 redaktørar i to ulike byar har fått eit tilbod om å bruke sorteraren. Dei har ikkje fått noko *påbod* om det, for i utgangspunktet skulle sorteraren dekkje eit uttalt behov hos nokre, men det var jo slett ikkje alle som hadde etterlyst ein digital sorterar. Etter knappe to år er det akkumulert 1446 artiklar som har sorterarar knytte til seg. Det er opplagt at redaktørane i ulik grad og på ulik måte har tatt dette verktøyet i bruk.

Artiklar med sorterar	færre enn 5	10–60	60–100	fleire enn 100
redaktørar	4	13	4	4

Figur 3. Redaktørane ordna i grupper etter kor mange artiklar dei har oppretta ein sorterar for.

Ved å gå databasen etter i saumane kan ein finne ut ganske mykje om bruken av sorteraren. I tillegg har vi intervjuet nokre av redaktørane i kvar gruppe for å få svar på desse spørsmåla:

- Når opprettar ein redaktør ein sorterar?
- Korleis blir sorteraren nytta?
- Sparar ein tid?
- Er kvaliteten på arbeidet blitt høgare?
- Er dataintegriteten blitt betre?

Når ein analyserer svara på desse spørsmåla, ser ein om intensjonane bak utviklinga av sorteraren er blitt realiserte.

#### 3.1 Når opprettar ein redaktør ein sorterar?

Som venta er det det slik at ein sorterar blir oppretta når ein redaktør skal redigere eit ”stort” ord, dvs. eit ord med stort materiale bakom seg i arkivet. Dette hadde vi rekna med, for ein når jo ei smertegrense for kor mykje informasjon ein kan greie å ha oversyn over i hovudet. På eit eller anna tidspunkt må ein lage seg ”hugselappar”. Men det vi ikkje hadde rekna med, er kor store forskjellar det er på kva redaktørane reknar som ”store” ord, med andre ord kor mykje materiale den einskilde tykkjer det er greitt å halde oversyn over utan hjelpemiddel. Nokre definerte eit stort ord som eit ord med meir enn 20 belegg attom seg i arkivet, medan andre definerte eit stort ord som eit ord med fleire enn 250 belegg attom seg.

Ein tilleggsgrunn for fleire var forventninga om kompleks tydingsstruktur. Om ein raskt ser at ordet kjem til å bli redigert med mange tydingsnummer fordi tydingsstrukturen er kompleks, opprettar redaktøren ein sorterar sjølv om ordet ikkje har svært mange belegg attom seg i arkivet. Det same gjeld for ord som må

homografseparerast, dvs. at belegga under eitt oppslag i arkivet faktisk skal fordelast på to forskjellige homografar og dermed to forskjellige oppslagsord.

Til sist nemnde fleire at typiske dialektord, dvs. ord med stor målførevariasjon, ofte får knytt ein sorterar til seg. Norsk Ordbok skal jo omfatte både den normerte skriftlege nynorsken og dei munnlege dialektane. Det gjer ein ved å ordne og føre opp dialektvariasjon under kvart oppslag. I mange tilfelle er formvariasjonen svært stor, og i slike tilfelle blir sorteraren nytta for å kartleggje denne typen av variasjonsmønster.

Ikkje uventa ser det altså ut som om ein sorterar blir oppretta for å behalde oversyn over komplekst materiale. Fleire seier eksplisitt at dei alltid vurderer tidsbruken ved å opprette ein sorterar. Der dei trur dei greier å halde oversynet utan sorterar, lèt dei vere å opprette ein, for dei trur altså at dei sparar tid på denne måten. Det må også seiast at dei som relativt rutinemessig opprettar ein sorterar for alle ord som har meir enn 20-30 belegg attom seg, har integrert sorterarbruken i den vanlege arbeidsrutinen sin som redaktør. Vi kjem tilbake til erfaringane desse redaktørane har gjort seg.

### 3.2 Korleis blir sorteraren nytta?

Svaret på dette spørsmålet var det ikkje så lett å lese ut av databasen, men etter å ha intervjuet ulike redaktørar, teiknar det seg eit bilete av to ganske forskjellige måtar å bruke sorteraren på. Interessant nok er det ikkje eintydig om desse to ulike måtane fordeler seg på anten storbrukarane eller dei som brukar dette verktøyet minst.

Vi har valt å karakterisere dei to bruksmåtene for ei induktiv tilnærming og ei deduktiv tilnærming. Den reinaste induktive metoden blir av ein av brukarane også kalla "notatblokkmetoden". Redaktøren går her til materialet utan førehandsoppfatningar om struktur, og opprettar ein sorterar "nedanfrå" så å seie. Det vil blant anna seie at kategoriane som vert oppretta får ganske innhaldstunge overskrifter, og at samstundes med at materialet blir gjennomgått, blir kategoriseringssystemet raffinert og justert ved at kategoriar blir delte eller at nye kjem til. Redaktøren har ikkje primært fokus på eit ryddig sluttprodukt, og det blir typisk nok ikkje eit ein-til-ein-forhold mellom artikkelstrukturen og strukturen i sorteraren. Om redaktøren skulle ønskje å kople relevant grunnlagsmateriale til kvart tydingsnummer i artikkeltreet, krev det ein eigen ryddeprosess til slutt. Dei redaktørane som arbeider slik, viser til fordelane ved å gå til materialet utan førehandsoppfatningar, for slike oppfatningar kan kanskje vere galne og stå i vegen for den analysen som høver beste med det materialet faktisk viser. Dei meiner nok også at metoden yter materialet størst rettferd.

Den reinaste deduktive metoden går ut på at ein i forkant har laga seg eit grovt kart over forventa relevant semantisk og dialektbasert variasjon, og at ein opprettar ein sorterar med kategoriar som speglar desse forventa variablane. Slike hypotesar om struktur blir etablerte på grunnlag av grunnmanuskriptet til Norsk

Ordbok (ein kompilasjon av alle dei eksisterande ordbøkene over nynorsk pr. 1940), og på bakgrunn av andre ordboksverk, grammatikkar og dialektoversyn.

Med denne arbeidsforma blir sorteraren nytta til å teste og justere hypotesar som er formulerte på førehand, og ein relaterer sin eigen materialgjennomgang meir eksplisitt til tidlegare arbeid. Metoden gjer det lettare å etablere eit ryddig sluttprodukt, og den gjev typisk nok eit større samsvar mellom sorterarstruktur og ordboksstruktur. I neste omgang blir det lettare å kople materialet i sorteraren til nodane i artikkelstrukturen.

Vi vil likevel understreke at i praksis er det sjølvstøtt oftast slik at metodane i nokon mon glid over i kvarandre. Dei som primært arbeider deduktivt får plutseleg auge på eit særtrekk i materialet som gjer at dei må justere den opphavlege strukturen. Typisk nok er det kanskje dei faste ordsambanda, dvs. dei konvensjonaliserte kollokasjonane som slik "materialiserer seg" undervegs. På den andre sida er ikkje dei som arbeider mest induktivt heller utan føresetnader, og det er sjølvstøtt slik at dei konsulterer tidlegare arbeid før dei set i gang. Dei lèt likevel tidlegare arbeid i mindre grad styre eigen analyse.

Det som er heilt klart er at intensjonane våre med å lage eit fleksibelt verktøy som opnar for metodisk mangfald, er oppfylde. Reint konkret er det slik at nokre nummerer tydingane i sorteraren med éin gong, t.d. "1", "2", "2a", "2b" osv. Andre gjev tydingane titlar i form av meir innhaldstunge stikkord, t.d. "lokativ", "temporal", "durativ" osv. Nokre plukkar berre ut målføresetlane og markerer dei t.d. med "Målf.", medan andre annoterer dei ulike dialektformene for seg. Somme gonger blir tekstdøme merkte med å føre opp kollokat, og såleis blir det lagt til rette for ei strengt kontaktbasert sortering. Nokre tilpassar sorteraren slik at dei kan føre inn alle heimfestingane, og alle belegg frå eldre kjelder, medan andre igjen markerer alle idiom. Nokre hakar av dei belegga dei tykkjer er gode kandidatar for å bli tatt med i den endelege ordartikkelen som sitat. Nokre brukar mange kolonner eller fleire sorterarar for å halde orden på alle informasjonstypene, medan andre heller lagar ulike innhaldstunge annoteringar innanfor same kolonne i sorteraren. Felles for alle systema er prinsippet om at all konsekvent annotering gjer det mogleg å søke fram dei aktuelle delmengdene av materialet etterpå.

Dette inneber ei raffinering av tilfanget som også aukar eigenverdien av ord-samlingane for seinare forskning. Om eit stort underlagstilfang for eit einskildord er merkt opp i sorteraren med kollokat, til dømes, blir det enkelt å plukke ut og bruke dei setlane til andre føremål seinare.

Den store fleksibiliteten systemet opnar for, er slik vi ser det ein stor føremon. Det gjer det mogleg for den einskilde redaktøren å vidareføre si eiga arbeidsform. På den andre sida kan det tenkjast at dette opne systemet kan framstå som krevjande for dei av redaktørane som ikkje, eller i svært liten grad, har tatt sorteraren i bruk.

### 3.3 *Sparar ein tid?*

Det er viktig å nemne at fire av redaktørane så å seie ikkje har tatt sorteraren i bruk. Dette er fire av dei eldste redaktørane (over 60 år), og sams for dei er at dei viser til at dei for lenge sidan har etablert gode rutinar som fungerer for dei. Dei har halde fram med å sortere papirsetlar på skrivebordet, og meiner det ville føre til unødig tidsbruk å forsere den tekniske terskelen som sorteraren ber med seg. Dei forventar altså ikkje at bruk av sorteraren skal gje nokon tidsgevinst. I samband med denne undersøkinga har vi sett at vi bør bruke litt meir tid på å gje innføring i bruk av sorteraren for dei som argumenterer med tidkrevjande teknisk terskel.

Bortsett frå desse fire redaktørane meiner alle dei andre at sorteraren gjev tidsgevinstar. Som alt nemnt gjeld dette for mange med visse atterhald. Sorteraren gjev tidsgevinstar for visse (store) ord, men ikkje generelt, og redaktørane gjer slike vurderingar av moglege tidsgevinstar frå ord til ord.

Tre tidssparingspunkt blir nemnde. For det første ligg det potensiell tidssparing i det å ikkje miste oversynet over ei stor materialmengd. Utan merking er faren der alltid for at ein må gjere operasjonar fleire gonger fordi ein ikkje finn ting att, eller fordi ein gløynde noko i førre runden. Sorteraren hjelper slik til med at ein raskt kan orientere seg i materialet og utan omvegar finne fram til relevante døme.

For det andre er det svært viktig for mange at sorteraren gjer det mogleg å annotere og kategorisere etter fleire kriterium samstundes. Når ein sorterer arkivbelegg i haugar på eit skrivebord, blir det eit problem om det finst fleire måtar å kategorisere materialet på som ikkje er synkroniserte. Typisk nok er dette tilfellet med dialektbasert, kjeldebaserst eller heimfestingsbasert kategorisering i møte med semantiske kategoriar. Papirløysinga er å sortere i fleire omgangar. Ulempen med denne metoden er at sorteringane blir borte etter kvart. Ein får ikkje lagra og slik rekapitulert sorteringane om ein har bruk for det. Med den digitale sorteraren kan ein sortere etter like mange kriterium som ein i utgangspunktet har brukt i annoteringa. Og ein kan samsortere etter fleire kriterium, t.d. etter heimfestingsstad og dialektform, eller etter dialektform og tydingsnummer. Ved dei store orda gjev dette vesentlege tidsgevinstar.

Det faktumet at desse sorteringane blir elektronisk lagra saman med artikkelen dei utgjer det empiriske grunnlaget for, blir no utnytta i opprettings- og kontrollfasen. Dette er det tredje punktet der mange redaktørar viser til tidssparing. Når ein kontrollør skal vurdere materialanalysen, eller ein i opprettingsfasen ikkje lenger hugsar kva grunnlaget var for vala ein har gjort, er det ein føremon at materialgjennomgangen faktisk er lagra. Både i funksjon av å vere redaktør som skal rette opp eigne manuskript eller i funksjonen som gjennomlesar som skal kvalitetssikre manusa til andre, peikar redaktørane på tidsgevinstane ein sorterar fører med seg.



Vi meiner då at det er riktig å konkludere med at bruk av sorteraren gjer at ein sparar tid. Men det føreset altså at ein brukar sorteraren slik som no, dvs. etter ei ganske gjerrig vurdering av tidsbruken når ein vel å opprette ein sorterar i utgangspunktet. Eit par av dei som i størst grad har integrert bruk av sorteraren i arbeidsrutinen sin, ser tidsgevinstane i opprettingsfasen som så viktige at dei også vel å opprette sorterarar til etter måten små ord.

### 3.4 Er kvaliteten på arbeidet blitt høgare?

Alle dei som brukar sorteraren meiner kvaliteten på arbeidet er blitt høgare. Dei viser til litt ulike element i kvalitetsomgrepet. Éin viktig grunn for mange er at dei kjenner at det empiriske grunnlaget for den analysen ordboksartikkelen er eit uttrykk for, er meir solid. Dei er tryggare på at dei har tatt omsyn til alle nyansane i materialet, og sikrere på at dei ikkje har gått glipp av vesentlege døme eller kollokasjonar. Éin redaktør viste til at sorteraren hjelper til med å etablere ei naturleg rekkjefølgd på tydingsnumra. Ein får lettare oversyn over kor prominente dei ulike tydingane er i materialet, og reint visuelt får ein indikasjonar på kva for tydingar som er primære og sekundære.

Dei som arbeider mest induktivt, held det fram som ein kvalitativ føremon at ein kan møte materialet utan hypotesar og slik yte det betre rettferd. Det kan kanskje lesast ut som at den induktive arbeidsforma gjer redaktørarbeidet til eit meir sjølvstendig vitskapleg forskararbeid, med resultat som er etablerte med uavhengig evidens i forhold til tidlegare ordboksarbeid. Vi vil understreke at det same gjeld for ei moderat deduktiv tilnærming. Om ein er open for å justere og endre dei hypotesane om struktur ein hadde då ein byrja å sortere materialet, vil ein ved hjelp av sorteraren gjere empirisk grunnlagsarbeid uansett.

Eit svært viktig punkt i høve til dataintegritet er det som har med kvalitetskontroll å gjere. Det er heilt opplagt at kontroll og oppretting av eigne manuskript er blitt betre fundert. Ein slepp å byrje på nytt med materialet når ein kvalitetskontrollerer manuskripta sine. I staden kan ein lett ta ein titt i sorteraren om det er noko ein lurar på. Det same gjeld for dei som kvalitetsikrar andre sine manuskript. Spesielt viktig er kanskje dette for redaktørar som er i ein opplæringsfase. Ein svært viktig del av opplæringa er jo å lære seg å analysere materiale rett. No kan dei som har ansvar for manuskripta til nybyrjarane, på ein lettvint måte også kvalitetssikre materialanalysen deira. Dette er blitt eit viktig punkt i redaksjonen vår, for så langt har vi i desse fem åra hatt nye redaktørar under opplæring kontinuerleg. Sjølv om vi no nærast har full bemanning, ser vi at det i ein stor redaksjon vil vere utskifting jamleg, og då gjeld det å gjere opplæringa av nye redaktørar mest mogleg effektiv.

Vi ser slik at det som i utgangspunktet var tenkt som ei løysing på eit praktisk problem, faktisk også har gitt kvalitetsgevinstar vi ikkje hadde sett så klart for oss då vi byrja med arbeidet. Vi *hadde* sett for oss at det ville bli lettare å skaffe seg oversyn og verre å gå glipp av relevant materiale, men vi hadde ikkje så klart sett

kva for gevinstar dei lagra sorteringane ville ha i kontroll- og kvalitetssikringsarbeidet.

### 3.5 Er dataintegriteten blitt betre?

Norsk Ordbok er ei vitenskapleg ordbok. Grunnen til det er at ein tek sikte på å kartleggje og dokumentere og analysere ord som anten ikkje er kartlagde før i det heile, eller som ikkje er kartlagde på denne måten før. Dokumentasjon er her eit nøkkelord. I ei dokumentasjonsordbok skal ein ved hjelp av materiale eller kjelder underbygge dei påstandane om ordtyding, utbreiing og formell variasjonsbreidd ein kjem med i ordartiklane. Eit krav til god vitenskap er at den er empirisk etterprøvbar. Det vil seie at dei påstandane eller konklusjonane ein legg fram, skal kunne prøvast mot grunnlagstilfanget, og avkreftast om tilfanget er brukt urett. Det skal vere mogleg å finne og leggje fram motprov mot den konklusjonen eller den påstanden ein kjem med.

I eit vitenskapleg dokumentasjonsordbok er påstandane eller konklusjonane som ein legg fram om språk, implisitt formulert. I den ferdige ordartikkelen vil tydingshierarkiet som ein har ordna tydingane til ordet i typisk nok vere ein slik konklusjon. Ein annan konklusjon vil vere det oversynet over distribusjonen av dialektvariantar som blir presentert. Ein tredje vil vere det sambandet som er etablert mellom oppslagsform, språkhistorisk opphavsform og dialektvariantar i artikkelen, det vil seie sjølv avgrensinga av lemmaet. I mange tilfelle vil dette sambandet bli etablert for første gang i ei ordbok av vår type som handterer både målføre, normert tale og skrift.

Spørsmålet blir då om sorteraren har gjort Norsk Ordbok meir *vitenskapleg* enn ho var før. Ein annan måte å stille dette spørsmålet på er å spørje seg om dataintegriteten er blitt betre. Med dataintegritet meiner vi her at det er eit godt samsvar mellom konklusjonane om materialet i ordboka og materialet sjølv. Og vi vil leggje til – dette samsvaret skal også kunne etterprøvast.

Det viktigaste einskildet elementet i den nye Norsk Ordbok-redaksjonen som betrar dataintegriteten, er faktisk ikkje sorteraren, men Metaordboka. At det digitaliserte arkivet gjennom Metaordboka er direkte kopla til den elektroniske ordboksdatabasen, gjer at ein kan gå fram og tilbake mellom ordartikkel og arkiv, og slik *etterprøve* ordartiklane. Dette er ein revolusjonerande ny måte å tenkje dokumentasjonsordbok på, og det er ikkje før Norsk Ordbok er tilgjengeleg for alle som elektronisk ordbok, at dette potensialet blir fullt ut realisert. For oss i redaksjonen er det likevel allereie ein realitet, vi har ei direkte kopling mellom empiri og ordartikkel, og vi utnyttar dette til etterprøving i det daglege kvalitetssikringsarbeidet.

Sorteraren representerer i denne samanhengen ei raffinering av denne koplinga mellom materiale og ordbok. Koplinga mellom materialet i Metaordboka og artiklane i ordboksdatabasen er jo ikkje nærare spesifisert og fortolka. Den som skal gå analysen etter i saumane må i og for seg gjere heile det leksikografiske

grunnlagsarbeidet om igjen. Sorteraren bøter på dette, for i sorterarane er det jo nettopp spesifiserte og fortolka analysar av tilfanget som blir lagra saman med artikkelen. Ved å kople sorterarane til ordboka, ja gjerne til dei enskilde nodane i artikkelstrukturen, er det etablert ei moglegheit for etterprøving på meir detaljert nivå. I dei tilfella der ein sorterar er kopla direkte mot kvar node i artikkelstrukturen, er kvar node empirisk fortolka mot eit større materiale enn det illustrerande dømet som før fekk ivareta denne funksjonen.

Før vi konkluderer på dette punktet, må vi understreke at undersøkinga vår har vist at denne eksplisitte oppkoplinga mellom nodane i ordartiklane og sorteraren til vanleg ikkje blir gjort. Med eitt unnatak er det ingen redaktørar som tek den ekstrajobben det er å kople sorteraren opp mot ordartikkelen, node for node. Hovudgrunnen har med tidsbruk å gjere. Slik oppkopling tek tid som ikkje vil gje gevinstar for produksjonen av papirutgåva. Omsynet til ei framtidig elektronisk utgåve er ein luksus redaktørane ikkje unner seg med dei krava til framdrift som dei har å forhalde seg til. Den andre grunnen er at dei som i størst grad brukar sorteraren som ei nedanfrå-og-opp notatblokk, ikkje har etablert ein så tydeleg struktur i sorteraren at ein utan vidare kan kople den opp mot ordboksdatabase. Her må det verkeleg eit reelt tidkrevjande ryddearbeid til før slik kopling kan gjerast.

Idealet om auka dataintegritet er først og fremst realisert på den digitale plattformen i og med Metaordboka. Opprettinga av sorteringar i samband med artiklane i databasen inneber at etterprøvarheita er realisert på eit meir spesifisert nivå. I og med oppkoplingsmoglegheita som ligg i systemet, kunne ein tenkje seg ei endå tettare og meir fortolka kopling mellom materiale og ordartikkel, men av grunnar som har med tidsbruk å gjere, er det berre éin redaktør som gjer slik oppkopling. Det er viktig å understreke at dette idealet om dataintegritet berre lèt seg realisere på den digitale ordboksplattformen.

#### 4. Konklusjon

For oss er det opplagt at ei digital kopling mellom materiale og ordbokslemma via Metaordboka gjev auka dataintegritet. Det er blitt mogleg å etterprøve det redaksjonelle arbeidet ved å gå fram og tilbake mellom ordboksdatabase og det digitalisert arkivet. I tillegg meiner vi at dei lagra sorterarane som no i mange tilfelle er hekta på ordboksartikkelen i databasen, gjer dei konklusjonane som ordboksartikkelen er eit uttrykk for, etterprøvbare på eit endå meir spesifisert nivå. I og med sorteringane er materialet fortolka, og desse fortolkingane er lagra. Det høgaste nivået av integrering mellom materiale og ordbok får ein ved å kople sorteringane til artikkeltreet node for node. Dette er ei moglegheit som redaksjonen så langt i liten grad nyttar seg av.

Det er slik eit stort arbeid som ligg framfor oss om den moglegheita vi no har fått til å integrere kjelder og ordboksdatabase skulle realiserast gjennomgåande.

I ein framtidig digitalisert versjon av ordboka kunne vi eksplisitt kople sorteringsresultatet til nodane i artikkeltreet slik at materialanalysen også blei tilgjengeleg for brukarane. Men dette set store krav til sorteringane. Dei må då vere ryddige og ferdiganalyserte med ein struktur som er i samsvar med den ferdige artikkelen. Dette inneber heilt klart eit stort meiarbeid. Om ein verkeleg skulle ha som ambisjon at brukarane av ordboka ved alle ord skulle få tilgjenge til oppkoplale materiale, ligg det dessutan ei kjempestor utfordring på metodesida. Etter vår meining måtte ein då uniformere sorterarbruken i større grad enn no, og dette ville gripe radikalt inn i arbeidsformene til den einskilde. Vi ser altså at dei ideelle krava til dataintegritet er i konflikt for det første med ønsket om effektiv tidsbruk og for det andre med ønsket om individuelt tilpassa verktøy og metodisk mangfald.

Vi vil likevel konkludere med at dei intensjonane vi hadde når vi sette i gang arbeidet med å utvikle ein sorterar, langt på veg er realiserte. Vi har fått eit fleksibelt metodisk hjelpemiddel, verktøyet har gitt tidsgevinstar, kvaliteten på ordboksproduktet er blitt høgare og til sist, dataintegriteten er blitt betre. Vi vil halde fram med å oppmuntre redaksjonen til å realisere det potensialet for betre metode som ligg i sorteraren, men vi vil av omsyn til metodisk fleksibilitet og effektiv tidsbruk førebels ikkje fullt ut realisere den optimale etterprøvbarheita for eksterne brukarar.

## Litteratur

Ore, Christian-Emil Smith 2000: Metaordboka. I: *Ord om Ord* 6, 30–32.

Kristin Bakken  
prosjektdirektør, f. 1959  
kristin.bakken@iln.uio.no

Oddrun Grønvik  
hovudredaktør, f. 1947  
oddrun.gronvik@iln.uio.no

Norsk Ordbok 2014  
Universitetet i Oslo  
postboks 1021 Blindern  
NO-0315 Oslo

ANNA BRAASCH

## Verber med fælles syntaktiske konstruktioner og semantisk slægtskab

This paper deals with the issue of observable relationships between the syntactic behaviour and semantics of selected verb groups. The aim is to investigate the feasibility of exploiting syntactic similarity for effective uncovering of semantic kinship. The investigation is inspired by the demand for large-scale lexical resources combining morphological, syntactic and semantic descriptions on the one hand, and on the other by the fact that semantic networks, like WordNet usually do not at all, or only to a very limited degree provide syntactic information. Both aspects show that there is a need for effective means that combine syntactic and semantic description methods and resources containing the two information types.

The pilot project presented here discusses verbs sharing selected syntactic descriptions; they are extracted from the Danish Lexicon for Language Technology Applications (STO) and grouped together according to semantic labels, which in the first run are intuitively selected. The idea is to derive underlying semantic information systematically from the syntactic surface structures described. Although this approach starts from the 'wrong side' compared to other, well-known analyzes of the semantics-syntax relationship, it might be exploited in prediction of verb senses.

### 1. Indledning

Denne artikel præsenterer en undersøgelse af i hvilket omfang man kan observere nogle fælles semantiske træk ved ord der har identiske eller stærkt lignende syntaktiske konstruktionsmønstre. Formålet er at undersøge om man kan berige ords beskrivelser i en leksikalsk resurse med betydningsrelaterede oplysninger ved at udnytte de eksisterende syntaktiske oplysninger. Metoden sigter på at finde frem til grupper af ord med fælles betydningsindhold for derefter at kunne beskrive dem gruppevis vha. leksikalsk-semantiske skabeloner.

Udgangspunktet for denne tilgang er en stor dansk leksikalsk resurse for sprogteknologiske anvendelser, den SprogTeknologiske Ordbase (herefter STO), der indeholder udførlige, formaliserede morfologiske oplysninger om mere end 81.000 lemmaer og syntaktiske beskrivelser af mere end 45.000 ord (jf. Braasch & Olsen, 2004). En mindre del af lemmaerne har også leksikalsk-semantisk beskrivelse, i alt ca. 10.000 betydninger – heraf ca. 1.000 verbbetydninger.

Semantikdelen er oprindeligt udarbejdet i SIMPLE-projektet (jf. Pedersen & Paggio, 2006) og er baseret på SIMPLE-ontologien, der er et flerdimensionelt semantisk beskrivelsessystem med 135 ontologiske overbegreber. Det er disse

overbegreber som de enkelte ords betydninger er relateret til. Grunden til at væsentlig færre opslagsord er forsynet med semantiske beskrivelser end det er tilfældet mht. morfologi og syntaks, er, at arbejdet med semantikken var særdeles tidskrævende og dermed omkostningstungt, derfor var der i første omgang fokus på en central delmængde af ordforrådet.

Ordbasen er i anvendelse i flere projekter, og brugerne har fortalt os at de meget gerne vil have semantiske oplysninger sammen med den omfattende syntaks. Da vi så præsenterede dem for den meget komplekse SIMPLE-baserede semantik, syntes de at selv om den er rigtig interessant fra en forskningsmæssig synsvinkel, virker den meget omfattende i forhold til mulige sprogteknologiske anvendelser. Der blev udtrykt behov for noget "mere enkelt", fx grundlæggende, formaliserede semantiske beskrivelser af verber/verbgrupper, der skulle være anvendelig i praktiske applikationer. I sidste ende stiller disse anvendelser nemlig nogle særlige krav til hvilke betydningsrelaterede oplysninger der skal beskrives – og hvordan.

Det blev den direkte anledning til at se på hvilke muligheder der er for at udbygge STO med oplysninger om ordenes betydning(er). I denne forbindelse overvejes også hvilken type semantik der bør vælges, hvilken ordklasse der var mest relevant at starte med, og hvordan man kan etablere forbindelse mellem STOs syntaktiske beskrivelser og den valgte type semantiske beskrivelse af ord.

Formålet med den her præsenterede undersøgelse er derfor at afprøve hvilke muligheder der er for at tilføje STOs ordforråd formaliserede betydningsbeskrivelser på en praktisk håndterlig og økonomisk måde. Antagelsen var at der med udgangspunkt i de enkelte ords syntaktiske beskrivelser, kan afgrænses grupper af ord der udviser (næsten) identiske egenskaber mht. deres konstruktionsmønstre, og at disse ordgrupper vil have en række fælles betydningskomponenter. Indsigterne mht. arten og graden af sådanne fællesskaber skal siden danne grundlag for arbejdet hen imod en leksikalsk resurse af en betydelig størrelse med integrerede morfologiske, syntaktiske og semantiske beskrivelseslag.

## 2. Baggrund og inspiration

Emnet er ikke nyt i leksikografiske sammenhænge, men forskningen tager som regel udgangspunkt i ords betydningsrelationer (fx hyponymi, meronymi, troponymi) og undersøger derefter ordenes syntaktiske konstruktionspotentiale med henblik på at beskrive eventuelle sammenhænge. Der findes mange store undersøgelser af relationerne mellem ords semantiske og syntaktiske egenskaber, ikke blot ud fra den leksikologiske/leksikografiske (Levin, 1993), men også andre, fx den datalingvistiske (Fillmore et al., 2003) eller psykolingvistiske synsvinkel (Fellbaum, 1998). Den generelle, fælles erkendelse er at der er en naturlig sammenhæng mellem ords meningsstruktur og dennes syntaktiske udtryk. Fellbaum skriver eksempelvis at "verbs whose LCSs (viz. lexical conceptual structures) are

identical in terms of specific meaning components tend to share syntactic behaviour” (Fellbaum, 1998:95).

Der er forskellige måder at beskrive disse relationer på, i traditionelle almene ordbøger sker det grundlæggende ved en betydningsbaseret inddeling af ordbogsartikler, ved hver betydning af et ord anføres også det relevante konstruktionsmønster og et eksempel. Anvendelsen af databaseteknologien har imidlertid banet vejen for store systematiske undersøgelser af ordforrådet ud fra forskellige synsvinkler, og enkelte af disse projekter er kommet frem til grundlæggende nye måder for sprogbeskrivelse.

De projekter der inspirerede min undersøgelse, er forskellige mht. deres formål, teoretiske grundlag, praktiske tilgang og beskrivelsesmetode. Disse inspirationskilder er vigtige netop fordi de forholder sig forskelligt til sammenhængen mellem semantik og syntaks (eller omvendt!), nedenfor gives et par eksempler på hvordan de behandler forholdet mellem semantik og syntaks.

Levins semantiske kategorisering og syntaktiske gruppering af engelske verber (*English Verb Classes and Alternations*, 1993) “...examines a large number of semantically based verb classes and shows how syntactic patterns systematically accompany the semantic classification” (Fellbaum op.cit.: 95). Dette værk har ikke blot dannet skole med hensyn til registreringen af systematiske relationer mellem semantik og syntaks, men også udfordret andre forskere til at gå videre med arbejdet og udbygge, forfine systemet (fx Kipper et al. 2006). Som Fellbaum (op.cit.:11) konkluderede “Levin has amassed impressive evidence showing that the semantic makeup of verbs is intimately linked with their syntactic behavior”.

De semantiske net, Princeton WordNet, (WN, jf. Fellbaum, 1998) og senere EuroWordNet (EWN, jf. Vossen, 1998) har fået ganske stor udbredelse i Europa idet der er lavet WordNets for en lang række europæiske sprog foruden engelsk (startende med udviklingen af Princeton WordNet), spansk, italiensk, tysk, ungarsk, tjekkisk osv., men også uden for Europa og USA udarbejdes der WN’s. De fleste semantiske net af WN-typen har kun meget få – hvis overhovedet nogen – syntaktiske oplysninger med. Dermed mangler de en vigtig parameter.

Det forholder sig noget anderledes med FrameNet (Fillmore et al., 2003), et semantisk net for engelsk, hvis grundlæggende ide omfatter en afbildning af ords indholdsside (semantiske argumentstruktur) og udtryksside (deres syntaktiske komplement) i de såkaldte “semantic frames”. Det er en konceptuel struktur der beskriver fx kommunikation og specificerer hvilke semantiske roller der er relevante i forhold til kommunikationen som handling, og hvordan de realiseres i tekster.

Det er derfor en oplagt ide at berige STO’s syntaksmateriale med semantiske beskrivelser og inddrage de erkendelser der er kommet ud af dette pilotprojekt mht. de relationer der kan betegnes som “syntaktisk lighed – semantisk nærhed”.

Sidst, men ikke mindst bør det nævnes det danske DanNet-projekt (se artik-

len af Bolette S. Pedersen, Sanni Nimb og Lars Trap-Jensen herom i dette bind). Projektet er i gang med at opbygge et semantisk net for dansk efter WordNet-modellen, under inddragelse af Den Danske Ordbogs materiale og SIMPLE/STO's formelle beskrivelser. På Center for Sprogteknologi, Københavns Universitet er der i 2007 nedsat en arbejdsgruppe til at undersøge mulighederne for at sammenkoble DanNets semantiske repræsentationer med STOs syntaktiske beskrivelser. En sådan leksikalsk resurse vil kunne imødekomme sprogteknologiens behov for en ordbase der dækker både syntaks og semantik (samt morfologi). Resultaterne af denne pilotundersøgelse vil så vidt muligt blive inddraget i arbejdsgruppens overvejelser og anbefalinger der vil ligge klar sidst på efteråret i 2007.

### 3. Arbejdsgrundlaget

Denne undersøgelse fokuserer på verber af følgende grunde. For det første er verbers syntaktiske konstruktionsmønstre ganske komplekse: De er en kombination af mange lingvistiske egenskaber, og dette forhold afspejles godt af STOs finmaskede syntaktiske beskrivelser. Derfor er denne ordklasse særlig interessant. For det andet er der et særligt forskningsmæssigt perspektiv i at arbejde med en formaliseret beskrivelse af verbers semantik set i sammenhæng med deres syntaks, da det er et aktuelt emne i forskellige projekter der vedrører andre sprog end dansk (jf. eksempelvis Kipper et al.)

#### 3.1 Den syntaktiske beskrivelse i STO

VALENSTYPE (aritet)	Antal KONSTRUKTIONSMØNSTRE (ledtype og -funktion, kontrol, styret præposition...)	Antal SYNTAKTISKE ENHEDER (verballæsninger)	Antal SYNT. ENHED pr. MØNSTER (gennemsnit)
Dv0 (zerovalent)	4	47	~11,8
Dv1 (monovalent)	95	1.164	~12,2
Dv2 (divalent)	479	5.722	~12,5
Dv3 (trivalent)	304	1.573	~5,17
Dv4 (tetravalent)	6	52	~8,66
<b>Total</b>	<b>888</b>	<b>8.558</b>	<b>~9,65</b>

Table 1. Oversigt over ca. 5.770 verbers syntaktiske beskrivelse og fordeling mht. deres konstruktionspotentiale

Arbejdsgrundlaget er de formaliserede beskrivelser af verbers syntaktiske opførelse i STO (se fx Braasch, 2006). STO er den største danske leksikalske database udarbejdet for sprogteknologiske anvendelser – med detaljerede syntaktiske beskrivelser for en relevant mængde af det danske ordforråd. De syntaktiske



beskrivelser er valensbaserede (tager udgangspunkt i ords syntaktiske komplementstruktur) med en del lingvistisk forskning som grundlag. Den syntaktiske komplementstruktur er sproglig realisering af ordets semantiske argumentstruktur, derfor er der en nær sammenhæng mellem et ords betydning (leksikalske semantik) og dets prototypiske syntaktiske omgivelser (syntaktiske konstruktionspotentiale/valensmønstre).

Tabel 1 giver en oversigt over verbernes syntaktiske beskrivelse i STO mht. antallet af konstruktionstyper og verbernes syntaktiske læsninger, fordelt på de fem valenstyper (ariteter). Det fremgår af oversigten at de 5.770 verballemmer har i alt 8.558 syntaktiske beskrivelser, hvilket afspejler at et verbum naturligvis godt kan have mere end et konstruktionsmønster. Der er defineret i alt 888 forskellige konstruktionsmønstre for verber ud fra STOs valensbaserede formelle beskrivelsessystem. Dette ganske omfattende materiale lægger op til den valgte metode, nemlig at nærme sig verbernes semantiske indhold fra deres syntaktiske beskrivelse ved at undersøge verber med identiske eller stærkt lignende syntaktiske egenskaber ud fra tre grundspørgsmål:

- Har de betydningsmæssige ligheder (“semantiske fællesnævner”), hvis ja, hvilke?
- Kan man ud fra de fundne ligheder opstille nogle grundlæggende betydningsgrupper?
- Kan man tildele de enkelte grupper nogle generaliserende semantiske etiketter?

### 3.2 Empirisk materiale

De syntaktiske oplysninger om de enkelte ord er baseret på ordets korpusforekomster og afspejler de almindelige syntaktiske forbindelser som ordet indgår i, dvs. deres autentiske brug. Naturligvis belyser ordets kontekst i korpusforekomsterne også dets semantiske egenskaber, såsom argumentstruktur, selektionsrestriktioner og betydning. Ved den korpusbaserede udarbejdelse af de syntaktiske beskrivelser er der også taget et vist hensyn til de semantiske læsninger af ordet.

### 3.3 Valensbaseret tilgang til syntaks

Den lingvistiske tilgang i STO bygger på valensteorien, dvs. ords *syntaktiske konstruktionspotentiale* er beskrevet med fokus på kategoriel og funktionel valens. Ords *semantiske argumentstruktur* bestemmer dets syntaktiske valens, idet komplementerne er argumenternes realisering i ytringer. I den valensbaserede tilgang beskrives syntaksen med udgangspunkt i ordets komplementstruktur og semantikken ud fra dets argumentstruktur (Grimshaw 1990, Levin 1993). Argumentstrukturen realiseres som ordets konstruktionspotentiale og er dermed bindeled mellem dets betydning og syntaks. Eksempelvis beskriver de fleste zerovalente verber vejrfænomener (*regne, tordne, sne...*); forholdet mellem syntaks og seman-

tik er tydeligt om end ikke entydigt, jf. undersøgelsen der refereres i Braasch (2006). Det forholder sig anderledes i tilfælde af verber med mere kompleks syntaks, fx konstruktioner med flere styrede led. Her er der naturligvis mange flere kombinationsmuligheder mellem syntaks og semantik. Det gælder fx for trivalente verber (fx *beskyldte/rose/straffe* eller *undersøge/behandle/helbrede...* nogen for noget); disse tager subjekt, objekt og præpositionalobjekt. Det sidstnævnte kan have forskellig syntaktisk form, dvs. et nominalsyntaxme, infinitiv eller en (in)fininit at-sætning. De ca. 1.570 trivalente konstruktioner i STO er beskrevet med 304 forskellige konstruktionsmønstre, disse kan samles i større grupper på grundlag af de enkelte komplementers syntaktiske egenskaber.

Den syntaktiske beskrivelse omfatter antallet og arten af styrede (valensbundne) led, og for hvert styret led angives følgende: *position* i den kanoniske rækkefølge, syntaktisk *kategori*, fx NP, PP (styret præposition og styrelse), at-sætning osv., syntaktisk *funktion*, fx subjekt, direkte, indirekte eller præpositionsobjekt, objektprædikat, infinitivkomplement, osv., og hvorvidt leddet er obligatorisk i konstruktionen eller ej. Andre registrerede egenskaber er refleksivitet, partikelverbum, subjekt/objektkontrol, mv. Hver unik kombination af disse egenskaber udgør et konstruktionsmønster. Egenskaberne formaliseres i STO i såkaldte *attribut/værdi-par* (fx kategori = NP), og deres unikke kombination kaldes *syntaks-mønster*. En sådan formaliseret beskrivelse er ofte en generalisering over flere af verbets konstruktioner, hvor det har den samme grundbetydning.

I eksemplerne nedenfor bruges følgende notation:

[ ]: omslutter et obligatorisk komplement

( ): omslutter et optionelt komplement

{ } : omslutter mulige realisationer af et obligatorisk eller optionelt komplement.

Et verbum kan naturligvis optræde med mere end et mønster, fx i tilfælde af polysemi som i

(1a) *anholde* [+ ngn] + (*for* ngt) i betydningen 'arrest'

(1b) *anholde* [*om* ngt ] i betydningen 'apply'

hvor (1a) er en trivalent konstruktion (subjektet er altid til stede, det er derfor underforstået), mens (1b) er divalent.

Et væsentligt spørgsmål i denne sammenhæng er, hvad der egentlig regnes som styret (valensbundet) led. I gængse fremstillinger (fx i tyske valensgrammatikker) anskues valensbegrebet binært, dvs. et led i en given konstruktion er enten valensbundet eller ej, hvorimod Somers (1987:27 & 266) beskriver det ud fra sin datalingvistiske synsvinkel som et kontinuum med seks forskellige grader af "bundethed". I Somers' kontinuum udgør de såkaldte "middles" en kategori der ligger mellem valensbundne led (også kaldet komplement) og de frie led (også kaldet adjunkter). Denne kategori udviser nogle særlige egenskaber, eksempelvis kan det ofte være svært at tildele et sådant led en af de traditionelle semantiske

roller (se fx Trask, 1993: 249), og det udtrykkes ofte med et præpositionssyntagma. Semantisk set er “middles” bestemt af det valensbærende ord (her af verberne *bo* og *gå*), hvorimod deres syntaktiske form eller kategori kan variere, ligesom de kan være obligatoriske eller optionelle, som det fremgår af eksemplerne (2a-2b) nedenfor.

- (2a) Han *bor* her/på gården/ikke ret langt herfra  
 – det semantiske argument angiver *stedet* (lokativisk), det syntaktiske komplement kan udtrykkes med et præpositionssyntagma eller et adverbial og er obligatorisk.
- (2b) Han *gik* ud på pladsen/ned til bækken/hen til landsbyen  
 – det semantiske argument angiver retningen (direktionelt), det syntaktiske komplement kan udtrykkes med et præpositionssyntagma eller et adverbial og er optionelt.

Den syntaktiske beskrivelse af verber i STO inkluderer som regel også “middles”, og det er i mange tilfælde en rigtig god hjælp i sprogteknologiske anvendelser – og i denne undersøgelse selv om kodningen desværre ikke er helt konsekvent gennemført.

#### 4. Undersøgelsen

Hvert verbum (samt substantiv og adjektiv) har således mindst et konstruktionsmønster, og hvert konstruktionsmønster er repræsenteret ved et antal verber. Der er mønstre som beskriver mange verbers konstruktionspotentiale som i tilfælde af almindelige transitive verber (*klippe*, *spise*), og der er andre der kun gælder for ganske få, fx de zerovalente vejrligsverber som *sne* og *hagle*. Kombinationen af lemma og et af dets syntaksmønstre udgør en såkaldt *syntaktisk enhed* i STO (disse kaldes også syntaktiske læsninger). Det er sådanne syntaktiske enheder af verber med fælles syntaksmønster der udtrækkes fra databasen, og materialet undersøges mht. relationen mellem deres syntaktiske og semantiske egenskaber. Verbet “tale” har eksempelvis otte syntaktiske enheder, så forskellige som bl.a. i (3a-c):

- (3a) Vi taler (*med dem*) (*om* {alt muligt/*at* komme på besøg})  
 med subjekt og to optionelle præpositionsobjekter  
 (STO-kode: *Dv3P0Pnw0-med-om*)
- (3b) Han taler [{*stille* og *dæmpet*}]  
 med subjekt og obligatorisk adverbiumssyntagma  
 (STO-kode: *Dv2ADVP*)
- (3c) S-SF-politikere talte [*mod* {KV-regeringen}]  
 med subjekt og et obligatorisk præpositionsobjekt  
 (STO-kode: *Dv2Pnti-mod*)

Syntaksmønstrer i (3a) beskriver 20 verber i STO, hvoraf ti er entydigt talehandlingsverber, fx *konferere*, *passiare*, *småsnakke*, *pladre*, og de resterende ti beskriver handlinger der inkluderer en form for kommunikation, ofte ad mundtlig vej, fx *forhandle*, *vadde*, *skændes*, *enes*, *toppes*, *strides*. Her er det altså ganske tydeligt at disse verber også semantisk er stærkt beslægtede, selv om ikke alle er (nær)synonymer. Det er denne type umiddelbare iagttagelser der er efterprøvet i undersøgelsen. I denne sammenhæng er det interessant at bemærke at Fellbaum (2002:8) diskuterer et lignende forhold mellem to beslægtede verbalgrupper. I hendes beskrivelse vil den første gruppe udgøre forskellige realisationer af verbet *tale* (det overordnede verbum) mht. måden ("manner") handlingen udføres på, den defineres derfor som *genuine troponymy*. I den anden gruppe har verberne derimod en klar semantisk komponent af kommunikationens funktion ("function"), og dermed står disse verber i en anden relation til *tale*, det kaldes *function relation*.

Der er desuden mange tilfælde hvor flere verber deler mere end ét konstruktionsmønster, men dette forhold bliver ikke behandlet her. Levins klassifikation af engelske verber tager sit udgangspunkt netop i dette og giver en meget omfattende redegørelse over alternationerne for næsten 4.200 verber.

#### 4.1 Valg af materiale til undersøgelse

Udvælgelsesprocessen omfatter tre trin. Det indledende trin er at udtrække alle syntaktiske konstruktionstyper for verber for at skaffe et overblik over det samlede potentielle materiale. Den talmæssige fordeling af syntaktiske enheder er som vist i tabel 1 (ovenfor): der er fem valenstyper defineret ud fra ariteten, og for hver valenstype er der registreret et antal syntaktiske konstruktionsmønstre der hver især beskriver en unik kombination af konstruktionsegenskaber.

Næste trin var at udvælge en delmængde af en valenstypes konstruktionsmønstre som den egentlige genstand for undersøgelsen. Der er udvalgt en delmængde (otte) af de (i alt 304) trivalente konstruktionsmønstre der beskriver verber med følgende styrede led: subjekt, direkte objekt og præpositionsobjekt (POBJ). Verber med denne konstruktionstype kaldes "ditransitive verbs" (Grimshaw, 1990:28) eller "double object constructions" (Levin, 1993:274). Subjektet og objektet er i de valgte konstruktioner et nominalsyntaxme (NP). Det er præpositionsobjektet der er særlig interessant ud fra forskellige aspekter. Valget af præpositionen er i mange tilfælde (men ikke altid?) bestemt af verbets semantik (i det præsenterede materiale er præpositionen *til* eller *for*). Styrelsen for præpositionen kan have forskellige syntaktiske egenskaber mht. kategori: i nogle konstruktioner kan det kun være nominalsyntaxme, i andre alternerer nominalsyntaxmet med en at-infinitiv (sætning); og sidstnævnte kan optræde med eller uden objektskontrol. Den sidste egenskab vedrører leddets tilstedeværelse i konstruktionen: det kan være optionelt eller obligatorisk. Oplysningen om hvorvidt leddet er obligatorisk eller ej har dog vist sig ikke at være nogen særlig indikator med hensyn til semantiske forskelle i forhold til de her undersøgte verber. I andre tilfælde kan det forholde

sig ganske anderledes, fx "manden taler sandt" er forskellig fra "manden taler". Det kan ofte dreje sig om en stærk valensbinding med strenge selektionsrestriktioner, og konstruktionen nærmer sig til kollokationsstatus. (Et sådant stærkt styret led betegnes hos Somers (1987:27) som "integral complement", denne type konstruktioner falder uden for undersøgelsen.)

Derefter fulgte udtræk af alle verber (216 stk.) med de valgte konstruktioner sorteret i otte lister, svarende til de otte konstruktionsmønstre. Tabel 2 viser en oversigt over fordelingen af verbers syntaktiske enheder på de enkelte konstruktionsmønstre. Disse er nært beslægtede selv om præpositionsobjektets egenskaber varierer noget: det kan være obligatorisk eller optionelt i konstruktionen, den styrede præposition kan være *til* eller *for*, og styrelsen kan være et NP, eller en at-infinitiv eller at-sætning med objektskontrol. Den talmæssige fordeling viser at der er en overvægt af syntaktiske enheder med optionelt præpositionsobjekt med NP uanset præpositionen.

KONSTRUKTIONSMØNSTER (STO-kode)	EKSEMPEL	ANTAL
med POBJ: til + NP, optionelt ( <i>Dv3NP0-til</i> )	genopstille ngn (til posten)	56
med POBJ: til + NP, obligatorisk ( <i>Dv3NP-til</i> )	udråbe ngn [til vinder]	31
med POBJ: til + NP/at-sætn. med objektskontrol, optionelt ( <i>Dv3NPnio0-til</i> )	besnakke ngn (til {et lån/at låne...})	20
med POBJ: til + NP/at-sætn. med objektskontrol, obligatorisk ( <i>Dv3NPnio-til</i> )	beordre ngn [til{march/at marchere}]	30
med POBJ: for + NP, optionelt ( <i>Dv3NP0-for</i> )	snynde ngn (for penge)	35
med POBJ: for + NP, obligatorisk ( <i>Dv3NP-for</i> )	tiltale ngn [for drab]	21
med POBJ: for + NP/at-sætn. med objektskontrol, optionelt ( <i>Dv3NPnio0-for</i> )	kritisere ngn (for {sjuskethed/ for at være sjusket})	16
med POBJ: for + NP/at-sætn. med objektskontrol, obligatorisk ( <i>Dv3NPnio-for</i> )	anse ngn [for {en god læge/at være en god læge}]	7
<b>TOTAL</b>		<b>216</b>

Tabel 2. Oversigt over de undersøgte trivalente syntaktiske enheder fordelt på konstruktionsmønstre

#### 4.2 Materialets behandling

Herefter fulgte den egentlige behandling af de enkelte lister, den blev udført i tre hovedtrin; disse skitseres nedenfor.

##### 4.2.1 Tildeling af semantisk etikette og gruppering af verber

Det første trin var at gennemse alle lister og tildele en foreløbig *semantisk etikette* til hvert enkelt verbum. Etiketterne er valgt intuitivt, dvs. ikke ud fra et på forhånd defineret system, men ud fra verbets hovedbetydning i den konkrete eksempelsætning med den konstruktion, som syntaksmønstret beskriver. De fungerer som en slags semantiske fællesnævner med udgangspunkt i de åbenbare betydningsmæssige ligheder mellem listens verber. Denne metode er i god overensstemmelse med andres erfaringer, som Fellbaum skriver (1998:70): “Dividing the verb lexicon into semantic domains initially on a purely intuitive basis (min fremhævelse, A.B.) might lead one to discover relations that organize verbs and concepts.”

De tildelte etiketter er overskrifter på engelsk (fx SELECT, ESTIMATE, CHANGE\_of <FEATURE>), og der er to grunde til dette valg. For det første er det for at undgå at skulle vælge et dansk overbegreb til verbet allerede fra starten, hvilket kunne fastlåse analysen til et bestemt begrebshierarki, da et sådant hierarki er mere velegnet til substantiver end til beskrivelse af verber. For det andet er formålet at forberede hovedtrin 3, sammenligning af de semantiske grupperinger med eksisterende beskrivelser for engelske verber (Levins og EuroWordNets klassifikationer), som kunne vælges senere som inspirationskilde og arbejdsgrundlag for den semantiske beskrivelse af danske verber. Verberne blev derefter sorteret i foreløbige semantiske grupper (lister) på grundlag af deres fælles semantiske etikette.

SEMANTISK GRUPPE	VERBUM (eksempler)	POBJ (præp + NP)	POBJ (præp + at-inf.)	ANTAL (tentativt)
FORCE	tvinge ngn beordre ngn	til denne handel til march	til at rejse til at lukke alle dørene	12
URGE/ INQUIRE	overtale ngn formane ngn	til en kop kaffe til større tolerance	til at melde sig til at holde aftalen	9
ASK/ADVISE/ INSPIRE	råde ngn motivere ngn lokke ngn	til modstand til træning til det tomme hus	til at gøre modstand til at lære noget nyt til at holde vinterferie	8

ELECT	udtage ngn nominere ngn/ngt	til kampen til en Oscar	til at spille mod Polen	14
RELATE	henregne ngt forbinge ngn	til en ... kategori til konflikten	Ø	10
QUANTIFY/ ESTIMATE	estimere ngt tidsfæste ngt	til 3 måneder til foråret 2000	til at ligge over 3 m	7
QUALIFY	uddanne ngn	til et job	til at kunne bestride jobbet	10
PROSECUTE	straffe ngn	for indbruddet	for at have begået indbrud	7
ACCUSE/EXCUSE	beskyldte ngn frikende ngn	for illoyalitet for ansvaret	for at have krænket loven for at have optrådt forkert	5
COMMUNI- CATE_ JUDGE- MENT	klandre ngn rose ngn	for beslutningen for sangen	for at tage fejl for ikke at sylte sagen	8
.....	.....	.....	.....	.....
CHANGE_of POSS. 1	støvsuge ngt	for sand og snavs	Ø	5
CHANGE_of POSS. 2	narre ngn	for et stort beløb	Ø	7
CHANGE_of POSS. 3	donere ngt	til kampagnen	Ø	4
.....	.....	.....	.....	.....
CHANGE_of LOC.	udskibe ngt	til Amerika	Ø	6
CHANGE_of STATE	forarbejde ngt	til gelé	Ø	9
.....	.....	.....	.....	.....

Tabel 3. Eksempler på tildelte semantiske etiketter og verber med angivelse af tentativ fordeling (udsnit)

Tabel 3 viser et udsnit af de mulige semantiske grupperinger, og det illustrerer at samme verbum kan indplaceres under forskellige etiketter – alt efter hvilket semantisk element/træk eller konnotation der fokuseres på, eksempelvis verberne *råde* eller *lokke*. Den kontekst verbet optræder i bidrager naturligvis betydeligt til at præcisere betydningen, fx hvorvidt der er tale om tvang eller falske forudsætninger i forhold til handlingen. (Dette er et generelt problem som bliver taget op i afsnit 5.1.)

## 4.2.2 Behandling af verbalgrupperne

Det andet trin var behandlingen af de enkelte semantiske grupper. Det viste sig nemlig at tildelingen af semantisk etikette ikke var helt konsistent, hvilket afspejlede fx at de enkelte verbers konnotationer i første omgang blev vægtet forskelligt, men når verberne var sorteret i grupper, blev denne inkonsistens mere tydelig. Der blev udført justeringer mht. tildelingen, og enkelte etiketter viste sig at være for snævre eller for brede med hensyn til deres betydningsomfang, så de blev erstattet med mere dækkende overskrifter. De enkelte gruppers præpositionsobjekter blev derefter gennemset, og det viste sig at de havde en større eller mindre grad af semantisk lighed. Deres prototypiske semantiske type er registreret i et første forsøg, til dette er foreløbig brugt nogle generelle overbegreber fx *genstand*, *person*, *beløb*, *ydelse*. Desuden er de semantiske roller noteret i form af etablerede betegnelser som fx PATIENT, BENEFICIARY, GOAL, MEASURE. (jf. eksempelvis hos Grimshaw, op.cit og Trask, op. cit). CAUSE hører egentlig ikke til de centrale roller, den betragtes som en perifer rolle, denne kan tildeles bl.a. komplementer af typen ”middles” (jf. diskussionen i Somers 1987:74 ff.). PATIENT-rolle er her delt op i to undertyper der refererer til forholdet mellem det første objekt i konstruktionen (*nogen* eller *noget*) og præpositionsobjektet (*for* + *nogen* eller *noget*) som i eksemplerne i (4a) og (4b):

- (4a) Postbuddet tømmer *bilen* (OBJ) *for pakker* (POBJ: PATIENT, LOCATED)
- (4b) Selskabstømmeren lænsede *virksomheden* (OBJ) *for aktiebeholdningen* (POBJ: PATIENT, POSSESSED).

GRUPPE m. SEMANTISK ETIKETTE	EKSEMPEL PÅ VERBER	EKSEMPEL PÅ POBJ: for + NP	POBJ 's SEM. TYPE	POBJ's SEM. ROLLE
CH-of Location (REMOVE)	tømme, rydde tappe, lænse_1	{pakker, passagerer}, {saft, vand}	genstand, person	PATIENT (LOCATED)
CH_of Possession (CLEAR/ REMOVE)	rippe, lænse_2	{spillere, aktie- beholdningen}	genstand (værdi)	PATIENT (POSSESSED)
CH_of Possession (CHEAT/ FOOL)	afpresse narre, blokke snyde	{penge} {et stort beløb} {en gave}	værdi (beløb) værdi (ydelse)	PATIENT (POSSESSED)
CH_of Possession (OFFER)	skænke, ofre	{gæsterne} {familien, sagen}	person	BENEFICIARY
CH_of Possession (SELL)	sælge videresælge	{sin bror} (=fri dativ) {500 kroner}	person beløb	BENEFICIARY MEASURE



CH_of Posses- sion (DEBET/ CHARGE)	debitere, bon'e	{reparationerne}  {400 kroner}	værdi (ydelse, vare) værdi (beløb)	CAUSE MEASURE
--	--------------------	--------------------------------------	--	------------------

Tabel 4. Trivalente verber med POBJ (præposition 'for' + NP) og deres semantiske egenskaber (STO-kode: *Dv3NP-for*) (udsnit)

Tabel 4 viser et lille udsnit af resultaterne for de undersøgte verber med den overordnede semantiske etikette *CHANGE\_of\_ <FEATURE>* efter at dette trin var afsluttet.

#### 4.2.3 Sammenligning af grupperingen med beskrivelse i to etablerede systemer

I det tredje trin udførtes sammenligninger af materialet med Levins verbalklasser og med EuroWordNets metode for tildeling af de såkaldte *situationskomponenter* og indordning i dets topontologi (EWN, 2002:61) på et overordnet plan. Dette betyder at undersøgelsen ikke går i dybden på dette punkt, fordi formålet er at få et første overblik over om de her forslåede syntaksbaserede semantiske grupperinger er kompatible med beskrivelsessystemer der har en anden tilgang.

Sammenligningen med Levins verbalklasser er udført på den måde at de enkelte semantiske grupper er identificeret i forhold til Levins klassifikation, og de relevante kategorier derfra er noteret for den pågældende gruppe. Det er indlysende at der er betydelige forskelle mellem verbernes syntaks på dansk og engelsk, og derfor er klassifikationen af syntaktiske alternationer også forskellig for de to sprog. Desuden medtager Levin som regel ikke præpositionsobjekter, især ikke med infinitivsætning. Derfor kan det være svært at indplacere enkelte verber med sådanne valensled. Dette problem skyldes formentlig forskellen mellem den binære (hos Levin) og den graduerede anskuelse af valensbundethed (i STO/hos Somers, op.cit.). På den anden side er der et ganske stort sammenfald med hensyn til den semantiske gruppering af verberne, og det er dette forhold undersøgelsen fokuserer på.

EuroWordNets system opererer med andre parametre, beskrivelsen omfatter tre væsentlige oplysningsniveauer, for verber er det "EntityType" (som altid er *2nd Order Entity*), "SituationType" som kan være *Dynamic* (herunder *BoundedEvent*: verber med tidsmæssig afgrænsning og et forventet resultat; jf. eksemplerne i tabel 5) eller *Static*, samt en – eller oftest en kombination af – flere semantiske egenskaber der kaldes "SituationComponents", fx *Communication*, *Location*, *Manner*, *Purpose*, *Quantity* og *Social*. Tildelingen af disse beskrivelseskomponenter er vanskelig, men også meget vigtig i sammenligningen. Vanskeligheden består i at disse semantiske etiketter efter min begrænsede erfaring dels ikke kan tildeles konsistent, dels at der i EWN i princippet kan tildeles et stort antal sådanne etiketter til samme semantiske verballæsning (fx *blow\_2*: Bounded Event+Agentive+Physical+Location+Purpose+Manner), og de kan optræde i

en hvilken som helst kombination, idet deres antal og rækkefølge ikke er fastlagt på forhånd (EWN, 2002:65 ff). Dette resulterer i overlapninger og krydsninger i klassifikationen.

Tabel 5 giver et lille indblik i at det godt kan lade sig gøre at indplacere de semantiske grupper i EWN på et ikke for specifikt niveau, samtidig med at de enkelte verbalgrupper også har deres pendant i Levins' klassifikationssystem.

SEMANTISK GRUPPE	LEVIN'S VERB CLASS	EWN-PLACERING BoundedEvent+Agentive+ <SituationComponent>
CH-of Location (REMOVE)	10: Verbs of Removing	+Physical+Location+Possession
CH_of Possession 1 (CLEAR/REMOVE)	10.1+3 Remove, Clear	+Physical+Location+Possession
CH_of Possession 2 (CHEAT/FOOL)	10.5+6 Steal, Cheat Poss. Deprivation	+Possession+Manner
CH_of Possession 3 (OFFER)	13.3 Future having (Ch_ Poss)	+Possession+Social
CH_of Possession 4 (SELL)	13.1 Give (Ch_Poss)	+Purpose+Possession
CH_of Possession 5 (DEBET/CHARGE)	54.5 Bill (Measure)	+Purpose+Possession+Social

Tabel 5. Eksempel på resultater af sammenligningen: Semantiske grupperinger, Levin's klassifikation og EWN's indplacering af udvalgte trivalente verber (*udsnit*)

Resultatet af denne sammenligning skulle give et fingerpeg om hvorvidt den her afprøvede metode kan bidrage til en systematisk og effektiv kodning af semantiske egenskaber i forhold til eksisterende syntaktiske beskrivelse af verber. En anden mulighed der er under overvejelse i skrivende stund, er en semiautomatisk sammenføring af to resurser for dansk: den færdige STO databases indhold og DanNets semantiske beskrivelser. Erfaringerne fra denne undersøgelse vil kunne bidrage dertil fx med en metode for selektionen af syntaktiske enheder. I denne sammenhæng vil en sammenligning mellem de her beskrevne semantiske grupper og DanNets kodninger af verber være en yderligere forudsætning.

## 5. Relevante iagttagelser vedrørende udnyttelse af STO-materialet

### 5.1 Generelle forhold

Det er indlysende at de divalente verber udgør den allerstørste gruppe (med 479 forskellige konstruktionsmønstre for 5.722 syntaktiske enheder), og at denne gruppe er meget kompleks mht. semantiske egenskaber. Det er også evident at forholdet mellem syntaks og semantik er mere entydig i tilfælde af zero- og te-

travalente verber (jf. Braasch 2006) fordi disse valensgrupper er mere homogene, dvs. de omfatter kun få konstruktionsmønstre (jf. tabel 1). En anden iagttagelse er at samme syntaktiske komplement (objekttype) kan være syntaktisk manifestation af mange forskellige semantiske argumenter, dvs. der er ikke nogen 1:1 relation på alle planer mellem argument- og komplementstruktur. Dette sætter en naturlig begrænsning for i hvor høj grad den beskrevne metode kan generaliseres.

Det graduerede syn på valensbundethed ved kodningen af "middles" sætter fokus på det problematiske i at tildele præpositionsobjekter semantiske standardroller, de tenderer til at have perifere og mindre entydige roller.

Det viste sig, som tidligere nævnt, i mange tilfælde at være svært at tildele en dækkende semantisk etikette til verber i en intuitivt set semantisk homogen verbalgruppe, fx verberne *indbyde*, *invitere*, *tilskynde*, *motivere*, *formane*, *lokke*, *råde* og lignende (jf. afsnit 4.2.1). Valget stod mellem etiketter som fx URGE, ASK, INQUIRE, ADVISE, INSPIRE, alt efter hvilken betydningskomponent der fokuseres på i kombinationen af denotation/konnotation, fx graden af intensitet i handlingen, eller formålet med handlingen. Et skift i dette fortolkningsfokus kan også medføre en ændring af selve grupperingen. Det er meget mere kompliceret at udnytte hierarkiske og semantiske relationer (i form af troponymi) til at organisere verber i semantiske domæner end for substantiver (ved fx hyponymi/hyperonymi); de semantiske relationer mellem verbers betydninger er mere komplekse.

### 5.2 Specielle forhold

Den systematiske sammenstilling af verber der er kodet med ens eller stærkt lignende konstruktionsmønstre gav mulighed for et konsistent stykke af kodningerne. Gennemgangen af materialet afslørede enkelte problemer/mangler i forhold til verbernes syntaktiske beskrivelse i STO som kan sammenfattes i følgende punkter:

- (a) Manglende data mht. opslagsord, fx *skose* [nogen] (*for* {noget/at}) eller relevant syntaktisk enhed inden for den undersøgte valensgruppe, fx for verberne *elske* og *hade* [nogen] (*for* {noget/at}).
- (b) Inkonsekvens i kodning mht. hvilke sætningsled der er registreret som valensbundet (fx *anholde* er beskrevet som divalent med eksemplet "Politiet anholdt ham, da færgen...", hvorimod verbet i de fleste korpusforekomster er trivalent med optionelt præpositionsobjekt, der fremgår af konteksten: *anholde* nogen *for* noget/at + infinitiv)
- (c) Inkonsekvens mht. hvilket præpositionsobjekt der er markeret som optionelt, fx *presse* [nogen] [*for* noget] vs. *afpresse* [nogen] (*for* noget)).

Disse iagttagelser er blevet noteret, og rettelser kan udføres ved lejlighed, hvilket bliver en sidegevinst for STO. Der skal naturligvis også tages hånd om disse problemer inden eller under en (semi)automatisk udnyttelse af STO ved at rette inkonsekvenser og evt. tilføje af manglende data.

## 6. Sammenfatning og perspektivering

De hidtidige erfaringer viser at der, i hvert fald for en del af verbernes vedkommende, kan foretages en sortering og gruppering ud fra verbers valensstruktur, og denne gruppering kan danne grundlag for en systematiseret semantisk beskrivelse. Ifølge Fellbaum (1998:95) "An exploration of the syntactic properties of co-troponyms occasionally provide the basis for distinguishing semantic subgroups of troponyms", pilotundersøgelsens resultater falder godt i tråd med dette.

De her præsenterede semantiske grupper kan revideres/tilpasses til at kunne bruges i forhold til forskellige beskrivelsessystemer, fx et semantisk net for danske verber. Til semantisk kodning kan der opstilles – eller vælges fra eksisterende semantiske klassifikationer – leksikalsk-semantiske kodningsskemaer, profiler eller skabeloner passende for de enkelte grupper og undergrupper. En sådan skabelon skal indeholde (generaliserede, formaliserede) oplysninger om grundlæggende betydningsindhold (fx semantisk type i den valgte klassifikation) og prototypiske semantiske argumenter specificeret med hensyn til deres betydningsmæssige egenskaber (formuleret som selektionsrestriktioner, igen i forhold til en på forhånd fastlagt typologi). Der er forskellige muligheder for valget af semantisk beskrivelsestype og formalisme; det væsentlige er fra sprogteknologiens synsvinkel at den leksikalske ressource ikke på forhånd må være dedikeret til en bestemt applikation, men må kunne anvendes bredt i forskellige sammenhænge, fx semantisk opmærkning af tekster, i systemer til avanceret/indholdsbaseeret informationssøgning, tekstresumering osv. Endvidere kan STO-materialet udnyttes i lingvistisk forskning vedrørende prædiktion om forholdet mellem syntaks og semantik, eksempelvis ved at undersøge verber der har flere syntaktisk identiske konstruktioner for at få overblik over alternationstyper i dansk og sammenligne dem med Levins klassifikation.

## Referencer

- Braasch, Anna & Olsen, Sussi (2004): STO: A Danish Lexicon Resource – Ready for Applications. In: *Fourth International Conference on Language Resources and Evaluation, Proceedings, Vol. IV*. Lisbon, pp. 1079-1082.
- Braasch, Anna (2006): Exploitation of Syntactic Patterns for Sense Group Identification. In: *Proceedings, XII. EURALEX International Congress*, pp.133-140. Edizioni dell'Orso, Alessandria.
- EWN= EuroWordNet: General Documentation, LE2-4003, LE-8328. Ed.: Piek Vossen.

- Elektronisk dokument ([www.vossen.info/docs/2002/EWNGeneral.pdf](http://www.vossen.info/docs/2002/EWNGeneral.pdf)).
- Fellbaum, Christiane (1998): A Semantic Network of English Verbs. In: *WordNet. An Electronic Lexical Database* (ed. Fellbaum). The MIT Press. Cambridge, Massachusetts, London.
- Fellbaum, Christiane (2002): On the Semantics of Troponymy. In: *The Semantics of Relationships: An Interdisciplinary Perspective*. R. Green, C. A. Bean, S. H. Myaeng (eds.) Kluwer. Dordrecht, Holland.
- Fillmore, Charles J., Christopher R. Johnson, Miriam R. L. Petruck (2003): Background to FrameNet. In: *International Journal of Lexicography*, Vol. 16 No.3. OUP, Oxford.
- Grimshaw, Jane (1990): *Argument Structure*. The MIT Press. Cambridge, Massachusetts, London.
- Kipper, Karin, A. Korhonen, N. Ryant, M. Palmer (2006): Extending VerbNet with Novel Verb Classes. Elektronisk publikation ([www.cl.cam.ac.uk/~alk23/lrec06.pdf](http://www.cl.cam.ac.uk/~alk23/lrec06.pdf)).
- Levin, Beth (1993): *English Verb Classes and Alternations*. A preliminary Investigation. The University of Chicago Press. Chicago and London.
- Pedersen B., Paggio P. (2004): The Danish SIMPLE Lexicon and its Application in Content-based Querying. In: *Nordic Journal of Linguistics Vol.27:1*, s. 97-127. United Kingdom.
- Pedersen, B.S., S. Nimb, N., J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen (2006). DanNet - A WordNet for Danish. In: *Proceedings from Third International Conference on Global Wordnets*, pp. 329-331. Jeju, South Korea.
- Somers, Harold (1987): *Valency and Case in Computational Linguistics*. Edinburgh University Press. Edinburgh.
- Trask, R.L. (1993): *A Dictionary of Grammatical Terms in Linguistics*. London and New York.
- Vossen, Piek (1998): Introduction to EuroWordNet. In: *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*. (ed. Vossen). Kluwer Academic Publishers. Dordrecht/ Boston/London.

**Hjemmesider på internetet** (pr. 12. december 2008)

- DanNet: <http://wordnet.dk>  
 EuroWordNet: <http://www.illc.uva.nl/EuroWordNet/>  
 FrameNet: <http://framenet.icsi.berkeley.edu/>  
 SIMPLE: <http://www.ub.es/gilcub/SIMPLE/simple.html>  
 STO: [http://cst.ku.dk/sto\\_ordbase](http://cst.ku.dk/sto_ordbase)  
 Wordnet, Princeton: <http://wordnet.princeton.edu/>

Anna Braasch  
 seniorforsker, f. 1946  
 Center for Sprogteknologi  
 Københavns Universitet  
 Njalsgade 80  
 DK-2300 København S  
 e-mail: [braasch@hum.ku.dk](mailto:braasch@hum.ku.dk)



LARS BRINK

## Leksikalske huller

Everyone thinks lexical gaps are interesting. But they do not appear in our dictionaries. Except for those more trivial cases where the very phenomenon behind a concept does not exist in a population, e.g. *Højnasbjerget*, the name of a steep hill in my home town for which there is no established name in any other language, or *Kildeskattedirektoratet*, the tax authorities in the same area. But an English concept like *cousin* has no established Danish counterpart, and in spite of this being a well-known fact, it is well concealed in our dictionaries. – I analyse a list of the 1000 most frequent Danish words and find as many as 27 instances where Spanish has a lexical gap, i.e. lacks an equivalent. I try to determine whether the gaps are due to culture or accidental. But whichever the reason may be, I appeal to all lexicographers not to obscure the gaps, for not only do we have an obligation to show them in good dictionaries, they are also of high interest for the learned as well as the unlearned.

### 1. Præsentation af leksikalske huller og deres behandling i ordbøger

Både lægmand og lærd synes, at leksikalske huller er spændende. Der verserer endda folkloristiske historier om, at dét og dét sprog mangler et ord for dét og dét. Eller – nok så hyppigt – at *vores* sprog er det eneste i verden, der har ord for dét og dét. I Danmark siger mange, at vi er de eneste, der har ord for begrebet *hygge*. Og svenskerne plejer at sige, at ingen andre sprog har noget, der svarer til *lagom*.

Også lingvister finder det højinteressant. Og når man har fundet et hul, går tanken naturligvis straks hen på årsagerne til det: Skulle vi danskere have en særlig veludviklet sans for hygge?

Leksikografer synes sikkert også, det er interessant, men dét kan man ikke se på deres ordbøger. Dé er nemlig rensset for huller. Undtagen én slags: Dé tilfælde, hvor selve *fænomenet* ikke forekommer hos et folk. Dét folks sprog mangler naturligvis en ækvivalent. Hvad der ikke kendes, kan naturligvis heller ikke *benævnes*. Hvis rødgrød og kravenisser og store bededag slet ikke findes hos andre folkeslag, så har disse selvfølgelig heller ikke noget *etableret udtryk* for det. Men dét er ikke spor interessant. Det kulturelt interessante er *sagforholdet*, at danskerne har en unik helligdag. Det er helt trivielt, at vi også har et sprogligt etableret udtryk for den, og det er helt trivielt, at de andre sprog mangler et etableret udtryk for den. De gode ordbøger markerer klart, at fx *skattedirektorat*, *Himmelbjerget*, *stegt flask med persillesovs* mangler ækvivalent i de andre sprog, ligesom de giver forslag til, hvordan man kan komme tæt på, eller de definerer dem

slet og ret ved hjælp af en uetableret definition. Udmærket. Men lingvistisk er det uinteressant.

Lingvistisk interessant er det kun, hvis et sprog mangler et etableret udtryk for et fællesmenneskeligt fænomen. Altså fx det fællesmenneskelige fænomen "hygge".

Sådanne huller er der flere af, end jeg for få år siden troede. Men først skal vi se på nogle huller, som alle lingvister kender. Først en slægtskabsbetegnelse. Svarende til dansk *bedstefar* har svensk et leksikalsk hul. Om man så står på hovedet, kan man ikke finde et etableret udtryk i svensk med samme betydning. Man kan sagtens forklare det på svensk (svensk er jo et *komplet* sprog): *farfar* eller *morfar*. Men der er ingen ækvivalent. – Jeg gennemgik på konferencen i Torshavn 2003 en række ord, kendt af alle lingvister, men knap så kendte af leksikografer, der mangler ækvivalenter i andre sprog: Dansk *by* og *skov* er uspecificerede mht. størrelse og mangler ækvivalenter i flere andre sprog. Tysk og svensk er nødt til at specificere bystørrelsen, engelsk er nødt til at specificere skovstørrelsen, hvis man da skal holde sig til etablerede udtryk. Til gengæld mangler dansk overbegrebet for 'morgen' og 'formiddag' svarende til engelsk *morning*. Det interessante er her ikke disse gamle traverses eksistens, men at leksikograferne nægter at gøre rede for dem. Leksikograferne har berøringsangst for ordbogsartikler uden en ækvivalent. Det fører dels til groteske fejl, dels fører det til, at leksikografen hverken selv erkender eller lader sin læser erkende, at her har vi en højinteressant leksikalsk situation. Jeg viste, hvordan eng. *cousin* i hæderkronede ordbøger bliver oversat til dansk *fætter*. Det er jo grotesk: *I've just got a new cousin – Jeg har lige fået en ny fætter*. Ja, god morgen. Selv det for alle indlysende ækvivalent-manglende danske *De* med stort får ikke redaktørerne til at bemærke, at engelsk her mangler noget. Man gengiver det uden videre med *you*. Dvs. man giver en *standard-oversættelse* i stedet for at bemærke, at der ikke findes ækvivalent og så i øvrigt *beskrive* de signaler, der ligger i *De*. En standard-oversættelse er noget ganske andet end en ækvivalent. En standard-oversættelse er "dét udtryk, de målsprogstalende *altovervejende* bruger, når de er anbragt i præcis dé situationer, som de udgangssprogstalende er anbragt i, når de bruger ordet." Det er noget rent statistisk. Det er meget muligt, at vi danske oftest vil sige *fætter* dér, hvor englænderne bruger *cousin* – frekvensmålinger viser jo, at der er flere mænd end kvinder i vore tekster – men *fætter* er ingen ækvivalent. *Cousin* betyder ikke 'fætter'. Der er heller ingen tvivl om, at lige præcis på dét sted, hvor vi i Norden siger *kilometer*, dér siger englænderne *altovervejende mile*, men det har dog intet med ækvivalens at gøre. Og hvis en stakkels dansk kostskoledreng skulle tage mod til sig og tiltale sin brutale lærer med et "*De har ikke lov til at slå mig!*", så vil de danske ordbøger med deres trivielle standard-oversættelse føre til "*You are not allowed to hit me!*" til trods for, at *De*-signalet her skal gengives: *You are not allowed to hit me, Sir!*, men dét siger ordbøgerne ikke noget om. Det lyder, som om jeg er vred på de danske ordbøger. Ingen-



lunde. De er gennemgående fremragende. Men dét teleskop, der burde rettes mod – ikke sorte huller, men – leksikalske huller, har de, som en anden admiral, sat for det blinde øje.

## 2. Undersøgelse af dansk-spansk

Jeg vil nu gennemgå en række hidtil upåagtede leksikalske huller i spansk. De er fremkommet under mit arbejde med en dansk-spansk ordliste over de 1000 almindeligste danske ord. Jeg er ikke ekspert i spansk, ikke engang fagmand, så der kan godt være fejl i min beskrivelse. Men dém får jeg forhåbentlig rettet af mere kyndige. Derved vil det også blive belyst, at man skal kende to sprog næsten optimalt for at konstatere leksikalske huller.

Det første hul møder vi ved ordet *fræk*. Det hedder *procaz* på spansk. Carl Bratli har eksemplet *una criada procaz ~ en fræk tjenestepige*; det lyder ækvivokt, men det betyder med garanti 'uforskammet, impertinent, næsvis', sådan en, der kan finde på – hvad der i 50'erne var en dødssynd – *at svare igen*. Fra 1965 fik ordet en ny betydningsvariant i dansk: 'utilsløret seksuelt præget, således at det går noget over stregen for god borgerlig opførsel'. "*Fræk pige vil gerne forkæles af fræk fyr*", *en fræk historie, frække blade* osv. Dén variant – der kunne se ud til at være lånt fra svensk – er totalt fraværende fra den store dansk-spanske ordbog. Men selvom den havde været der, tror jeg, at vi ville have stået over for en manglende ækvivalens, et hul. Det specielt skandinaviske er nemlig, at ordet er positivt værdiladet. Dermed er der heller ingen engelsk ækvivalent. For eng. *naughty* og *dirty*, som angives af ordbøgerne er ikke entydigt positive; de kan både bruges med deres negative grundbetydning og med ironisk afstandtagen; *fræk*, derimod er entydigt positivt. Den victorianske mor, der finder *frække blade* på sin søns værelse, vil garanteret ikke bede ham om at *fjerne sine frække blade*. Hun må kalde dem *beskidte, modbydelige* eller endnu stærkere ting. Her har vi et leksikalsk hul i en variantbetydning af et ord.

Skulle dette uoversættelige skandinaviske *fræk* have noget med skandinavisk kultur at gøre? Selvfølgelig har det dét! Skandinavien, specielt Danmark og Sverige, var i 1965 Vestens førende antivictorianske område. Der er stadig mange amerikanere, der smiler skævt, når man siger *Danish*. Og det er ikke wienerbrød, de tænker på. Det er 100% naturligt, at vi i 1960'erne havde brug for et positivt ord, der kunne afløse *sjofel, lummer, slibrig, klam*.

Det næste hul er *frokost*. Det nærmeste spanske ord er *almuerzo*. Men dét er typisk et noget større og fremfor alt *varmt* måltid, der indtages kl. 14–15, altså flere timer efter den danske frokost. Her har vi naturligvis et hul betinget af virkeligheden. Der findes i spaniernes verden ingen frokoster, og derfor har de selvfølgelig heller ikke noget fast udtryk for det.

De fleste huller er af dén art, at det danske ord udgør et underbegreb eller et overbegreb til det nærmeste spanske ord.

Man kan således på spansk ikke sige *bage*. Man kan “gøre brød, gøre kager, gøre tærter”: *hacer pan/pasteles/tortas*, men ikke slet og ret bage.

På dansk er et *bor* mange ting, fx minebor og tandlægebor. Spansk har *barreno* og *fresa* herom, foruden det almindelige bor, *la barrena*, men signalerer aldrig slet og ret ‘bor’ med et hævdvundet udtryk.

Det danske *bondegård* ligger skævt for spansk, idet en spansk *granja* er en mindre gård – her er *bondegård* altså overbegreb. Til gengæld overlapper det *hacienda*, der omfatter både store gårde og godser. En landbrugskyndig kan sikkert finde nogle grunde til, at grænserne ikke falder nogenlunde sammen.

På dansk kommer der både *dug* på græsset og på vinduesruden. Men spansk “ser” det som to forskellige ting: *rocío* (afledt af det opr. ord for dug, lat. *ros*) henholdsvis *empañamiento*. Kulturel forskel? Naturligvis ikke. Der er tale om to synsmåder, der er omtrent lige naturlige: den umiddelbare, at det *ikke* er *det samme*, og den mere forstandsmæssige, at det *er* det samme. Spansk har altså ikke noget ord for ‘dug’, kun to underbegreber.

Et dansk *hegn* svarer omtrentligt til spansk *cerca*, men kan dog umuligt være en mur. Det kan en *cerca* udmærket. Her er dansk altså underbegreb til den nærmeste spanske ækvivalent.

På dansk skelner vi skarpt imellem *at klæde sig ud* (til karneval fx) og *at forklæde sig* (som spion fx). Men det gør spansk ikke. *Disfrazarse* bruges om begge dele, dvs. overbegrebsligt om at iklæde sig noget illuderende, hvad enten det er en leg, eller nogen for alvor skal vildledes. Dansk har kun underbegreberne, ikke noget etableret overbegreb.

Det mest forbløffende “hul” i det spanske leksikon er *tyv*. Man skulle jo tro, at alle sprog havde en ækvivalent til *tyv*. Men spansk *ladrón* er overbegrebsligt for ‘tyv’ og ‘røver’! Man kan præcisere ‘et tyveri’: *hurto* (afledt af det lat. ord for tyv, *furt-*), selvom det er almindeligere at bruge *robo*, der er overbegreb for tyveri og røveri. Man har også et præcist ord for ‘stjæle’: *hurtar*. Men de stakkels tyveknægte bliver altså af hundreder af mio. mennesker hver dag slået i hartkorn med røvere!

I alt finder jeg 27 huller i spansk blandt de 1000 almindeligste danske ord. Jeg må sikkert trække et par af dem tilbage. Til gengæld har jeg uden tvivl óverset nogle.

### 3. Hvor stor er ulykken?

Men hvilken *praktisk* rolle spiller det nu, at vore ordbøger ikke er opmærksomme på dem? Hvad kan det gøre, at en oversætter lader den spanske detektiv indtage en frokost med sin elskerinde? Jeg vil kalde det kulturelt undergravende arbejde. Oversætteren trivialisere kulturforskellene mellem Danmark og Spanien; han globaliserer, hvor han burde lókalisere. Han bedriver *interpretatio romana*. De gamle romere yndede at identificere fremmede folks institutioner med romerske.

Oversatte fx fremmede gudenavne med romerske. Tacitus var ikke bleg for at kalde germanske sagnhelte som Sigurd eller de to Helge'r for *Hercules*. Det var skønt, for dét, læserne elskede, var naturligvis at høre om det *fremmede*.

Og hvis den spanske oversætter skriver *procaz* om "en fræk pige", så har han misforstået sin tekst; han er måske klog nok til at skrive *indecente*, *immoral*, dvs. uanstændig; men også her misforstår han, for disse ord er negativt ladede. Han er nødt til kreativt at udfylde hullet i det spanske sprog med et uetableret udtryk.

Det siger også sig selv, at hvis man bruger et fremmedsprogligt overbegreb, fx oversætter *hegn* med spansk *cerca*, så bliver det elementært galt. Det ville svare til, at man i danske tekster erstattede *sofa* med *møbel*, eller *kjole* med *tøj*. Her går der oplysninger tabt. Derimod vil man i mange tilfælde kunne gengive et overbegreb med et underbegreb, hvis konteksten i øvrigt peger på et sådant. Og det vil den meget tit gøre. *København er en smuk by* kan gengives ved *Copenhague es una ciudad hermosa*, men dansk *by* dækker også *pueblo* og *aldea*. Så en sætning som *Der er ca. 20.000 byer i Danmark* kan ikke oversættes simpelt til spansk. Men det er heller ikke en ordbogs opgave at oversætte konkrete tekster. Ordbøgernes emne er sproget, og derfor må ordbøgerne nøjagtigt formidle, hvad sprogets ord betyder: det gøres bedst med ækvivalenter – hvis de vel at mærkes *findes*.

#### 4. Hvordan skal det gøres?

Hvordan skal nu hullerne se ud i ordbøger? Der skal simpelthen stå: **MANGLER!** og så skal de nærmeste ækvivalenter og deres begrænsning beskrives. Man kan også udmærket forsøge sig med en definition af det manglende begreb. Fx skrive, at ital. *bambino* er et "barn på ca. 0–11 år". Enhver snak om over- og underbegreber vil gå hen over hovedet på læseren, parentes vil blive misforstået, og skråstreg bør bandlyses. Hvis man fx skriver, at svensk *kusin* eller eng. *cousin* på dansk betyder 'fætter/kusine', så vil læseren slet ikke forstå den specielle sproglige situation. Han vil tværtimod som alle amatører tro, at *kusin* 's o m m e t i d e r' betyder 'fætter' og sommetider 'kusine', skønt sandheden er, at *kusin* aldrig betyder 'fætter'. Ganske som *hund* aldrig betyder 'hanhund'. Og *menneske* aldrig 'kvinde'. Hvis der er noget i konteksten, der viser, at referenten er en han, så er det med garanti ikke *kusin*, *menneske* el. *hund*. Disse ord kan kun signalere et overbegreb.

Disse betragtninger gælder ikke deiktiske ord. Under diskussionen mente en deltager, at fx det finske pronomen for mennesker, *hän*, godt kan signalere en konkret kvinde. Selvfølgelig kan det det. De deiktiske ord i 3. person betegner præcist, hvad man peger på – eller subjektet i den foregående sætning. Derved fremkommer en kamæleon, der – efter helt faste regler – udpeger et eksemplar blandt et overbegreb. *Hän* betyder derfor snart 'Marjatta Kuusinen', snart 'Erik den 14.'. Ganske som *her* snart præcist betyder 'i mit redskabsskur' snart 'på Hawaii'.

## 5. Hygge

Lad os slutte med det svære spørgsmål: Har vi danske patent på begrebet 'hygge'? Umiddelbart nej, for vi har tyvstjålet betydningen fra norsk. Ordet betød i oldnordisk 'tanke, fornuft', jf. *hu*, men drejes både i dansk, svensk og norsk hen til 'beskyttelse', hvad det jo er yderst *fornuftigt* at ty til for småfolk. De norske dialekter drejer det så hen til den nuværende betydning, og dén tyd optager dansk i 1700-tallets slutning som en ny variant til *hygge* 'beskyttelse', som endnu levede, men snart efter uddøde. Lånet befordredes vist især af de mange nordmænd i København, der skrev på dansk, men med diskrete norvagsmer. Hvordan *er* så den nye betydning? Det er så heldigt, at ODS har en glimrende definition af ordets nuværende prototypiske betydning. (Dén bør man her som så ofte holde sig til, så undgår man en mængde udvandringer):

tilstand ell. følelse af velbefindende (især: ved at sidde lunt og fredeligt inden døre), af hjemligt velvære, og ODS har endda en supplerende kunstnerisk definition af mesteren Jeppe Aakjær:

Der er over Stuen det, som vi kalder Hygge, noget der ligesom besjæler Tingene, saa de staar og sender Ynde fra sig, som Blomster udsender Duft.

Min eneste anke er, at ODS ikke opretter en variant-betydning for interiører. En stue kan være nok så hyggelig, men hvis man betragter den ensom, fx på et museumsbillede, er selve den kollektive stemning jo typisk ikke til stede. Aakjær-citatet har denne nuance. Stuen giver én det indtryk, at der *kunne* være megen hygge (i grundbetydningen) her.

Svensk fik ikke denne videreudvikling. Standard-oversættelsen er *trevnad*, men det er ingen god ækvivalent. Det betyder 'velbefindende', 'at man har det rart og afslappet'. Momentet 'indendørs tryk, afslappet, kollektiv stemning' mangler. Men skulle svenskerne, disse Vestens kollektivismekspertter virkelig have et hul her? Næppe. For i nyere tid er udtrykkene *mysig* og *ha det mysigt* dannet. Først om interiører, men så også om selve stemningen.

Nuvel, danskerne kender kun ét fremmedsprog, engelsk, så det er vel i forhold til dét, at vort fortrin skulle gælde. Ordbøgerne nævner *comfort* både om følelsen og det interiør, der fremkalder den. Det duer ikke, for det kollektive mangler. Men så *snug* adj.? Det kommer tæt på. Det gengives i sin nærmeste variant i ved 'comfortable, cosy' med Thackeray-eksemplet "He liked snug dinners of all things in the world." Og *cosy*? Dét gengives *snug, comfortable* med association til beskyttelse og varme. Endelig har eng. *fireside cosiness*, men det kræver ifølge min hjemmelmand en pejs, så det er kun et – udmærket – underbegreb. Jeg må konkludere, at engelskmanden har en række muligheder til præcis gengivelse af *hygge*. Ja, det er mit skøn, at den *dyrkelse* af hjemlig hygge, der i dansk opstår i 1800-tallet og kulminerer i slutningen af århundredet, i virkeligheden stammer fra England. Engelsk kultur var idealet for den danske, og vi fik i Imperiets glans-

tid begreber som *gentleman, fair play, sport og liberalisme*. For englænderne var centrum den åbne kamin med hjemmets klenodier på kaminhylden. For os var det petroleumslampen.

Skulle *hygge* da være et kultur-fænomen? Jeg tror det. I hvert fald dyrkelsen af det, for selve følelsen af indendørs tryghed og åbenhed i familiens skød er sikkert fællesmenneskelig. De gamle romere har da vel kendt *dén*, men de dyrkede den ikke. Og derfor har latin intet ord for *hygge*. De kendte *securitas*, tryghed, varme, god mad og slægts-kærlighed, men ikke så meget familie-kærlighed, for kernefamilien var ikke opfundet. Hjemmets centrum var arneilden, *focus*, men de dyrkede den med religiøs andagt og anbragte deres husguder, lærer, ved den, men *hygge* er ingen religiøs følelse. Og hvordan skulle de kende noget videre til *hygge*, når de velafrettede børn for længst var lagt i seng og slaver vimsede ind og ud af stuen, i et samfund fuldt af hemmeligt politi og stikkere, og hvor man skulle op ved daggry?

### Ordbøger

Bratli, Carl 1947–1951: *Spansk-dansk ordbog*. København: Forfatterens Forlag.

ODS = *Ordbog over det danske sprog* 1919–1956. København: Gyldendal.

SOED = *The Shorter Oxford English Dictionary* 1978. Clarendon.

Windfeld Hansen & Gawinsky 1999: *Dansk-spansk ordbog*. 5. udg. København: Gyldendal.

### Anden litteratur

Brink, Lars 2003: Hvad har leksikografien lært af semantikken? Intet! I: Hansen & Johansen (udg.): *Nordiske Studier i Leksikografi* 6. Torshavn.

Lars Brink  
 professor, f. 1945  
 Elmevænget 30  
 DK-2880 Bagsværd  
 labri@webspeed.dk



TORBEN CHRISTIANSEN

## Fra redaktørens værksted Om færdigredigeringen af stor Tysk-Dansk Ordbog

**From the editor's workshop – the final editing of Stor Tysk-Dansk Ordbog.** After the authors had handed in the final data files for the dictionary, it fell to me as publishing editor to carry out a cross check in a number of areas in collaboration with the authors. We went through the dictionary data in a transverse fashion to eliminate errors and ensure the necessary consistency. We used as our basis a log file which served both as a to-do list and a memory aid for the editor and the project. In my paper I present some methods to survey and edit a dictionary of this large size, and I present some representative examples: 1. Statistics of fields and their content. 2. Spell check of all Danish words. 3. Check of nouns and their endings and gender within individual entries. 4. Comparison of entries: 50 semantic fields and super-/subordinate concepts such as chemical elements, the grading scale, and Greek letters. 5. Alphabetization of bold parts of idioms.

### 1. Sigtet med færdigredigeringen

I slutningen af 2005 udkom Gyldendals store *Tysk-Dansk Ordbog*. Den er på 2.125 sider og indeholder 153.000 opslagsord og 42.000 faste udtryk og vendinger. Henrik Bergstrøm-Nielsen og Preben Späth har udarbejdet artiklerne, der er gennemlæst og kommenteret af Peter Jürgensen og Birgit Lohse.

Efter at forfatterne havde afleveret alle artiklerne i den foreløbigt færdige version, kontrollerede jeg som forlagsredaktør i samarbejde med forfatterne ordbogen på en række områder. Formålet med dette arbejde, der strakte sig over ca. halvandet år, var at forbedre ordbogens kvalitet og forhindre senere dyre og besværlige rettelser. I denne artikel har jeg søgt at udvælge nogle af de redigeringsmetoder, der kan have interesse for andre. Jeg koncentrerer mig om ordbogens *indhold*, idet forskellige ordbøgers struktur vil afhænge af det anvendte ordbogsredigeringsprogram mv.

Ordbogen er redigeret i Compulexis, der er et program med en ret "flad" tagtekst-struktur, der består af et feltnavn plus et feltindhold. Artiklerne er lagret i en database. Hvis man skønner, at det er tidsbesparende, kan man eksportere data til WordPerfect, foretage gennemgående ændringer ved hjælp af søg og erstat, makroer el.lign. og genimportere data i ordbogen.

## 2. Kontrol på tværs af artiklerne

Ved afleveringen af artiklerne var der læst korrektur. Man kan imidlertid ikke gå ud fra, at alt er i orden, og forlagsredaktøren kan roligt være lige så mistænksom som en fængselsbetjent. Det skyldes, at der er mange ting i spil samtidig, når forfatterne gennemgår en artikel: Der er mange forskellige oplysningstyper som opslagsord, bøjningsformer og oversættelser, og der er både tysk og dansk tekst. Det er umuligt at overskue det hele på en gang, og det er svært at være systematisk og huske, at artikler som fx ugedagene bør være ensartede. Når man sammenligner artikler og felter “på tværs”, vil man uvægerlig få øje på fejl og inkonsekvenser. I den sidste fase gælder det følgelig om at isolere ét fænomen ad gangen og undersøge det.

Færdigredigeringen skulle derfor ikke bestå i endnu en gennemlæsning af ordbogen fra a til z. I stedet skulle ordbogens tekst angribes på tværs af artiklerne ved hjælp af metoder og ressourcer, som forlaget stillede til rådighed.

Mine vigtigste værktøjer i færdigredigeringen var: en feltopdelt artikelstruktur, statistikfunktioner, gode søgemuligheder i basen, fleksible printformer, gennemgående ændringer i tekstbehandling samt en logfil, hvor jeg minutiøst registrerede alt kontrolarbejde.

## 3. Logfil

“Erfaring er evnen til at genkende en fejl, når man begår fejlen for anden gang,” siger man. Men for at undgå fejl kan man notere sine erfaringer i en logfil, der både fungerer som en *to do*-liste og som hukommelse for redaktøren og projektet – nu og fremover. Også andre ordbogsprojekter kan have gavn af søgninger, der enten er generelle eller som kan tilpasses de aktuelle sprog. I figur 1 er der eksempler på generelle søgninger, mens figur 2 indeholder nogle af søgningerne vedr. opslagsord. Jeg har lavet ca. 600 tjek og foretaget godt 2.000 søgninger.

GEN01	Typografi
GEN02	Duplicates check
GEN03	Field code count
GEN04	Forkortelser generelt
GEN05	Forkortelser – enkelte
GEN06	Forkortelser: jm, jn, js
GEN07	ng./ngn./ngt./nt.
GEN08	Enkeltord
GEN09	Stavekontrol – dansk
GEN10	Retskrivningsordbogens nye udgaver 1996 og 2001
GEN11	Stavekontrol – tysk
GEN12	Danske bogstaver i tyske felter
GEN13	Tyske bogstaver i danske felter
GEN14	Bogstavet l i stedet for tallet 1



GEN15	Tal (i stedet for bogstaver)
GEN16	Romertal med punktum
GEN17	Spatium ved bindestreg mellem tal
GEN18	Manglende bindestreg mellem tal og "årig"
GEN19	Ubalancerede parenteser i felter
GEN20	Brøker
GEN21	Bindestreg ændres til tankestreg
GEN22	trykstreg ændres til accentueret bogstav (é)
GEN23	... el. ... ændres til ... (udeladelsestegn, ellipse)
GEN24	Konkurrerende ord
GEN25	??
GEN26	Trykstreger (hovedtryk og bitryk) i tyske ord
GEN27	Trykstregers placering ved -tion
GEN28	personligt subjekt: flere angivelsesmåder
GEN29	"" og »«, replikker og citater
GEN30	Komma
GEN31	overdreven mangler -t
GEN32	Sekvenser af 3 ens bogstaver: aaa/.../ååå
GEN33	Manglende mellemrum efter a. kolon, b. komma, c. lighedstegn, d. punktum
GEN34	Mellemrum ved parentes, spørgsmålstegn og udråbstegn
GEN35	Entities

Figur 1. Nogle generelle tjek, der ikke vedrører specifikke felter.

HWD01	begynder med -
HWD02	,
HWD03	() eller /
HWD04	Komma mellem HWD og REF
HWD05	Feltrækkefølge i artiklens hoved
HWD06	*
HWD07	Uregelmæssige verber
HWD08	Trykstreg i sammensat verbum
HWD09	Trykstreg i sammensat verbum
HWD10	Manglende tryk på -lei
HWD11	Manglende tryk på -rei

Figur 2. Nogle tjek vedr. opslagsord.

#### 4. Overblik over ordbogens indhold

Ud fra en statistik over antal felter og deres indhold grupperede jeg felterne således:

- Felter med tysk indhold, der egner sig til en tysk stavekontrol.
- Felter med dansk indhold, der egner sig til en dansk stavekontrol.

- Felter med indhold, der tjekkes af programmet, fx ordklassebetegnelse og stilværdi. Er der nogle, der bør nedlægges eller slås sammen?
- Felter med en vis uniformering, som har et begrænset antal indhold. Det er fx homografnumre, bøjningsendelser, grammatiske oplysninger samt henvisningsformler.
- Tegnene i udtalefeltet, der skal sammenholdes med listen over lydtegn i vejledningen.
- Små ordklasser som præpositioner og pronomener, der bør gennemgås samlet.

Forkortelser uden for de automatisk tjekkede felter bør man lave statistik over og sammenholde med brugervejledningens forkortelsesliste. Har man eksempelvis brugt *osv.* eller *etc.*, eller er nogle forkortelser så sjældne, at man vil skrive dem helt ud og dermed spare en forkortelse?

## 5. Stavekontrol af alle danske ord

Vi kontrollerede stavningen af alle 715.000 løbende danske ord i ordbogen ved at køre dem op mod en standardliste med 178.000 grundord og bøjningsformer, der især er udvalgt på grundlag af frekvens. 72.000 ord, altså ca. 10 %, blev ikke genkendt. Det faldt i min lod at kontrollere de ikkegenkendte ord manuelt. Fidusen ved den manuelle stavekontrol er, at det er nemmere at finde fejl, når man tvinges til at se på ét suspekt ord ad gangen, end når man læser løbende tekst.

Jeg burede mig inde, rullede gardinet ned og bevæbnede mig med ordbøger og faglige opslagsværker. På en rigtig god dag kunne jeg kontrollere 2.400 ord, så det tog i alt omkring 35 dage. Jeg gjorde brug af makroer for at bevæge mig hurtigt frem gennem filen og inddelte ordene i tre typer (se figur 3):

- åbenlyse fejl eller eventuelle fejl (dvs. slåfejl, stavefejl, forkert felt, ikkeeksisterende ord, forældet ord eller ord, som jeg ikke kunne få verificeret), fx *\*afdelig*;
- de "måske egnede" ord, fx sammensætninger eller ad hoc-dannede afledninger, som vi har brug for til en oversættelse, fx *baseler*, *basistysk*, *besættelsesstatut*, *opfyldelighed*, *opinionsforskning*, *opsejlet*, *opstigningssted*;
- gode gængse ord, fx *advokatfirma*, *affaldstræ*, *afisolere*, *aktiehandel*, *aktiens*, *orddannelse*, *ordgåde*.

suspekt ord	TC's kommentar	feltnavn + indhold	opslagsord
afdelig	-ING	TSL neurologisk afdelig	<i>Neurologie</i>
afdragelse	FORFATTER TJEK	TSL afdragelse	<i>Tilgung</i>

afdragning	DO.	TSL afdragning	<i>Abtragung</i>
afdæmpetr	-PET	EXB hans kritik var særdeles afdæmpetr	<i>zahn</i>
affaldsafgit	-IFT	TSL affaldsafgit	Abfallabgaben
affejelse	TC RETTER TIL AFFEJNING (GOOGLE 78, MOD AFFEJELSE 1)	TSL affejelse	<i>Abfertigung</i>
affileret	AFFILIERET	TSL affileret selskab	<i>Konzerngesellschaft</i>
affindelses-godtgørelse	IKKE PÅ GOOGLE. AFFINDELSESSUM FINDES, SKAL DET BRUGES?	TSL affindelses-godtgørelse	<i>Abfertigung</i>

Figur 3. Stavekontrol. De suspekterede ord, der er vist både isoleret og i sammenhæng, er forsynet med forlagsredaktørens kommentarer.

Resultatet af gennemgangen var følgende:

- Ca. 10.000 ord kunne umiddelbart rettes eller sendes videre til forfatterne til kontrol. Der blev rettet eller ændret i ca. 4.000 artikler.
- 38.000 ord var "måske egnede". Dem gjorde vi ikke mere ved.
- Og endelig var der 24.000 gode gængse ord, som blev lagt til standardlisten. Vi passer på ikke at putte for mange ord ind i listen, for at den ikke skal miste sin skarphed.

## 6. Artikelinterne tjek

I en tysk grammatik kan man finde regler for sammenhæng mellem substantivers endelser og deres køn. Eksempelvis skal tyske substantiver, der er afledt med *-heit*, være femininum. Jeg søgte derfor på opslagsord på *-heit*, der ikke stod som femininum, og fandt bl.a. *Birkenscheit* (neutrum). Men det er faktisk ikke en fejl, da ordet er sammensat med *Scheit* (neutrum). Hvis vi havde oplysninger om, hvilke morfemer der indgår i opslagsordene, ville *Birkenscheit* være klassificeret som en sammensætning. Moralen er, at jo flere, evt. skjulte oplysninger man har om sine opslagsord, jo mere præcist kan man søge.

Nogle bøjningsformer kan være uforenelige med bestemte bogstaver sidst i opslagsordet. Eksempel: Hvis et opslagsord, der ender på *-e*, har en bøjningsendelse, der starter med *-e*, kan det være en fejl. Jeg fandt ca. 40 artikler, der skulle rettes, fx **Farbenprobe die** - *-en* i stedet for *-n*. Det samme kan man søge efter i ordbøger med danske opslagsord.

## 7. Semantiske områder og over-/underbegreber

Vi har arbejdet med ca. 50 semantiske områder og over-/underbegreber, i alt knap 3.000 opslagsord. Hver gruppe indeholder et kvantitativt begrænset antal ord, der kan være alt lige fra ugedage, grundstoffer, græske bogstaver og landnavne til den tyske karakterskala. Formålet med denne redigering er dels at sikre, at vi har optaget alle de relevante opslagsord, dels at redigere dem ensartet. Til klassifikation af disse opslagsord har jeg brugt et skjult felt, hvor vi frit kan skrive et overbegreb (fx 'karakter'), så vi kan søge på artiklerne, printe dem og behandle dem samlet.

Et eksempel: Den tyske karakterskala går fra *Eins, Einser* op til *Sechs, Sechser*. Vi skal her sikre os en overensstemmelse med trinnene på den danske skala. Desuden skal vi være opmærksom på, at eksempelvis *Eins* oversættes "elleve, ellevetal; tretten, trettental", mens *Einser* oversættes "ellevetal; trettental".

Denne ordbog skal indeholde alle de græske bogstaver. Vi fandt de relevante tyske betegnelser ved at søge på indhold i *Duden – Deutsches Universalwörterbuch* (se figur 4). En søgning viste imidlertid, at nogle af dem simpelthen manglede som opslagsord i vores ordbogsbase, og dem, der var optaget, stod en lille smule forskelligt, idet der ved nogle stod "græsk bogstav", ved andre blot "bogstav" eller slet ingenting. Jeg vil påstå, at hvis man laver den slags søgninger i ældre ordbøger, så vil man finde mange ting, der ikke er redigeret konsekvent.

### OMR12 Græske bogstaver

Status: o.k.

#### Baggrund

TC har konstateret, at vi har Alpha og Beta. Vi skal sikre 1) at vi har alle de græske bogstaver, 2) at de er opstillet ensartet.

TC har indsat CLC:

CLC .græsk.bogstav.

Vedlagt udtræk fra Dansk Fremmedordbog.

\*\* griechischen Alphabets" gefunden: 24

Sucheinstellungen:

Suchbereich: Buch "Duden - Deutsches Universalwörterbuch"

Suchoptionen: Volltext (inkl. Stichwörter)

\*Alpha

\*Beta

\*Chi

\*Delta

\*Epsilon

\*Eta

\*Gamma

\*Jota  
 Kappa  
 \*Lambda  
 \*My  
 Ny  
 \*Omega  
 \*Omikron  
 Phi  
 \*Pi  
 Psi  
 Rho  
 Sigma  
 Tau  
 Theta  
 \*Xi  
 \*Ypsilon  
 Zeta

\* = er med.

Mangler: Kappa, Ny, Phi, Psi, Rho, Sigma, Tau, Theta, Zeta.

Print til HBN 11.12.2003 (15 artikler). Print retur 5.1.2004, TC har rettet 13.1.2004.  
 Modtaget print 4.2.04, rettet 6.2.2004.

*Figur 4.* Uddrag af logfilen: kontrol af de græske bogstaver.

Samme behandling fik sammensætninger, der starter med tal, fx *-stöckig* (‘-etages’/‘etagers’). Vi skulle her fastsætte et rimeligt tal-interval (i dette tilfælde medtog vi fra 1 til 10 etager), der måtte desuden ikke være nogen umotiverede huller, og artiklerne skulle være ensartede.

## 8. Tjek ved hjælp af tekstbehandling

Et ordbogsprogram vil aldrig kunne tage højde for alle de tjek, som man kan finde på at foretage. Derfor bør man kunne eksportere sine data i en form, hvor man får feltets navn plus indholdet, så man kan massere indholdet på alle mulige måder. Et eksempel:

Foden i artiklerne indeholder idiomer, der kan have et fremhævet ord (“blikfang”), som idiomerne skal være alfabetiseret efter. Da ordbogsprogrammet ikke kan alfabetisere efter blikfangene, trak jeg idiomerne ud. Dernæst stillede jeg blikfangene op under hinanden. I artiklen **aber** stod to blikfang i forkert alfabetisk rækkefølge:

... *sehr* ...  
 ... *schon* ...

Jeg alfabetiserede blikfangene i tekstbehandling, sammenlignede de to filer automatisk og fik derved en markering af forkerte rækkefølger.

Torben Christiansen  
cand.mag., redaktør, f. 1955  
Børglumvej 12  
DK-2720 Vanløse  
tc@post.tele.dk

SILVIE CINKOVÁ

## “Movement towards Structure” Basic Verbs in Learners’ Lexicons

I denna artikel beskrivs en syntaktiskt grundlagd semantisk skiftning hos den svenska verbala pseudosamordningen *hålla på och* som används som progressivetsmarkör. I vissa kontexter får den en närmast motsatt betydelse, nämligen ‘hela tiden’. Sökningar i den svenska PAROLE-korpusen samt i en liten svensk-tjeckisk parallellkorpus visar att den nya betydelsen framför allt är kopplad till negativa imperativyttranden men att det samtidigt finns en spridningstendens till andra syntaktiska strukturer. Undersökningens resultatet jämförs med innehållet i några svenska referensverk och relateras till Paul Hoppers (1987) idé om ”emergent grammar”. Den observerade semantiska skiftningen förklaras med en allmän mekanism som Heine, Claudi och Hünnemeyer (1991) har definierat och kallat ”context-induced reinterpretation“.

### 1 Emergent Grammar

#### 1.1 Introduction

Very common lexical verbs, especially verbs of motion and physical action are in many languages used to denote abstract events and relations. This results in semantic generalization, which in extreme cases makes these verbs (semi-)auxiliary or modal verbs. In numerous other cases they undergo less apparent semantic shifts that are too context-dependent to be noted by grammars, but at the same time they are too vague to be recognized as separate lexical units in lexicons. Drawing on the ideas of *grammaticalization* and *context-induced reinterpretation*, this paper describes an ongoing semantic shift in the Swedish auxiliary verb *hålla på* in its use as progressivity marker, suggesting that cases of context-induced reinterpretation should be systematically identified and observed by lexicographers in general.

#### 1.2 “Movement towards Structure”

The title contains a quotation borrowed from Hopper (1987). Hopper does not conceive grammar as an abstract set of rules filled by words at given communicational occasions, but he rather believes that grammar is what emerges from words, collocations and phrases: “Structure, or regularity, comes out of discourse and is shaped by discourse as much as it shapes discourse in an ongoing pro-

cess” (Hopper 1987). Making no difference between what is traditionally called ‘grammaticalization’ and ‘lexicalization,’ he encourages linguists “to study the whole range of repetition in discourse” and to “seek out those regularities which promise interest as incipient sub-systems”. This in effect requires searching large corpora for formulations that are considered especially useful by many speakers, and therefore they are used (and hence re-assessed and re-shaped) in many different contexts. To fit the immense variety of contexts (i.e. to increase their cross-textual consistency), the formulations extend their collocability, which results in their increasing vagueness.

### 1.3 Context-Induced Reinterpretation

Heine, Claudi and Hünemeyer (1991) as well as Lakoff (1987), Lakoff and Johnson (1980) and others have listed a hierarchy of semantic domains, such as SPACE – TIME etc. Many grammatical structures have developed by a metaphorical transfer from a more concrete semantic domain to a more abstract one. For instance, the English volitional-future marker *to be going to* (as in *Peter is going to write a letter*) is a case of metaphorical use of the verb *go*, which primarily denotes a spatial change, in the temporal domain. However, the process of metaphorical shift is a continuous one. Many uses are ambiguous as to which semantic domain they relate to. To follow Heine, Claudi and Hünemeyer’s example with *to be going to*, the following continuum can be observed (cf. Heine, Claudi and Hünemeyer 1991):

- |   |                               |
|---|-------------------------------|
| (1) <i>Henry is going to town.</i>                          | SPATIAL MOVEMENT              |
| (2) <i>Are you going to the library?</i>                    | SPATIAL MOVEMENT/INTENTION    |
| (3) <i>No, I am going to eat.</i>                           | INTENTION/SPATIAL MOVEMENT?   |
| (4) <i>I am going to do my very best to make you happy.</i> | INTENTION/NO SPATIAL MOVEMENT |
| (5) <i>The rain is going to come.</i>                       | PREDICTION                    |

For such indistinct semantic shifts Heine, Claudi and Hünemeyer (1991) introduce the term *context-induced reinterpretation*. It is described in three idealized stages:

- I. A form F has a focal meaning A. It comes in use with a context-specific side meaning B.
- II. F can be used as B only in unambiguous contexts.
- III. F/B becomes conventionalized and F becomes polysemous.

The case of *to be going to* has reached stage III. Thus a sentence like

- (6) *Peter is going to leave the hospital on Monday.*



has become ambiguous, expressing either Peter’s intention to leave the hospital or the speaker’s prediction about Peter’s future, and other contextual clues must be added to disambiguate it.

## 2 Verbs in the Vocabulary

### 2.1 Swedish ‘Lexical Profile’

Nouns and verbs are believed to be the only universal parts of speech in human languages; i.e. all known human languages have verbs and nouns, though their syntactic profile in general might be very different from that of verbs and nouns in Indoeuropean languages. Verbs as well as nouns are considered to be open parts of speech, which obtain new members during the development of any given language. Frequency analyses of various European languages show an interesting fact: there are significantly fewer verbs than nouns. As Hanks (2003) puts it: “... it seems almost as if all the other parts of speech (verbs and function words) are little more than repetitive glue holding the names in place”. For Swedish, Viberg (1990) notes that there are about 8.5 times fewer verbs than nouns. This implies that verbs, which by their definition denote relations between entities, must have the ability to fit many more different contexts than nouns do. Besides that, almost one half (45.5 %) of Swedish verb occurrences is represented by the 20 most frequent verbs. Almost every second verb usage belongs then to the top twenty! The most frequent verbs (which are not just auxiliary and modal verbs) must evidently have the ability to fit a wide range of different contexts, and some of their usages really “promise interest as incipient sub-systems” (Hopper 1987).

### 2.2 Basic Verbs

Not specifically for Swedish, many of the most frequently used verbs are lexical (full) verbs that are stylistically neutral and denote basic-level categories of events, i.e. events identified and classified with just the level of granularity that is first and intuitively best perceived when acquiring the common knowledge or a foreign language. (For instance, *go* expresses a basic-level event, while *march* would be too specific and *move* too general.) For the purpose of this paper, verbs denoting basic-level categories of events will be called basic verbs.

Basic verbs primarily denote motion/position, like *stå* ‘stand’, *ställa* ‘put vertically’, *komma* ‘come’, *gå* ‘go’, or actions of physical control, such as *få* ‘get’, *ta* ‘take’, *ge* ‘give’ or *hålla* ‘hold, keep’. They often undergo semantic shifts to denote abstract events and states; e.g. they often act as support verbs (*fatta beslut* – lit. ‘grasp a decision’, *ställa en fråga* – lit. ‘put a question’, *ge/få/ha besvär* ‘give/get/have problems’).

Sometimes metaphorical uses of basic verbs gain such a significant cross-textual consistency that they develop into auxiliary verbs. To name but a few, *have*

and its equivalents have in many languages developed into perfect tense markers and the English *go* as well as the Swedish *komma* express different types of future tenses. With Hopper's commitment in mind, any frequent uses of basic verbs enabled by minor meaning generalizations deserve attention as they are unpredictable for foreign learners while they are neither universal enough to be noticed by grammars, nor do they form collocations stable enough to be put on a list of idioms in a lexicon. The following case study is tracing the context-induced reinterpretation (see 1.2) in a well-established grammaticalized use of *hålla* 'hold, keep'.

### 3 The Case of *hålla på och*

#### 3.1 Language Resources

My reasoning on an ongoing semantic shift within the construction *hålla på och* is based on the evidence from two corpora:

- 1) the Swedish PAROLE corpus, which comprises about 20 million tokens and has morphosyntactic tagging. It belongs to the Språkbanken corpora set maintained by the Gothenburg University (PAROLE).
- 2) the Swedish-Czech parallel corpus belonging to the Czech National Corpus, which is being built within a larger project of parallel corpora (INTERCORP). It has no tagging and comprises approx. 1.5 million tokens at the moment. The texts are mostly taken from fiction, with Swedish and Czech being the source as well as the target language respectively.

#### 3.2 *Hålla på* in Lexicons and Grammar

The corpus evidence was first confronted with *Svenskt språkbruk* (SS), a large Swedish monolingual dictionary, which is partly designed as a production dictionary.

The lexicon entry *hålla på* in SS captures:

- 1) the lexical meaning of the construction *hålla på med* – 'to be busy with', 'to pursue', 'to be engaged in' etc. (*Han höll på med deklARATIONEN när det ringde på dörren.*) (...)
- 2) information on the constructions *hålla på att/hålla på och/hålla på med att* as progressivity markers, even with the implicit warning that the last only has marginal use (*även*). (*Han höll på att stryka när telefonen ringde. Jag håller på och diskar. Han höll just på med att deklarerera när det ringde. Landskapet håller på att förändras.*) (...)
- 3) information on the construction *hålla på att* in the sense 'to be on the verge of'.
- 4) "pragmatic phrases/idioms" – *Så här kan du inte hålla på!* (lit. 'this way can you not hold on!') with the explanation "sägs när ngn el.ngra inte kan

fortsätta på samma vis längre” (...) (‘said when someone cannot go on in the same way any more’)

Another large monolingual dictionary, *Norstedts stora svenska ordbok* (NSSO), provides roughly the same information, except that it does not list the idiom *Så här kan du inte hålla på!*.

The largest Swedish grammar, *Svenska Akademiens grammatik* (SAG), makes remarks on the different distribution of the three constructions *hålla på att/hålla på och/hålla på med att*. It notes that *hålla på att/hålla på med att* used to be the only correct form until the late 20th century, though the competing *hålla på och* had occurred centuries ago. Nowadays *hålla på och* is also recognized as standard. Yet *hålla på att* is strongly preferred with inanimate subjects while *hålla på med att* is going out of use. The construction *hålla på och* can only be used as a progressivity marker (except in verbs denoting states) while *hålla på att* can, when used with telic (perfective) verbs (i.e. verbs denoting events with an inherent terminal point), also be used in the sense ‘to be on the verge of’, e.g.

(7) *Jag höll på att svimma* ‘I nearly fainted’.

### 3.3 Analyzing the Corpus Evidence

The corpora basically confirm the statements given by the grammar and the two dictionaries. They have returned a few outstanding concordances, however, for which the reference books give no explanation. It is to be stressed that all the titles are comprehensive and up-to-date, and that they have been willingly chosen as the ultimate choice for every advanced foreign learner. The most striking concordance comes from the parallel corpus:

(8) ... *proto se taky náš farář musel v jednom tahu modlit, aby nebyl tak zlej...*  
lit. ... and that’s why also our priest **had to pray all the time** in order not to be that vicious...

... *därför måste också vår präst hålla på och be stup i ett, så att han inte skulle vara så elak...*

lit. ... and that’s why our priest had to also **hold on and pray all the time** in order not to be that vicious...

The sentence comes from a Czech novel (written in a low-standard spoken Czech), translated into Swedish. The Czech original uses a simple infinitive of the reflexive verb *modlit se* ‘pray’ and a temporal adverbial (*v jednom tahu*). The Swedish translation, however, puts the infinitive *be* ‘pray’ in syntactic coordination with *hålla på*.

The general understanding of the progressive tense suggests that it is to be used under one of the following conditions:

- 1) the speaker focuses on one given moment, making a “snapshot” of the situation, as in:
- (9) *Jag håller på och diskar.*  
I am washing up.
- 2) discourse backgrounding: the speaker reports on an event or a state that provided the background for a more important event, e.g.:
- (10) *Han höll på att stryka när telefonen ringde.*  
He was ironing when the phone rang.

The use of a progressive marker in a sentence like (8) is unexpected. The sentence (8) may express anything but a situation snapshot. Moreover, it combines the progressive tense with a semantically incompatible adverbial that expresses constancy. Neither can the progressive tense be used for discourse backgrounding as the context does not provide any event which could serve as background for the priest’s praying. The translator said that he had used *håller på och* both to indicate the low language standard and as intensification of *all the time*. Neither of these two additional functions of *hålla på och* has been explicitly mentioned by SAG, SS or NSSO. SS, though, contains an idiom that can be paraphrased with “you cannot go on like this any more”. It possibly shares the semantic feature of constancy with the sentence (8).

A closer search in the PAROLE corpus gave more than 40 occurrences of *hålla på och* in contexts that suggested the presence of the constancy feature. Most of them actually fitted the syntactic pattern of what SS had classified as idiom, e.g.:

- (11) *Ni ska inte hålla på och larva er sådär, för jag har ingenting att skämmas för.*  
You shouldn’t keep acting like this because I have nothing to be ashamed of.

However, some occurrences lack the imperative:

- (12) *Men i längden så kan vi ju inte hålla på att bara försvara oss.*  
But in the long run we cannot just keep defending ourselves.

Some occurrences even lack both the imperative and the negation:

- (13) *I princip tyckte hon att det verkade botten att hålla på och knega mellan nio och fem.*  
Actually she meant that it appeared miserable to keep working from nine to five.

The construction *hålla på och* in the constancy sense could appear in questions,

- (14) *Det vet jag inte heller. Varför ska du hålla på och fråga så där?*  
I don't know myself. Why should you keep inquiring like this?

and even one positive declarative sentence like (8) can be found in PAROLE:

- (15) *“Du förnekar det fortfarande. Det är otroligt.”*  
”Det är otroligt att du fortfarande håller på och ältar det. Jag gillade henne aldrig.”  
“You are *still* denying it. It is incredible.”  
“It is incredible that you still keep agonizing over that. I never liked her.”

In accordance with the translator's statement, the “constancy-*hålla på*” mainly occurred in direct speech or in free indirect speech, which both imitate spoken language. With the reservation that both the corpora used are written-language corpora it is likely that the constancy-*hålla på* is a construction typical of or (at the moment) confined to spoken language. Another conclusion that can be drawn from the concordance is that the construction mainly occurs in sentences with low facticity. The example sentences (8) and (15) suggest that high facticity may require a disambiguating adverbial as no positive declarative sentence without a disambiguating adverbial was found either in the parallel corpus or in PAROLE.

How is it that a progressive construction has acquired precisely the opposite meaning? The progressive *hålla på* is the default interpretation of *hålla på* with atelic verbs. It appears in positive as well as in negative declarative clauses, questions etc. in all tenses. Unlike that, the constancy-*hålla på* seems to appear almost exclusively in negations, questions and infinitives. It is the negation that might give the clue for this semantic shift. In a negated progressive sentence, it is not just a single moment of the given event that is negated, but it is the entire event. For instance, the sentence

- (16) *De håller på att bråka.*  
They are fighting.

focuses just one moment of the ongoing event. The same goes for the progressive tense as discourse backgrounder:

- (17) *De höll på att bråka när jag kom.*  
They were fighting when I arrived.

The negation of a sentence predicate, however, says that the entire event does not take place (at the moment of reference) rather than that a single moment of the event does not take place at the moment of reference. This is best perceived in imperative; for instance, by saying:

- (18) *Don't be doing something!*

the speaker necessarily means:

- (19) *Stop doing what you have been doing just long enough to annoy me!*

Implicitly, the event really must have been taking place.

More to say, the relation between progressivity and facticity also works the other way round: when the constancy-*hålla på* is employed in a negative utterance with an event, it suggests that the given event is taking place and should be stopped. This implication has been effectively exploited for rhetorical purposes: at many occasions, especially when reading a text or listening to a speech, the audience naturally has no way of immediately being certain that the mentioned event is taking place at all, mainly when the commented event has been underspecified and expressed by a more or less evaluative paraphrase. By employing the constancy-*hålla på* the speaker adds some kind of asserting modality. A random search in Google returned a few interesting examples of this type:

- (20) *Vi ska inte hålla på att keynesianskt försöka mota konjunkturer.*  
We are not supposed to keep catching conjunctures in the Keynesian fashion.
- (21) *De latinamerikanska, asiatiska och afrikanska staterna ska inte hålla på och blanda sig i USA's och Europas affärer hela tiden!*  
The South American, Asian and African states should not get permanently involved in the USA's and Europe's business!

#### 3.4 Constancy-*hålla på* and Context-Induced Reinterpretation

The close association of the constancy reading of *hålla på* with negation, together with its sparse occurrence in positive utterances, points to the context-induced reinterpretation (see 1.2) as the explanation of this spectacular semantic shift of *hålla på* from progressivity into constancy, which could even have been supported by the semantics of the related construction *hålla på med något* ('to pursue something'). The focal sense "progressivity" came in use in negative utterances with the context-specific side meaning 'stop doing'. As shown above, the constancy sense of *hålla på* is unambiguous in negative sentences, and thus convenient to use. Therefore it has gradually become a truly grammatical means of forming a new semantic shading of event negation. From there the constancy-*hålla på* has been spreading even to contexts in which it is ambiguous. Additional disambiguators must be added to make sure that progressivity, which is the primary sense of *hålla på*, remains inactive. This is achieved with adverbials semantically incompatible with the progressivity sense, such as *all the time*, *permanently*, etc.

### 3.5 Constancy-*hålla på* and Reference Books

The constancy-*hålla på* has neither been mentioned in SAG, nor in NSSO. It can only be traced in the idiom

- (22) *Så här kan du inte hålla på!*  
lit. This way can you not hold on!

captured by SS. Capturing this construction is evidence of corpus-based approach applied to this entry. Unlike in SAG and NSSO, the semantically deviant use of *hålla på* was identified among the enormous amount of the progressive uses. However, the few occurrences of constancy-*hålla på* that were not in negative imperatives were ignored as marginal, and it was described as bound to negative imperative and classified as an idiom (or a *pragmatic phrase*, to stick to SS’s terminology). This is likely to be too strong a statement, though. Idioms are supposed to have their fixed patterns. Representing the constancy-*hålla på* as an idiom implies that only the elements expressed with dummies (indefinite pronouns etc.) can be modified. There was no way for the user to find out that this construction can appear outside the negative imperative.

## 4 Lexicographical Relevance of Context-Induced Reinterpretation

The approved practice of today’s corpus-based lexicography has been to sort out and to display prototypical uses of lexical units found in a large corpus. One reading of a lexical unit in a corpus-based lexicon is often determined by its syntagmatic behavior as well as by its most typical collocates (here it was the negative imperative and the constancy-*hålla på*). Uses of a lexical unit that are supposed to be confined to certain collocations and syntactic patterns are to be presented as phraseological units, while those that show tendencies for syntagmatic variability should be rather presented as separate readings. Heine, Claudi and Hünneymeier’s multilingual evidence for context-induced reinterpretation suggests a standard way to examine how a marginal syntactic pattern of a potential new reading of a lexical unit is associated with the prototypical one. The context-induced reinterpretation helps us to be alert to regular syntactic transformations.

## 5 Conclusion

Large corpora and tools for collocation analysis alone enable us to identify many lexical patterns we would not be aware of when using only introspection, which makes the lexicographical description of lexical units more comprehensive than ever. So far, the approved lexicographical practice in the corpus-based description of lexical units has been to pick out the prototypical uses out of the corpus evidence, sorting others aside as exploitations of the prototypes or simply oddi-

ties. This approach is mostly efficient but it carries the risk of ignoring systematic spreading of lexical patterns out of limited contexts towards general usability. Including the idea of context-induced reinterpretation into the lexicographical reasoning makes it easier to trace structural tendencies in the discourse in the Hopperian sense.

### Acknowledgements:

This work has been supported by the Grant Agency of the Czech Republic (GA-CR 405-06-0589/2006-2008) and by the Czech Ministry of Education (joint project 0021620823/2005-2011).

### Lexicons, Grammars and Corpora:

INTERCORP = Intercorp: Swedish-Czech Parallel Corpus. Joint Research Project 2005-2011, Nr. 0021620823, Czech Ministry of Education. <https://trnka.ff.cuni.cz/ucnk/intercorp/>.

NSSO = Allén, Sture m.fl. (red.). *Norstedts stora svenska ordbok*. Stockholm: Språkdata & Norstedts Förlag 1995.

PAROLE = Språkbanken. <http://spraakbanken.gu.se>.

SAG = Teleman, Ulf m.fl.: *Svenska Akademiens grammatik*. Stockholm: Svenska Akademien 1999.

SS = Clausén, Ulla m.fl. (red.): *Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Stockholm: Svenska Språknämnden & Norstedts Ordbok 2003.

### Bibliography

- Hanks, Patrick 2003: *Norms and Exploitations: Corpus, Computing, and Cognition in Lexical Analysis*. MIT Press, forthcoming. Manuscript, obtained 2003 from the author.
- Heine, Bernd, Ulrike Claudi and Friederike Hünemeyer 1991: *Grammaticalization. A conceptual Framework*. Chicago and London: The University of Chicago Press.
- Hopper, Paul 1987: Emergent Grammar. I: *Berkeley Linguistics Society* 13, 139–157.
- Lakoff, George 1987: *Women, Fire and Dangerous Things. What categories reveal about the mind*. Chicago: Chicago University Press.
- Lakoff, George, and Mark Johnson 1980: *Metaphors We Live By*. Chicago: Chicago University Press.
- Viberg, Åke 1990: Svenskans lexikala profil. *Svenskans beskrivning* 17, 391–408.

Silvie Cinková  
 research assistant, f. 1975  
 Institute of Formal and Applied Linguistics  
 Faculty of Mathematics and Physics  
 Charles University in Prague



Malostranske nam. 25  
CZ-118 00 Praha 1  
cinkova@ufal.mff.cuni.cz



## Norrøn metaordbok – en rapport fra en øvelse

The term Old Norse usually denotes the (written) languages used in Iceland until 1540 and in Norway until 1370. Old Norse was a weakly normalised language with a wide variety of forms. The scholarly study of Old Norse has long traditions and several dictionaries have been published. *The Dictionary of Old Norse Prose* (ONP) is a current major initiative but so far covering only letters A-E. J.Fritzner's dictionary, FRI, (1883–1896,1974) and F.Jónsson's *Lexicon Poeticum*, LP, (1931) still remain the major dictionaries. The dictionary group of the Medieval Nordic Text Archive (Menota) has as one of its main objectives to establish a framework for linking the disparate Old Norse lexicographic resources. The fragmented state corresponds closely to the state of the *New Norwegian* lexicography 10 years ago. Thus it was a natural choice to try to apply the methods and tools to Old Norse developed for the *Norwegian Dictionary 2014* project (NO2014), the so called Meta Dictionary, a tool for systematizing lexicographical material for weakly normalised languages. In its basic form the Meta Dictionary is an electronic form thesaurus linking realizations of the lemmas. In this pilot project we have focused on the letter B and merged the headword lists of ONP, FRI, LP and a word list based on five lemmatised texts and 200 lemmatised charters. The process and the structure of the resulting Meta Dictionary are divided in three levels of normalisation. All the original forms are kept and made searchable.

1. *Normalisation*: The first step towards merging was simply to ensure one common system for normalised spelling.
2. *Citation form*: The wordlists follow different principles in choice of the form of the cited headword, e.g. nom.sg versus nom.pl. Consequently we had to make sure that all the wordlists use the same forms.
3. *Meta level, further merging of related word forms*: The last step is to connect the multitude of related word forms, both between and within lists, in meta lemmas. Such related forms can be nouns differing in gender, or small differences in word formation.

We also briefly discuss the linking of the medieval material to the modern dictionary projects, like NO2014. We conclude that such a linking at least for Norwegian will involve considerable manual processing. The realization of an Old Norse Meta Dictionary should also rely on a close cooperation with the ONP project.

## 1 Bakgrunn

Norrønt er det språket vi finner i Norge og på Island i middelalderen. Det nedfeller seg i skriftlige kilder fra rundt 1150. I Norge pleier en å regne den norrøne periode til ca. 1370 og på Island til første halvdel av 1500-tallet. Språket var ikke statisk i denne perioden. Der var dialektale forskjeller og det skjedde endringer over tid. Det overleverte materialet viser både ortografisk og språklig variasjon. Norrøn og nordisk filologi har utviklet seg som disiplin gjennom nærmere 200 år. Tekstene i utgaver og ordbøker for nordiske middelalderspråk preges av et stort mangfold med hensyn på normering, antatt leksikalisering og ortografi. Slik sett er det en dobbel varians for norrønt, originaltekstene er svakt normert og metadisiplinen har utviklet seg. Situasjonen er dermed ikke ulik den vi finner for nynorsk skriftradisjon og leksikografi.

De to store pågående ordbokprosjektene for norrønt og nynorsk, *Ordbog over det norrøne prosasprog* (ONP) og *Norsk Ordbok* (NO), har også mange likhetstrekk. Begge ble grunnlagt på 1930-tallet, og for begge prosjektene var det på slutten av 1990-tallet åpenbart at den daværende redaksjonskapasiteten ville kreve ytterligere flere tiårs redigering før verkene var komplette. For begge prosjektenes vedkommende er det i de senere år blitt foretatt en omfattende digitalisering av kildematerialet. NO har dessuten blitt fullfinansiert med en planlagt ferdigstillelse i 2014.

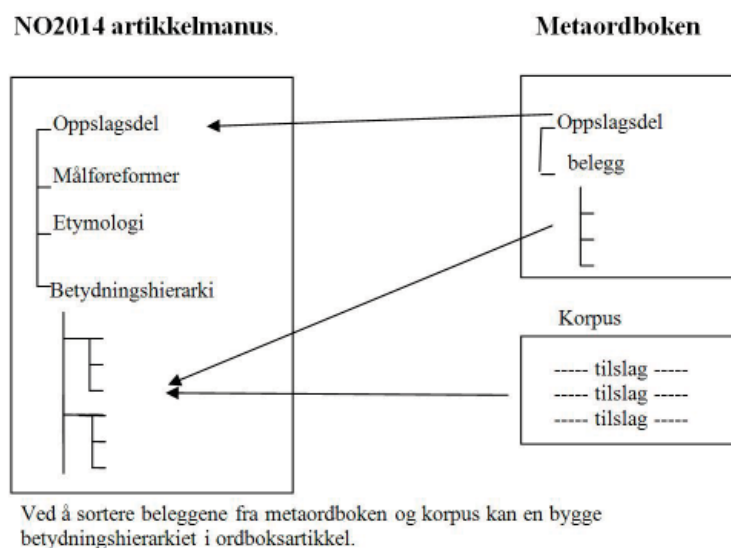
Selv om situasjonen for norrøn leksikografi minner om situasjonen for den nynorske leksikografien på slutten av 1990-tallet, er rammebetingelsene noe annerledes og det organisatoriske atskillig mer komplekst. Norrønt er ikke et levende språk og heller ikke bare knyttet til ett land. Samtidig som norrønt og norrøn tekstutgivelse også er et internasjonalt forskningsfelt, er språket en vesentlig del av Islands og Norges kulturarv.

ONP ble startet på 1930-tallet da Island var underlagt Danmark og må sann sett sies å være et dansk prosjekt. Norge har siden 1940-tallet hatt et eget Gammelnorsk Ordboksverk (GNO) med det formål å dokumentere alle sider av det gammelnorske språket. GNO har mellom annet utgitt et supplementsbind til J. Fritznors ordbok og arbeider med å få realisert *Mellomnorsk ordbok* som skal beskrive norsk i den perioden ONP ikke dekker, 1370–1580. De leksikografiske hovedverkene for norrønt er fortsatt J. Fritznors *Ordbog over det gamle Norske sprog* (1883–1896, 1972) og F. Jónssons *Lexicon Poeticum* (1931). I tillegg kommer de tre første bindene av ONP. Hele verket vil neppe bli ferdig redigert på flere tiår. Det finnes altså ingen fullstendig og samordnet leksikografisk beskrivelse av norrønt.

I løpet av de siste 15 årene har det likevel skjedd en rekke positive ting. Det har vært en tiltagende produksjon av elektroniske utgaver av nordiske middelalderhåndskrifter, og standardordbøkene er blitt gjort elektronisk tilgjengelig. Ved ONP har en også digitalisert store deler av grunnlagsmaterialet. I 2001 ble det

nordiske nettverkssamarbeidet Medieval Nordic Text Archive (Menota, [www.menota.org](http://www.menota.org)) etablert. Menota er et nettverk av nordiske forskningsinstitusjoner og -selskaper som har til hensikt å bevare, formidle og utgi middelaldertekster i elektronisk form. Menotahåndboken har gitt oss en utmerket standard for koding av nordiske middelalderhåndskrifter. Samarbeidet om publisering av elektroniske nordiske middelaldertekster har kommet i godt gjenge. Det har kommet på plass et elektronisk tekstarkiv med et tilhørende lemmatiseringsverktøy og et rammeverk for et elektronisk korpus. Menotas gruppe for leksikografi, Nordiske middelalderordbøker, ([www.edd.uio.no/menota](http://www.edd.uio.no/menota)) ser på mulig samarbeid og deling av leksikalske ressurser mellom de ulike ordboksprosjektene for middelalderspråk i Norden. Arbeidet som er beskrevet i denne artikkelen er gjort i den forbindelse. Vi har i vår øvelse sett på hvordan en kan knytte sammen de leksikalske ressursene i et felles rammeverk for søk i de ulike digitale ordboksressursene og tekstkorpusene.

## 2 Hva er norrøn metaordbok?



Figur 1: Norsk Ordboks redigeringssystem, Metaordboken og korpus

Middelalderspråk som norrønt er svakt normerte språk med et stort utvalg av ulike ordformer. For å optimalisere søk og analyser bør en størst mulig del av ordene i de løpende tekstene være lemmatisert og påført bøyninginformasjon. Det finnes ingen utviklet morfologisk analysator (tagger) for norrønt. Som et viktig ledd i å sikre god funksjonalitet trenger korpuset å ha en formtesaurus koblet mot eksisterende ordbøker for norrønt. Denne formtesauren eller metaordboken vil være ryggraden i sammenkoblingen av de leksikalske enkeltdatabasene. Den vil også kunne danne forbindelsen til de moderne ordboksverkene for islandsk og

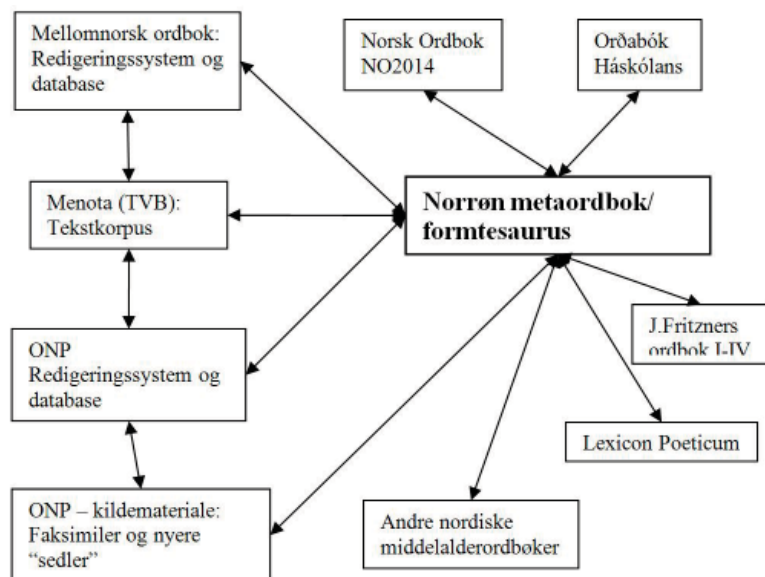
norsk, *Orðabók Háskólans* og *Norsk Ordbok*, som beskriver språkene fra etterreformatorisk tid til i dag. Vi mener at et slikt rammeverk kan bygges over samme lest som Metaordboksystemet for Norsk Ordbok. Metaordboken er et velprøvd verktøy som har blitt brukt av NO2014 de siste 6 årene. Det er kanskje ikke så velkjent, så det er naturlig å bruke litt plass på å forklare bakgrunnen og hva det går ut på.

Nynorsk leksikografi var på 1990-tallet i samme situasjon som norrøn leksikografi med det unntak at norrøn leksikografi er spredt over flere land og prosjekter. I den nettbaserte virkeligheten spiller det siste en mer organisatorisk enn praktisk rolle. Som et ledd i revitaliseringen av de store norske ordboksprosjektene gjennomførte en på 1990-tallet en storstilt digitalisering av tekstsamlinger, kortarkiv og ordbøker, og det ble lagd systemer for morfologisk merking av løpende tekst. Dette arbeidet har siden fortsatt. Det revitaliserte Norsk Ordbok prosjektet, NO2014, har nå også et tekstkorpus på mer enn 30 millioner ord. I løpet av siste halvpart av 1990-tallet ble det utviklet en modell for hvordan disse ressursene, både det gamle materialet og det nye elektroniske tekstkorpuset, kunne bindes sammen i et rammeverk for leksikografisk arbeid for å understøtte redigeringsarbeidet, men også for å gjøre materialet tilgjengelig for andre forskere og den interesserte allmennheten. Dette rammeverket ble kalt en *metaordbok* og grunnideene er beskrevet i Ore 1999.

Metaordboken er et verktøy for å samle belegg og normere svakt normerte språk. Hver artikkel i Metaordboken er en samling referanser til kilder som beskriver det aktuelle ordet eller gir brukseksempler. Hver artikkel kan ha et eller flere normerte oppslagsord med tilhørende bøyninginformasjon. Disse oppslagsordene kan være konstruert i henhold til en gitt rettskrivningsnorm. De brukes som bindeledd til de egentlige ordboksartiklene i Norsk Ordbok. For norrønt vil en kunne ha oppslagsord til meta-artiklene både i henhold til GNOs standard og til ONPs.

Fullt utbygd og korrekt anvendt skal Metaordboken til NO2014 avspeile prosessen med å skape ordboksartikler helt fra ubehandlede samlinger, via omstruktureringer og systematiseringer til en eksplisitt beskrivelse av hva de rå samlingsdataene betyr, se figur 1. I prinsippet skal man ha tilgang til alt materiale som er brukt i redigeringsprosessen, også det som ikke er kommet med i ordboksartikkelen. På den måten vil man komme et steg videre med å gjøre en ordbok vitenskapelig etterprøvable.

Når vi nå har sett på mulighetene av å lage en metaordbok for norrønt, er det uten intensjoner om å lage noen felles nordisk middelalderordbok. Hensikten er å få til et system for å koble sammen alle elektroniske tekster og leksikalske ressurser. Den norrøne metaordboken er ment som et hjelpemiddel både for de norrøne ordboksprosjektene og for tekstdelen av Menota-samarbeidet, se figur 2. Den vil også være en måte for elektroniske versjoner av de to store ordbøkene for nyere tid, *Orðabók Háskólans* og *Norsk Ordbok*, å hekte seg på til det førreformatoriske materialet.



Figur 2: Noen mulige forbindelser gjennom en norrøn metaordbok

### 3 Grunnlagsmaterialet

Metaordboken kan i sin enkleste form sees som en samindeksering av en rekke kilder og belegg. Vi har foreløpig begrenset oss med visse unntak til denne enkle formen og kun sett på ord som begynner på bokstaven *B* for å finne ut mer om arbeidsmengde og problemer. Kildene vi ønsker å koble er hovedordbøkene for norrønt. Det vil si ONP, J. Fritzners ordbok og F. Jónssons *Lexicon Poeticum*. For å få til en kobling med korpusmaterialet har vi også tatt med en liste over ordforrådet i fem større tekster<sup>1</sup> og 200 norskspråklige diplomer fra før 1308. De elektroniske tekstene er i xml-format og lemmatiseringen er basert på GNOs ordseddelsamlinger. Tekstene svarer til om lag 420 000 løpende ord.

Tabellen i figur 3 gir en oversikt over forholdet mellom løpende ord (tokens), forskjellige ordformer (types på formnivå) og grunnord (types) på lemmanivå i de fem tekstene og diplomene. I snitt utgjør lemma på bokstaven *B* om lag 6,2 prosent av hele lemmautvalget, mens på token-nivå ligger *B*-ordformer nede på 3,3 prosent. Det skyldes antagelig at det ikke finnes så mange høyfrekvente funksjonsord på bokstaven *B*. Men det er verdt å legge merke til at det relative antallet oppslagsord på bokstaven *B* i de tre ordbøkene er i godt samsvar med den faktiske fordelingen i tekstutvalget. For ONP må spesielt totaltallene tas med en klype salt. De er hentet fra nettordlisten. Tallene i raden markert med "hovedord" er

<sup>1</sup> Konungs skuggsjá, AM 243 b α fol; Óláfs saga ins helga, Upps DG 8; Strengleikar, Upps DG 4–7; Þiðriks saga af Bern, Holm perg 4 fol; Thómass saga erkibiskups, Holm perg 17 4to.

de ord som ikke er markert som poetisk, navn, fremmedord, eller yngre islandsk eller norsk. De aller fleste oppslagsordene i *Lexicon Poeticum* er derfor ikke med i ONPs hovedordliste slik den fremstår på nettet. Tilsvarende er ikke navn eller latinske ord i tekstene med der. Yngre norske og islandske ord i Fritznors ordbok er heller ikke med i ONPs hovedordliste.

Tekst	Forskjellige grunnord			Forskjellige ordformer			Løpende ordformer		
	B	alle	%	B	alle	%	B	alle	%
Thóm.	186	3 709	5,0	699	14 284	4,9	2 430	64 380	3,8
Kgs.	280	4 116	6,8	951	14 194	6,7	1 780	56 226	3,2
Þiðr.	235	4 210	5,6	1 305	24 334	5,4	4 139	147 766	2,8
Str.	147	2 597	5,7	487	9 889	4,9	1 100	42 085	2,6
Ólaf.	183	3 234	5,7	603	9 738	6,2	1 218	41 060	3,0
Dipl.	312	4 806	6,5	1 235	20 111	6,1	3 050	66 175	4,6
Alle	<b>841</b>	<b>13 598</b>	<b>6,2</b>	<b>4 544</b>	<b>75 567</b>	<b>6,0</b>	<b>13 717</b>	<b>417 692</b>	<b>3,3</b>

Ordbok	B	Hele	%
Fritzner, artikler	2 400	38 500	6,2 %
Lexicon Poeticum, artikler	1 150	1 8000	6,4 %
ONP, ordliste, hovedord	3 900	72 000	5,4 %
ONP, ordliste, alle ord (antall artikler a-em)	5 600	91 700	6,1 %

Figur 3: Lemma- og ordformfrekvenser i de ulike tekstene og i de to hovedordbøkene for norrønt.

#### 4 En tre-trinns prosess for samkjøring av ordlister

Et konkret mål for vår øvelse var å samkjøre de fire ordlistene slik at like lemma i hver av listene ble koblet sammen i en formtesaurus. For å gjennomføre dette satte vi opp følgende tre-trinns prosess:

1. felles normalisering
2. valg av oppslagsform og grammatisk klassifisering
3. metanivå, en videre sammenslåing av relaterte ordformer

##### 4.1 Trinn 1: normalisering

Normalisering innenfor norrønt er konstruert i nyere tid etter mønster av islandsk ortografi. Det finnes flere normaliseringsvarianter. Ingen av de fire ordlistene våre følger nøyaktig samme prinsipper for normalisering. En oversikt over vesentlige forskjeller gis i figur 4.

Det kan være snakk om rent ortografiske ulikheter, som at ONP gjengir lang 'ø' med aksent for markering av vokallengde, mens de andre listene bruker 'œ'-ligatur. Ulik normaliseringspraksis kan også gjenspeile geografisk variasjon, som



når GNOs ordformer viser gammelnorske trekk ved at 'h' ikke står foran 'l', 'n' og 'r' i framlyd.

Fenomen	ONP	GNO	Fritzner	LP
æ	ǣ	æ	æ	æ
langø	ø	œ	œ	œ
kort ø	ø	ø	ö	ø
o med kvist	o	o	ö	o
lango med kvist	á	á	á	ȝ
øy/ey	ey	øy	ey	ey
konsonantisk i	j	i	j	j
pt/ft	ft	pt / (ft)	pt / (ft)	pt
langK + ny K	ggð, lld	gð, ld	gð, ld	ggð, ld
hl, hn, hr	hl, hn, hr	l, n, r	hl, hn, hr	hl, hn, hr

Figur 4: Normaliseringsforskjeller mellom ordlistene

I tråd med Menotas prinsipper valgte vi å følge ONPs standard for normalisering. Mye av samnormaliseringen kunne utføres ved kontrollerte søk-erstatt rutiner. Men enkelte ulikheter mellom listene måtte lukes ut manuelt. Det siste gjaldt f.eks. for ordlistenes ulike realisering av vokalen 'o med kvist'. Dette var en lav, bakre, rundet vokal som oppstod ved overgangen fra urnordisk til norrønt. Den ble ikke værende i språket. I islandsk falt den etterhvert sammen med kort 'ø' og i norsk hovedsakelig med kort 'o'. ONP og LP bruker et entydig spesialtegn for 'o med kvist'. Fritzner følger nyere islandsk ortografi med bokstaven <ö>, mens GNO bruker bokstaven <o>. Fritzner har dermed sammenfall mellom etymologisk kort 'ø' og kort 'o med kvist', mens GNO har sammenfall mellom etymologisk kort 'o' og kort 'o med kvist'. Denne typen variasjon i listene måtte lukes ut manuelt og føres over til ONPs standard.

Til hjelp i arbeidet med å identifisere avvikende ordformer, foretok vi jevnlig flettinger av ordlistene og kjørte ut *diskrepanslister* over de ordformene i Fritzner, GNO og LP som ikke falt sammen med noen ordform i ONP. Ved fletting fulgte vi følgende prinsipper: GNO-listen er ikke homografseparert, så vi skiller ikke mellom homografer av samme ordklasse. I ordlistene er forledd og etterledd egne innførslers, så vi har valgt å skille mellom oppslagsord markert som selvstendige ord og oppslagsord markert som for- eller etterledd. Former som *bakki* m. og *-bakki* m. ble dermed ikke koblet sammen.

Ved samkjøring av ordformer beholdt vi de ulike listenes opprinnelige oppslagsform. For hvert trinn i prosessen, ble det i hver liste lagt til et ekstra felt for lemma og grammatisk kategori, og alle endringer på trinnet ble foretatt i de nye feltene. Figur 5 viser trinn-1 felt etter samnormalisering av ordet *banahogg*. Listenes opprinnelige form står i feltene til høyre.<sup>2</sup>

<sup>2</sup> For feltet 'Gram-1' i tabellen i figur 5 har vi ikke fulgt ONPs notasjon. Vi vil etterhvert følge Menotas grammatiske notasjon som baserer seg på anbefalinger fra Eagles (<http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>). I tabellen er det brukt en midlertidig arbeidsnotasjon.

Liste_ID	Trinn-1	Gram-1	Oppslagsord	Gram
GNO	banahøgg	n	banahøgg	n
ONP	banahøgg	n	banahøgg	sb. n
Fritzner	banahøgg	n	banahögg	n

Figur 5: Trinn-1 felt for banahøgg med 'o med kvist' ført inn.

Etter fullført samnormalisering, trinn 1, kjørte vi ut nye diskrepanslister og regnet ut avviksprosent for hver av listene i forhold til ONP. Av Fritzners totalt ca. 2400 ord på B, kom det 180 ord på diskrepanslisten, hvilket tilsvarer avvik på 7,5 %. For LP og GNO var avviket på henholdsvis 5 % og 16 %. Vi kom dermed ganske langt bare ved å samnormalisere listene.

#### 4.2 Trinn 2: valg av oppslagsform og grammatisk klassifisering

##### 4.2.1 Ulike prinsipper for valg av oppslagsform

De ordene vi fikk ut på diskrepanslisten etter trinn 1, falt i en del klart definerbare kategorier som igjen har sammenheng med ulike prinsipper for valg av oppslagsform. De mest frekvente av disse er vist i figur 6. Eksempelene her er hentet fra Fritzners avvik i forhold til ONP, men de samme kategoriene går igjen også i de andre listene.

Fritzner		ONP		
Substantiv nom. pl.	<i>barir</i> <i>bróðrasynir</i>	Substantiv nom. sg.	<i>bör</i> <i>bróðrasonr</i>	'(lik)bære' 'brødresønner'
Verb refleksiv	<i>bakverpast</i>  <i>baðast</i>	Verb infinitiv	<i>bakverpa</i>  <i>baða</i>	'forkaste - vende ryggen til' 'bade'
Adverb	<i>berbakt</i> <i>bátlaust</i>	Adjektiv	<i>berbakr</i> <i>bátlauss</i>	'med bar rygg' 'uten båt'
Adjektiv (=perf.part.)	<i>borinn</i>	Verb infinitiv	<i>bera</i>	'bære'

Figur 6: Ulike prinsipper for valg av oppslagsform

Fritzner har ført opp flertallsformen av substantiv som han i kildematerialet utelukkende har funnet belagt i flertall, mens ONP har ført opp entallsformen. Enten har ONP funnet belegg for entallsformen, hvilket vil framgå av ordboksartikkelen, eller også er entallsformen konstruert. Fritzner har likeledes ført opp refleksivform av verb han ikke har funnet belagt i aktiv, mens ONP har ført opp infinitiven av verbet, enten de har funnet belegg for bruk i aktiv eller ikke. De to verbene i figur 6 er forøvrig begge funnet belagt i aktiv. Tilsvarende har Fritzner ført opp adverb der han ikke har adjektivformen belagt, mens ONP har ført opp adjektivet, enten det er en reell eller konstruert form.

Det er klart det kan diskuteres hvorvidt man skal konstruere oppslagsformer

som kanskje aldri har eksistert. Men vi hadde som overordnet prinsipp å følge ONP, og da ble det å identifisere denne typen diskrepans og føre den over til ONPs valg av oppslagsform, en relativt enkel oppgave.

#### 4.2.2 Adjektiviske partisipper

Et eksempel på en mer problematisk samkjøring av ordformer på dette trinnet, er adjektiviske partisipper. I ordlistene finner vi blant annet følgende oppslagsformer med utgangspunkt i verbet 'bære':

Fritzner:	<i>bera</i> v.	<i>borinn</i> adj.	<i>arfborinn</i> adj.
ONP:	<i>bera</i> v.	<i>-borinn</i> adj.	<i>arfborinn</i> adj.

Hos Fritzner er slike partisippformer ført opp som selvstendige lemma og klassifisert som adjektiv. I ONP er de samme partisippformene behandlet under artikkelen til verbet. Ut fra vårt prinsipp om å følge ONPs valg av oppslagsform, burde vi føre Fritzners adjektivpartisipper til verbets infinitiv, dvs. *borinn* adj. føres til lemma *bera* v. Men dersom partisippformen inngår i en fast sammensetning i norrønt, har både ONP og Fritzner klassifisert sammensetningen som adjektiv, f.eks. *arfborinn* adj, 'arveberettiget'. I ONP er etterleddet i tillegg ført opp som adjektiv, med henvisning til verbet. Siden vi har valgt å skille mellom selvstendige ord og ledd, førte dette til følgende to lemma på trinn 2, med tilknyttede ordformer:

Lemma 1, <i>bera</i> v.:	<i>bera</i> v. (ONP)
	<i>bera</i> v. (Fri)
	<i>borinn</i> adj. (Fri)
Lemma 2, <i>-borinn</i> adj.:	<i>-borinn</i> adj. (ONP)

En kompliserende faktor her er at Fritzner ikke nødvendigvis opplyser at en oppslagsform utelukkende opptrer som del av en sammensetning. Vi kan derfor gå ut fra at flere av de adjektiviske partisippene i Fritzner i realiteten er ledd og dermed identisk med adjektivleddene i ONP av typen *-borinn* adj. Men det vil neppe gjelde alle partisippformene i Fritzner, og det virket lite gunstig å behandle adjektivpartisippene ulikt på dette stadiet i prosessen. Derfor valgte vi på trinn 2 å føre alle slike partisipper hos Fritzner til verbets infinitiv.

#### 4.2.3 Oppsummering trinn 2

Diskrepanslistene vi kjørte ut avslutningsvis på trinn 2, viste et ganske oppløftende resultat, særlig for Fritzners vedkommende. Det var nå bare 0,6 % av ordformene i Fritzner som vi ikke hadde klart å føre over til en form i ONP. GNO hadde en noe høyere avviksprosent på 12 %. Dette skyldes hovedsakelig at GNO-listen inneholder mange navn og navneledd, og slike former er ikke tatt med i ONP.

### 4.3 Trinn 3, en videre sammenslåing av relaterte ordformer

#### 4.3.1 Relaterte ordformer

Etter fullført trinn 2 har vi etablert en formtesaurus som tilsvarer ONPs lemmaliste pluss særformer fra de øvrige listene. ONP er den mest omfangsrike av de norrøne ordlistene, blant annet fordi den har med langt flere variantformer i form av nært relaterte ordformer der de andre listene økonomiserer ordutvalget, se figur 7. En slik ordrikdom er positiv sett fra en bruker av ordbokens side. Når man skal slå opp et ord man støter på i en norrøn tekst, vil man stort sett finne det i ONP. Men denne overflod av nært relaterte ordformer er ikke nødvendigvis like gunstig i en metaordbok, f.eks. med tanke på søk i korpus, eller generering av statistikker over ordforekomster. Dersom man søker etter *barnafaðir*, kan det være av interesse også å få informasjon om relaterte former som *barnfaðir* og *barnsfaðir*. Den ene formen bør ikke utelukke de to andre. Det fører oss til trinn 3 i samkjøringsprosessen, hvor vi går ut over ONPs lemmaliste og slår sammen nært relaterte ord i metalemma, som de tre sammensetningene for 'barnefar' i figur 7. Disse sammensetningene er behandlet i separate oppslag i ONP. Fritzner har kun formen med nullfuge, *barnfaðir*. Ordet forekommer ikke i GNOs materiale eller i LP.

Fenomen	ONP	Fritzner
Substantiv, sterk og svak form	<i>bjartleikr</i> sterk m. <i>bjartleiki</i> svak m.  <i>bjarnfell</i> sterk n. <i>bjarnfelli</i> svak n.	<i>bjartleikr</i> sterk m. = 'lys, glans'  <i>bjarnfell</i> sterk n. = 'bjørneskinn'
Substantiv med variasjon i grammatisk kjønn	<i>bið</i> f. <i>bið</i> n.	<i>bið</i> n.pl. = 'venting, ventetid'
Svake adjektiv på -a og -i	<i>blóðrisa</i> <i>blóðrisi</i>	<i>Blóðrisa</i> = 'blodsprengt'
Sammensetninger med morfologisk variasjon i 1. ledd	<i>barnfaðir</i> m. <i>barnafaðir</i> m. <i>barnsfaðir</i> m.	<i>barnfaðir</i> m. = 'barnefar'

Figur 7: Eksempler på typer av variasjon i ONP utelatt hos Fritzner

Vi trengte imidlertid kjøreregler for en slik grovere sammenslåing av ordformer. Vi var i den heldige situasjon for bokstaven B at ONP allerede har behandlet den. I ONPs artikler henvises det alltid mellom beslektede ordformer. ONP er dessuten digitalisert og gjort tilgjengelig via internett. Vi kunne derfor hente disse krysshenviisningene fra nettutgaven og generere lister over relaterte former. Men siden ONP ikke skiller mellom henvisninger som angir semantisk slektskap og de som angir formvarians, måtte det en manuell gjennomgang til.

Valg av metalemma har vært gjenstand for en del diskusjon, og vi har til dels også eksperimentert med mulige tilnæringsmåter. Vi kunne velge den etymologisk mest korrekte formen, eller den mest frekvente formen, eller en bestemt fuge i sammensetninger. Men disse framgangsmåtene var ofte svært tidkrevende og passet som regel aldri for alle typer av variasjon mellom relaterte ordformer. I dette prøveprosjektet var det også et poeng å få dataene samkjørt relativt raskt slik at vi kunne få et overblikk over materialet. Derfor valgte vi en enklere og kanskje noe uortodoks metode: Metalemma er den formen som kommer først alfabetisk. Dette er en metode som lar seg klart definere, den er etterprøvbar og kan utøves på alle ordformene. Det må også understrekes at noe av ideen i metaordboken er at artiklene kan gis flere metaoppslag, det vil si kan indekseres etter flere ordlister.

#### 4.3.2 Sammensetninger med variasjon i fugen

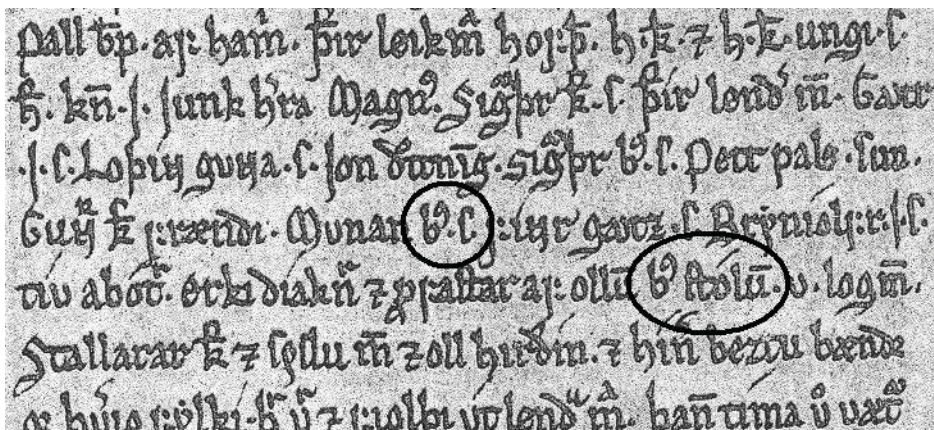
For sammensetningers vedkommende vil den alfabetiske tilnæringsmåten føre til at form med g.pl. i førsteleddet, såkalt a-fuge, dersom den er belagt, blir valgt som metalemma:

metalemma:	<i>barnafaðir</i>		
relaterte former:	<i>barnafaðir,</i>	<i>barnfaðir,</i>	<i>barnsfaðir</i>

I en ferdig metaordbok vil ordlistenes opprinnelige ordformer, samt alle samkjøringsfeltene fra trinn 1 og trinn 2, være søkbare, ikke bare det normaliserte metalemma. Det innebærer at et søk på en form som *barnfaðir*, med nullfuge, vil gi direkte treff på Fritznere og ONPs artikler med denne formen, men også gi tilgang til de to alternative formene i ONP, med a-fuge og s-fuge.

Det faktum at Fritznere i stor grad slår sammen denne typen ordformer, er et argument for at det er forsvarlig også for oss å gjøre det i metaordboken. Et annet argument i samme retning er at ordbøkene, særlig de eldre, dels ekserperer sine ordbelegg fra trykte tekstutgaver av håndskriftene. I middelalderen praktiserte man en utstrakt bruk av forkortede skrivemåter. I trykte utgaver av håndskriftene vil utgiver ofte løse opp forkortelsene. I nyere diplomatariske, dvs. bokstaverette, tekstutgaver er slike oppløste forkortelser som regel tydelig markert. I eldre diplomatariske tekstutgaver er ikke dette nødvendigvis tilfelle, og man kan som bruker av teksten ikke vite hva som er originalens skrivemåte og hva som står for utgivers regning. Som et eksempel kan vi se på følgende skrivemåter for *byskup(s)son* og *byskup(s)stólum* i håndskriftet *Frisbók*<sup>3</sup> fra begynnelsen av 1300-tallet, figur nr. 8.

<sup>3</sup> AM 45 fol, Codex Frisianus, ca 1300–1325, her gjengitt etter Halldór Hermannsson 1932.



Figur 8: Frisbók s. 123v spalte 452l. 12 *byskvp(s) son* og l. 13 *byskvp(s) stolum*. Understreking markerer oppløste forkortelser.

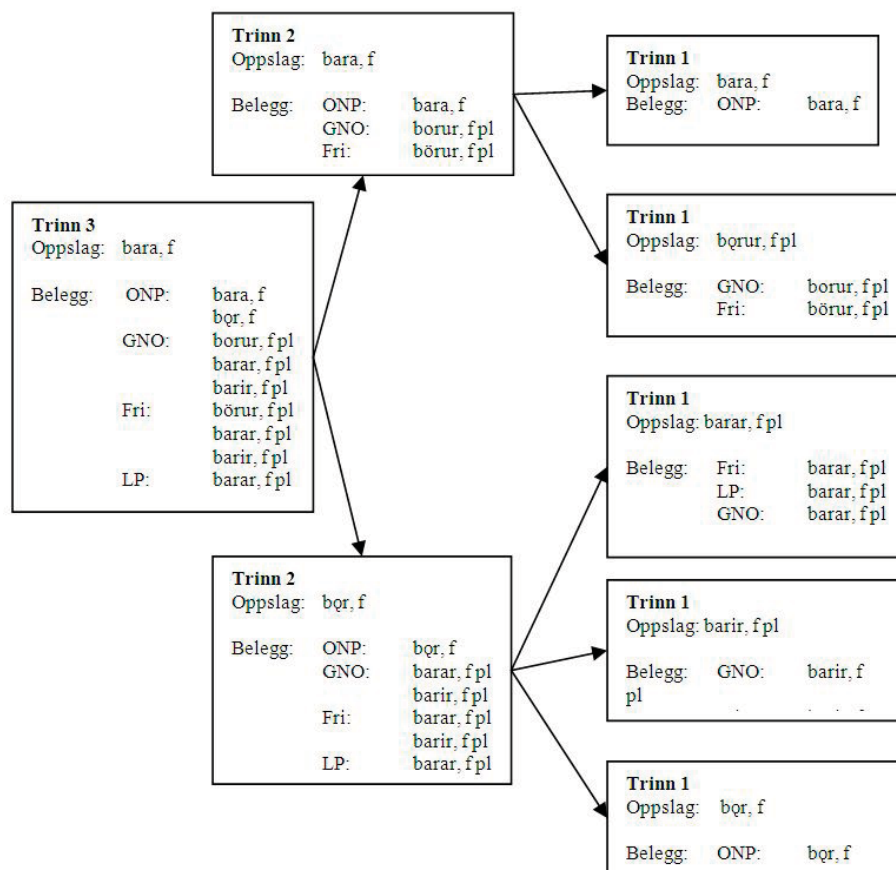
Teksten er her såpass forkortet at vi ikke ut fra skrivemåten kan se hvilken fuge skriveren mon hadde i tankene. I tekstutgaven av Frisbók, Unger 1871, er disse forkortelsene oppløst uten markering som henholdsvis *byskvpsson*, med s-fuge, og *byskvpstolum*, med nullfuge. Utgiver følger det han oppfatter som håndskriftets norm ved oppløsning av forkortelser. ONP er med alle sine ordformer en deskriptiv ordbok, mens Fritzner kan sies å være en mer normativ ordbok. Men grunnlaget for beleggene i ONPs deskriptive samling av ordformer er i enkelte tilfeller allerede normert i og med utgivers valg av oppløsning. Metaordboken viderefører denne normeringstendensen i den grovere sammenslåingen av ordformer i trinn 3.

#### 4.3.3 Metalemma og relaterte former; et eksempel

I det norrøne kildematerialet finner vi flere relaterte ordformer for det som tilsvarer vårt '(lik)bære'. Begrepet forekommer hovedsakelig i flertall, med tre belagte flertallsformer, *barar*, *barir* og *børur*. Disse finner vi som tre separate innførsler i GNO og Fritzner, mens LP bare har formen *barar*. I ONP er de to første formene, *barar* og *barir*, begge ført til en konstruert entallsform *bør f*, mens den siste flertallsformen, *børur*, er ført til en belagt entallsform *bara f*. Dette viser at ONPs tendens til å ta med et stort antall relaterte ordformer til en viss grad dempes noe av ordbokens strengere regler for valg av oppslagsform.

ONPs to oppslagsord er en svak og en sterk form av et feminint substantiv. Det er flere paradigmer knyttet til den sterke formen (nom. pl. *barir* og *barar*), og det er også sammenfall i enkelte bøyningsformer mellom sterk og svak form (f.eks. dat.pl. *børur*.) ONP krysshensier mellom de to oppslagsordene. På et metanivå kan vi se ONPs to oppslagsord som varianter av samme lemma, med ulike paradigmer knyttet til seg. På trinn 3 blir alle de ni innførslene for 'bære' ført sammen i én metaordboksartikkel, se figur 9. I metaordboken vil hver ordform

på hvert trinn i prosessen være søkbar og gi tilgang til alle formene tilknyttet metaartikkelen.



Figur 9: Metalemma bara f. og relaterte ordformer i listene.

## 5 Noen betraktninger om linjene fram til de moderne ordbøkene

Den norrøne metaordboken skal være en felles inngang til ordbøkene over islandsk og norsk i norrøn tid ved å gruppere sammen ulike skriftrealiseringer av det en ønsker å oppfatte som ord eller lemma. Norrøn metaordbok vil gjennom de ulike ordbøkene koble sammen ordbelegg over en periode på rundt 500 år. Til sammenligning dekker *Norsk Ordbok* norske dialekter i drøye 300 år og *Orða-bók Háskólans* islandsk i 460 år. Det er lange spenn, men ikke oppsiktsvekkende lange. *Oxford English Dictionary* dekker engelsk over en 1000 års periode.

For de norske historisk-orienterte leksikografiske miljøene er en norrøn metaordbok interessant fordi den vil danne grunnlag for arbeidet med en Mellomnorsk ordbok for perioden 1370–1580 og kunne brukes til å lage en forbindelse

mellom *Norsk Ordboks* beskrivelse av tidlig-moderne og moderne norsk. Det er i så fall et spørsmål hvordan en slik forbindelse skal etableres. I de etablerte norske ordbøkene er det relativt få ord som er markert med gammelnorsk eller gammelislandsk opphav. A. Torps nynorske etymologiske ordbok behandler om lag 15 000 stort sett usammensatte ord. Bare 5 000 av dem har henvisning til norrønt eller gammelislandsk. For bokstaven B har de tre nynorske ordbøkene, *Norsk Ordbok*, A. Torps etymologiske ordbok og *Nynorskordboka*, tilsammen drøye 900 ord med slik henvisning. *Norsk Ordbok* har 18 400 oppslagsord på bokstaven B, mens håndordboken *Nynorskordboka* har 7200. Det er pussig om bare fem prosent av nynorske ord skulle ha norrønt opphav. Motsatt vei er det maksimalt 20 prosent av oppslagsordene i ONP og maksimalt 40 % av oppslagsordene i Fritzner som er referert fra de moderne norske ordbøkene. Dette er kanskje mer naturlig, men er verd et nærmere studium. Denne raske kvantitative testen sier oss at en etter all sannsynlighet må utføre et omfattende leksikografisk arbeid for å få på plass en ordentlig kobling mellom *Norsk Ordbok* og det norrøne materialet. Det foreligger ikke nok ordhistorisk informasjon i de eksisterende ordbøkene til å gjøre en særlig god mekanisk kobling.

## 6 Oppsummering og videre arbeid

Både den alfabetiske tilnæringsmåten og bruk av krysshenvisingene fra ONP, er nyttige hjelpemidler i arbeidet med å føre sammen relaterte ordformer på trinn 3, men de er ikke absolutte prinsipper. De genererte listene over krysshenvisinger bør derfor brukes til manuell gjennomgang av de i ONP relaterte formene.

Den alfabetisk utvalgte metalemmalisten kan tilsvarende sees som et utgangspunkt for videre bearbeiding. I de digitaliserte ordlistene er hovedform og sideform splittet i separate innførsler for å muliggjøre søk på alle former. Ved femininer på *-an* og *-un* (eks: *skipan*, *skipun*) fører ONP formen med u-vokalisme i endelsen som hovedform, mens den alfabetiske tilnæringsmåten vil velge ut formen med a-vokalisme. En manuell gjennomgang av en alfabetisk generert metalemmaliste er derfor både ønskelig og nødvendig. Men i og med at alle ordformer på alle trinn i samkjøringsprosessen vil være søkbare, er valg av metalemma mindre viktig enn valg av former knyttet til dette metalemma. Det vesentlige er at både *skipan* og *skipun* knyttes sammen, ikke hva vi velger å kalle enheten på toppen. Gjennomgang og redigering av den alfabetisk genererte metalemmalisten kan derfor gjøres på et senere stadium.

Sammenslåingen i trinn 3 har i vårt prøveprosjekt basert seg på informasjon som finnes i ONPs ferdigredigerte artikler på nettet. Disse finnes så langt ONP er kommet med artikkelredigeringen, altså til midten av bokstaven E. For resten av alfabetet er der ingen krysshenvisinger å hente ut fra nettversjonen. I ONPs database for resten av alfabetet er slike krysshenvisinger bare satt inn sporadisk. Det vil si at her må materialet gjennomgås manuelt for å identifisere tilsvarende



relasjoner mellom ord, noe som vil bli en langt større og mer ressurskrevende jobb enn vi har utført i dette prøveprosjektet. Det vil også være en forutsetning at dette kan gjøres i samarbeid med ONPs redaksjon.

## Litteraturliste og nettsteder

- Halldór Hermannsson 1932 (ed. facs.): Codex Frisianus (Sagas of the Kings of Norway): MS. No. 45 Fol. in the Arnamagnæan Collection in the University Library of Copenhagen, *Corpus codicum Islandicorum medii aevi* 4, Copenhagen 1932.
- Jónsson, Finnur 1931: Lexicon Poeticum antiquæ linguæ septentrionalis, Ordbog over det norsk-islandske skjaldesprog oprindelig forfattet av Sveinbjörn Egilsson, S.L.Møllers bogtrykkeri, København, 1931.
- Fritzner, Johan 1886–1896: Ordbog over det gamle norske Sprog I–III, 4. utgave Universitetsforlaget, Oslo, Bergen Tromsø. 1972, ([www.dokpro.uio.no](http://www.dokpro.uio.no), sjekket 270907).
- Hødnebo, Finn 1972: Rettelser og tillegg til Johan Fritzner, Ordbog over det gamle norske Sprog, Universitetsforlaget, 1972, Oslo.
- Menota – Medieval Nordic Text Archive ([www.menota.org](http://www.menota.org), sjekket 270907)
- Norrøn metaordbok [www.edd.uio.no/menota/metaordbok.html](http://www.edd.uio.no/menota/metaordbok.html)
- Nynorskorboka, tredje utgave, Det norske samlaget, Oslo 2001.
- Norsk Ordbok 1–6, Det norske samlaget, Oslo 1950–2007, ([www.no2014.uio.no](http://www.no2014.uio.no) (sjekket 270907)).
- Ordbog over det norrøne prosasprog, 1–3, København, [www.onp.kua.dk](http://www.onp.kua.dk) (sjekket 270907).*
- Orðabók Háskólans ([www.lexis.hi.is/indexny.html](http://www.lexis.hi.is/indexny.html), sjekket 270907).
- Ore, Christian-Emil 1999: *Metaordboken – et rammeverk for Norsk ordbok?* Nordiske studier i leksikografi 6, NFL, Göteborg, 1999.
- Torp, Alf 1919: Nynorsk etymologisk ordbok, Kristiania 1919.
- Unger, Carl Richard 1871 (ed.): Codex Frisianus: En Samling af norske Konge-Sagaer, Christiania, 1871.
- Font brukt for norrøne tegn: *Andron Scriptor Web*, designet av Andreas Stötzner, Leipzig, og distribuert av Medieval Unicode Font Initiative (Mufi) [www.mufi.info/fonts](http://www.mufi.info/fonts) (sjekket 270907).

Mette Gismerøy Ekker  
f. 1963  
Kjeldeskriftavdelingen, Norsk Riksarkiv  
[mettegek@ulrik.uio.no](mailto:mettegek@ulrik.uio.no)

Christian-Emil Ore  
førsteamanuensis, f. 1957  
Institutt for lingvistiske og nordiske studier/EDD  
Humanistisk fakultet  
Universitet i Oslo, boks 1123 Blindern  
NO-0317 Oslo  
[c.e.s.ore@edd.uio.no](mailto:c.e.s.ore@edd.uio.no)



ELI JOHANNE ELLINGSVE

## Digitale vegar til Norsk Ordbok

In 2014, *Norsk Ordbok* will be complete in 12 volumes. At a time when more and more people are getting knowledge and information via the Internet it is natural to consider developing a digital version of the dictionary as well. A digital version will present the dictionary makers with opportunities as well as challenges. The digital version will make it possible to add audiovisual supplements to the traditional written presentation of entries. Coordination and possible synergy effects in relation to other digital media and the Internet will involve finding technological solutions and a strategic location. The lecture will consider two possible ways to address the location on the web. Linguistic adaptation and strategic thinking related to the use of New Norwegian are also elements which should be considered along the route towards a digital dictionary.

### 1 Innleiing

I 2014 skal *Norsk Ordbok* liggje føre, trykt i 12 band. Det knyter seg ei viss spenning til om verket vil vere ferdig til rett tid. Det knyter seg òg spenning til *om* og *korleis* ordboksverket vil bli nytta etter den tid.

Eit storverk som *Norsk Ordbok* (NO) fortener brei utnytting. Det kan etter mitt syn best skje i ei digital utgåve, via Internett og verdensveven. For det første har digitale medier bruksmulegheiter og ressursar som er unike og særigne i høve til andre medier. For det andre er Internettet døgnope, og geografisk grenselaust. For det tredje er Internettet i ferd med å bli den viktigaste kjelda for informasjon.

NO er ingen pioner på nettfeltet. Fleire, kanskje dei fleste, andre større ordbøker som det er naturleg å samanlikne seg med, ligg alt på nettet (i ei eller anna form) eller er på veg ut på det. Det same gjeld andre store datasamlingar, private så vel som offentlege. Og sjølvsagt er NO alt digitalisert – både kjelder, redigeringsverktøy og arbeidsmetodar. Det er alt open tilgang til ordsamlingane attom verket. Utfordringa blir å gjere den digitale ordboka attraktiv og brukarvenleg, slik at mange menneske tek ho i bruk.

Eg ser to hovudoppgåver i arbeidet med å gjere NO til eit attraktivt digitalt ordboksverk. Den eine oppgåva inneber eigenutvikling av ordboksdatabasen. Den andre oppgåva inneber utvikling i samspel med andre digitale databasar. Eg skal kome stutt inn på nokre punkt under eigenutvikling, og så bruke noko meir tid på samspel og felles løysingar.

## 2 Eigenutvikling av NO

Med eigenutvikling av NO meiner eg ei utvikling som gjer NO i seg sjøl lett tilgjengeleg og attraktiv for alle brukarar som vil ha informasjon om ord i norsk språk. Eigenutvikling gjeld både tilgang, form og innhald. Utvikling av digitalversjonen av NO har skote monaleg fart sida årtusenskiftet. Gevinsten til no har likevel lege meir på avsendarsida enn på mottakarsida. Det er fleire nyvinningar for redaktørane enn for brukarane så langt. Men i alle fall – vi er undervegs mot nettet.

### 2.1 Tilgang

Innleiingsvis skal stutt eit par ting nemnast om tilgang – søkjeprosedyrar og rettskriving.

Søkjeprosedyrane bør vere enkle. Brukargrensesnitta bør vere intuitive, beint fram ”menneskelege” om ein kan nytte eit slikt ord. Svara bør helst ikkje vere meir enn 2–3 tastetrykk borte for ein vanleg internettbukar. Det er eit mål eg trur det er klokt å prioritere både høgt og snøgt.

Rettskrivinga i NO fylgjer i prinsippet den nynorske rettskrivingsnormalen frå 1938. Så vel målforma som rettskrivingsnormalen kan vere eit problem for tilgangen til og bruken av boka. I røynda har eit stort fleirtal av nordmenn bokmålet som si målform, og nyttar intuitivt denne målforma når trongen for å slå opp i ei ordbok melder seg. Og jamvel om nynorskbrukarar har ein føremon framfor bokmålsbrukarar her, vil ein nynorsk rettskrivingsnormal frå 1938 kunne hindre bruken også for nynorskbrukarane. Løysinga kan vere å leggje inn ei skjult metaordbok mellom søkjesida og sjølve ordboka, med alle ord i moderne nynorsk- og bokmålsform, såleis at ein får tilslag i alle høve, om ein berre kan skrive rett moderne norsk. Større språkkrav bør ein ikkje stille til brukarane.

### 2.2 Innhald

Den norske kulturministeren Trond Giske sa i ein tale i mars 2007 dette om kultur: ”Det viktigste med kultur er ikke at den skal forklares, eller defineres. Kultur skal oppleves.” Dette er ikkje noko sidespark til *Norsk Ordbok*, og alle andre ordboksverk, som har den verbale ordforklaringa eller definisjonen som fremste verkemiddelet i arbeidet med å spreie kunnskapen om ord og kultur. Det kan likevel vere ei påminning om at det er råd å oppleve noko gjennom andre verkemiddel enn ord, og andre sansar enn intellektet. Eg vil derfor nemne nokre opplagte ønske eg har til NO.

#### 2.2.1 Visualisering på kart

Digitale kart er kraftige verkemiddel når det gjeld å spreie geografisk informasjon, om heimfesting av ord og dialektformer. Eg sluttar meg til ei utsegn frå *Jysk Ordbog* i 2002: ”Netop for en dialektordbog er det en fordel at kunne frådse med

kort, der langt bedre opplyser brukeren om geografisk udbredelse, end det kan gøres ved opremsning av geografiske områder.”

Mange menneske har lite kunnskap om norsk geografi og stadnamn. I tillegg er mange av dei geografiske nemningane som NO opererer med, gått ut av bruk for fleire tiår sia, som ei følgje av samanslåing av administrative område som herad (minste referanseining i NO for heimfesting av førekomst og uttale). Her vil kart faktisk kunne vere heilt naudsynleg for å kunne forstå ord- og lydgeografiske tilhøve.

Innanfor NO arbeider ein med å utvikle ein ny versjon av innskrivingsprogrammet for nye ordbokssetlar. Dette er ei prioritert oppgåve i 2008. Den nye versjonen vil mellom anna innehalde ein komponent for kart.

### 2.2.2 Illustrasjonar (bilde, film)

Ein ordboksillustrasjon legg tyngd til den informasjonen som definisjonane gjev. Og ettersom vi veit at moderne menneske gjerne ser bilde, men ikkje alltid tek seg tid til å lese ein tekst, vil det vere klokt å tilby ei illustrert ordbok, om ein vil ha brukarar – ikkje minst unge menneske.

Mange omgrep kan vere vanskeleg å definere. Det gjeld mellom anna farge. *Ordbog over det Danske Sprog* definerer fargen *blå* såleis: ”den af hovedfarverne, der i regnbuen og spektret ligger ml. grønt og violet”. Samstundes tek ein i bruk adjektiv som ”mørkfarvet” og ”sort” og ”blaaligbleg” for å karakterisere bruken av *blå* i samband med hudfarge. Bilde kan vere ein god stønad for definisjonen av slike omgrep.

Systemteikningar (til dømes av ei kvern, med 50 omgrep med tilknytning til kverna og verksemda kring ho) kan syne form, bruk og relasjonar vel så bra som definisjonar, ikkje minst i ei verd der det etter kvart er langt mellom dei fysiske installasjonane av slike innretningar.

Det kan òg vere greit å illustrere definisjonar av handlingar og rørsle gjennom stutte videosnuttar, i tillegg til å definere dei. Eit døme kan vere verbet *slå*, med undertyding i (det maritime) uttrykket *slå ein knute*, i motsetnad til hovudinnhaldet i verbet. I samband med denne undertydinga kan ein også ønske å gjere greie for at ein slår på ulike vis, alt etter kva for resultat (type knute) ein vil få fram. Resultative omgrep som *pålestikk* og *båtmannsknop* let seg knapt definere fullt ut anna enn gjennom illustrasjonar som syner både framgangsmåten og resultatet.

I den nye versjonen av innskrivingssetlar for NO som er ei prioritert oppgåve i 2008, vil det også vere ein komponent for bilde.

### 2.2.3 Lyd

Lydspor høyrer naturleg heime i ei digital ordbok for språk og språklydar. Det finst alt mykje digitalt tilfang som på glimrande vis syner både ord og språket i bruk. Nemnast skal her, i samband med ei folkemålsordbok som Norsk Ordbok, tilfang som er samla inn serskilt for å dokumentere talemålsskilnader.

Også andre lydlege fenomen og omgrep, innanfor song, musikk, og støy, let seg teknologisk sett greit illustrere i ei elektronisk ordbok, som eit supplement til den verbale definisjonen. Som døme på ein verbal definisjon av en viss lyd kan ein trekkje inn *Svenska Akademiens Ordbok* definisjon av ordet *bjäff*: ”uttryck som afser att återgifva en liten hunds skällande”. Eit eksempel på bruken av ordet er det òg funne plass til: ”Humlegården är full av unga glada hundar, som ... då anledning erbjuder sig, rusa upp och säga vov, vov, om de äro stora, eller vaff, vaff om de äro små; men de allra minsta säga bjäff, bjäff, bjäff.” Her kunne eit lydspor ha vore eit interessant tillegg.

NO finansierer digitalisering av gamle lydband med relevant ord- og taletilfang, som etter planen skal utnyttast i ein nettversjon av ordboka.

#### 2.2.4 Samandrag om innhald

Eg trur bruken av kart, bilde og lyd vil fremme kvaliteten på og bruken av ordboka. Eg vonar at vi ordboksfolk vil la oss inspirere til å tilpasse oss den nye nett-røyndomen så snart som råd er, også såleis at digitale verkemiddel og strukturar naturleg vil inngå allereie i den primære ordboksutviklinga, og ikkje bli eit post-editerings-fenomen og -verkty. Eg vil altså med andre ord gjerne at vi raskt når det nivået som Bakken og Ore (2005: 11) nemnde i ei artikkel om ei muleg elektronisk *Norsk Ordbok*, nemleg at vi kan ”konseptualisere ordbøker direkte for det elektroniske mediet”.

### 3 NO gjennom nettportal – samliv med anna digitalt tilfang

Eg nemnde innleiingsvis eit par punkt som ein bør ta omsyn til dersom ein ønsker at *Norsk Ordbok* skal bli ein bruksting (rask tilgang, tilgang gjennom moderne bokmål og nynorsk). Eg ser for meg nok ei oppgåve som eg meiner ein bør vurdere for å sikre bruken av NO, nemleg det å gjere den digitale ordboka tilgjengeleg saman med andre digitale samlingar, innanfor ein nettportal, med sams eller relaterte emne og mål.

#### 3.1 Data gjennom nettportal

Med nettportal forstår eg ein eittpunkts adgang – ein virtuell portal eller korridor – til informasjon som er tilgjengeleg via Internett, og som består av forskjellige, men emnemessig eller på anna vis logisk relaterte datasamlingar, som er av interesse for ulike brukargrupper.

Gjennom portalløysingar kan ein gje og få effektiv tilgang til mange databasar og dokument med særleg relevans for visse område. Ein kan få rask tilgang til fleire sett av aktuelle digitale data. Den mentale så vel som den fysiske vegen mellom datasamlingane blir korta monaleg ned (til få tastetrykk). Felles adgangsprosedyrar og gode brukargrenseflater, saman med felles teknologiske løysingar for struktur, visning og kopling, minkar hindringar og tidsbruk for innhenting og vidare bruk av kunnskap. Dette vert ofte vist til – noko overflatisk – som ”saumlause” løysingar.

Om ein vil synleggjere *Norsk Ordbok*, for at ho skal bli brukt, kan det vere ein god idé å samordne verket med andre relaterte digitale samlingar. Eg skal gje eit par døme på slike samlingar.

### 3.2 "Humaniora digitalt"

Ei muleg portalløysing kunne vere ein humanioraportal, ein portal til datasamlingar for humanistiske vitenskapsfag eller -felt.

Målet med ein offentlig humanioraportal kan vere – i kortform – å vere leiande når det gjeld å formidle og organisere forskingsbasert humaniora på nettet, ved å vere ein garantist for kvantitet og kvalitet på humaniorafeltet.

Nettsidene til det norske Dokumentasjonsprosjektet viser døme på mulege deltakarar innanfor ein humanioraportal. Prosjektet var eit samarbeidsprosjekt, i åra 1992–1997, mellom dei humanistiske miljøa ved dei (då) fire universiteta i Noreg. Informasjon som på den tida låg i ulike arkiv og samlingar – i manuskript, katalogar, hefte, bøker eller på arkivkort – fekk elektronisk form. Det vart bygt opp eit informasjonssystem kalla "Universitetenes databaser for språk og kultur. Dette har gjeve dei einskilde fagmiljøa betre verkty for forskning og studiar. I tillegg har databasen auka den allmenne tilgangen til arkivmaterialet, samstundes som ein tek hand om opphavsrettar og tryggleiken til samlingane.

På nettsidene til Dokumentasjonsprosjektet finn ein i dag data frå mange humanistiske forskingsfelt, som arkeologi, etnografi, folkloristikk, historie, leksikografi, mynter/numismatikk, middelalderdiplom og stadnamn. Denne nettstaden, og samlingane, vert halden ved lag av Universitetet i Oslo. Eg kjenner ikkje til dei konkrete planane for vidareutvikling av nettstaden. Eg ser det likevel slik at det er naturlege innhaldsmessige relasjonar mellom NO og samlingane frå Dokumentasjonsprosjektet. Om ein meiner at ein felles nettportal kunne styrke og effektivisere bruken av digitale humanistiske data, bør ein drøfte om desse samlingane kanskje kunne vere eit utgangspunkt, eller iallfall om desse samlingane burde finnast innanfor same portalen.

Innanfor ein humanioraportal kan NO dekkje ein funksjon som leverandør av særnorske ord og uttrykk, med definisjonar, for forskarar og formidlarar som skal presentere data, funn og gjenstandar med tilknytning til norsk kultur. Det vil vere praktisk å få dekt dette behovet gjennom to tastetrykk og enkle kopi- eller koplefunksjonar mot ein substansiell portalnabo som NO. Det vil heller ikkje skade omdømet til NO å bli brukt og referert i slike høve.

Når det gjeld NO sin gevinst ved å ha saumlause overgangar til andre humaniorasamlingar som det kunne vere interessant å gå inn i informasjonssamarbeid med, er han fleirsidig. Under utviklinga av ordboksartiklar vil det mellom anna kunne vere behov for informasjon om målføre, ordgeografi, faglege omgrep og termar innanfor andre forskingsfelt, eller (eldre, eldste) førekomstar av ord i eldre kjelder (t d diplom og stadnamn). Ein vil kanskje òg gjerne illustrere artiklane med bilde eller teikningar (t d frå ein museumsdatabase eller database for arkeologiske funn).

Eg vil oppfatte dei datasamlingane som eg har nemnt her som positive supplement til kvarandre, for mange føremål. Eg meiner derfor at ein bør leggje teknologisk til rette for effektiv utnytting av humanistiske data, mellom anna med tanke på samankopla eller samanstøypte presentasjonar, for ein gjensidig synergieffekt.

### 3.3 "Norge digitalt"

Eg skal også nemne ein annan interessant dataportal – som alt finst, og som er i rivande utvikling. Det gjeld den offentlege portalen "Norge digitalt", som vert definert som "ein geografisk infrastruktur" for offentlig administrasjon. Denne portalen inneheld ei mengd geodata (:data som gjev informasjon om objekt (vatn, hus, vegar, fyr osv), hendingar og forhold der posisjonen på jorda er ein vesentleg del av informasjonen).

"Norge digitalt" er ein viktig del av norske styresmakter sin IT-politikk. Portalen skal, i følge Stortingsmelding nr. 30 (2002 – 2003), vere ein sterk og effektiv reiskap i forvaltninga av areal, naturressursar og kulturminne.

Ryggraden i denne portalen er det som kallast Sentralt stadnamnregister (SSR), den nasjonale databasen for norske stadnamn. Per 1. januar 2007 inneheldt denne databasen om lag 740 000 (739 566) namn. Namna reknast for primære data i portalen. Sekundære data (fagdata) er data om t d naturressursar, befolkning, miljøtilstand, kulturminner og arealbruk.

Setelsamlinga til NO er eit av hovudverktøya i arbeidet med å fastsetje skrivemåten av norske stadnamn. NO har altså alt i dag ein viktig støttfunksjon i høve til arbeidet med det som er primære referansedata innanfor portalen "Norge digitalt".

Kvalitativt vil eit nærare samband mellom NO og namnedatabasen SSR kunne medføre ei styrking av innhaldet i NO. Rett nok er *propria* (herunder stadnamn) ikkje oppslagsord i NO, men det er likevel høve til å dokumentere bruk av appellativ i stadnamn, som *døme*. Det finst også mange *døme* på at eit appellativ berre er belagt i stadnamn. Om ein vil freiste gje eit heilskapleg bilde av førekomstane og utbreiinga av ord i norsk folkemål, bør ein altså også sjå på stadnamna, i meir systematisk grad enn det som vert gjort i dag.

For NO, som òg har som mål å spreie kunnskap om og styrke bruken av det nynorske skriftmålet, kan det også verte ei stor språkpolitisk vinning å kome inn i eit portalfellesskap som "Norge digitalt". Det er grunn til å tru at dei fleste datasamlingane i dette fellesskapet, og andre nasjonale digitale fellesskap, vil ha bokmål som målform. Det å vere til stades i eit slikt fellesskap vil derfor i seg sjølv vere ei språkpolitisk markering og ei hevdning. Det kan vere målpolitisk viktig og riktig for NO å freiste vinne seg ein plass innanfor ein nasjonal geodaportal, og ikkje berre søkje saman med andre nynorske datasamlingar på ein rein nynorsk nettstad, eller gå i bås saman med anna humaniora- eller kulturstoff innanfor ein smalare kultur- eller humanioraportal. NO kan sjølvsagt



også vere innanfor slike portalar, men ein bør kan hende ikkje nøye seg med å vere berre der.

Kjem NO først på plass innanfor den nasjonale geodataportalen "Norge digitalt", vil bruken kunne verte vidare enn det å levere temadata for stadnamn. Men støttefunksjonen for stadnamna i SSR er likevel kan hende NOs sterkaste argument for å kome på plass innanfor denne portalen. Portalen kan synest som eit sær s ambisiøst mål, sjøl for ein nasjonal ordboksdatabase. Eg meiner likevel at ein bør vurdere også ei slik plassering i det fremtidige utviklingsarbeidet for NO på nettet.

#### 4 Etterord

Etter snart 8 tiår og 7 papirband er tida inne for å leggje meir substansielle planar for ein nettversjon av *Norsk Ordbok*. Dei mange sidene som dette inneber, tilseier at det krevst både strategisk tenking og handling, på både kort og lang sikt, i einarom og saman med andre digitale vener, for at utviklinga av denne versjonen skal kunne skje på ein målretta, samordna og effektiv måte, med eit brukarvenleg og attraktivt resultat. For dette nasjonale storverket vil ein rådføre seg med dei fremste kunnskapsmiljøa innanfor netteknologi og løysingar, for å kunne gå trygt ut på nettet i fremtida.

For utviklinga av ein nettversjon av NO har det òg vore avgjerande å kome fram til ei avklaring med forlaget bak den trykte versjonen, Det Norske Samlaget, om løysingar og rettigheter. Det ligg no føre ei skriftleg intensjonsavtale mellom NO og Det Norske Samlaget om korleis nettpubliseringa skal skje. Avtala syner at det er semje om at spørsmålet om eigarskap ikkje skal seinke utviklingsarbeidet med nettversjonen. Nettpubliseringa skal vere eit ikkje-kommersielt tiltak, og databasen bak nettversjonen skal driftast av NO. Nettpubliseringa vil vere ei prioritert sak for den nye leiinga i Norsk Ordbok.

#### Kjelder:

- Bakken, Kristin og Christian-Emil Ore 2005: Norsk Ordbok – også ei elektronisk ordbok? I: *Lexico Nordica* 2005, 7–18.
- Engebreetsen, Martin 2004: Teknologi til glede og besvær. Om undervisningsrelatert nettpublisering i lærerutdanningen. Rapport fra en delstudie innen prosjektet *Nettportal HiA* (2003 – 2005).
- McClean, Shilo T. 2007: *Digital Storytelling*. Cambridge, Massachusetts: The MIT Press.
- Stortingsmelding nr. 30 2002 – 2003: "Norge digitalt" – et felles fundament for verdiskaping". Det kongelige miljøverndepartement.

112 ELI JOHANNE ELLINGSVE

Eli Johanne Ellingsve  
Dr.Art./Ordboksredaktør, f. 1952  
NTNU, INL, Norsk Ordbok  
Dragvoll  
NO-7491 Trondheim  
eli.ellingsve@hf.ntnu.no

## Hvad er en videnskabelig flersproget idiomordbog?

We address the question of what might deserve the name "a scientific multilingual idiom dictionary". Few interlingual idiom dictionaries actually call themselves so or try to leave the impression that they are scientifically based. Why is that so? It may have to do with the very character of the object itself, idioms, and the special interest many people show in them. We want to demonstrate that there are no principal barriers against such a project: A catalogue of lexicological and lexicographical measures and methodologies already exist. What could be questioned, however, is the *lexicographical relevance* of the matter. On the other hand, there may still be pragmatic arguments in favour of such a dictionary. A metalexicographical discussion of this dilemma will be given.

### Indledning

Denne artikel er ikke en fremlæggelse af et konkret ordbogsprojekt, men en mindre metaleksikografisk udredning. Vi har prøvet at stille os selv spørgsmålet, hvad der skal til, for at man i 2007 kan kalde en publikation for en "videnskabelig flersproget idiomordbog". Videnskaben er i de senere år kommet et stort skridt videre på flere områder med relevans for vores genstand. Det gør på én gang opgaven nemmere og vanskeligere. Nemmere, fordi der er kommet flere holdepunkter og retningslinjer for idiomleksikografer, men samtidig vanskeligere, fordi kravene til produktet sideløbende hermed er skærpet. Det er i hvert fald godt for én gruppe, nemlig brugerne, hvem de så end er. Netop dét spørgsmål vil vi også se nærmere på, for det er ikke helt så trivielt, som det lyder.

Vores undersøgelse veksler mellem teoretiske overvejelser i relation til det stillede spørgsmål og empirisk analyse af eksisterende flersprogede idiomordbøger, både i print- og i elektronisk form. Herudover leverer vi eksempler på kontrastive korpusanalyser af idiomer og et enkelt bud på en artikelløsning på de skitserede leksikologiske problemer.

Det er klart, at der kan være flere svar på det spørgsmål, vi har stillet. Vi ville derfor glæde os, hvis andre kom med et andet svar efter seriøs overvejelse af vores. Det afgørende er, at spørgsmålet overhovedet bliver stillet, for det er vores kvalificerede fornemmelse, at problemets kompleksitet overses. At det overhovedet er tilfældet, prøver vi også at formulere en forklaring på.

Bidraget bygger i høj grad på Aarhuskolens lære om funktionel leksikografi (Tarp 2006, se også anmeldelsen Farø/Gottlieb 2007) og de ideer om idiomatik og leksikografi, som er blevet fremlagt i Farø (2005a) og (2006), se også Almind et al. (2006).

## 1. Videnskabelig I: sprogvidenskabelig fraseologiforskning

“Vi bruger idiomer, ordsprog og faste vendinger i sproget, som vi bruger salt i maden. Sproglblomsterne er med til at krydre sproget. De gør festsangen sjovere, talen skarpere og kritikken mildere.”<sup>1</sup> (pressemeddelelse og indledning til et interview med Henning Bergenholtz på [www.asb.dk](http://www.asb.dk) i anledning af lanceringen af Ordbogen over Faste Vendinger, primo 2007 – jf. anmeldelsen i Farø 2005b). Dette citat genspejler en gængs teoretisk opfattelse af idiomer, men ikke nødvendigvis deres reelle rolle i sprogbrugen. Derfor gælder den første diskussion betydningen af “videnskabelig” i relation til den sprogvidenskabelige fraseologiforskning med dens definition af idiomer. Vi bruger her vores egen teoretiske tilgang til fraseologi som grundlag (jf. fx Farø 2006, Hallsteinsdóttir 2001, Hallsteinsdóttir/Farø 2006), selv om det ikke skal være nogen hemmelighed, at der findes andre (jf. Burger et al. 2006).

“Fraseologi” bruges traditionelt både som overbegreb for leksikaliserede flerordforbindelser i et sprog og som betegnelse for den videnskabelige disciplin, der undersøger flerordsforbindelser. Sidstnævnte er i de sidste 40 år vokset betydeligt, med mange, til dels meget forskellige, teorier om delgenstanden idiomatik til følge. Det samme gælder systematiseringsforsøg og beskrivelse af idiomer. I stedet for at gengive de mange forskellige idiomdefinitioner, der findes, præsenterer vi vores eget koncept (“idiomatik i snæver forstand”), som har været genstand for omfattende overvejelser:

- Idiomer er normale leksikalske enheder og dermed en del af ordforrådet. De består af komponenter, der (næsten altid) selv fungerer som selvstændige leksikalske enheder. Idiomer har en holistisk leksikaliseret idiomatisk betydning, der ikke stemmer overens med en potentiel kompositionel betydning.
- Idiomer er latent flertydige, de er på én gang en leksikaliseret enhed og en potentiel kompositionel ordgruppe.
- Idiomer er ikonografiske (Farø 2006b), dvs. de har en billedlig dimension.
- Idiomer er semiotisk upåfaldende, idet de ikke umiddelbart kan identificeres som en leksikalsk enhed via formen.

Idiomers latente flertydighed kan demonstreres med det danske idiom *have ben i næsen* fra en autentisk tekst (Jyllandsposten 27. januar 2006, s. 2), jf. Figur 1:

1 Det er problematisk at sideordne “faste vendinger” med de to andre begreber, idet der må være tale om et hyperonym til dem.



Figur 1: Ben i næsen

Idiomet selv er en flertydig leksikalsk enhed. Der angives tre forskellige betydninger i Ordbogen over faste vendinger ([www.idiomordbogen.dk/](http://www.idiomordbogen.dk/))

- (a) 'have en stærk vilje (og vide, hvordan man bruger den)'
- (b) 'have gåpåmod'
- (c) 'være fornuftig'

Ud fra den minimale kontekst i eksemplet er betydningerne (a) og (b) fuldt acceptable, og hvis intentionen med teksten er at formidle, hvor fornuftigt det er at rejse med lige netop "Explorer", så er betydning (c) også acceptabel.

En anden form for flertydighed er forankret i idiomers ydre form som flerordsenheder. De enkelte komponenter har selv en (eller flere) betydninger som altid kan indgå i en betydningskonstruktion med en kompositionel betydning som resultat:

- (have) *ben + i + næsen*

Problematiske ved denne betydningskonstruktion er den hyppigt forekommende flertydighed hos de enkelte komponenter. På den ene side bliver bestemmelse af én og kun én kompositionel betydning i mange tilfælde så umulig. På den anden side er der frihed til at vælge den kontekstuel passende betydning, som det rent faktisk sker i fig. 2:



Fig. 2: Latent flertydighed: eksempel på en potentiel kompositionel betydning af (have) *ben + i + næsen*

Figur 2 viser derudover en mulighed for realisering af ikonografi i idiomers.

## 2. Videnskabelig II: leksikografi

I dette kapitel omtaler vi skitseagtigt Funktionslærens forskellige leksikografiske funktioner og deres relevans for interlingval idiomatik. Herefter kommer vi ind på eksisterende idiomordbøger, idet vi ser på eksempler på to- og flersprogethed i eksisterende idiomordbøger. Dernæst diskuterer vi forskellige forudsætninger og muligheder for flersproget idiomleksikografi (jf. yderligere i Hallsteinsdóttir 2006).

### 2.1 Kommunikative funktioner og idiommer

Funktionslæren går ud fra en hovedopdeling af de leksikografiske funktioner i hhv. kommunikative og kognitive. Her behandler vi på listeform først de kommunikative funktioner i relation til idiomatik:

- **Reception** af tekster i fremmedsprog (fremmedsprog → modersmål):
  - formidling af idiomatisk betydning
  - betydning ≠ idiomatisk ækvivalens
- **Produktion** af tekster i L2 (L1 → L2):
  - instruktion af brugeren i hvordan han/hun anvender idiomet
- Oversættelse (L2 → L1 og L1 → L2):
  - *tegn*baseret: ækvivalente idiommer ("oversættelse" i fremmedsproglige receptions-/produktionsprocesser)
  - *tekst*baseret: funktionelt ækvivalente sproglige tegn (oversættelse af tekster i traditionel forstand)

Hvordan de kommunikative funktioner mere konkret kan løses i forhold til idiommer, er et helt kapitel for sig, som det vil føre for vidt at diskutere her (jf. Farø 2006). Det vigtigste er, at hverken receptions-, produktions- eller oversættelsesfunktionen nødvendigvis løses bedst ved angivelse af et L2-idiom. I en del tilfælde vil det endda være en kontraproduktiv løsning. Spørgsmålet afhænger desuden i høj grad af "de leksikologiske kendsgerninger" (se ndf.), som ikke kan tages for givet uden en nærmere undersøgelse.

### 2.2. Kognitive funktioner

Århuskolen opererer desuden med en række kognitive,<sup>2</sup> dvs. vidensorienterede funktioner (Tarp 2006):

- "**Ønsket om at vide noget mere**"<sup>3</sup>
  - underholdningsfunktionen (*lexicotainment*)

<sup>2</sup> Man kan overveje det hensigtsmæssige i termen, eftersom den kognitive lingvistik (fx Croft/Cruise 2004) netop ikke skelner skarpt mellem viden og kommunikation, men gør kognition til et integreret element af og grundlag for sproglig aktivitet.

<sup>3</sup> Brugervejledning til "Den Danske Idiomordbog" <http://netdob.asb.dk/idiom/guide.html>.

- indlæringsfunktionen
- **Metaviden om sprog**
  - “sproglige billeder”
  - historisk-etymologisk motivering
  - “falske venner”
  - litteralisering af L2-idiomer
  - hvad er et idiom?

Relevansen af de kognitive funktioner i forbindelse med idiomatik i en videnskabelig idiomordbog er åbenlys. Der er stor forskel på rent kommunikative opslagsbehov og så mere metasproglige interesser, som tilsyneladende spiller en stor rolle i forbindelse med idiommer. Idiomatikken er dermed med til at demonstrere relevansen af den principielle opdeling af de leksikografiske funktioner.

I det følgende ser vi nærmere på en række forskellige eksisterende idiomordbøger.

### 2.3. Flersprogede idiommer i etsprogede idiomordbøger

Den første overraskelse var konstateringen af, at der forekommer idiommer på flere end et sprog i etsprogede idiomordbøger, dvs. idiommer i etsprogede ordbøger forsynes delvis med formparallelle idiommer i andre sprog til yderligere forklaring af udgangsideomet. I figur 3 gives et eksempel på en “etsproget” dansk idiomordbog med angivelse af – må man gå ud fra, se dog ndf., – ækvivalente engelske, tyske og franske idiommer:

... **pr. næse**  
*(uformelt)* ... for/til hver person  
*(især om betaling)*. XV  
 • pro Nase • par tête de pipe.

**stikke sin næse (el. næsen) i ngt**  
 blande sig i ngt, som ikke kommer  
 én ved. XV  
 • poke/push/stick one's nose into  
 something (*f.eks. someone else's  
 affairs/business/concerns/things  
 that don't concern one*); (*lad være  
 med at stikke din næse i mine affæ-  
 rer*) curiosity killed the cat • die/  
 seine Nase in etwas stecken; (*stik-  
 ke sin næse i andres sager*) (jeman-  
 dem) in seinen Kram hineinren-  
 den; (*stikke sin næse i alting*) die/  
 seine Nase in alles (*el. in jeden  
 Dreck/Quark/Topf*) stecken; sich  
 um jeden Quark kümmern; sich  
 in jeden Quark einmischen  
 • fourrer/mettre son/le nez dans  
 quelque chose; (*udvise indiskret  
 nysgerrighed, der nærmer sig det  
 voyeuragtige*) ne pas avoir les yeux  
 dans sa poche.

Figur 3: Stig Toftgaard Andersen (2001): Talemåder i dansk.

Figur 4 viser en idiomordbog på esperanto med angivelse af ækvivalente tyske, engelske og tjekkiske idiommer. Udgangsideomerne forklares på esperanto og udgangspunktet ved bestemmelsen af ækvivalensen tages i idiomernes *form* og deres billedlighed – ikke deres brug:

**Balai ion sub la tapiŝon**


**Klarigoj:**

= kaŝi aŭ ne trakti temon; forpreni ĝin el la tagordo; ne findiskuti ĝin, ĉar ĝi estas malŝatata aŭ komplika

**Ekzemploj:**

El sub la tapiŝo  
La aŭtoro ofte atentigas redaktorojn de periodaĵoj pri erara lingvouzo, por ke en sekva numero ili faru ĝustigojn. Sed - li plendas - "anstataŭ danki min, la plej multaj kutime strutmaniere kaŝas sian kapon en paper-korbo aŭ *balaas la erarojn sub la tapiŝon*, kie ili ŝime kaj putre akumiliĝas kaj stinkigas la tutan redakciejon." (ESPERANTO 1/96: 18)

Ĝis nun miaj bonintencaj konsiloj por plibonigi la kvaliton de la periodaĵo estis senkomente *forbalaitaj sub la tapiŝon* de la redaktejo. (MONATO 8/96: 6)




**Similaj esprimoj en aliaj lingvoj:**

*germane:* etw. unter den Teppich kehren  
*angle:* to sweep sth. under the carpet/rug  
*ĉeĥe:* zamést něco pod koberec


Figur 4: Fiedler, Sabine/Rak, Pavel (2004): Ilustrita frazeologio.

#### 2.4. Idiommer i tosprogede idiomordbøger

Beskrivelse af idiommer i tosprogede idiomordbøger har forskellige udgangspunkter; i det følgende giver vi eksempler på nogle få muligheder. I det første eksempel tages der udgangspunkt i ækvivalente idiomers forskellige ikonografi i tysk og fransk (jf. Figur 5):



Faire son beurre



Sein Schäfchen  
ins trockene bringen  
Emmener son petit mouton au sec

Figur 5: Blum, Geneviève/Salas, Nestor (1989): Les Idiomatics français-allemand.



Ifølge brugervejledningen skal den frit tilgængelige tysk-engelske idiomordbog Phrasen.com (www.phrasen.com) bruges til hjælp ved forståelse og udvidelse af ordforrådet i det fremmede sprog:

Phrasen.com er en ordbog med engelsk-tyske og tysk-engelske faste vendinger (ordsprog, floskler, bevingede ord, etc.). Phrasen.com hjælper til at forstå det fremmede sprog. For at slå oversættelser op til en bestemt fast vending og udvide dit ordforråd kan du indsætte stikord i søgevinduet eller bruge det alfabetiske register.<sup>4</sup>

Det viser sig dog, at ordbogen – som eneste leksikografisk information – indeholder én eller flere ækvivalente idiommer i de to sprog, fx “oversættes”<sup>5</sup> det tyske idiom *die Ohren steif halten* med de engelske idiommer *to keep a stiff upper lip* og *to keep one’s chin up*.

Ordbogens brugere må altså selv finde ud af idiomernes eventuelt forskellige brugsregler og indbyrdes semantiske divergenser.

Et eksempel fra en tosproget tysk-engelsk ordbog viser, at idiommer her ligeledes “oversættes” med idiommer, uden yderligere forklaringer vedrørende betydning eller brug (jf. Figur 6):



Figur 6: Langenscheidt e-Handwörterbuch Deutsch-Englisch

Ud fra disse og mange andre eksempler kan vi konstatere, at der er et stort fokus på idiomernes *form* i den leksikografiske praksis. Dette gælder såvel formen som kategori, altså det at idiommer forklares med idiommer, som formen forstået som tegnets udtryksside, dvs. den danner grundlaget for bestemmelsen af ækvivalens mellem sprog.

Leksikografiske informationer om betydning og brug er derimod meget sjældne. En ægte kontrastiv leksikografisk beskrivelse i form af sammenligning af idiomernes betydning og funktion har vi ikke fundet endnu.

4 Oversættelse KF.

5 Termen er problematisk i leksikografien, fordi den skaber forvirring i forhold til den egentlige oversættelsesfunktion.

### 2.5. *Idiomer i flersprogede idiomordbøger*

Færdiggjorte flersprogede idiomordbøger findes der ikke mange af, mens en del er under udarbejdelse (jf. fx Pacholski/Laskowski (2007) og Arsentyeva (2006)). AUTOPHRAS er en spansk-fransk-engelsk-italiensk-portugisisk-tysk-ukrainsk-slovakisk-rumænsk-tjekkisk idiomdatabase udarbejdet af et forskerhold fra de pågældende lande (jf. Pamies et al.). En søgning på indholdet 'FALSENESS' i en i forvejen ordnet liste med semantiske "koncepter" fører videre til en liste med subkoncepterne 'CRIME' og 'TRUTH', som så ender med muligheden for valg af en betydning ud af fire mulige:

- (a) '(good) appearances of people can be deceiving',
- (b) '(good) appearances of things can be deceiving',
- (c) 'though it seemed unbelievable, it is real (though nobody could have imagined it, it happened)',
- (d) 'with real generosity'.

Til hver betydning er der knyttet en opremssning af fraseologismer på de 10 sprog, efterfølgende gengives de franske, engelske og tyske fraseologismer, der svarer til betydning (a):

- *l'habit ne fait pas le moine; les apparences sont trompeuses; il ne faut pas se fier aux apparences; les apparences sont bien souvent trompeuses; ce n'est pas à la barbe qu'on connaît le philosophe; il ne faut pas juger de l'arbre par l'écorce; il ne faut pas juger du bois par l'écorce*
- *the cowl does not make the monk; clothes do not make the man; appearances are deceiving, it's not the beard that makes the philosopher, you can't judge a book by the cover, all are not hunters that blow the horn; all are not soldiers that go to the wars; all are not thieves, that dogs bark at*
- *die Kappe macht den Mönch nicht aus; es sind alle Jäger, die das Horn gut blasen; das Kleid macht nicht den Mann*

Ud fra disse eksempler kan man skitsere følgende tendenser inden for to- og flersprogede idiomordbøger: Der er generelt en meget sparsom omgang med leksikografiske informationer i de undersøgte to- og flersprogede ordbøger. Som regel er der tale om, at L1-idiomer forklares med angivelse af ækvivalente idiommer i L2 uden yderligere informationer om betydning eller brug (idiomer *skal* åbenbart "oversættes" med idiommer (jf. Farø 2006 om idiomoversættelser og Hallsteinsdóttir 1997 om mulig oversættelse af idiommer). Det skyldes delvist, at ordbøgernes konception tager udgangspunkt i selve sprogtegnet idiom; leksikografens opgave anses for at være beskrivelse af sproglige tegn – her et eksempel på denne holdning fra den monolingvale metaleksikografiske diskurs:

The task of a lexicographer analysing a corpus is to identify all normal uses of words (**not** all possible uses), to group them into categories, and to offer

an explanation (a ,definition') of each word in each of its categories. (Hanks et al. 2006: 444).

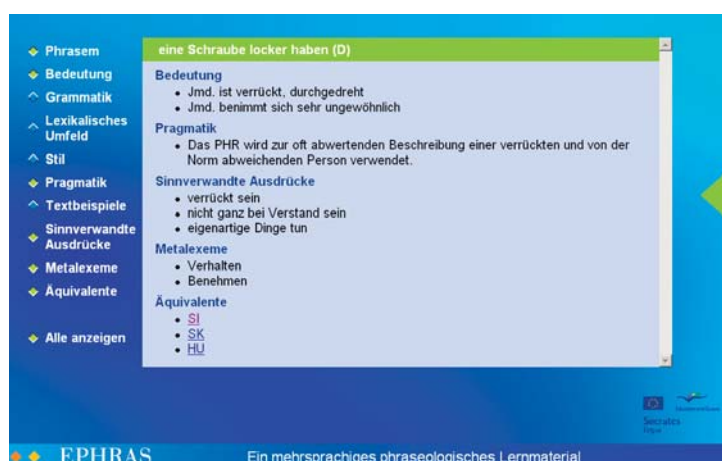
Selve sproget er altså det leksikografiske udgangspunkt og mål:

In this paper we have presented entries for a new treatment of German light verbs (*verblasste Verben*). The aim of such a dictionary is to bring out the subtle phraseological norms with which each light verb is associated and to explain and exemplify the normal collocations in which such verbs occur. (Hanks et al. 2006: 448)

Ordbogsbrugere og deres behov spiller åbenbart ingen rolle her. Denne tilgang til leksikografien gælder fx også konceptet for AUTOPHRAS (Pamies et al. EUROPHRAS 97 i SL), Pacholski/Laskowskis (2007) koncept for en engelsk-tysk-polsk idiomordbog (præsenteret på EUROPHRAS 05 i Slovenien) og Arsentyevas (2007) koncept for en russisk-engelsk-tysk-tyrkisk idiomordbog (præsenteret på EUROPHRAS 04 i Schweiz). Det konkrete leksikografiske resultat af en sådan konception er i værste fald ikke-brugerorienterede lister med "ækvivalente" idiommer i to eller flere sprog der stilles ved siden af hinanden uden nogen som helst forklaringer om betydning, brug, betydnings- eller brugsforskelle.

### 2.6. Case: EPHRAS

Den flersprogede lærerdatabase EPHRAS er resultatet af et større fraseologisk forskningsprojekt. Dens leksikografiske koncept er en modulær opbygning med monolingvale moduler (tysk, slovensk, slovakisk, ungarsk), der forbindes via hyperlinks (jf. figur 7). Udgangspunktet er stadig det sproglige tegn, ikke brugerbehov, dog sættes der sekundært fokus på de behov, lærere har ved forståelse og brug af idiommer. Derudover sættes dogmet om idiom:idiom delvist ud af kraft med ordbogens angivelse af ækvivalente enkeltord.



Figur 7: EPHRAS: [www.ephras.org](http://www.ephras.org)

Spørgsmålet om EPHRAS er en videnskabelig idiomordbog eller ej, kan besvares med et ja! Den har et solidt sprogvidenskabeligt, metaleksikografisk og sprogdidaktisk grundlag (jf. Jesenšek 2006). Til gengæld må man konstatere, at der ikke er tale om en flersproget idiomordbog. Forbindelsen mellem to sprog foregår via idiomernes form (delvist også deres betydning), dvs. der skiftes mellem monolingvale beskrivelser. Men der er ingen ægte flersproget leksikografisk beskrivelse.

Et andet problem hænger sammen med selve ordbogens grundtanke, det multilingvale koncept: Reelt bruges nemlig altid kun to sprog ad gangen ude i sprogvirkeligheden, også i forhold til deciderede flersprogede idiomordbøger. Der er simpelthen ikke *kommunikative* brugersituationer for flere end to sprog. Til gengæld kan man muligvis forestille sig et behov under vidensfunktionen, som kunne dækkes af denne type ordbøger.

### 3. Videnskabelig III: Leksikologi

Leksikologiens opgave er dokumentation og beskrivelse af leksikalske enheder, såvel enkeltord som fraseologi. Ud fra en holistisk orienteret sprogbeskrivelse kan en opdeling af ordforrådet i enkeltord og fraseologi kun være baseret på en metodisk beslutning – som i denne artikel –, og den har intet at gøre med selve sproget.

Som i alle andre leksikografiske arbejder bør grundlaget for idiomordbøger være en solid empirisk leksikologisk beskrivelse. Dermed får teorier om idiomers stor leksikografisk relevans, afhængigt af brugernes behov for informationer i brugssituationer.

En ideel deskriptiv ensproget leksikologisk analyse bygger på en opfattelse af et abstrakt sprogsystem, der forbinder

- frekvens og manuelt analyserede empiriske korpusdata (ikke kun introspektion) og
- viden om mental repræsentation, receptions- og produktionsprocesser (i relation til korpusdata) (jf. Hallsteinsdóttir 2007).

Der findes allerede en del ensprogede videnskabelige idiomordbøger, som i virkeligheden er leksikologisk beskrivelse af idiomers og med fordel kan bruges som grundlag for leksikografisk arbejde ved flersprogede ordbøger, fx i tysk og dansk:

- “Wortverbindungen online” (<http://hypermedia.ids-mannheim.de/elexiko/wv-online/>; Steyer/Lauer 2006) og
- “Kollokationen im Wörterbuch” (<http://kollokationen.bbaw.de/>; Fellbaum et al. 2006).
- „Ordbogen over Faste Vendinger“ ([www.idiomordbogen.dk](http://www.idiomordbogen.dk); Bergholtz/Vrang 2007; Almind et al. 2006).

Sammenlignelige leksikologiske data og beskrivelsesmetodik vil kunne udgøre et brugbart materiale for kontrastive analyser. I form af en database med en “fuldstændig” beskrivelse af idiomer vil sådanne beskrivelser være et egnet grundlag for kontrastive og leksikografiske sammenhænge:

The multilingual idiom database provides important information for linguistic, computational linguistic and psycholinguistic use, and allows for the comparison of different phenomena in different languages. (Villavicencio et al. 2004, 1130).

Der er altså ikke tale om idiomordbøger for almindelige ordbogsbrugere, men snarere for sprogspecialister.

#### 4. Eksempler på kontrastivt-leksikologiske undersøgelser

Vi skal nu se på, hvordan man kan arbejde, når man vil undersøge idiomer videnskabeligt kontrastivt. To cases, som er hentet fra Farø (2006), skal demonstrere principperne og arbejdsgangen.

Det første eksempel er en korpuslingvistisk undersøgelse af et hypotetisk idiompar (HIP), det tyske *im Keller sein* og det danske *være (helt nede) i kulkælderen*, jf.:

- (1) Kartoffelpreis ist im Keller ([www.abendblatt.de](http://www.abendblatt.de))
- (2) Jeder Knochen tut weh, der Puls ist total im Keller und die Beine absolut leer ([www.rad.swiag.at](http://www.rad.swiag.at))
- (3) Den ene dag kan man være helt nede i kulkælderen ([www.blog.tv2.dk](http://www.blog.tv2.dk))

De 2 x 100 realiseringer, som er blevet analyseret kontrastivt, viser overraskende brugsforskelle, trods den meget parallelle komposition. Skemaet demonstrerer, at der er stor forskel på, hvilke typer subjekter, de to idiomer tager.

subjektstyper ----- idiom	person	krops- tilstand	stemning, humør	priser	andet
<i>im Keller sein</i>	0%	24%	10%	12%	54%
<i>være (nede) i kulkælderen</i>	86%	0%	8%	2%	4%

Påfaldende er i første omgang spredningsforskellen: Det danske idiom bruges i meget høj grad med personer som subjekt (jf. eksempel 3.). Denne brug mangler helt for tysks vedkommende. Til gengæld har *im Keller sein* en pæn repræsentation af “kropslige” subjekter, som (2.) er et eksempel på (“puls er i kælderen”). Denne variant er ikke registreret i dansk. Herudover har tysk en broget samling af subjekttyper, som modsvares af en meget lille dansk restgruppe.

Undersøgelsen viser mao., at der er tale om to distinkte idiomer med hver sine brugsbetingelser – og ikke om “samme” idiom i en lidt anden indpakning.

Ganske vist gælder i begge tilfælde den “konceptuelle metafor” NED ER DÅRLIGT (jf. Lakoff/Johnson 1980), men det er så grov en indholdsbeskrivelse, at den højst kan bruges i meget primitive receptionssammenhænge og ikke fx til hverken produktion eller oversættelse. Hvordan denne viden skal omsættes leksikografisk, er ikke givet på forhånd, men må overvejes ud fra de funktioner, ordbogen skal opfylde (se ndf.).

Det andet eksempel, vi skal se på, er en undersøgelse af HIP'en *holde ørerne stive – die Ohren steif halten*, som i endnu højere grad end det forrige er udtryksmæssigt konvergent. Imidlertid viser det sig, at sprogbrugen ikke understøtter indholdsmæssig ækvivalens, jf. disse få eksempler:

- (4) Hendes oplæg var det sædvanlige, luftige New Age-hokuspokus, og jeg må indrømme, at det var hårdt arbejde at holde ørerne stive ([www.skeptika.dk](http://www.skeptika.dk))
- (5) Man skal holde ørerne stive, men så kan man nyde det flotte sprog og glæde sig over, at der bliver skrevet underholdningsromaner i Robert Goddards klasse ([www.gyldendal.dk](http://www.gyldendal.dk))

Det danske idiom bruges åbenbart om situationer, hvor nogen skal lytte godt efter eller simpelthen bare koncentrere sig. Hvad med det tyske? Det fremgår allerede af disse belæg:

- (6) Wünsche dir alles gute, und halt die Ohren steif. Lieben Gruß Reiner ([www.fotocommunity.de](http://www.fotocommunity.de))
- (7) Ich drücke dich mal lieb und drücke dir weiterhin die Daumen, das [sic!] die nächste Zeit besser für dich verläuft. Alles liebe und halte die Ohren steif ([www.patientenfragen.net](http://www.patientenfragen.net))

Ikke alene er der semantisk uoverensstemmelse mellem de to idiommer: ‘koncentrere sig’ vs. ‘holde sig munter/have det godt’; der er også pragmatisk divergens imellem dem, idet det tyske idiom har en markant pragmatisk funktion som terminativ hilseformel, typisk som del af et større hilsemæssigt afslutningsforløb. Tekstlingvistisk er der desuden den forskel, at det danske idiom i princippet kan optræde overalt i teksten, mens det tyske naturligt står i slutningen af den.

Dette idiompar vil være udgangspunkt for en leksikografisk modelartikel i næste kapitel.

## 5. Forslag til artikel

Hvad det videnskabelige *leksikografiske* aspekt angår, så er det i praksis afhængigt af den kontrastivt-leksikologiske analyse. Det er bare ikke nok. For herudover kræves et leksikografisk koncept. Er man fx funktionalist, vil man fokusere på, at der findes forskellige brugertyper, som igen har forskellige behov, som kan tilfredsstilles med forskellige leksikografiske data. Desuden kan man foretage en skelnen mellem kommunikative og kognitive funktioner i overensstemmelse med Aarhuskolens forslag. Ikke bare de enkelte elementer leksikologi og leksikografi

kan altså defineres videnskabeligt, men også forholdet mellem dem.

For at anskueliggøre mulighederne for den videnskabelige idiomleksikografi vil vi vise et eksempel på en ordbogsartikel for danskere, som vi mener lever op til prædikatet “videnskabelig” i både leksikologisk og leksikografisk forstand. Strukturen følger opdelingen i først kommunikative funktioner (til og med “Mulig dansk pendant”), sidenhen vidensfunktioner:

***die Ohren steif halten***

- **Betydning:** ikke tabe modet
- **Kommunikativ funktion:** især som afskeds- og opmuntringsformel
- **Eksempel:** *also damn tschüß – halt die Ohren steif!*
- **Mulig dansk pendant:** *hyg dig/jer!, hav det godt!, pas på dig/jer selv!*
- **Direkte oversættelse:** „holde ørene stive“ (se ndf.)
- **Advarsel:** ≠ *holde ørerne stive*; **JF:** tysk *die Ohren spitzen*
- **Historisk information:** det danske idiom betød tidligere omtr. det samme som det tyske
- **Motivation:** når dyrs ører rejser sig, er de opvakte og energiske. Det kan både forklare den tyske og den danske betydning af idiomet.

Først gives der altså en dansk parafrase, som nogenlunde gengiver det tyske idioms semantiske betydning, nemlig “ikke tabe modet”. Dernæst gives en pragmatisk forklaring, som hjælper brugeren til selv at producere idiomet. Det samme gør det autentiske eksempel, som viser samspillet med andre terminative formler. De danske udtryk *hyg dig/jer!, hav det godt!, pas på dig/jer selv!* kan opfattes som “oversættelser” i modsætning til parafrasen, som typisk ikke ville stå i den samme kontekst som det tyske idiom. Herefter følger en litteralisering, som mange mennesker efterspørger. Næste punkt er en vidensorienteret advarsel om, at det danske idiom med parallel komposition mod forventning ikke er nogen ækvivalent til det tyske idiom. Herfra henvises så videre til *die Ohren spitzen*, som har funktioner til fælles med det danske. Dernæst gives en sproghistorisk oplysning om, at de to idiommer tidligere tilsyneladende *har* været ækvivalente. Endelig får brugeren afslutningsvis en motiverende forklaring på, at denne komposition har fået dette indhold.

Det er ikke den typografiske opsætning, der er i centrum her, men udelukkende de informationer, brugeren kan have glæde af ud fra den leksikologiske analyse og de leksikografiske behov. Vi har med vilje ikke lagt os fast på en bestemt brugertype, se dog sammenfatningen.

Som vores ordstyrer under foredraget i Akureyri, Anna Helga Hannesdóttir, med rette påpegede, har vores eksempler hidtil været rent semasiologiske. En videnskabelig tosproget idiomordbog med produktion for øje kan imidlertid desuden indeholde et tesauruslignende indholdsregister med opslagsbegreber på brugerens modersmål, som henviser til fraseologiske (herunder kollokationelle) og monoleksikalske data i hovedlisten (jf. fx Schemann 1991). Eksempler for en

dansk bruger kunne være 1. VREDE, 2. ANGST, 3. DØ, 4. PINLIGHED, 5. TJENE PENGE, 6. FÅ SEX, 7. HAVE PROBLEMER, 8. VÆRE BERUSET, 9. SNYDE, 10. GÅ I SKOLE etc. Til disse indholdsbegreber, som skulle differentieres yderligere, hører bl.a. følgende fraseologismer: 1. *auf die Palme steigen*, 2. *Schiss haben*, 3. *das Zeitliche segnen, keine Steuern mehr zahlen, den Löffel abgeben*, 4. *rote Ohren haben*, 5. *Geld machen*, 6. *eine Nummer abziehen*, 7. *in der Zange sitzen, unter Druck stehen*, 8. *einen Affen sitzen haben, eine Fahne haben*, 9. *jn. über den Tisch ziehen, jm. das Fell über die Ohren ziehen*, 10. *die Schulbank drücken, die Schule besuchen* etc.

## 6. Sammenfatning

Vi har set nærmere på, hvilke kriterier, man med en vis rimelighed kan anvende som målestok for at kalde en publikation "en videnskabelig flersproget idiomordbog". Vi har understreget, at vi i udgangspunktet er kritiske over for brugerrelevansen af en sådan selvstændig ordbog, fordi idiommer efter vores opfattelse er et ordforrådsfænomen og derfor bør integreres i en almen tosproget ordforrådsordbog, sådan som det heldigvis også er blevet mere og mere almindeligt. Endnu mindre brugerrelevans har polylingvale idiomordbøger, hvis man ser bort fra leksikotainmentfunktionen.

På den anden side er idiommer af en sådan kompleksitet og subtilitet, at der er gode *pragmatiske* argumenter for et tosproget idiomopslagsværk. Man kan nemlig realistisk set ikke forvente, at bilingvale almenordbøger kan løfte den tunge leksikologiske opgave, det er at undersøge og om muligt sætte to sprogs idiominventarer i relation til hinanden og derefter formidle resultaterne leksikografisk. Derfor kan den grundforskningsmæssige værdi af et sådant projekt alligevel være højt, fordi det kan smitte af på resten af den leksikografiske fødekæde.

Set i det lys får en bilingval idiomordbog imidlertid en anden brugerprofil end den gennemsnitlige fremmedsprogsrecipient og -producent med kommunikative behov. I stedet bliver det fire alternative hovedgrupper: (1) andre leksikografer, (2) professionelle sprogbrugere med særlige kommunikative problemer, (3) sprogforskere, (4) sprogvidenskabeligt skolede personer med særlig interesse for idiomatik. Denne væsentlige indskrænkning af brugertypologien er ikke noget problem, men snarere nyttig for udviklingen af et bæredygtigt leksikografisk koncept. Det er nemlig en stor fordel ikke at skulle betænke lørnere og andre brugere uden særlige forkundskaber.

Vi er nået frem til en række punkter, som efter vores opfattelse kendetegner den videnskabelige flersprogede idiomordbog:

- den har en klar brugerprofil, forankret i et klart leksikografisk koncept,
- den er udarbejdet af grundigt skolede leksikografer med solid fraseologisk indsigt,



- den bygger på velfunderede kontrastive korpuslingvistiske undersøgelser i begge sprog,
- den har kontrastivt udarbejdede leksikografiske data, der er baseret på et ækvivalensbegreb med klare kriterier for ækvivalens, herunder en specifikation af analysens granularitet,
- dens primære mål er klart: enten kommunikationsunderstøttelse (udgangspunkt: sprogtegnets funktion) eller udtryksorienteret filologi (dvs. i videst mulige at matche ækvivalente idiommer i de to sprog) – eller begge dele.

At forskellige leksikografiske teams nok vil nå frem til forskellige løsninger på opgaven er ikke mærkeligt, de mange leksikografiske fremstillingsmuligheder taget i betragtning. Til gengæld burde den leksikologiske side af arbejdet kunne opnå en høj grad af intersubjektivitet, hvis grundlaget er i orden.

Videnskabelig leksikografi er en kompleks størrelse, videnskabelig flersproget idiomleksikografi endnu mere kompleks. Fokuseringen på idiomets ikonografi og komposition i leksikografien er udbredt og forståelig, og den betyder også noget nogle gange, men den er ikke nødvendigvis og til enhver tid hovedsagen. De afgørende instanser for en videnskabelig flersproget idiomordbog – i det omfang genren er relevant – er sprogbrugen og de leksikografiske funktioner i et subtilt samspil med hinanden.

## Litteratur

### Ordbøger

- Andersen, Stig Toftgaard 2001<sup>2</sup>: *Talemåder i dansk. Ordbog over idiommer*. København: Gyldendal.
- Bergenholtz, Henning/Vrang, Vibeke 2007. *Ordbogen over Faste Vendinger*. Aarhus. [www.idiomordbogen.dk](http://www.idiomordbogen.dk)
- Blum, Geneviève/Salas, Nestor 1989: *Les Idiomatiques français-allemand*. Paris: Seuil.
- EPHRAS = Jesenšek, Vida m.fl. (red.): *EPHRAS - Ein mehrsprachiges phraseologisches Lernmaterial auf CD-ROM*. Maribor. [www.ephras.org](http://www.ephras.org).
- Fiedler, Sabine/Rak, Pavel 2004: *Ilustrita frazeologio*. Dobrichovice: KAVA-PECH.
- Langenscheidt-Redaktion (udg.): *Langenscheidt e-Handwörterbuch Deutsch-Englisch. Phrasen.com – Wörterbuch der Redewendungen*. [www.phrasen.com](http://www.phrasen.com)
- Schemann, Hans 1991: *Synonymwörterbuch der deutschen Redensarten*. Stuttgart: Klett.
- Steyer, Kathrin/Lauer, Meike 2006: *Wortverbindungen online*. <http://hypermedia.ids-mannheim.de/elexiko/wv-online/>

### Anden litteratur

- Almind, Richard/Bergenholtz, Henning/Vrang, Vibeke 2006: Theoretical and Computational Solutions for Phraseological Lexicography. I: *Linguistik-online* 27, 2/06, 159–181. [www.linguistik-online.de/27\\_06/almind\\_et\\_al.pdf](http://www.linguistik-online.de/27_06/almind_et_al.pdf).

- Arsentyeva, Elena 2007: Ways of Translation of Phraseological Units in a Multilingual Phraseological Dictionary. I: Annelies Häcki Buhofer og Harald Burger: *Phraseology in Motion II. Theorie und Anwendung*. Baltmannsweiler: Schneider Verlag Hohengehren, 287–292.
- Burger, Harald/Dobrovolskij, Dmitrij/Kühn, Peter/Norrick, Neal (udg.) 2006: *Phraseologie/Phraseology*. (HSK). Berlin/New York: de Gruyter. Bd1.
- Cruse, D. Allan/Croft, William 2004: *Cognitive Linguistics*. Cambridge: CUP.
- Farø, Ken 2005a: Hvornår går man over åen efter vand? Idiomaticke ækvivalensproblemer i leksikologi og leksikografi. I: *LexicoNordica* 11, 85–108.
- Farø, Ken 2005b: Idiommer på nettet: Den Danske Idiomordbog og fraseografien. I: *Hermes* 32, 201–235.
- Farø, Ken 2006a: *Idiomatizität - Ikonizität – Arbitrarität. Beitrag zu einer funktionalistischen Theorie der Idiomaquivalenz* (upublic. ph.d.-afhandling). København: Københavns Universitet.
- Farø, Ken 2006b: Ikonographie, Ikonizität und Ikonizismus. Drei Begriffe und ihre Bedeutung für die Phraseologieforschung. I: *Linguistik online* 27, 2/06, 57–71. [http://www.linguistik-online.de/27\\_06/faroe.pdf](http://www.linguistik-online.de/27_06/faroe.pdf).
- Farø, Ken/Gottlieb, Henrik 2007: Socioleksikografi, eller: „Det sekundære og afledte er ordbøgerne“. Ny disputats om leksikografi. I: *LexicoNordica* 14.
- Fellbaum, Christiane/Kramer, Undine/Neumann, Gerald 2006: Corpusbaserede lexicographische Erfassung und linguistische Analyse deutscher Idiome. I: Annelies Häcki Buhofer og Harald Burger (udg.): *Phraseology in Motion I – Methoden und Kritik*. Baltmannsweiler: Schneider Verlag Hohengehren, 43–56.
- Fellbaum, Christiane/Kramer, Undine/Neumann, Gerald 2006: Corpusbaserede lexicographische Erfassung und linguistische Analyse deutscher Idiome. I: Annelies Häcki Buhofer og Harald Burger (udg.): *Phraseology in Motion I – Methoden und Kritik*. Baltmannsweiler: Schneider Verlag Hohengehren, 43–56.
- Földes, Csaba 1994: Mehrsprachige phraseologische Wörterbücher - als Herausforderung für die Lexikographie. I: Barbara Sandig (udg.): *EUROPHRAS 1992. Tendenzen der Phraseologieforschung*. Bochum: Universitätsverlag Brockmeyer, 175–201.
- Gréciano, Gertrud 2005: *Phraseographische Prioritäten, erfüllt und unerfüllt*. I: Henrik Gottlieb, Jens Erik Mogensen og Arne Zettersten (udg.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen*. Tübingen: Max Niemeyer, 59–72.
- Hallsteinsdóttir, Erla 1997: Aspekter der Übersetzung von Phraseologismen am Beispiel Isländisch - Deutsch. I: Eberhard Fleischmann, Wladimir Kutz og Peter A. Schmitt (udg.): *Translationsdidaktik: Grundfragen der Übersetzungswissenschaft*. Tübingen: Gunter Narr Verlag, 561–569.
- Hallsteinsdóttir, Erla 2001: *Das Verstehen idiomatischer Phraseologismen in der Fremdsprache Deutsch*. Hamburg: Verlag Dr. Kovac.
- Hallsteinsdóttir, Erla 2006: Phraseographie. I: *HERMES* (36) 2006, 91–128.
- Hallsteinsdóttir, Erla 2007: Wörtliche, freie und phraseologische Bedeutung. Eine korpusbaserede Untersuchung des Vorkommens von freien und phraseologischen Lesarten bei deutschen Idiomen. I: Erika Kržišnik (udg.): *Phraseologie in der Sprachwis-*

- senschaft und anderen Disziplinen (Akten der Europhras-Tagung in Strunjan/Slowenien, 19.–22. September 2005)*. Ljubljana: Universitet Ljubljana, 107–121.
- Hallsteinsdóttir, Erla/Farø, Ken 2006: Neue theoretische und methodische Ansätze in der Phraseologieforschung. Vorwort zu *Linguistik online* 27, 2/06. I: *Linguistik online* 27, 2/06, 3–10. [http://www.linguistik-online.de/27\\_06/einleitung.pdf](http://www.linguistik-online.de/27_06/einleitung.pdf).
- Hanks Patrick/ Urbschat, Anne/Gehweiler, Elke 2006: German Light Verb Constructions in Corpora and Dictionaries I: *International Journal of Lexicography* 19, 439–457.
- Jesenšek, Vida 2006: Phraseologie und Fremdsprachenlernen. Zur Problematik einer angemessenen phraseodidaktischen Umsetzung. I: *Linguistik online* 27, 2/06, 137–147. [http://www.linguistik-online.de/27\\_06/jerensek.pdf](http://www.linguistik-online.de/27_06/jerensek.pdf).
- Jyllandsposten*, 27.02.06, 2.
- Lakoff, George/Johnson, Mark 1980: *Metaphors we live by*. Chicago: University of Chicago Press.
- Pacholski, Robert/ Laskowski Marek 2007: EAR, OHR AND UCHO AS PHRASEOLOGICAL ELEMENTS IN ENGLISH, GERMAN AND POLISH PHRASEOLOGY. Descriptive, Contrastive, Psycholinguistic, Lexicographical Approaches and their Educational Implications. I: Erika Kržišnik (udg.): *Phraseologie in der Sprachwissenschaft und anderen Disziplinen (Akten der Europhras-Tagung in Strunjan/Slowenien, 19.–22. September 2005)*. Ljubljana: Universitet Ljubljana.
- Pamies, Antonio/Balmacz, Martina/Iñesta, Eva 1997: *Autofras. An Onomasiological Tool for Multi-Lingual Phraseology*. I: Peter Durco (udg.): *Europhras 97. Phraseology and Paremiology. International Symposium September 2–5, 1997*. Bratislava: Akadémia PZ, 267–275.
- Schemann 1991
- Steyer, Kathrin/Lauer, Meike 2006: Elektronische Mehrwortartikel. Korpusgesteuerte Erarbeitung, Artikelstruktur und Online-Präsentation. I: *Wortverbindungen online*. [www.elexiko.de](http://www.elexiko.de).
- Tarp, Sven 2006: *Leksikografi i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særligt henblik på lærerleksikografi. Doktorafhandling ved Handelshøjskolen i Aarhus*.
- Villavicencio, Aline/Baldwin, Timothy/Waldron, Benjamin 2004: A Multilingual Database of Idioms. I: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon, 1127–30.

Ken Farø  
 forskningsadjunkt ph.d., f. 1972  
 Institut for Engelsk, Germansk og Romansk, Københavns Universitet  
 Njalsgade 120  
 DK-2300 København  
 kenfaroe@hum.ku.dk

Erla Hallsteinsdóttir  
 dr. phil., f. 1970)  
 Institut for Sprog og Kommunikation, Syddansk Universitet Odense  
 erla@language.sdu.dk



RUTH VATVEDT FJELD

## Plan for leksikalsk dokumentasjon av moderne bokmål

The lexicon of the modern Norwegian bokmål standard needs a better description and documentation than what is the situation today. The article presents a plan for building a modern lexical database based on a balanced corpus of 40 million words of modern bokmål. This base should serve as a source for a traditional scientific dictionary as well as a dictionary for language technological applications.

### 1. Bakgrunn

Ordforrådet i moderne norsk bokmål er ikke godt dokumentert. Det nynorske ordforrådet blir godt dokumentert, både det historiske og det samtidige, i prosjektet *Norsk Ordbok: Ordbok over det norske folkemålet og det nynorske skriftmålet*, med Alf Hellevik som hovedredaktør for band I og II, fra band III Lars S. Vikør, fra band VI utvidet med Oddrun Grønvik og Laurits Killingbergtrø. Første band i prosjektet ble publisert i 1966, og planen er at siste band skal komme til det norske tohundreårsjubileumet i 2014. Med over tretti redaktører og god finansiering er det sannsynlig at det målet nås. Begrunnelsen for et så stort prosjekt for denne minoritetsvarieteteten av norsk språk er blant annet at nynorsk må dokumenteres ut fra tekster og språkbruk der nynorsk er brukt (Skard 1932, Hageberg 1992). Tidligere er det gitt ut flere ordbøker over nynorsk, først og fremst Ivar Aasens ordbøker fra 1850 og 1873, senere utvidet og oppdatert med ordbøkene til Hans Ross (1895) og Steinar Schjøtt (1914), som var tillegg til Aasens verk. Disse ordbøkene danner grunnstammen i Norsk Ordbok, sammen med et stort innsamlet dialektmateriale og et moderne elektronisk korpus.

Ordbokstradisjonen for bokmålet er magrere, og det er synd, siden det brukes av ca. 85-90 prosent av alle nordmenn ifølge Språkrådets beregninger. Bakgrunnen er bl.a. at bokmål/riksmål lenge var omtrent det samme som dansk, og det danske ordforrådet var godt dokumentert i *Ordbog over det Danske Sprog* og andre ordbøker. En av de første ordbøkene som forsøker å dokumentere norsk bokmål, er skrevet av Knud Knudsen, han som også kalles bokmåletts far. Ordboka *Unorsk og norsk, eller Fremmedords Afløsning* fra 1881 er imidlertid en svært spesiell ordbok, der puristen Knudsen selv lager mange nye norske ord som avløser for internasjonale ord og også for tyske lån i norsk. Svært få av disse ordene er kommet i vanlig bruk, og ordboka er nærmest å regne som en kuriositet i dag.

Norsk riksmål og offisielt bokmål skilte lag i 1939 som følge av den offentlige

språkrøktens mål om å føre bokmål og nynorsk sammen i ett skriftspråk. Mange motsatte seg en slik styring av språket og lanserte *riksmål* som en uoffisiell variant ved siden av det offisielle bokmålet. Riksmål er en konservativ skriftnorm fastlagt av Det Norske Akademi for Sprog og litteratur, og det er den som leksikografisk sett er best dokumentert. Først med *Norsk Riksmålsordbok* av Trygve Knudsen, Alf Sommerfeldt og Harald Noreng m.fl., som kom ut i perioden 1937–57, fikk vi en egen norsk allmennordbok som ikke dokumenterte nynorsk. Ordboka dokumenterer først og fremst det som nå kalles riksmålstradisjonen slik den er anvendt hos de kjente norske dikterne som brukte riksmål. Men også en del avis-språk og annet skriftspråk ble dokumentert. Men det er helt klart en selektiv og normativ ordbok der store deler av det offisielle bokmålet ble utelatt. Den er utgitt av Riksmålsvernet. Nesten førti år seinere kom Harald Norengs to tilleggsbind, *Norsk riksmålsordbok, tilleggsbind I og II* (1995), som i seg selv er et storverk laget av én mann alene, og som dokumenterer det moderne riksmålet noe bredere og mindre normativt enn den tradisjonelle riksmålsstandard. Men de ble gitt ut av det språkideologiske Norsk Akademi for Sprog og Litteratur, og lemma- og redigeringspråket er derfor riksmål.

I mellomtiden hadde Universitetet i Oslo i 1986 gitt ut de to håndordbøkene *Bokmålsordboka* og *Nynorskordboka* i samarbeid med Norsk språkråd. De er gode dokumentasjoner av det sentrale ordforrådet i moderne bokmål og nynorsk, men dekker selvsagt ikke hele ordforrådet. De er klart normative.

## 2. Behov

Ordforrådet i et språk er i stadig utvikling, og det trengs kontinuerlig oppdatert utforskning og dokumentasjon av ordenes betydning, bøyning, kombinasjonsegenskaper og annen bruk. En forskningsbasert ordboksbase kan både fylle behovet for ordkunnskap som understøtter kommunikasjon mellom brukerne, og dokumentere den faktiske utviklingen og bruken, slik at man har et faglig velfundert utgangspunkt for den normering og språkplanlegging som foretas av Språkrådet, samt gi grunnlag for annen språkforskning innen semantikk, morfologi og syntaks. Særlig er det interessant for et språk der normen gir stor valgfrihet, slik som i norsk, da man ikke vet noe systematisk om hvordan denne vide normen utnyttes. I tillegg til Norsk Ordbok trenger vi i Norge derfor en vitenskapelig basert dokumentasjon av hele ordforrådet i moderne bokmål, både slik det er normert av Språkrådet og vedtatt av Kulturdepartementet, og slik det faktisk brukes både i offentlig sammenheng og i privat språk som ikke er underlagt den offisielle normen, bygd på observasjon av faktisk språkbruk. Språkrådets melding ”Norsk i hundre!” fra 2005 sier da også: “Norge mangler store ordbøker over samtids-språket som tilsvarende man har i de andre nordiske landene” (2005:145).

I tillegg er det et dokumentert behov for leksikalsk folkeopplysning om normert bokmål. Det vet vi blant annet fordi det i gjennomsnitt hver dag er ca. 25 000

søk i Bokmålsordboka på Internett. En enda bedre og mer utfyllende ordbok ville rimeligvis bli brukt av enda flere.

Videre er det et stort behov for å utvikle norsk datamaskinell leksikografi. For automatisk språkanalyse er det nødvendig med et veldefinert og automatisk lesbart leksikon. Det er blant annet nødvendig for automatisk syntaktisk analyse og automatisk ordklassemerking av løpende tekster. Maskinlesbare ordbøker brukes i utvikling av automatiske oversettingsprogram og flerspråklig informasjonssøking. Det er viktig å sikre norsk som nasjonalspråk ved å delta i den internasjonale utvikling av oversetting av mellomstatlig regelverk og andre språkfestede standarder. Men det krever veldefinerte, maskinlesbare leksikon, og det er nødvendig at en utforskning av norsk ordforråd tilpasses internasjonale språkteknologiske produkter. Systematisk gjennomgang av ordforrådet for en større ordbok gjør det mulig samtidig å trekke ut informasjon for en språkteknologisk leksikalsk beskrivelse uten store merkostnader. Arbeidet med de to ordboksproduktene vil gi en synergieffekt. EU-prosjektet SIMPLE (Semantic Information in Plurilingual Lexicons, jf. Lenci et. al. 2000) er et internasjonalt prosjekt for utvikling av et datamaskinlesbart leksikon over 12 europeiske språk til bruk i språkteknologi. En norsk variant av det danske SIMPLE-leksikonet er utviklet som et forsøksprosjekt, og et slikt leksikon kan videreutvikles og utvides i samsvar med den internasjonale forskningen innenfor LBD-prosjektet (Leksikografisk bokmålsdatabase). SIMPLE-leksikonet bygger på Pustejovskys semantiske modell for leksikalsk beskrivelse av nominale ledd ved hjelp av kvaliastruktur (Pustejovsky 1995) og Levins semantiske klassifisering av verb (Levin 1993). Disse teoriene er videreutviklet og tilpasset beskrivelse innenfor SPINN-nettverket (Pedersen, B, R.V.Fjeld, M.Toporowska Gronostaj 2002). Denne leksikalske beskrivelsen er imidlertid fortsatt på utprøvningsstadiet, da mer fullstendig dekning av det norske ordforrådet krever større ressurser enn det som er tilgjengelig.

### 3. Ressurser

Det er utviklet en rekke bokmålsressurser for leksikografisk utnyttning og videre forskning:

#### *Ekserptsamling og elektronisk tekstsamling:*

Store deler av den norske litterære arven fra perioden 1550–1950 er dokumentert i ekserptsamlingen for det ufullførte prosjektet *Det norske litterære ordboksverk*. Deler av dette er gjort elektronisk tilgjengelig ved at man i Dokumentasjonsprosjektet skannet inn mange av de samme verkene, i tillegg til ikke-ekserperte verk fra samme periode. Det elektroniske materialet utgjør 46 840 sider tekst, og er tilgjengelig på <http://www.dokpro.uio.no/litteratur/>.

*Nyordsmaterialet*

Nyordsmaterialet er en elektronisk base av ekserpter fra norske aviser for perioden 1972–2000. Ekserptene er kodet med ordklasse og med en rekke tilleggs-koder for etymologiske, syntaktiske og semantiske egenskaper. Det består av over 500 000 ekserpter og er tilgjengelig på [http://www.dokpro.uio.no/bokmaal/nyord/nyord\\_ramme.html](http://www.dokpro.uio.no/bokmaal/nyord/nyord_ramme.html).

*Bokmålsordboka*

Bokmålsordboka er den største norske definisjonsordboka som også dokumenterer hele den offisielle bokmåls morfologien. Den blir stadig oppdatert med nye ord hentet fra moderne avistekster og annet materiale, og formverket er justert i forhold til de siste vedtak som er godkjent av Kirke- og kulturdepartementet. Den er tilgjengelig på <http://www.dokpro.uio.no/ordboeker.html>.

*Nyord i norsk*

Dette er en ordbok som følger opp *Nyord i norsk 1945–75* med nyord fra perioden 1975–2003. Den bygger særlig på innholdet i basen *Nyordsmaterialet*. Ordboka er under utgivelse og er et samarbeidsprosjekt med Språkrådet i Norge.

*Leksikalsk bokmålskorpus (LBK)*

LBK er det første balanserte tekstkorpus for moderne norsk og består av litterære tekster og sakprosa, samt normert og unormert litteratur fra perioden 1985 til i dag. Korpuset er søkbart i en database etter grunnord/lemma (leksikaliserte ord eller uttrykk), og gir opplysninger om forfatter, teksttype, kontekst og ordenes grammatiske egenskaper. Korpuset inneholder 40 mill. ord. Det er balansert slik:

- Periodika            20%
- Sakprosa            45 %
- Skjønnlitt        25%
- TV-tekst            5%
- Upublisert        5%

Disse kategoriene er inndelt i spesifiserte teksttyper som også holdes best mulig balansert. Beregningsgrunnlaget for balanseringen er hentet fra Norsk Mediebarometer, som angir folks lesevaner fordelt omtrent slik:

- Internett            35%
- Bøker                15%
- Avis                 40%
- Tidsskrift        5%
- Tegneserie        5%



Korpuset skal dermed dekke omtrent det en gjennomsnittsnordmann møter av ord i vår tid.

#### *Norsk ordbank*

Norsk ordbank er utviklet i samarbeid med Språkrådet, Dokumentasjonsprosjektet/EDD og Tekstlaboratoriet ved Humanistisk Fakultet. Det er en leksikalsk database der lemmalisten med ortografi og morfologi fra Bokmålsordboka utgjorde grunnstammen og seinere ble utvidet, bl.a. med ordlister utarbeidet av IBM (jf. Engh 1992). Lemmaene er tillagt formalismer slik at ordbanken i dag framstår som en maskinleselig base over de aller fleste norske bokmålsordene. Alle ordene har formalisert beskrivelse av morfologiske og ortografiske egenskaper.

#### *Norsk aviskorpus, bokmålsdelen*

Norsk aviskorpus er det største tekstkorpuset for norsk. Det ble påbegynt i 1998 og består av tekstmateriale fra nettutgavene av utvalgte riks- og regionaviser. Det svært omfattende materialet har i 2008 passert 500 millioner ord og er det største i sitt slag i Norge. I prosjektet bygges det også opp en nyordsdatabase som vokser i omfang hver dag. Dagsaktuelle nyordslister klassifiseres automatisk og gjøres tilgjengelige for forskere og leksikografer. Hjemmeside: <http://avis.uib.no/>

#### *NoTa- Oslo (Norsk talespråskorpus, Oslo-delen)*

Ved Tekstlaboratoriet ved Universitetet i Oslo er det utarbeidet et talespråskorpus for flerbruk. Dette er landets største og best merkede korpus for moderne norsk talemål ca. 2005. Oslo-delen inneholder ca. 1 mill. ord som er ortografisk transkribert med lenker til video- og lydfiler. Ordene er grammatisk tagget. Dette korpuset er et verdifullt nytt grunnlag for nærmere utforskning av bokmålets leksikon. Korpuset ligger på: <http://www.tekstlab.uio.no/nota/>

#### *Den Danske Ordbog (DDO)*

Den Danske Ordbog er en ordbok i seks bind over det danske nåtidsspråket (utgitt 2003–05). Ordboken gir en grundig beskrivelse av ordforrådet i moderne dansk i perioden fra ca. 1950 frem til i dag og er en etterfølger til den store *Ordbog over det danske Sprog*. Den Danske Ordbog inneholder systematiserte opplysninger om oppslagsordene og bygger på et moderne dansk korpus. Ordboken regnes som den mest nyskapende og fullstendige beskrivelsen av et skandinavisk språk. Det norske og det danske språket har så mye felles at denne ordboken fungerer som et nyttig sammenlikningsgrunnlag og som en idéskaper for et prosjekt for moderne norsk bokmål.

Av ressurser som er under utvikling, er følgende særlig relevant for et prosjekt som LBD:

*Datamaskinelt lesbart leksikon*

På basis av det danske SIMPLE-leksikonet er det gjort forsøk med en norsk parallellkopi ved å legge inn norske ekvivalenter for de danske oppslagene, såkalte SemU-er, som tilsvarer delbetydninger, og legge til brukseksempler fra norske tekster. Alle oppslagsordene i SIMPLE er koblet til samme lemmas delbetydning i Bokmålsordboka i en database, slik at den norske morfologien og definisjonene der kan hentes inn for hvert enkelt oppslag. Ordene i leksikonet er markert med fire ontologiske strukturer (jf. Pedersen 2005 og Pedersen & al. 2002). Dette skal gjøre at en datamaskin med ganske stor nøyaktighet kan koble f.eks. et engelsk eller spansk ord til ekvivalenten i et av de skandinaviske språkene. Likeledes kan den koble rett ekvivalent mellom de skandinaviske språkene. Arbeidet har ført til en grundig ekvivalensanalyse mellom norsk og dansk for de ferdige postene, og viser falske venner mellom de to språkene som neppe har vært registrert tidligere.

Å skrive slike ordboksartikler kan være tidkrevende. Det forutsetter kunnskap om semantiske strukturer og plassering i ontologisk hierarki som ikke er utforsket før. Dessuten må alle definisjoner ha et stramt strukturert format som gjør det nødvendig å ta inn informasjon som for menneskelige lesere er "selvfølgeligheter", f.eks. at en rose er en blomst, og at en blomst er en plante. Fullføring av dette prosjektet forutsetter mer kunnskap om det norske ordforrådet og utvikling av et formalisert beskrivelsesapparat. Et språkteknologisk leksikon vil gi grunnlag for en systematisk og enhetlig beskrivelse av det norske bokmålets semantikk, som er en forutsetning for de fleste former for moderne språkteknologi.

Hittil har det norske leksikonet 13 434 innganger som er lenket til danske og svenske oppslag. I Danmark har man investert i en mer fullstendig språkteknologisk ordbok STO (<http://cst.dk/cgi-bin/defisto/defisto>), som kan vise seg å være enda bedre egnet som modell for en norsk språkteknologisk ordbok.

*Norsk Ordbok 2014*

Det leksikografiske fagmiljøet ved Universitetet i Oslo har fått et betydelig løft i og med særbevilgningene til Norsk Ordbok 2014. Mye av det arbeidet som legges ned i dette prosjektet, kan gjenbrukes i beskrivelsen av bokmålet, og et større bokmålsprosjekt kan gi synergieffekt tilbake til Norsk Ordbok. Samarbeidet med redaksjonen i Norsk Ordbok har hele tiden vært godt, og det er tjenlig å fortsette dette samarbeidet. Det er en klar fordel at ordforrådet i nynorsk og bokmål utforskes og dokumenteres ved samme enhet, slik at fagmiljøet i dokumenterende leksikografi er samlokalisert og kan samarbeide mest mulig rasjonelt.

*Nationalencyklopedins ordbok (NEO) og Svensk ordbok utgiven av Svenska Akademien*

Ved Språkdata i Göteborg ble Nationalencyklopedins ordbok utviklet i perioden

1995–96. Med utgangspunkt i den arbeides det videre for å utvikle Svensk ordbok utgiven av Svenska Akademien. Den skal bygge på et egenutviklet korpus, men den har også historisk informasjon med belegg helt tilbake til runetiden. Prosjektet skal ha mer systematisk angivelse av kollokasjoner og idiomer og fullstendige valensangivelser. Dette arbeidet vil ha stor relevans for utarbeiding av en omfattende ordbok for norsk bokmål.

#### 4. Anvendelse

Den leksikalske databasen LBD skal danne grunnlag for forskjellige ordbøker, i første omgang er det planlagt en stor elektronisk ordbok som dokumenterer moderne bokmål, og en språkteknologisk ordbok.

##### *Norsk elektronisk bokmålsordbok (NEBO)*

NEBO skal gi utfyllende informasjon om norske ords ortografi, bøyning og betydning slik det er normert, og en publisering av Språkrådets vedtak om den offisielle normen. Målgruppen er journalister, forfattere, oversettere, lærere og andre språkinteresserte. Ordboka skal være normativ og til en viss grad deskriptiv, slik at den gjør rede for mye brukte former som ligger utenfor normen, med henvisning til de normerte formene. NEBO skal også dokumentere faste uttrykk eller flerordslemmaer (leksikaliserte kollokasjoner), noe som mangler i norsk leksikografi, men som man har utforsket og dokumentert i de fleste andre språk.

##### *Språkteknologisk bokmålsordbok (STEBO)*

STEBO vil bli en enspråklig maskinlesbar ordbok med formalisert beskrivelse av de semantiske relasjonene mellom ordene, samt de syntaktiske egenskaper enkeltordene har. En språkteknologisk ordbok med god dekning av hele ordforrådet er hittil ikke utviklet for norsk. Som enspråklig base har den mange anvendelsesmuligheter:

- Til informasjonsfinning ved hjelp av søkemotorer er en ontologisk ordnet ordbase viktig.
- I dataprogrammer som genererer naturlig språk er en maskinlesbar leksikalsk komponent uunnværlig. Det gjelder særlig for taleteknologiske programmer der ordbøker med variantuttaler er maskinlesbart tilgjengelig.
- I dataprogrammer som lager skriveveiledning er det nødvendig med veldefinert leksikon.
- I dataprogrammer for T9-språk i mobiltelefoner er frekvensberegnet leksikon en forutsetning. De eksisterende leksikon har for dårlig kunnskap om frekvens i norsk, og kan forbedres betraktelig med mer nøyaktig kunnskap om bruken av norske ord.
- I rettskrivingsprogrammer (allmenne rettskrivingsprogrammer og spesielt

tilrettelagte for lese- og skrivesvake) trengs kunnskap om vanlige variantformer/feil i rettskriving og bøyning for å gi veiledning om rett form.

#### *Andre ordbøker*

Ut fra en utfyllende og godt organisert leksikalsk database der alle informasjonstyper er sortert og merket, kan man lage en lang rekke mer spesialiserte ordbøker, f.eks. rettskrivingsordbok, fremmedordbok, lånordbok, skoleordbok, barneordbok, slangordbok m.v.

En leksikalsk database vil også være et viktig grunnlag for tospråklige ordbøker der bokmål er det ene språket. Databasen er dessuten viktig som grunnlag for annen leksikalsk forskning, både innen morfologi, orddanning, syntaks og semantikk. Det vil særlig gjelde revisjonen av *Norsk referansegrammatikk (NRG)*. Det vil gi en ny mulighet til innsikt i leksikalske og syntaktiske beskrivelser av moderne norsk språk.

## 5. Prosjektplan

### *Ontologi*

Som et første skritt i selve ordboksarbeidet skal det genereres en ontologi ut fra Bokmålsordboka. En ontologi er et begrepssystem der alle ordene er ordnet etter hvordan det som ordene betegner (referentene), forholder seg til hverandre. Innenfor SIMPLE-prosjektet er ordene allerede ordnet ontologisk, og til en viss grad kan det bygges videre på den. Det er dessuten gjort forsøk på automatisk generering av en ontologi for norsk bokmål (DOXA, Nygaard 2006) som et ordnett ekstrahert ut fra ordklassetaggning av definisjonstekstene i Bokmålsordboka. Dette programmet skal prøves ut i full skala for hele lemmautvalget i Bokmålsordboka. En ontologi som er produsert automatisk, må etterprøves manuelt, og det gjelder også jamføringen med SIMPLE-ontologien, men dette er et nødvendig arbeid for at ordboksarbeidet skal kunne gjennomføres systematisk og semasiologisk og ikke i den tilfeldige og alfabetiserte onomasiologiske rekkefølgen. En videre utbygging av ontologien skal også jamføres med det danske OntoQuery-prosjektet (<http://www.ontoquery.dk/index.php>), slik at den norske beskrivelsen harmonerer med internasjonale beskrivelser.

### *Ordboksbasis: Leksikografisk bokmålskorpus LBK*

LBK er et balansert korpus over bokmålstekster fra 1985 til nåtiden. Den inneholder 40 mill. ord, og hvert ord er merket med ordklasse av en automatisk ordklassetagger. Tekstene er manuelt merket med teksttype, sjanger, domene og forfatter, samt sosiologiske variabler knyttet til forfatter. Korpuset vil således gi en god oversikt over ordbruken i forskjellige sammenhenger og gi grunnlag for en utfyllende dokumentasjon av hvordan normen anvendes eller brytes. Dermed

vil den leksikalske databasen også være godt grunnlag for normeringsarbeid og for språksosiologisk forskning.

I tillegg til LBK-korpuset skal talespråkskorpuset NoTa, Norsk aviskorpus AVIT, Norsk riksmålsordbok og Norsk Ordbok brukes som ordboksbasis for dokumentasjon av bruken av norske ord og uttrykk.

#### *Lemmutvalg*

Grunnlaget for de ordene som skal defineres og beskrives i prosjektet, hentes fra Bokmålsordboka. Den har en ordboksbasis som er stadig utvidet og oppdatert fra førsteutgaven i 1986, og er således den mest utfyllende beskrivelse vi har av det standardiserte bokmålet. Dette ordforrådet skal suppleres med de ordene som fins i LBK-korpuset og i andre kilder og korpus som er tilgjengelig, og som ikke er med i Bokmålsordboka. Ut fra korpuset lages det en konkordans (alfabetisert liste over alle ord som fins i tekstsamlingen). Ordlista i Bokmålsordboka kjøres mot konkordansen fra LBK, slik at vi får en liste over de leksikalske enhetene som fins i korpuset, men ikke i Bokmålsordboka, og omvendt. Det samme vil vi gjøre med andre tilgjengelige korpus, som NoTa og Norsk Aviskorpus. Disse lemmakandidatene blir undersøkt i forhold til sekundærkildene og eventuelt lagt inn i databasen. Utvalgelse for videre behandling og presentasjon i NEBO er bl.a. avhengig av bruksfrekvens og plass i ontologien. Alle lemmakandidatene skal i første omgang lemmatiseres. Etter nøyere overveielser kan grenseverdier og hapax legomenon strykes, men i prinsippet skal alt ordmateriale som forekommer i LBK-korpuset dokumenteres i den leksikalske databasen.

Med denne metoden får vi også se hvilke lemmaer som fins i Bokmålsordboka men ikke i korpus, og således kartlagt eventuelle lakuner i korpuset. I tillegg til lemmautvalget fra LBK skal det vurderes om det trengs et tilleggsutvalg fra Sekundærkildene, NoTa og Norsk Aviskorpus pga. åpenbare lakuner i korpuset.

#### *Lemmasbeskrivelse og informasjonstyper*

Hvert enkelt lemma i lemmautvalget beskrives med utgangspunkt i definisjonen i Bokmålsordboka, sammenliknet med den i DDO og tilgjengelige ordbøker for norsk. I tillegg søkes eksempler fra LBK og fra suppleringsmaterialet i ordboksbasen.

Lemmasbeskrivelsen skal bestå av artikkelhode, artikkelkropp og artikkelfot.

Artikkelhodet angir lemmategnform og id-nummer for SIMPLE-basen. Deretter følger forskjellige normerte former, samt unormerte, belagte former med kildehenvisninger. Det samme gjelder forskjellige bøyingsformer av den lemmatiserte formen, både normerte og unormerte.

Hver delbetydning skal lemmatiseres som eget oppslag, men slik at sammenhengen mellom delbetydningene går klart fram. Dette er nødvendig for den semantiske beskrivelsen og for at kopling til SIMPLE-beskrivelsene skal bli mulig.

Artikkelkroppen angir definisjonen fra Bokmålsordboka, som da blir delt opp som egne lemmaer for hver deldefinisjon, såkalte Semantic Units (SemU). Den første betydningen skal være grunnbetydningen, det vi kaller den bokstavelige eller konkrete betydningen, deretter følger de forskjellige metaforiske eller avledede betydningene etter et fast mønster. Delbetydningene fra DDO legges inn der det er ekvivalens mellom dansk og norsk for å sikre konvergens med SIMPLE-definisjonene. Alle SemU-ene plasseres i et system som svarer til den ontologiske beskrivelsen med samling av kohyponymer under felles hyperonym. I tillegg angis antonym, synonym og meronym. Videre skal argumentstruktur og kvaliastruktur angis, noe som i praksis ofte vil bli en formalisert gjengivelse av den tradisjonelle delen av artikkelkroppen. For flerordslemmaer angis en henvisning til ettordslemma og kollokasjonell betydning, for øvrig gis de samme informasjonstyper som ved ettordslemmaene.

I artikkelfoten angis lemmaets avledningsmuligheter, leksikaliserte sammensetninger det kan inngå i, og etymologisk informasjon.

#### *Redigeringsverktøy:*

Det fins flere redigeringsverktøy det kan være aktuelt å bruke.

Prosjektet Norsk Ordbok 2014 har sammen med EDD utarbeidet et redigeringsprogram som har vist seg effektivt for det prosjektet. Det er spesielt tilrettelagt for det dokumentasjonsarbeidet som gjøres med nynorsk skrift- og talemål, men det er åpenbart at mye av dette kan gjenbrukes i et bokmålsprosjekt.

I 2004 ble det utarbeidet et redigeringsverktøy og et program for lenking av forskjellige SemU-er i SIMPLE-leksikonet, dvs. en editor som gjør det forholdsvis enkelt å koble de skandinaviske språkene sammen. Det vil da vise seg hvilke delbetydninger/SemU-er for hvert flertydig lemma som er felles for de tre språkene, og det vil gi et godt utgangspunkt for videre utbygging for et mer anvendbart maskinlesbart semantisk leksikon for skandinaviske språk. I denne editoren er også Bokmålsordboka lagt inn for direkte utnyttelse i redigeringsarbeidet. Derved er sammenlikning og avdekking av falske venner mellom de nordiske språkene forholdsvis lett å få med. Den norskutviklede editoren har vakt interesse både i Norden og internasjonalt (Fjeld et. al. 2004).

## 6. Kvalifikasjoner

Leksikografmiljøet ved Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo er det største i Norge og har gode faglige referanser. Miljøet er det eneste i landet som arbeider med teoretisk leksikografi og gir undervisning i emnet. Det er derfor her det er rimelig å forvente teoretisk innsikt og vilje til å gjennomføre et så stort leksikografisk prosjekt. Miljøet har et internasjonalt faglig nettverk og deltar i den generelle fagutviklingen. Ved samme institutt er

Tekstlaboratoriet, som har et godt datateknologisk fagmiljø for språkbehandling og vil være en viktig samarbeidspartner.

## 7. Publisering

LBD skal ligge gratis tilgjengelig på Internett for søk i ordbasen og i korpuset, jf. slik man har valgt for det danske *Ordnet.dk*. Det må utarbeides forskjellige grensesnitt for forskjellige brukergrupper og forskjellige behov.

### *Arbeidskraft til redigering og datamaskinell ekspertise*

Omfanget av databasen er beregnet til ca. 65 000 ordartikler. Om man regner et snitt på tre arbeidstimer pr. oppslagsord, vil det bli 195 000 timeverk. Tidsbruken settes så lavt da man regner med at mye kan hentes fra definisjonene i Bokmålsordboka, DDO og andre ordbøker. Hovedarbeidet vil bestå i å samordne og systematisere tidligere beskrivelser og komplettere og oppdatere dem ut fra funn i korpuset. Den datamaskinelle beskrivelsen vil sannsynligvis gå raskt når man får rutine i å systematisere de nødvendige informasjonene ut fra de tradisjonelle definisjonene. Det trengs datamaskinell ekspertise for redigering og utvikling av databasen og for drift av databasene.

## 8. Litteratur

- Engh, Jan 1992: Leksikografi i IBM Norge. I: Fjeld, R.V. (red.) *Nordiske studier i leksikografi*. Oslo: Nordisk forening for leksikografi, Skrift nr. 1, 409–422 .
- Fjeld, Ruth Vatvedt, Wik, Preben and Nygaard, Lars 2004: Managing complex and multilingual lexical data with the Simple-editor. I: *The Eleventh EURALEX 2004 Proceedings*, Bretagne, 693-69.
- Hageberg, Arnbjörg 1992: Norsk Ordbok – i 1950 eit nybrottsverk, i 1991 ein anakronisme? I: R.V. Fjeld (red.). *Nordiske studier i leksikografi*. Oslo: Nordisk forening for leksikografi, Skrift nr. 1, 235–243.
- Lenci, Alessi et. al. 2000: SIMPLE – A General Framework for the Development of Multilingual Lexicons. I: T. Fontenelle (ed.): *International Journal of Lexicography*. Vol 13, 249–263.
- Levin, B. 1993: *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago
- Norsk i hundre! Norsk som nasjonalspråk i globaliseringens tidsalder. Et forslag til en strategi 2005. Språkrådet.
- Nygaard, Lars (2006): *Frå ordbok til ordnett. Cand.philol.-oppgåve, Institutt for lingvistiske og nordiske studium*. (<http://wo.uio.no/as/WebObjects/theses.woa/wa/these?WORKID=42189>)
- Pedersen, Bolette, V.Fjeld, Ruth, Toporowska Gronostaj, Maria 2002: Harmonisering og sammenkædning af sprogteknologiske ordbaser med særligt henblik på informationssøgning – en rapport fra SPINN-netværket I: *Nordisk Sprogteknologi*. København: 2002, 233–256.

- Pedersen, Bolette 2005: Datamatisk leksikografi i Norden – status og visioner. I: *Nordiske studiar i leksikografi* 7, Volda: Nordisk foreining for leksikografi, 302–314.
- Pustejovsky, James 1995: *The Generative Lexicon*. Massachusetts: 1-298
- Skard, Sigmund 1932: *Norsk Ordbok. Historie – Plan – Arbeidsskipnad*. Oslo

## 9. Omtalte ordbøker og ordboksprosjekt

- Bokmålsordboka* (BOB). 2006. Oslo
- Den Danske Ordbog* (DDO) 2003-2004. København
- Knudsen, Knud. 1881: *Unorsk og norsk eller fremmedords avløsning*. Kristiania
- Nationalencyklopedins ordbok* (NEO). 1995-96. Göteborg
- Norsk Ordbok* (NO). 1966- . Oslo
- Norsk Riksmålsordbok* (NRO). 1937-1957. Oslo
- Norsk Riksmålsordbok* (NRO). 1995, Tilleggsbind I og II. Oslo
- Ordbog over det danske Sprog* (ODS). 1919-1954. København
- Ross, Hans. 1895: *Norsk Ordbog*. 1895-1913. Christiania
- Schjøtt, Steinar. 1914: *Norsk Ordbok*. Oslo
- Aasen, Ivar 1850: *Ordbog over det norske folkesprog*. Christiania
- Aasen, Ivar. 1873: *Norsk Ordbog med dansk forklaring*. Christiania

Ruth Vatvedt Fjeld  
professor, f. 1948  
ILN  
Universitetet i Oslo  
Boks 1001 Blindern  
N-0315 Oslo  
r.e.v.fjeld@iln.uio.no



ALEXANDRA GRANSTRÖM

## Ordböcker och informationsbeteende

### En gränsöverskridande studie i ordböksanvändning

**Dictionaries and information behaviour – a cross boundary study in dictionary use.** *User studies* is a dynamic and well developed field in Library and Information Science (LIS). The research deals principally with the phenomena of *information behaviour* especially concerning information needs, information seeking and information use. This article examines the relation between user studies in LIS and in lexicography by analysing the common user aspects need, seeking and use. A study in dictionary use among librarians forms the empirical foundation for a theoretical discussion about the information behaviour process in relation to dictionary use. The interdisciplinary approach gives a deeper understanding of the dictionary user as an actor in a complex user context. User studies in LIS is an inspiring field for further studies in dictionary use also among other user groups.

#### 1. Inledning

Det lexikografiska arbetet syftar i huvudsak till att dokumentera språket och presentera det i olika slags ordböcker på ett så användarvänligt sätt som möjligt. Således finns en ofrånkomlig komponent att räkna in i projektplanen: ordböksanvändaren. Genom att bedriva forskning kring ordböksanvändning kan det praktiska arbetet med ordböcker utvecklas och ge välbehövlig kunskap om hur användaren tar till sig och uppfattar den information som ordböcker förmedlar. Med andra ord skulle man kunna tala om ordböksanvändning som ett *informationsbeteende*. Detta fenomen undersöks särskilt inom biblioteks- och informationsvetenskapens användarstudier (user studies), ett välutvecklat forskningsfält som för fram många intressanta aspekter på informationsanvändaren i olika kontexter. I föreliggande artikel diskuterar jag ordböksanvändning i relation till informationsbeteende och pekar därmed på de tvärvetenskapliga möjligheterna för den lexikografiska användarforskningen.

#### 2. Vedertagna begrepp

Ett tvärvetenskapligt resonemang innebär en del nya begrepp för respektive fält och föranleder viss definiering av centrala termer. De definitioner jag väljer att lyfta fram är på inget vis allenaordande; inom biblioteks- och informationsvetenskapen (BoI) florerar en mångtydig terminologi kopplad till olika teoretiska och

metodologiska infallsvinklar. Ingången till det vetenskapliga området är därmed inte lätt att begränsa, men ger samtidigt en givande inblick i en dynamik något särskild den lexikografiska användarforskningen.

### 2.1 Informationsbeteende

Att definiera vad informationsbeteende är blir problematiskt redan vid förleden information; att det är ett svåravgränsat begrepp avspeglas också i den BoI-relaterade litteraturen. Enligt Case (2002: 43) är det centrala problemet att samma term används för att beteckna flera olika koncept inom vitt skilda discipliner och därmed inte går att beskriva enligt ett universellt mönster. Det är, som Case påpekar, "important to note that defining information in an absolute and final sense is not entirely necessary for the study of information phenomena to proceed" (2002: 59). Den definition som jag utgår ifrån är hämtad från *Online Dictionary for Library and Information Science* (ODLIS 2004), med följande innebörd av informationsbegreppet:

Data presented in readily comprehensible form to which meaning has been attributed within the context of its use. In a more dynamic sense, the message conveyed by the use of a medium of communication or expression. Whether a specific message is *informative* or not depends in part on the subjective perception of the person receiving it.

Även helhetsbegreppet *informationsbeteende* anges med litet olika termer. Enligt Wilson (1999: 249) förstås informationsbeteende som "those activities a person may engage when identifying his or her own needs for information, searching for such information in any way, and using and transferring that information", en definition som jag understöder. Mer precist syftar mitt val av efterleden *beteende* att förstå användarsituationen som ett komplext agerande i en psykolingvistisk process.

Den definition av information som ODLIS presenterar indikerar särskilda kärnområden för forskningen inom informationsbeteende: informationsbehov, informationssökning och informations-användning. Dessa förklarar Savolainen (2003: 74) som en beteendekedja med "antagandet att fenomen kallade informationsbehov utlöser en informationssökningsprocess som utmynnar i användning eller utnyttjande av anskaffad information". Det forskningen vidare undersöker är frågor som berör de olika komponenterna i processen: vilka faktorer ger upphov till informationsbehov och på vilka sätt tar de sig uttryck? Till vilka källor eller kanaler vänder sig användaren för att tillfredsställa sina behov och på vilka grunder väljs dessa? Hur utnyttjas källorna och den anskaffade informationen? Innan jag närmare går in på dessa frågeställningar ger jag en presentation av det fält som är grundläggande för all forskning av informationsbeteende.

## 2.2 Användarstudier

Användarstudier ger “modeller av och kunskap om människors informations-sökningsprocesser” (Limberg 1998: 15). Denna kunskap har som forskningstradition kontinuerligt utvecklats under 1900-talet, men inleds på allvar först under seklets mitt. Ett flertal forskningsöversikter (jfr t.ex. Case 2002; Wilson 1994) beskriver tiden fram till 1948 som i huvudsak koncentrerad till undersökningar av bibliotekssystem och olika informationskanaler. Med påverkan från *Royal Social Scientific Information Conference* 1948 vidgades fältet till att även gälla informationsbeteende generellt. Men det är 1960-talet som blir ett avgörande decennium då man nu börjar inse den komplexitet och svårfångade karaktär som kännetecknar informationsbeteende, en insikt bevisad av de modeller för olika informationsfenomen som samtidigt utvecklas (Savolainen 2003: 81 f.). Forskningsansatsen förhöll sig dock övervägande strukturinriktad med användarstudier som “syftade till att utforska hur och i vilken utsträckning olika informationssystem används av olika grupper” (Sundin 2003: 22). Som Sundin fortsätter innebar det till största delen kvantitativa forskningsmetoder med survey-undersökningar där användargrupperna i huvudsak bestod av akademiker och forskare.

Småningom uppstod kritiska röster kring den systemorienterade forskningen och i fokus ställdes det som Case (2002: 6) belyser: “What was actually studied were the information *sources* and how they were used, rather than the individual users, their needs (as they saw them), where they went for information, and what kind of results they experienced.” I en välciterad artikel kategoriserar Dervin & Nilan (1986) den traditionella forskningens problemområden och beskriver den förändrade forskningsinriktningen som ett paradigmskifte, mot individcentrerade ansatser. Det nya synsättet menar Sundin (2003: 23) “har haft stor betydelse för synliggörandet av användaren utifrån dennas egna perspektiv”, i praktiken ett förhållningssätt starkt präglad av kognitiva influenser. Litteraturen från denna tidsperiod fram till 1990 uppvisar en mängd olika modeller och teorier kring informationsbeteendeprocesser, vilka närmare beskrivs i t.ex. Wilson (1994), eller Limberg (1998) ur ett nordiskt perspektiv. Den kanske mest inflytelserika forskaren att nämna är Brenda Dervin (1986; 2003) med sin Sense-Making-teori, men även Wilson (1999) har utvecklat modeller värda att uppmärksamma.

Det senaste decenniet har inneburit ett vidgat perspektiv på användaren som aktör i en social kontext inom vilket informationsbeteendet konstrueras i socio-kulturella sammanhang. Skapandet av den internationella konferensen *Information Seeking In Context* (ISIC 1996) kan ses som ett led i den kritik som riktades mot paradigmskiftets alltför rationella och individcentrerade synsätt (Sundin 2003: 25). Denna kontextorienterade ansats har vid sidan av en kognitiv linje befästs och vidareutvecklats främst inom den nordiska BoI-forskningen med fenomenologiska (Limberg 1998) och domänanalytiska (Hjørland 2002) inslag.

### 3. Ordboksanvändning som informationsbeteende

Min teoretiska diskussion bygger till stor del på den undersökning jag gjorde hösten 2006 i samband med skrivandet av min D-uppsats i Praktisk svenska vid Uppsala universitet (se Granström 2006). Att undersöka hur en informationsutbildad yrkesgrupp som bibliotekarier använder ordböcker i sitt dagliga arbete föreföll mig mycket intressant och givande. Att förmedla information på ett lättbegripligt sätt kräver ett medvetet språkbruk vilket i problemsituationer ofta underlättas av ordböcker. Bibliotekarier är dessutom många gånger språkförebilder för låntagare, inte minst då svenska är ett andraspråk.

Den grundläggande frågan jag sökte svar på var hur användarbeteendet kan se ut bland bibliotekarier och vidare vilken roll ordboken spelar på bibliotek i allmänhet. Mitt material bestod av en gruppintervju med anställda vid ett universitetsbibliotek (UB) och en enkätundersökning bland folkbibliotekarier (FB). Informanterna svarade på frågor utifrån frågekategorier gällande användning och tillgång av ordböcker, ordbokens roll i arbetet och på bibliotek samt elektroniska ordböcker och hjälpmedel. Ur svaren kunde jag utkristallisera tre typer av användarsituationer som jag valt att nämna *individuellt bruk*, *låntagarbruk* och *bruk i referenssituationer* (informationsförmedling mellan bibliotekarie och låntagare). Dessa dimensioner varierar i intensitet och omfattning mellan informanterna och är i hög grad kontextbundna, men lika mycket baserade på kunskapsnivå och tidigare erfarenheter av ordboksanvändning. De olika bruksformerna inbegriper ordboksanvändning i både passiva (läsning och lyssnande) och aktiva (skrivning och tal) språkaktiviteter, med tonvikt på den aktiva formen.

Stoffet i varje svarsdimension kopplade jag samman med de tidigare nämnda komponenterna behov, sökning och användning. På så vis bildas en informations-beteendeprocess i en yrkesrelaterad kontext där olika faktorer inverkar på bibliotekarien som ordboksanvändare. I det följande diskuterar jag varje komponent var för sig.

#### 3.1 Behov

Den lexikografiska användarforskningen ställer en rad nödvändiga frågor kring förståelsen av ordboksanvändarens behov, vilket även kan uppfattas som att området utgör ett visst intresse för undersökning. Att så är fallet bekräftas i den lexikografiska översiktslitteratur jag bekantat mig med, och Svensén (2004: 534) förklarar aspekten som "en av de viktigaste" eftersom forskningen strävar efter kunskapen om "hur ordböckerna skall utformas för att på ett bättre sätt tillgodose dessa behov". Ingenstans har jag dock i lexikografilitteraturen funnit svar på den kanske mest grundläggande frågan att ställa: vad är ett "behov"? Jag vänder mig då till BoI-litteraturen och finner följande, något nedslående, svar från Case (2002: 67): "Few investigations of information seeking delve very deeply into the issue of what human 'needs' really are. Not many even question the notion of 'in-

formation needs'. Rather, most writers assume that information needs exist and are relatively unproblematic." Min avsikt är inte heller att ge något uttömmande svar på frågan med tanke på artikelns omfång, utan jag vill öppna upp för reflektion över vedertagna begrepp inom lexikografin som därmed är betydelsefulla för min undersökning. Den diskussion som finns är aktiv inom BoI-fältet (se t.ex. Case 2002; Savolainen 2002) och jag nöjer mig här med en definition av *behov* ur ODLIS: "A gap in a person's knowledge that, when experienced at the conscious level as a question, gives rise to a search for an answer. If the need is urgent, the search may be pursued with diligence until the desire is fulfilled."

Informationsbehovet är det fenomen som "utlöser en informationssökningsprocess och som i hög grad styr hur denna process går till" (Limberg 1998: 45). Hur uttrycks då detta behov bland bibliotekarierna? Genom att se på i vilka slags situationer informanterna upplever ett tillstånd av bristande kunskap kan behovet av ordböcker tydligare förklaras. På ett individuellt plan är det främst i skriftliga sammanhang som både FB och UB behöver ordböcker och mer specifikt då det uppstår osäkerhet kring lexikalisk information som betydelse, stavning eller synonym. Inte lika ofta tycks läsande av texter eller muntlig kommunikation föranleda ordböksbehov, förutom vid mer informella situationer kollegorna emellan. Med ett samlat begrepp ger det individuella bruket uttryck för ett *problemorienterat informationsbehov*, dvs. "krav i anslutning till lösandet av något problem eller utförandet av någon uppgift" (Savolainen 2002: 76).

Att olika typer av ordböcker ska kunna tillfredsställa olika behov för användaren är bibliotekarierna visserligen medvetna om, men i praktiken är det nästan enbart SAOL som används för det individuella bruket, särskilt bland FB. Därmed bekräftas Bogaards (2003: 29) konstaterande att "[f]or many people the only thing that exists is 'the' dictionary". En möjlig konsekvens av att förlita sig på en enda informationskälla blir ett icke tillfredsställt behov och en sökprocess som utmynnar i andra hjälpmedel än ordböcker. För en majoritet av informanterna löser Google de flesta informationsproblem, vilket motiveras av att internetverktyget är snabbt och smidigt, om än inte alltid tillförlitligt. Det är tryckta referensverk som sist och slutligen föredras och i de fall varken konventionella ordböcker eller internetalternativ är behjälpliga täcker olika uppslagsverk upp för aktuellt informationsbehov.

Ett betydligt mer påträngande behov av ordböcker uppstår i referenssituationer. Bibliotekarierna omtalar användarkontexter som inte enbart syftar till att förmedla önskad språkinformation till låntagaren, utan även som auktoritetsknuten kompetenskontroll, dvs. säkerställning av den egna språkkänslan genom att kontrollera sakförhållanden i ordböcker. Språkkompetens ställer dock krav på bibliotekarien att kunna förmedla relevant information, varför behovet av ordböcker i vissa fall ökar. Särskilt gäller det i situationer då låntagaren har ett annat modersmål än svenska och inte behärskar språket tillräckligt för att precis kunna formulera sin sökfråga. För UB:s del är förhållandet mindre komplext eftersom

kommunikationen med låntagargruppen "utländska studeranden och forskare" nästan uteslutande sker på engelska och dessutom underlättas av en stencil för bibliotekstermer på engelska. Ordlistan är intressant ur ett behovsperspektiv eftersom den visar på en tidigare lucka i utbudet av referensverk som nu täppts igen och som informanterna upplever som ett nödvändigt hjälpmedel. Samtidigt uppvisar ordlistan ett gott exempel på tillvaratagen yrkeskunskap på arbetsplatsen eftersom den sammanställts av en engelsktalande kollega till informanterna.

Utgångspunkten för FB är något annorlunda eftersom engelska som möjligt samtalspråk inte kan förutsättas bland den betydligt mer heterogena låntagargruppen i fråga om faktorer som t.ex. ålder, sysselsättning och etnisk bakgrund. Därmed intensifieras rollen som god svensktalande språkförebild, vilket bekräftar av det något oväntade behovet av ordböcker i referenssituationer. Genom att utnyttja ordböcker, däribland *Synonymordboken*, kan bibliotekarien tydligare ringa in låntagarens sökfråga, finna ersättningsord eller helt enkelt försäkra sig om korrekt stavade titlar i efterfrågad litteratur.

Det faktiska behovet av ordböcker bland låntagarna antyds inte bara i samtal med bibliotekarien utan framkommer mer konkret i form av lånestatistik på folkbiblioteket. Trots att det inte går att utläsa vilka låntagarna är, anser sig FB ändå uppfatta ett behov bland elever och andraspråksinlärare som utnyttjar ordböcker som del i undervisningsmaterial. Knappt någon bland FB omtalar låntagarnas eventuella behov av elektroniska ordbokskällor, vilket kan hänga samman med att folkbibliotek inte primärt gynnar datoranvändning, annat än till att söka i kataloger. En mer hypotetisk förklaring vore det faktum att majoriteten av FB säger sig föredra tryckta ordböcker framom elektroniska varianter och därför inte i lika hög grad reflekterar över t.ex. stavningskontroller eller internetbaserade ordbokskällor. För de studenter och forskare som befinner sig på universitetsbiblioteket utgör datorn ett naturligt arbetsredskap, vilket skulle kunna innebära ett lika naturligt behov av datorstödda ordboksfunktioner. Intressant nog hävdar en av informanterna att så inte alls behöver vara fallet – den tryckta ordboken används alltså jämt vid sidan om datorn.

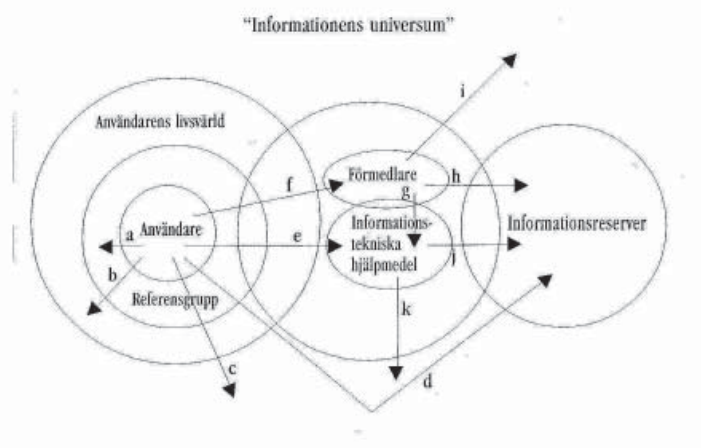
De varierande behov av ordböcker som uppstår i olika kontexter på bibliotek kan sägas spegla behov i ett bredare perspektiv, som socialt och kulturellt betingade under påverkan av samhällets värderingar (Savolainen 2002: 76). Här menar bibliotekarierna att (den tryckta) ordboken innehar en oförändrat stark ställning som hjälpmedel och inte nämnvärt låter sig påverkas av yttre faktorer, som sökmotorn Google "där man hittar det mesta", eller av bibliotekarierna förmodad minskad aktivitet inom ordboksbranschen. Ur demokratisk synpunkt är tillgången till ordböcker på bibliotek viktig att upprätthålla, eftersom "rik som fattig" ska kunna ha tillgång till det utbud som uppfattas utgöra "stommen i ett bibliotek". En sådan inställning kan även sägas spegla gruppens normer, dvs. "vad man bör kunna inom ett visst yrke" (2002: 76). Informanterna är så väl medvetna om den egna yrkesrollen som informations- och språkförmedlare att de i vissa fall

tangerar gränsen till ett pedagogiskt förhållningssätt till låntagaren. Användarundervisning i ordbokskunskap sker visserligen inte på biblioteken, men det hindrar ändå inte bibliotekarien att t.ex. "hänsynsfullt" korrigera låntagarens språkbruk eller direkt uppmana låntagaren att själv anskaffa ordböcker. I linje med bibliotekets roll som demokratisk institution uppfattar bibliotekarierna språket som ett viktigt medel att kunna tillgodogöra sig samhällsinformation, varmed ett korrekt språkbruk blir behjälpt av ordböcker.

### 3.2 Sökning

Till skillnad från den bredare innebörden av informationssökning, tolkas informationsuppsökning i en snäv, instrumentell bemärkelse där tonvikten ligger på ett intresse för "de metoder med vars hjälp man kommer åt 'informationsbärande' dokument" (Savolainen 2003: 79). Det är dessutom av vikt att särskilja sökning *av* ordböcker från sökning *i* dem, där det sistnämnda beteendet snarare skulle benämnas användning (se avsnitt 3.3).

Efter att ett behov hos användaren uppmärksammas leder så följande steg i informationsbeteendeprocessen till ett uppsökande av information för att tillfredsställa nämnda behov. I en språkligt problematisk kontext övervinner bibliotekarien "a gap in knowledge" (Case 2002: 72) genom att vända sig till de referensverk hon har att tillgå och däribland finna en relevant ordbok för situationen. Detta kan illustreras av en modell i figur 1 gjord av Tom Wilson (1981). I modellen konkretiseras bibliotekariernas uppsökningsmetoder i form av *informationssökningsstigar*<sup>1</sup>, "information-seeking paths", vilka varierar i betydelse beroende på användarkontexten. Modellen visar därmed på Wilsons strävan att föra in en holistisk syn på användaren som främst delaktig i en *livsvärld* "defined as the totality of experiences centred upon the individual as an information user" (1981: 6).



Figur 1. Informationssökningens stigar (Wilson 1981: 6 hämtad från Savolainen 2003: 85)

1 Samtliga svenska begrepp återfinns i Savolainen (2003: 84 f.).

Innanför denna livsvärld existerar den miljö vari information anskaffas och används, exempelvis i en arbetsmiljö som biblioteket. De olika referensgrupperna som användaren, i detta fall bibliotekarien, identifierar utgörs av arbetskollegor och låntagare, viktiga medaktörer i det stora alltet, *informationens universum*. Wilson poängterar att modellbeskrivningen är starkt förenklad av en i realiteten komplex användartillvaro, men jag finner insatsen intressant och givande för mitt vidare resonemang kring bibliotekariernas uppsökningsbeteende i anslutning till ordboksanvändning. Det bör påpekas att jag ur mitt material inte kan identifiera samtliga sökstigar som modellen uppger, men hänvisar till Savolainen (2003: 84) som förklarar att sökstigarna "*exemplifierar* [min kursivering] hur informations-sökningen kan konkretiseras till val och utnyttjande av olika källor och kanaler".

Modellen återger alltså ett antal möjliga informationssökningsstigar som Wilson sammanför till fyra olika grupper. Sökstigarna a, b, c och d bildar den s.k. A-kategorin och innebär "search strategies by a user independent of any information system" (1981: 6). För bibliotekarierna betyder det situationer då ordböcker av olika anledningar inte används utan då språkinformation inhämtas på annat vis, t.ex. genom att rådfråga kollegor eller enbart förlita sig på språkkänslan.

Utnyttjande av stigar i kategori B, dvs. e och f, involverar en *förmedlare* och/eller *informationstekniska hjälpmedel* vilka ingår i det omgärdande området *informationssystem*. Hit kunde utnyttjande av konkordanser inräknas liksom sökande i uppslagsverk. Förmedlaren, "generally a living system, i.e. a human being" (Wilson 1981: 6), gestaltas rätt naturligt av bibliotekarien för låntagarens del, men det fungerar likväl omvänt eftersom referenssituationer innebär en förmedlande samtalskontext per se. Även språkexperter, vid t.ex. Språkrådet, fungerar som förmedlare i detta fall.

Kategori C, sökstig g, h och i, omfattar "search strategies employed by a mediator to satisfy a user's demand for information" (1981: 6). Med biblioteket och dess olika aktörer som "informationens universum" tolkar jag C-kategorin som funktionell inom låntagardimensionen. Sålunda är det låntagaren som uttryckligen ber bibliotekarien att leta fram ordboksrelaterad information, så som i fall med korsordslösningar eller lexikaliska problem. Bibliotekarien väljer då för situationen relevant ordbokskälla och kombinerar elektroniska och tryckta format vid behov. Dessa representeras i modellen av *informationsreserver*, "embodiments of knowledge".

Sökstigarna j och k bildar den fjärde och sista kategorin D. Enligt Wilson identifierar kategorin en användare som självständigt utnyttjar olika informationstekniska hjälpmedel, vilket i relation till bibliotekarierna avser bruk av elektroniska ordbokskällor. Bibliotekariens val av SAOL och SAOB i elektroniskt format är ett exempel på informationsreserver som D-kategorins sökstigar leder fram till, och i den mån man uppfattar internet och Google som "ordbokskällor" kunde även dessa inräknas. Modellen är de facto publicerad 1981 och sedan dess har den tekniska utvecklingen frambringat en mängd olika elektroniska resurser



som berikar utbudet av informationsreserver. Huruvida dessa förenklar för bibliotekarien att hitta relevant ordboksinformation eller gör sökningen snårigare skulle undersökningsfältet för informationsanvändning kunna ge svar på.

### 3.3 Användning

Ordboksanvändningens motsvarighet, konsultationsbeteende, innebär undersökningar om den enskilde användarens metoder och strategier för att finna den information i en specifik ordbok hon är ute efter. Detta kräver andra metoder och frågeställningar än vad jag fokuserat på och har därför inte behandlats i undersökningen. Delområdet har jag ändå valt att ange med egen rubrik för att visa på dess lika stora betydelse som komponent i informations-beteendeprocessen.

Lexikografen har hittills främst bedrivit forskning kring inlärningsordböcker och på så vis endast lyft fram en sida av olika användares beteende vid konsultation, även det beskrivet som en process (se Bogaards 2003). Området är, som Svensén (2004: 537) påpekar, besvärligt att kartlägga eftersom "[m]an har att ta hänsyn till en mängd olika typer av användare, ordböcker och användningssituationer och till de närmast oändliga variationsmöjligheter och de komplicerade psykolingvistiska processer som ryms inom konsultationsprocessen". Intressant vore dessutom att vidare undersöka frågan om vad "konsultationsfärdighet" består i. Att då fortsättningsvis hämta inspiration från BoI-fältet tror jag skulle vara fruktbart, särskilt i fråga om informationsanvändningskontexter och styrande faktorer som bl.a. affektiva aspekter (se Savolainen 2004: 94 ff.). Informationsanvändning är som synes ett lika komplext fenomen som de två övriga komponenterna informationsbehov och informationssökning, men är lika nödvändigt att utforska för att till fullo förstå användaren och dennes ordboks-beteende.

## 4. Slutsats

Jag har i min artikel velat påvisa ett samband mellan lexikografisk användarforskning och användarstudier inom biblioteks- och informationsvetenskap. Bryggan över till studier av informationsbeteende menar jag ger en djupare analys av ordboksanvändaren som delaktig i en komplex användarsituation. Ser man ordboksanvändaren i ljuset av informationsbeteende och utgår ifrån den kontext användaren befinner sig i kan även mer specifika områden som behov, sökning och användning av ordböcker tydligare ringas in.

Bibliotekarier som ordboksanvändare är en viktig grupp att undersöka; de är offentliga språkförmedlare och hanterar information (även) på ett professionellt plan. Att ordböcker kommer till användning på olika sätt visar den del av min undersökning jag tidigare återgett. Detta bruk föranleds av ett lexikaliskt informationsbehov som ger upphov till uppsökning av ordbokskällor individuellt eller tillsammans med en låntagare. På motsvarande sätt skulle även andra användargrupper kunna analyseras. I första hand tänker jag på lärare och elever, då ytter-

ligare en aspekt tillkommer: ordboksanvändning i ett pedagogiskt och didaktiskt sammanhang. Övriga grupper av intresse är t.ex. journalister och översättare.

För att komplettera min undersökning och gå vidare med hur konsultation av ordböcker går till krävs mer detaljerade närstudier av ett sådant användarbete-ende. För närvarande får vi nöja oss med att ha fått en uppfattning om vägen till ordbokskällan – en väg fylld av valmöjligheter för att tillfredsställa användarens informationsbehov.

## Litteratur

- Bogaards, Paul 2003: Uses and users of dictionaries. I: Sterkenburg, Piet van (ed.): *A Practical Guide to Lexicography*. Philadelphia: John Benjamins Publishing Company, 26–33. (18.09.2006) <http://site.ebrary.com/lib/uppsala/Doc?id=10040043&ppg=38>
- Case, Donald O. 2002: *Looking for information: a survey of research on information seeking, needs and behaviour*. San Diego, Calif.: Academic Press.
- Dervin, Brenda 2003: Human studies and user studies: a call for methodological interdisciplinarity. Paper 166. *Information Research* 9:1. (11.4.2006) <http://InformationR.net/ir/9-1/paper166.html>
- Dervin, Brenda & Michael Nilan 1986: Information needs and uses. I: Williams, Martha, E. (ed.): *Annual Review of Information Science and Technology (ARIST)* 21, 3–33.
- Granström, Alexandra 2006: *Ordböcker och informationsbeteende – en studie i ordboks- användning bland bibliotekarier*. Uppsala universitet: Institutionen för nordiska språk. <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7809>
- Hjørland, Birger 2002: Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation* 58, 4: 422–462.
- Limberg, Louise 1998: *Att söka information för att lära: en studie av samspel mellan informationsökning och lärande*. Skrifter från Valfrid 16. Borås: Valfrid.
- ODLIS = *Online Dictionary for Library and Information Science* 2004. (8.10.2006) [http://lu.com/odlis/odlis\\_i.cfm](http://lu.com/odlis/odlis_i.cfm)
- Savolainen, Reijo 2003: Informationssökning och informationsbehov. I: Mäkinen, Ilkka, & Katja Sandqvist (red.): *Introduktion till informationsvetenskapen*. 2 uppl. Tammerfors: Tampere University Press, 69–102.
- Sundin, Olof 2003: *Informationsstrategier och yrkesidentiteter – en studie av sjuksköterskors relation till fackinformation vid arbetsplatsen*. Skrifter från Valfrid 25. Borås: Valfrid.
- Svensén, Bo 2004: *Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm: Norstedts Akademiska Förlag.
- Wilson, T. D. 1981: On user studies and information needs. I: *Journal of Documentation* 37, 1: 3–15.
- 1994: Information needs and uses: fifty years of progress? I: Vickery, B. C. (ed.): *Fifty years of information progress: a Journal of Documentation review*. London: Aslib, 15–51. (12.4.2006) <http://informationr.net/tdw/publ/papers/1994FiftyYears.html>
- 1999: Models in information behaviour research. I: *Journal of Documentation* 55, 3: 249–270.

Alexandra Granström  
f. 1980  
Institutionen för nordiska språk  
Box 527  
SE-751 20 Uppsala  
alexandra.granstrom@nordiska.uu.se



## Fast redigerings skjema for verb? Semantikk versus grammatikalisering

The purpose of this article is to clear the ground for a general and common editorial schema for lexicographic treatment of all verbs in the Norwegian language. Therefore, claims that verbs should be classified into groups with special qualities and functions are examined. In grammatical tradition, verbs are content words, excepting several small subsets of verbs used in constructing multi-word verb forms, which according to the grammars require different treatment. In Norwegian lexicography another classification has emerged, counterpoising "big verbs" with multiple functions and a multitude of derived phrasal verbs against the rest. These classifications are examined against the defining practice shown in entries on individual verbs, and a general maximum model for ordering and defining verbs, including derived phrasal verbs, is presented.

### 1. Innleiing

Utgangspunktet for denne artikkelen er handsaminga av ordklassen verb i Norsk Ordbok. Men sjølv om utgangspunktet er behandlinga av verb i Norsk Ordbok, er intensjonen med denne artikkelen generell, i høve til leksikografisk behandling av ordklassa verb i norsk og nordisk leksikografi.

Alle større ordbøker har ein vedteken artikkelstruktur, med fleire nivå ordna etter rang og eit nummereringssystem som syner plassering i rekkje. Såleis har artiklane i Norsk Ordbok ein struktur med fire nivå<sup>1</sup> og i tillegg eit eige artikkelformat for såkalla **underoppslag**, det vil seia sublemma av faste, leksikaliserde fleiordige ordsamband.

Norsk Ordbok har ein redigeringsapplikasjon med eit innebygt artikkelformat som set faste rammer for artikkelstrukturen når det gjeld tydingsnivå og ulike typar einingar. Når det gjeld bruken av artikkelformatet til å organisera tydingar, set programmet ingen grenser. Da er det redigeringsreglane som gjeld, og redaktørane har i oppgåve å bruka dei.

### 2. Kva er meint med redigerings skjema?

Norsk Ordbok har få eksplisitte redigeringsreglar som gjeld tydingsordning for

---

1 Stor bokstav for hovudinndelingar, arabisk tal for tydingsnummer, liten bokstav for nivå under tydingsnummer, og endeleg dobbel skråstrek for undertyding. (Jfr Grønvik 2005: 160.)

ordgrupper med sams eigenskapar eller kjenneteikn. Det finst likevel ein del redaksjonell praksis i byggja på. Denne praksisen er ikkje heilt straumlinjeforma, men kan vera temmeleg fast. I somme tilfelle er praksis så fast at han kan seiast å utgjera eit implisitt ”redigerings skjema”, som redaktøren då realiserer i høve til kvar einskild artikkel.

Med ”redigerings skjema” meiner eg heretter ein eksplisitt instruks i redigeringsreglane til redaktørane om å ordna tilfanget under ein viss type lemma på ein viss måte og i ei viss rekkjefylgd når det gjeld tydingar og fraseologi. Eit redigerings skjema er då ikkje eit automatisert eller obligatorisk skjema som er bygt inn i innskrivingsapplikasjonen. Eit redigerings skjema vil vera ei formalisering, men ikkje ei automatisering, av redaksjonell praksis.

### 3. Status for redigeringspraksis i Norsk Ordbok

Norsk Ordbok er ei utprega dokumentasjonsordbok. Ordboka skal ikkje berre syna fram tyding og bruk i ord som alt er innanfor rettskrivinga, men også normera talemålstilfang som tidlegare ikkje er identifisert på lemmanivå.

Det sterke fokuset på å granska tilfanget som ligg føre, og framstilla det rett, har skapt ein tilsvarende sterk tradisjon i Norsk Ordbok for å redigera eitt og eitt ord nedanfrå og opp. Grammatiske tilhøve, fraseologi og metaforikk blir vist fram for kvart lemma. I høve til tyding og bruk finst det få eksplisitte konvensjonar for å handsama einskildord som medlem av ordgrupper ut frå meir generelle lingvistiske kriterium. Det er utarbeidd nokre definisjonsformlar og gjevi nokre råd om tydingsordning, men denne delen av redigeringa er stort sett overlata til det redaksjonelle skjønet.<sup>2</sup>

Eg skal i det følgjande sjå på om eit sams redigerings skjema for verb vil vera ei tenleg løysing, både i høve til stringens i språkbeskrivinga og i høve til redaksjonell konsistens.

### 4. Kva er eit verb? Terminologiske skilnader mellom grammatikk og leksikografi

Her er det nødvendig å seia noko om skilnader i termbruk og termforståing mellom leksikografi og grammatikk.

For leksikografar er grunneininga einskildord (sometid også fleirordige frasar) med eigen identitet og ordartikkel, og termen for denne grunneininga er *lemma*. Om ein i leksikografisk litteratur nemner ”verb”, ”substantiv” eller ei anna ordklasse, er det sjølv sagt for leksikografen at ein refererer til kvart verb som lemma, om ikkje noko anna blir sagt.

Slik er det ikkje i dei undersøkte norske grammatikkane. Ordet ”verb” blir

<sup>2</sup> Redigeringshandboka for Norsk Ordbok har nokre få sider (avsnitt 4.3.3.4) om definering og definisjonstypar.

helst nytta om ei ordform i høve til berre éi tyding eller berre éin (eller nokre få) (grammatiske) bruksmåtar av ordklassen (jf. t.d. Faarlund et al 1997: 526), og då er det det grammatiske fenomenet som er emne for drøfting. Ordforma er ein illustrasjon på det omtala grammatiske fenomenet, ikkje eit komplett objekt i seg sjølv, slik eit lemma er rekna for å vera.

Ein annan leksikografisk banalitet er den at eit lemma kan knyta til seg fleire tydingar, og altså vera **polysemt**. I fleire av dei aktuelle grammatikkane finn ein generelle kommentarar som syner at forfattaren er merksam på at verb som representerer grammatikalske funksjonar, også kan opptre som innhaldsverb, men det er ikkje viktig i samanhengen (jf. t.d. Faarlund et al 1997: 517, Beito 1986: 269). Det som er viktig er den gjevne grammatiske funksjonen og omgjevnader som kjenneteiknar han.

Generelt kan det verka som om grammatikkane er lite merksame på polysemi som fenomen, og ikkje reknar med det når t.d. verb blir klassifiserte eller grupperte. Korkje "lemma" eller "polysemi" er stikkord i registra i nokon av dei norske grammatikkane eg har sett på.

Det går derfor ikkje bra for leksikografen om han lit på at eit verb "er" eit modalt hjelpeverb og berre det, av di det står i grammatikkane. Grammatikaren har ikkje teki stilling til heile tilfanget som skal liggja til grunn for ein ordartikkel om eit gjevi lemma. Likeins kan ein ikkje ta for gjevi at den grammatiske funksjonen som er omtala, berre finst hos dei verba som grammatikkane nemner i gruppeinndelingane sine.

I denne artikkelen er perspektivet leksikografisk. Føremålet er å drøfta føresetnadene for å laga eit generelt redigerings skjema for verb. Men sidan leksikografane lit på grammatikkane som kunnskapskjelder, og til dels tek sine ordningskriterium for tydingar frå grammatikken, må dette også bli ei drøfting av grammatiske omgrep og perspektiv i høve til dei leksikografiske.

## 5. Kva slag samband blir tilordna verbet som lemma?

Den syntaktisk rolla til verbet er å knyta saman andre setningsledd til å utgjera heile setningar. Denne eigenskapen ligg til grunn for den vanlege leksikografiske handsaminga av verb. Tydingsinndelingane er oftast knytt til synlege faktorar i omgjevnadene til verbet. Dette ser ein av døma som er tilordna dei ulike definisjonane, sjølv om det ikkje er gjort eksplisitt i definisjonsteksten. Tydingsinndelingane kan kvila på subjekt eller subjektstype, objekt eller objektstype, adverb eller adverbtype, eller kombinasjonar av desse. Desse sambanda av verb og kollokat syner seg i tilfanget som konvensjonaliserte på alle nivå, frå verb pluss kollokattype (t.d. eit fritt valt substantiv i subjektrolla) til verb pluss ei einaste mogleg anna ordform (t.d. *drikka jabbesaft*, *gnika naudeld*). Dersom slike samband er både faste og leksikalisererte, blir dei handsama som sublemma i Norsk Ordbok, med status som underoppslag.

Verb kan i nordiske språk danna fraseverb med funksjonsord som pronomen og adverb (irekna adverbielle preposisjonsuttrykk) og med kombinasjonar av slike ord (døme: *gjeva ut, ha seg, bera over med, hiva seg rundt, taka seg til rette*). Slike fraseverb er ofte polysem. I somme nyare ordbøker får viktige fraseverb status som lemma, med eigen ordartikkel (Svensén 2004: 126). I dei store dokumentariske ordbøkene for nordiske språk er det tradisjon for å handtera fraseverba som sublemma under verbet, og altså som del av verbartikkelen. Kvart ordboksverk har meir eller mindre gjennomarbeidde redigeringsreglar for ordning og handtering av fraseverb. I Norsk Ordbok er dette systemet blitt til over tid og er enno ikkje heilt konsistent.

Innanfor verbartikkelen skal leksikografen såleis handsama heile den semantiske og kombinatoriske skalaen, frå verbet som isolert ordform mellom utskiftbare kollokat, til verbet som lekk i faste fraseverb som igjen tek si tydingsinndeling ut frå dei same kontekstkriterium som den isolerte verbforma.

## 6. Prinsipp for tydingsordning

Dersom ein skal vera sikker på at ei tyding er knytt til ei ordform, må tydinga halda seg uendra enda om omgjevnadene blir endra. Dess større endring i omgjevnadene ei kopling ordform – omgrep toler, dess sterkare er omgrepet knytt til sjølve ordforma.

For nynorsk manglar det dokumentasjon for tydingsutvikling for ein lang periode, og ei strengt historisk tilnærming til tydingsordning er derfor ikkje mogleg. Tydingsordninga i Norsk Ordbok 2014 er derfor i større grad knytt til det som verkar som ei logisk ordning av omgrep. Dette blir ei ordning som ofte (men ikkje alltid) fell saman med både alder (så langt vi kan prova det) og med frekvens i tilfanget. Da kjem den mest innhaldstunge (og kontekstfrie) tydinga først, og så aukar bindinga til kontekst (og spesifisering av tyding) utover i artikkelen.

Når det gjeld fraseverba, varierer handteringa i Norsk Ordbok. Etter gjeldande redigeringsreglar blir fraseverb med minst ei leksikalisert (og dermed ikkje føreseieleg) tyding handsama som underoppslag, og kan såleis redigerast med fleire tydingar og eigne døme til kvar tyding. Men plasseringa av underoppslaget i systemet av tydingsbolkar innanfor artikkelen kjem an på om verbet i seg sjølv blir rekna som eit ”stort verb”, eller eit vanleg (mindre frekvent) verb.

## 7. Redigerings skjema for ”store verb” i Norsk Ordbok

I Norsk Ordbok har det utvikla seg ein tradisjon for å gruppedela verba, der ein set opp ”store verb” mot resten av ordklassa.



Dei ”store” verba er verb som

- høyrer til det sentrale ordtilfanget
- er høgfrekvente
- finst i mange faste (og leksikaliserte) samband med andre ord
- har til dels svært lange artiklar i større morsmålsordbøker

Døme på slike verb er *ha, eiga, fara, fylgja, gå, koma, liggja, stå, vera*. Desse verba ligg også under eit utal avleiingar og samansettingar i andre ordklassar, som igjen kan vera produktive ordlagingselement (t.d. *tiltak, utlegg*).

Den redaksjonelle praksisen for ”store verb” i Norsk Ordbok tek utgangspunkt i at desse verba tek i bruk ein stor del av det potensielle artikkelskjemaet. Desse verba dreg med seg svært mange konvensjonaliserte samband med andre ordklassar, og dei dannar ei mengd fraseverb. *Norsk Ordbok* band 1–4 (under redaksjon frå 1946 til 2002) viser ei gradvis rørsle i leksikografisk praksis frå å redigera som om verbet aleine var tydingsberar, til ei meir kontekstorientert ordning av tilfanget. Praksis i dag er å bruka stor bokstav-nivået til å dela inn dei store verbartiklane i hovudbolkar. Alle tydingar der verbet står sjølvstendig til subjekt og eventuelle objekt, kjem under stor A. Refleksiv-sambanda kjem under stor B og sambanda med adverb/preposisjon kjem under stor C. Denne redaksjonelle praksisen med alfabetisk ordning av fraseverba sist i verbartikkelen finn ein også i andre større ordboksverk, som *Oxford English Dictionary* og *Svenska Akademiens ordbok*.

For dei andre, ikkje-store, verbartiklane er praksis mindre fast når det gjeld artikkelstruktur og tydingsordning. Til vanleg blir dei minst kontekstbundne tydingane handsama tidleg i artikkelen. Det kan også vera underoppslag for faste samband med *seg* eller preposisjon / adverb, men desse har ikkje fast plassering i artikkelskjemaet.

## 8. Fast plassering av sublemma for alle verb? Pluss, minus og praksis.

I redigeringa av Norsk Ordbok ligg det som nemnt føre ei glidande utvikling over tid mot fastare plassering av sublemma for verb, særleg dei store verba. Dermed er det grunn til å sjå på om fast plassering kunne ha vori gjort gjeldande for alle verb, som eit maksimumsskjema.

Dei praktiske føremonane kan synast openberre. Redaktøren kjem til eit ferdig ”kart” for verbartikkelen, og kan innplassera belegg og forma innhald utan å tenkja ut grovstruktur i kvart tilfelle.

Føremonen for ordboksbrukaren er at det same skjemaet blir brukt mange gonger. Ordboka får ein attkjenneleg modell for plassering av fraseverb som tek utgangspunkt i noko alle kan sjå, nemleg typar fraseord i fast rekkjefølgje som tilhøyr til verbet. Det vil gjera det lettare for brukaren å orientera seg i dei store verbartiklane i sær — desse kan vera svært lange og gje eit uover-

siktleg intrykk. Men også allment kan det vera ein føremon å kjenna orden og rekkjefylgd på dei viktigaste kategoriane, så ein kan gå rett på det ein er interessert i.

Den største veikskapen ville vera om eit skjema for fast plassering av fraseverba fører til at informasjon som redaktøren ønskjer å formidla om samanhengen mellom verbtyding og fraseverb ikkje kjem fram, eller blir forvanska. I ordbøker er artikkelstrukturen i seg sjølv ein viktig tydingsberar. Norsk Ordbok 2014 har no attom seg eit heilt band som er redigert inn i databasen. Vi kan derfor sjå på praksis for ein tolvtedel av ordboka, med høve til å bruka underoppslaget for fraseverb, og med pålegg om korleis underoppslag skal plasserast ved store verb.

Band 6 (bokstaven k) har 166 verb med i alt 564 underoppslag og 994 definisjonar — av desse har 6 store verb 174 underoppslag med 539 definisjonar. 80 verb har 2 til 9 underoppslag og 80 har eitt underoppslag.

462 av underoppslaga er av typen verb + adverb (*kasta ut*), verb + adverb + *seg* (*kanta på seg*), verb + *seg* (*klaga seg*)<sup>3</sup> eller verb + *seg* + adverb (*kle seg ut*). Det store fleirtalet av desse er ordna alfabetisk, eventuelt i ulike grupper, dersom verbet har meir enn eitt underoppslag av same type. Dei større verbartiklane har samla (og gruppert) underoppslag av desse typane sist i artikkelen. I tillegg kjem ein handfull parataktiske samband av typen *koma og gå* (som til vanleg blir handsama som sublemma til det første verbet i sambandet).

162 av underoppslaga har meir enn ei tyding, og av desse høyrer 132 til dei typane fraseverb som er nemnde ovanfor.

Dei typane underoppslag som er spesifiserte ovanfor, er dei det ville høva å gruppera og samla sist i verbartiklane etter skjema.<sup>4</sup>

Ut frå praksis i band 6 er det ikkje urimeleg å laga eit redigeringsskjema for Norsk Ordbok som inneber at ein samlar alle sublemma av dei typane som er nemnde ovanfor, i underoppslag sist i verbartikkelen. Ein kan så gruppera dei og ordna dei alfabetisk innanfor kvar gruppe etter behov.

Innvendinga vil vera at mange verbartiklar — i dette bandet er det 80 — har berre eitt underoppslag, ofte med berre éi tyding. Da kan det verka som overadministrasjon å krevja at det eine underoppslaget skal stå for seg sjølv sist i artikkelen. Men halvparten av dei einslege underoppslaga i band 6 har eige tydingsnummer, så dei er alt utskilde. Da er det ikkje så stor endring om dette underoppslaget står sist.

Vidare kan det innvendast at om ein samlar alle fraseverb sist i verbartikkelen, bryt ein sambandet mellom verbtyding og fraseverb. Men dette sambandet er alt

3 Det er etter måten få underoppslag av typen verb + *seg*, av di alternativet, å nemna i definisjonen at den aktuelle tydinga føreset refleksivt objekt, er føretrekt (nyttå 435 gonger i verbartiklar på k-).

4 Dei siste drygt hundre underoppslaga på verb i band 6 av Norsk Ordbok er av ulike slag. Dei aller fleste inneheld substantiv, og det er såleis er diskusjonsspørsmål om dei er best plasserte i verbartikkelen eller om dei heller burde ha stått under substantivet.

broti i høve til dei større verbartiklane, og sambandsbrotet når langt inn i dei mindre. Dersom eit fraseverb har fleire tydingar, er det uansett frikopla frå ei gjeven tyding av grunnordet. Så spørsmålet sams redigeringskjema dreiar seg ikkje lenger for Norsk Ordbok sin del om framvising av semantisk kontekst versus skjematisk oppstilling. Det dreiar seg om konsistens i behandlinga av kategoriar.

## 9. Føresetnad – eit lingvistisk forsvarleg redigeringskjema

Føremålet med eit sams redigeringskjema for verb er å gje redaktørane eit tenleg verktøy i materialanalysen, og brukarane eit attkjenneleg leseskjema. Men føremålet må ikkje gå på akkord med dei språklege rammevilkåra. Føresetnaden for å laga eit sams maksimumsskjema er at det kan gjerast på indre lingvistiske premisar.

Ovanfor har eg argumentert for at eit fast redigeringskjema er forsvarleg og tenleg i høve til plassering og handsaming av fraseverba. Neste punkt gjeld den inndelinga av verb som vi finn i morsmålsgrammatikkane. Korleis står denne gruppedelinga i høve til ei leksikografisk framstilling av verb?

Eit redigeringskjema for verb må spegla ordklassens faktiske eigenskapar. Føresetnaden er med andre ord at det går an frå lingvistisk synsstad å sjå på alle verb under eitt når ein formar ut eit generelt redigeringskjema. Det vil seia at dei sams eigenskapane ved verb som ordklasse er viktigare enn eventuelle gruppeinndelingar innanfor ordklassa i høve til redigering av kvart lemma.

## 10. Grammatikkane – felles eigenskapar for verb

Alle større grammatikkar for norsk<sup>5</sup> er samde om at alle verb

- kan binda til seg subjekt og kan binda saman subjektet med andre setningsledd (ei evne som også blir kalla valens)
- kan syna tempus, modus og aksjonsart
- kan stå i samanhengar som syner aspekt (anten som verb i tempusform, eller som kollokat til eit verb i tempusform)
- inngår i eit avleiingssystem som femner om adjektiv, adverb, substantiv (verbets adjektiviske og substantiviske former) og andre verb (t.d. *-st-*avleiingar).

Dette er mange og sterke sams eigenskapar. Korleis står dette i høve til gruppeinndelinga av verb i grammatikkane?

## 11. Gruppedeling av verb i grammatikkane

Gruppedeling av verb i grammatikkane skjer med referanse til fleirordige verb-

<sup>5</sup> Næs (1972: 265 ff), Beito (1986: 262 ff), Venås (1990: 84 ff), Faarlund et al (1997: 538 ff.)

konstruksjonar, der eit verb i finitt form ber verb-funksjonen i høve til tid eller modus, medan verbhandlinga ligg i ei infinitt verbform (partisipp eller infinitiv) (t.d. *har gått, kjem gåande, skal gå*). I tillegg kjem *og*-samband mellom to verb i finitt form (t.d. *står og ventar*).

Verb som kan bera verbfunksjonen i slike konstruksjonar blir tilordna undergrupper med nemne som "hjelpesverb" eller "usjølvtendige verb" (t.d. Beito 1986: 268 og 269), eller ein nyttar "hjelpesverb" som overomgrep, inndelt i t.d. "temporale hjelpesverb", "passivdannande hjelpesverb" og "modale hjelpesverb" (Faarlund et al 1997: 516). Desse gruppene blir sette opp i motsetnad til "innhaldsverb".

Desse inndelingane blir i grammatikkane fylgde av inventarlistar for gruppe-medlemmer, og verbgruppene blir avmerkte mot andre verb slik at det ser ut som om funksjonane dei dekkjer, også er avgrensa til dei opplista verba.

Dersom det er sant at ordklassa "verb" er strengt delt i desse gruppene, kan ein ikkje bruka same redigeringskjemaet på alle verblemma. Da trengst det fleire ulike redaksjonelle skjema eller sett med retningslinjer, eller kanskje ikkje noko redigeringskjema utover den strukturen som er førehandsdefinert gjennom sjølve artikkelskjemaet.

## 12. Handsaming av grammatisk funksjon i Norsk Ordbok

Norsk Ordbok har ikkje nokon gjennomarbeidd prosedyre for å handsama grammatikaliserte funksjonar ved einskildverb (hjelpesverb i høve til tempus, modalitet, aspekt). Tanken er vel helst at den overordna framstillinga av slike funksjonar skal finnast i grammatikkane, av di det gjeld eigenskapar som går over lemmanivå. Som ordklasse er verb rekna til innhaldsorda. I ein viss mon blir grammatisk funksjon likevel omtala, anten som lekk i ein definisjon (for å grunngje at visse sitat blir grupperte saman) eller for å syta for at målføredrag blir dokumenterte, også når det gjeld å markera grammatisk funksjon.<sup>6</sup>

Grammatikalske kommentarar til definisjonane i Norsk Ordbok er per dato lite standardiserte og truleg frå konsekvenssynsstad overvariable (same grammatiske fenomen kan mangla kommentar eller ha mange ulike former eller ulikt omfang på kommentaren).

Grammatikalisert bruk av verb er utan fast plassering eller fast omtale i redigeringskjemaet for verb i Norsk Ordbok 2014. I artikkelar om modale hjelpesverb treng den modale funksjonen ikkje vera nemnt eksplisitt, men kan vera synt i ulike tydingar til ulike grupper infinitivar, altså ein vanleg kontekstbasert semantisk analyse. Det manglar ein stø praksis for framstilling av verbbruk i perifrastiske konstruksjonar, i uttrykk for tempus, aspekt osfr. I artikkelane om eit par store verb (*hava, koma*) er såkalla hjelpesverb-funksjonar samla under stor bokstav (D).

6 Jfr t.d. underinndelingane i *Norsk Ordbok VI få A2b*

### 13. Hjelpeverb

Dei såkalla hjelpeverba har ei lang søge som problemkategori i lingvistisk beskriving innanfor mange språk og språkgrupper (Mberu 2006: 28). Her er ordet brukt som det vanlege kategori-nemnet for verb som regelmessig blir nytta saman med infinitiv eller partisipp av andre verb til å laga fleirordige (perifrastiske) verbalformer.

Dei fleste av hjelpeverba som til vanleg blir førde opp, høyrer til den sentrale ordtilfanget og er høgfrekvente i visse typar grammatiske konstruksjonsmønster, men leksikografisk beskriving viser at tydingar og bruk går vidare enn hjelpeverbsbruken.<sup>7</sup> Men dersom desse verba ikkje er berre hjelpeverb, men også har andre tydingar og bruksmåtar, bør dei kunna gå inn i eit vanleg verbskjema. Den leksikografiske utfordringa er då å koma fram til gode beskrivingsmodellar for hjelpeverbfunksjonane.

Motargumentet ville eventuelt vera at dei har eigenskapar som ikkje finst hos nokon andre verb. Vi skal sjå på om dette stemmer.<sup>8</sup>

#### 13.1. Temporale hjelpeverb

Dei verba som vanleg blir lista opp som temporale hjelpeverb i nynorsk, er *bli, få, ha, muna,*<sup>9</sup> *skula, vera, verta, vilja* (Beito 1986: 269). Fleire av desse fungerer også som hjelpeverb i høve til modus og i høve til vekslinga aktiv – passiv.

I tillegg finst det andre verb som har funksjon som temporale hjelpeverb. Det mest kjende i norsk er kanskje *bida*, som blir nytta i fleire perifrastiske verbkonstruksjonar. Eit anna døme finst under verbet *koma*, i det høgfrekvente ordsambandet *koma til*, som med infinitivsubjekt gjeld framtid.

*Nynorskordboka* har per dato den grundigaste handsaminga av fleire av desse verba for nynorsk. *Nynorskordboka* nemner eksplisitt i artiklane sine funksjonen som temporalt hjelpeverb for *bli, ha, mune, vere, verte* men ikkje for oppslagsorda *få, skulle, vilje*, som elles har det sams at dei er tekne med i fleire undergrupper av hjelpeverb i grammatikkane.

Den leksikografiske handsaminga av desse verba er snau i nynorske ordboksverk før *Nynorskordboka*. Det er forståeleg ut frå den tids arbeidsreiskapar, men ikkje noko førebilete for leksikografar med tilgjenge til korpus og andre hjelpemiddel for handtering av store tilfangsmengder.

Av verb med temporale hjelpeverbfunksjonar er *ha* etter måten nyleg redigert i Norsk Ordbok (oppslagsform *hava*). Då redigeringa byrja, var det meininga å gje hjelpeverbfunksjonane til *hava* minimalt med plass, av di desse funksjonane skulle vera adekvat dekte i standardgrammatikkane. Til slutt tok denne delen av artikkelen om *hava* to fulle spaltar. Det kom av at Norsk Ordbok måtte omtala

7 Sjå t.d. artiklane i *Nynorskordboka* for *ha, bli, kunne, måtte, vere*.

8 Opplysningane nedanfor byggjer på søk i databasane for *Nynorskordboka* og Norsk Ordbok. Eg har søkt etter "ordklasse=v" og med definisjonsfelt som inneheld "pass.", "modal", "hjelpev" osv.

9 *muna* er lite brukt (helst foralda) og finst no berre med infinitivsubjekt.

det som standardgrammatikkane ikkje har med, som eldre målbruk og målførebruk. Dermed måtte ein også ta med det normale, som ramme. Skilnaden frå ei grammatikk-beskriving er nettopp at fokus er på å visa bruksmåtane til lemmaet, ikkje på grammatisk funksjon i seg sjølv.

### 13.2. *Hjelpeverb som dannar passiv*

Verba *bli*, *vera* og *verta* er nemnde som verb som blir nytta i fleirordige passivkonstruksjonar. Dette er omtala som ein av fleire moglege funksjonar ved desse verba i Norsk Ordbok<sup>10</sup> og *Nynorskordboka*.

Søk i ordboksbasen og lesing i grammatikkane gjev elles eit inntrykk av at diatese er eit lite utforska fenomen i norske målføre. Det er absolutt ikkje overflødig om Norsk Ordbok kan bidra til å gje eit meir fullstendig bilete av passivkategorien slik han ovrar seg i verbinventar, konstruksjonsmåtar, tydingar og tydingskifte, ved systematisk omtale av einskildlemma.

### 13.3. *Modale hjelpeverb*

Gruppa *byrja*,<sup>11</sup> *kunna*, *lyta*, *måtta*, *skula*, *vilja* er verb som tek infinitivsobjekt (oftast utan infinitivpartikkel *å*), og som tilfører infinitiven eit tydingselement som gjeld måten noko skal skje på.

Men dette gjeld ikkje berre for den gruppa som har fått merkelappen modale hjelpeverb. Svært mange verb tek infinitivsobjekt, og det er kjenneteiknande for konstruksjonen verb + infinitiv at det finitte verbet då seier noko om korleis den handlinga som infinitiven nemner, blir utført.<sup>12</sup>

Dersom dei modale hjelpeverba var aleine om kvar sin modalfunksjon, skulle det ikkje vera mogleg å finna synonym til dei. Da burde substituttdefinisjonen, som er kjerneteknikken for leksikografar, vera systematisk erstatta av funksjonsbeskrivingar.

Men det er fullt mogleg å bruka substituttdefinisjonar på modale verb.<sup>13</sup> Synonyma har ofte ein kjenslekomponent i tillegg – ei eller anna form for forsterking – men slett ikkje alltid, jfr døme:

Modalt verb	Andre verbuttrykk med same funksjon <sup>14</sup>
byrja	vera oppmoda om, vera venta, ha skyldnad til
kunna	orka, greia, klara, makta
måtta, lyta	få, verta tvinga til, verta nauda til; få lov til
skula	vera pålagd, vera sett til
vilja	ynskja, vona, freista, søkja

10 Gjeld berre *bli*.

11 *II byrje* i *Nynorskordboka*.

12 Sjå t.d. Aasen (1965) § 324.

13 Sjå t.d. artiklane for dei aktuelle verba i *Nynorskordboka*, *Bokmålsordboka*, eller *Norsk Riksmålsordbok*.

14 I definisjonane for modale verbydingar finn ein ofte andre modale verb. Ekvivalentane i tabellen er mine, utfrå døme som er funni i Norsk Ordbok-korpuset.

Det er elles interessant at fleire av dei modale hjelpeverba har passivkonstruksjonar som leksikalske motsvar, jfr. gruppedelinga av hjelpeverba som er nemnt ovanfor.

Dei verba som i grammatikkane blir framheva som modale og som er redigerte i Norsk Ordbok, er ikkje karakteriserte som modale i verbartiklane.<sup>15</sup>

I *Nynorskordboka* – som byggjer på dei same samlingane som Norsk Ordbok – er alle dei modale hjelpeverba redigerte utan nokon omtale av den grammatikalske særstillinga, det vil seia at dei har fått vanleg leksikografisk behandling. Dei ulike nyansane av modalitet kjem fram i inndelinga i tydingsbolkar.

Ein anna indikator på at modalfunksjonen går vidare enn til dei få verba som blir grupperte som modale i grammatikkane, er om dei same verba blir brukte som motsvar i definieringa av andre verb. I Norsk Ordbok er stoda førebels denne:

Modalt verb	Verb som har modalt verb som motsvar på definisjons plass
kunna	begå, byrga, duga, greia, kynna
måtta	byrja, greia
lyta (ljota)	bli
vilja	gidda, intendera, lika

#### 14. Verb med infinitivobjekt

Verb som kan fylgjast av infinitiv, tilfører ofte aspekt eller modus i høve til den (abstrakte, potensielle) handlinga som infinitivforma representerer. Nedanfor er ei liste med verb som finst i nynorskorpuset framfor infinitiv, grovsorterte etter tyding:<sup>16</sup>

*slutta, opphøyra*  
*nehta, avvisa, unngå*  
*lika, elska, hata*  
*bruka, pla*  
*nytta (i tyd 'føra fram, lukkast')*  
*lata, tillata*  
*tora, våga*

Skiljet mellom desse verba og gruppa modale verb skal mellom anna syna seg i at infinitivobjekt skal ha med seg infinitivmerket:

*han likte å danse*  
*det nyttar ikkje å stola på posten*

15 Ordet "modus" finst i Norsk Ordbok berre i definisjonen av grammatikktermane *imperativ, indikativ, infinitt, konjunktiv* og *bydemåte*, medan "modal" er nytta i definisjons-tekst til verba *gerast* og *kny*, og i funksjonsbeskriving for some preposisjonar og adverb.

16 Det at verb er grupperte saman, tyder ikkje at dei er fullsynonyme eller tek dei same infinitivane etter seg.

Men i praksis blir infinitivsmerket sløyfa ved mange fleire verb,<sup>17</sup> og det er vel også ein viss målføregeografisk distribusjon både i korleis infinitivsmerket ser ut, og når det blir brukt. Motsvarande ser vi at somme av dei verba som er peika ut som modale, kan ta infinitiv med infinitivsmerket på plass:

Han kan å dansa, han der

Når verb som kan ha infinitiv etter seg, vert handsama leksikografisk eitt for eitt, må også modal og aspektuell bruk handsamast, og da blir skiljet mellom verb som tek infinitivsobjekt og ”modale hjelpeverb” langt på veg utviska.

### 15. Gruppedeling av verb i leksikografien

Den uformelle leksikografinnndelinga av verba er som nemnt ”store verb” mot resten. Men dei store verba femner om verb som frå lingvistisk synsstad er svært ulike. Mellom dei nemnde store verba (*ha, eiga, fara, fylgja, gå, koma, liggja, stå, vera*) finn ein såleis både transitivitet og inntansitivitet, hjelpeverb- og innhaldsverb -funksjonar, fleire ulike aspekt osb.

Kva slag verb er det då som står i motsetning til eit stort verb? Fleire av dei verba som blir rekna som store verb, er også hjelpeverb i tempussystemet, og skulle såleis stå i motsetnad til andre store verb. Men det er alt synt gjennom redigering at det går bra å bruka same redigeringskjema for svært ulike store verb.

Verb som ikkje er store i den forstand at dei har mykje tilfang attom seg, kan likevel ha svært stort tydingsspekter. Analysen attom kvar tyding kan bli mindre sikker av di eit mindre tilfang ikkje syner tydeleg kva som er konvensjonalisert språkbruk, og kva som er individuell eller tilfeldig språkbruk. Det finst mange døme på dette mellom verb med meir talemålstilfang enn skriftleg tilfang attom seg.<sup>18</sup> Likeins finst det nok av døme på at verb med eitt stort tilfang attom seg i samlingane, kan syna få leksikaliserte tydingar eller funksjonar. Det kan også vera noko tilfeldig kva for verb som har eit stort eller lite tilfang attom seg i samlingane.<sup>19</sup> Men dette er ikkje noko argument mot eit sams redigeringskjema for verb.

### 16. Leksikografisk versus grammatikalsk omtale av einskildord

Leksikografisk handsaming i ordartiklar er i prinsippet eit portrett av kvart lemma, sett saman av tydingsbolkar som syner tydingar og bruksmåtar. Desse større ordboka er, dess meir fintmerkande er inndelinga.

Definisjonsprinsippet er å føra opp ein eller fleire (ofte fleirordige) ekvivalen-

17 Faarlund et al (1997: 996) har døme på dette.

18 272 verb i basen har 6 eller fleire definisjonar (a-l).

19 Talemålstilfanget er ujamnt både i høve til geografi og storleiken på dei lokale samlingane. Korpus og skriftspråklege ekserperte kjelder har ujamn sjangerdekning og dekkjer heller ikkje alle tidbolkar like godt.



tar som kan fungera som substitutt. Praksis i Norsk Ordbok peikar mot at det blir lagt mykje arbeid i å skriva gode definisjonar og gruppedela (subklassifisera) døme slik at ekvivalenten er dekkjande.

Når ein finles verbartiklar og ser på dei i samanhang med døma, ser ein ofte at det ligg grammatiske kriterium til grunn for sorteringa, dvs at type og orden på kollokat avgjer grupperinga som ligg under definisjonane. Men omtalen av dei grammatiske rammevilkåra ymsar mykje, frå ordbok til ordbok og også innanfor større ordbøker og ordboksverk.

Grammatisk behandling av språkdrag rettar seg mot sams åtferdsdrag ved grupper av ord. Dette gjev eit anna fokus enn leksikografien har. Grammatikarar veit sjølvsagt at same ordet kan opptre i fleire grupper og illustrera ulike språkdrag, men dette er ikkje viktig så lenge fokus er på sjølve funksjonen.

Når grammatikaren tek for seg ei lita gruppe ord, kan den grammatikalske beskrivinga bli galen frå leksikografisk synsstad, av di ordet blir identifisert med ein viktig grammatikalsk funksjon. Men dei grammatiske gruppedelingane gjeld for tydingar og bruksmåtar som er knytte til verb, ikkje for heile lemma. Det er likevel lett å sjå at ein velordna ordboksdatabase kan vera ei viktig støtte for identifisering og beskriving av grammatiske drag i språket.

## 17. Kvifor skal ordbøker handsama grammatikaliserte funksjonar ved verb?

Med fare for å gjenta meg sjølv: Ordbøker bør handsama grammatiske funksjonar ved verb av di desse i praksis ikkje let seg skilja tydeleg frå den meir leksikaliserte delen av tydingsspekteret. Det er ein illusjon å tru at semantikk let seg isolera frå grammatikk, at grammatikalsk funksjon ikkje har tyding, eller at grammatikalske konstruksjonar ikkje verkar inn på tydingsinndeling. Språkbruk er jo tydingsformidling — korleis ein organiserer det innanfor kvart språk er eit spørsmål om funksjonsfordeling på grammatikk og leksikon.

For ei dokumentasjonsordbok som ønskjer å gje full dekning av talemålet, kjem det i tillegg at verbparadigma i målføra har både former og konstruksjonar som manglar i normalmålet, og som kan vera høgfrekvente, til og med dominerande, innanfor sine område. Slike bruksmåtar kan vera leksikaliserte, slik at dei kjem fram ved somme verb men ikkje andre. Alt dette må vi ha med, og det må med i den normalspråklege konteksten. Noko anna er at behandlinga nødvendigvis må gjerast stutt, og at til dømes ulike grammatiske forklaringsmodellar ikkje kan koma med.

## 18. Redigerings skjema for verb

### 18.1. Eitt eller fleire redigerings skjema?

Ovanfor er spørsmålet reist om det ut frå lingvistiske kriterium er mest forsvarleg med eitt eller fleire redigerings skjema for verb. Eg meiner å ha vist at det rettaste er å ha eitt sams maksimumsskjema. Gruppedelingane som finst i grammatikkane gjeld ikkje verb som lemma, altså med fullt dokumentert tydings spekter. Gruppedelingane er knytte til visse verbfunksjonar, eksemplifiserte med små grupper verb som er høgfrekvente i den aktuelle funksjonen. Gruppemedlemskap i ein gjeven funksjon tyder ikkje at eit verb ikkje kan ha andre tydingar eller funksjonar, eller at ikkje andre verb kan fylla same rolla.

Drøftinga ovanfor kan dragast saman slik:

- Dei generelle eigenskapane til verb legg til rette for polysemi (jf avsnitt 9)
- Dersom eit verb er frekvent, er det sannsynlegvis fleirtydig
- Konstruksjonsdrag ligg oftast attom tydingsindeling, anten det er gjort eksplisitt (med grammatisk kommentar til definisjonen) eller ikkje
- Faste verbsamband blir til vanleg skilde ut og får separat behandling under verbet
- Det er leksikografisk tradisjon for å byrja ordartikkelen med dei tydingane der verbet har mest eigentyding<sup>20</sup>

### 18.2. Føremonar og ulemper med fast redigerings skjema

Ovanfor har eg drøfta føremonar og ulemper med å ha eit fast redigerings skjema i høve til leksikaliserte sublemma (jfr. avsnitt 8). Ulempa med eit redigerings skjema som gjev fast rekkjefylgd av konstruksjons- og tydingstypar i høve til sjølve verbet, må vera at hovudfunksjonane til verbet ikkje blir topikaliserte på same måte som når redaktøren står fritt til å ordna tydingsrekkjefylgd. Står redaktøren fritt, kan han setja først det han meiner er viktigast.

Men når ein ser på dei ordartiklane for verb med hjelpefunksjonar i *Nynorskordboka* og *Norsk Ordbok*, ser ein at artiklane tek til med dei tydingane der verbet har mest eigentyding, og at dei grammatikalisererte funksjonane kjem lenger ut. Da vil eit eksplisitt fastare redigerings skjema meir fungera som ei formalisering og opprydding enn som noko brot med gjeldande praksis.<sup>21</sup>

Føremonen med eit fast redigerings skjema er at det blir lettare å planleggja redigeringa, innordna tilfang, og ikkje minst jamføra ulike verb. Slik kan eit redigerings skjema på sikt også bidra til å betra underlagstilfanget for handsaminga av modus, aspekt osv ved verb i grammatikkane, av di desse verbeigenskapane får ei ryddigare og meir føreseieleg framstilling i til dømes *Norsk Ordbok*.

<sup>20</sup> I døme som *ho har ny bil* (til *ha*), *dei kom seint* (til *koma*).

<sup>21</sup> Det same gjeld som hovudregel for ordbøkene til Ivar Aasen, og for *Grunnmanuskriptet* til *Norsk Ordbok*.

### 18.3. Krav til redigeringskjema for store ordboksverk

Redigeringskjema som er utarbeidde til bruk i store ordboksverk må stetta nokre viktige funksjonskrav. Det må vera fagleg robust, oversiktleg og attkjenneleg. Det må opplevast som rimeleg logisk av både ordboksredaktøren og ordboksbrukaren, og det må bruka dei sorteringssystema som ordboka generelt brukar.

Kva vil det seia at skjemaet må vera fagleg robust? Dette handlar om tid og tidsbruk. Store ordboksverk tek lang tid å redigera og skal vara lenge — mykje lenger enn svingningane i lingvistisk teori til vanleg gjer. Når ein ser attover eit par generasjonar, ser ein at lingvistisk teori ikkje utviklar seg frametter i rett line, men går ganske kraftig i sikksakk — jfr. t.d. Chomsky-perioden versus dagens korpuslingvistikk. Derfor kan ikkje eit ordboksverk gå for tett innpå aktuelle lingvistiske og grammatiske modellar, ein må halda seg nær skulegrammatikken<sup>22</sup> og den terminologien som har synt seg å vera etter måten stabil.

Vidare skal strukturen vera lett å læra for brukarar med ei anna lingvistisk skoloring enn den redaksjonen har. Dette kan ein oppnå om ein held terminologien enkel og syter for at tilfanget er ordna og definert slik at det får tala for seg. Då vil strukturen isolera og framheva fenomenet, medan definisjonen gjev det ramme og substitutt i språkleg kontekst. Er artikkelen ordna med slike mål for augo kan brukaren tola at den lingvistiske beskrivinga er knapp eller utdatert, og likevel forstå poenget med omtalen.

### 18.4. Korleis bør eit skjema for verbredigering sjå ut?

Spørsmålet om ei skjema for redigering av verb kan handterast meir radikalt om ein tek til med blanke ark, i planlegginga av eit nytt ordboksprosjekt. Norsk Ordbok er midt i seilassen, med seks band utgjevne og seks att. Vi må derfor byggja på dei gode tradisjonane vi har, og ta utgangspunkt i dei eksisterande konvensjonar for store verbartiklar. Kan ein få til kvalitetsforbetring utan å bryta med innøvd og tilvand skikk, så er det også ein klar føremon.

Dersom det skal vera mogleg for mange redaktørar å praktisera eit slikt skjema i høve til ei større ordgruppe, bør ordningskriteria så langt som mogleg vera ikkje-semantiske, det vil seia knytte til form og kollokat. Tilfangsordninga bør gå

- frå sterkast mogleg eigen tyding til bunden fraseologi og grammatisk funksjon
- frå maksimal variasjon i kollokat og nedover til meir og meir føreseieleg kontekst

Da får vi fylgjande redigeringskjema:

- A. verb + type kollokat (inndeling etter valens)
- B. verb + "seg" (+ adverb) (Alfabetisk ordning av underoppslag)

<sup>22</sup> Skulegrammatikken kan også bli endra, jf. ordklassenemna i lærebøker for norsk skule, som skal endrast så dei kjem meir i samsvar med Faarlund et al (1997).

C. verb + adverb (faste samband) (Alfabetisk ordning av underoppslag)

Når det gjeld dei mest grammatikaliserte funksjonane, må dei koma etter eller sist i A, eller som ein eigen seksjon D til slutt.

## 19. Konklusjonar

Ordklasse verb har eit tydingsspekter som strekkjer seg frå sterkt og eintydig innhald, knytt til sjølve verbet, til ymse grammatikaliserte funksjonar. Dei såkalla store verba syner dette spekteret i størst mangfald, men svært mange mindre belagde verb dekkjer heile skalaen.

Som ordklasse har verb sterkare sams drag enn gruppedrag, slik at eit sams redigeringsskjema er fagleg forsvarleg. Førmonen med eit maksimumsskjema er at det ikkje hindrar delutnytting, ein verbartikkel blir korkje stuttare eller lengre av å stø seg på eit førehandsdefinert skjelett.

Svært mange verb utnyttar heile redigeringsskjemaet ovanfor, men i ulik grad på ulike punkt. Skjemaframlegget byggjer på etablerte komponentar, dvs sorteringssystemet og den logiske ordninga som alt er i bruk.

Ein mogleg ulempe kunne vera svakare topikalisering av hovudfunksjonar ved einskildverbet. Men slik topikalisering blir ikkje nytta. Det er til dømes slik at same kor dominerande dei grammatikalske funksjonane ved eit verb er i tilfanget i høvet til frekvens, så blir dei alltid plassert etter den meir leksikaliserte bruken.

Førmonen med eit meir eksplisitt redigeringsskjema at det blir lettare å jamføra nærståande verb. For redaktørar vil det bli lettare å redigera nytt tilfang. For brukarane av Norsk Ordbok er førmonen eit meir oversynleg produkt, som er lettare å finna fram i og referera til.

## Litteraturliste

- Beito, Olav T. 1986: *Nynorsk grammatikk. Lyd- og ordlære*. Oslo: Det Norske Samlaget.
- Bokmålsordboka* 2005. Definisjons- og rettskrivningsordbok. 3. utg., 2. rev. oppl. Redaktør: Boye Wangenstein. Oslo: Kunnskapsforlaget.
- Faarlund, Jan Terje, Svein Lie og Kjell Ivar Vannebo 1997: *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Grønvik, Oddrun 2005: Redigeringsprogrammet for Norsk Ordbok 2014. I skjeringspunktet mellom menneskeleg skjønn og automatisering. I: *Nordiske studiar i leksikografi* 7. Rapport frå Konferanse i leksikografi i Norden. Volda 20.–24. mai 2003. Oslo: Nordisk foreining for leksikografi, 156–165.
- Guttu, Tor 2005: *Norsk ordbok med 1000 illustrasjoner* : riksmål og moderat bokmål. 2. utg. Oslo: Kunnskapsforlaget.

- Mberi, Nhira Edgar 2006: *The Categorical Status and Functions of the Auxiliary in Shona*. Harare. Oslo. The ALLEX-project, Department of Linguistics and Scandinavian Studies, University of Oslo Unipub.
- Norsk Ordbok*. Ordbok over det norske folkemålet og det nynorske skriftmålet 1966–. Hovudredaktør: Alf Hellevik ... [et al.]. Oslo: Det Norske Samlaget.
- Norsk riksmålsordbok 1937–1957* (med tilleggsbind 1995). Trygve Knudsen, Alf Sommerfelt [og Harald Noreng; under medvirkning av K.E. Bødtker ... [et al.]. Oslo: Kunnskapsforlaget.
- Nynorskordboka* 2006. Definisjons- og rettskrivingsordbok. Redaktør: Marit Hovdenak ... [et al.]. Oslo: Det Norske Samlaget.
- Næs, Olav 1972: *Norsk grammatikk*. Elementære strukturer og syntaks. Oslo: Fabritius.
- Svensén, Bo 2004: *Handbok i lexikografi*. Stockholm: Norstedts Akademiska Förlag.
- Venås, Kjell 1990: *Norsk grammatikk*. Nynorsk. Oslo: Universitetsforlaget.
- Aasen, Ivar 1850: *Ordbog over det norske Folkesprog*. Kristiania.
- Aasen, Ivar 1870–73: *Norsk Ordbog*. Kristiania.
- Aasen, Ivar 1965: *Norsk Grammatik*. Etter fyrsteutgåva frå 1864. Oslo: Universitetsforlaget.

### Ordbøker på nettet

- Grunnmanuskriptet for Norsk Ordbok: <http://www.edd.uio.no/perl/search/search.cgi?tabid=993&tappid=59>
- Oxford English Dictionary*: <http://www.oed.com/>
- Svenska Akademiens ordbok*: <http://g3.spraakdata.gu.se/saob/>

Oddrun Grønvik  
hovudredaktør, f. 1947  
Norsk Ordbok 2014  
Universitetet i Oslo  
postboks 1021 Blindern  
NO-0315 Oslo  
[oddrun.gronvik@iln.uio.no](mailto:oddrun.gronvik@iln.uio.no)



## Båndoptagelser som kilde til Ømålsordbogen

**Use of tape recordings as empirical basis at Ømålsordbogen.** The aim of Ømålsordbogen (The Dictionary of the Insular Dialects) is to embrace the whole vocabulary of the dialects, dialecticisms as well as words in common with standard Danish. The dictionary is based on a collection of approximately 3 million slips, excerpted from various sources. The major portion of the material, however, has been collected by the staff itself. Despite a thorough work trying to cover all of the vocabulary, a number of words and expressions are poorly attested in the collection of slips. This goes especially for function words and frequent words with no striking difference in use or meaning between standard and dialect – words of this kind have often been left unnoticed both when recording data and when excerpting. A corpus of transcribed tape recordings has therefore turned out to be a useful supplement. And the corpus has not only permitted a more adequate description of function words etc., it has also provided the dictionary with more authentic quotations.

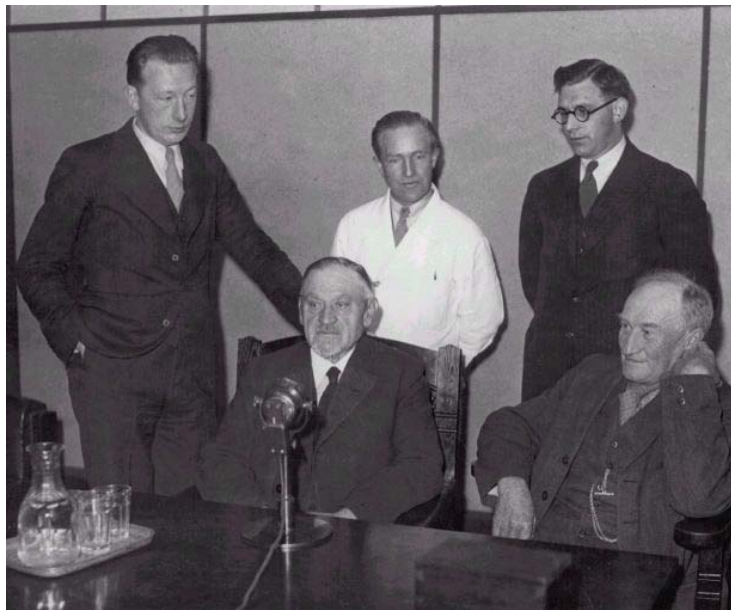
### 1. Indledning

Ømålsordbogen er en dialektordbog der beskriver dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. Den er projekteret til at omfatte 18 bind og giver altså en bred og detaljeret beskrivelse af sproget på øerne. Den medtager således ikke kun de særdialektale ord, men også fællesdanske ord, herunder hjælpeverber, præpositioner osv. Ordbogen bygger først og fremmest på en seddelsamling. Samlingen er stor, omkring 3 millioner ord, men det er noget svingende hvor velbelyste ordene er. Fx er der omkring 150 sedler med den sjællandske betegnelse på formiddagsmåltid *halvgåenmiddag*, selv om ordet var ved at gå af brug på indsamlingstidspunktet, mens derimod *altså* brugt for at udtrykke sammenhæng med det foregående oprindelig kun var repræsenteret med to belæg i seddelsamlingen (jf. Pedersen 1983:133). De meget almindelige ord og betydninger overses ofte ved optegnelse og excerpering – og i øvrigt kan brugen af funktionsord ofte være svær at fastholde uden lydoptagelser. Det korpus af båndudskrifter, som ordbogen har fået opbygget i de sidste 30 år, er derfor blevet et værdifuldt supplement til seddelsamlingen, men som vi vil vende tilbage til i slutningen af artiklen, er båndoptagelserne af flere grunde ikke helt uproblematisk som kilde.

### 2. Forhistorien

Indsamlingen af lydoptagelser begyndte inden båndoptageren kom frem i 1950'erne. På billedet (fig. 1) fra 1938 ses til venstre Poul Andersen, daværende

leder af Udvalg for Folkemaals kontor og hermed også lederen af indsamling af materiale til Ømålsordbogen. Forrest sidder to langelandske meddelere, og bag ved dem står to teknikere fra Statsradiofonien, som Danmarks Radio hed dengang. Man havde tidligere forsøgt sig med optagelser med fonografvalser af voks. Det havde den fordel at optagelserne kunne foregå i meddelernes egne hjemlige omgivelser, men man blev bekymret for at valserne ikke ville være tilstrækkeligt holdbare, og gik så over til optagelse på lakplader i et radiostudie.



Figur 1. Lakpladeoptagelse i Danmarks Radio 1938.

Det er jo ikke en ideel optagelsessituation at hive et ældre menneske ind i et radiostudie, men i almindelighed gik det forholdsvis godt, bl.a. fordi den generation af meddelere havde dialekt som deres eneste daglige og naturlige udtryksform og ikke normalt havde brug for at slå over i rigsmålstilnærmede sprogformer. Metoden havde yderligere en ulempe: Lakpladerne kunne kun rumme ganske korte optagelser af 4-5 minutters varighed. Indholdet var derfor i store træk aftalt på forhånd, og optagelserne er i reglen helt eller næsten monologiske. Optagelsen indledes ofte med en replik fra dialektforskeren, fx *Vi skulde snakke om de her gamle Skikke I havde der ved Nytårstid* (Danske Dialekttekster 4 1976:23), men herefter fortæller meddeleren stort set uden afbrydelser. Det var derfor af flere grunde et stort fremskridt da man kunne tage båndoptageren i brug i 1950'erne. Ikke alene kunne man nu optage meddelerne i hjemlige omgivelser; man kunne også lave lange optagelser der nærmer sig en almindelig samtale. Den mest intensive optagelsesperiode var i 1970'erne, og størstedelen af indtalerne er født i perioden 1890–1910.

Hovedformålet med optagelserne var at bevare prøver på dialekterne for efter-



tiden, og det er vel stadigvæk hovedformålet. Men i 1970'erne blev man opmærksom på at båndoptagelserne også kunne bruges som kilde til Ømålsordbogen. Det var på et forholdsvis sent tidspunkt i ordbogens historie.

I 1909 blev Udvalg for Folkemaal nedsat, og udvalgets vigtigste opgave var at tage initiativ til en ordbog over de danske øsmål. Reelt går ordbogens historie dog kun tilbage til 1922, hvor arbejdet blev reorganiseret: Udvalget fik egne lokaler, og den første fuldtidsmedarbejder blev ansat, nemlig Poul Andersen. De første mange år var især koncentreret om indsamlingsarbejdet, selv om man dog også efter 1940 begyndte at lave en foreløbig redaktion af visse saglige emner, fx høst, pløjning, bagning og lege (jf. Andersen 1932 og 1939, Ømålsordbogen Tillægsbind:1f.). Omkring 1975 blev ordbogens format endeligt fastlagt, og ordbogen gik ind i den egentlige redaktionsfase. Det første bind udkom i 1992, og herefter er der udkommet et bind hvert andet år med bind 8 (*hundepisk-isrør*) fra 2006 som det hidtil sidste. Det sprog der beskrives er de traditionelle landdialekter som de taltes til hen imod 2. verdenskrig.

### 3. Ømålsordbogen og båndoptagelserne

#### 3.1 Etableringen af korpusset

Afdeling for Dialektforskning har omkring 4000 optagelser med dialekttalende, men det er kun en lille del af dem der udnyttes ved ordbogen, nemlig dem der er udskrevet under en eller anden form. Det er jo forudsætningen for at de er praktisk anvendelige i ordbogsarbejdet.

Den mere systematiske udskrivning af båndoptagelserne blev indledt i slutningen af 1960'erne, hvor man begyndte at udgive serien *Danske Dialektttekster*. Det er en serie hæfter med båndoptagelser transskriberet i Danias lydskrift og med parallelgengivelse på rigsmål. Det var imidlertid ikke primært med ordbogen i tankerne at udgivelsen blev sat i gang. Formålet var, skriver Poul Andersen i forordet til bind 2, "dels at sikre det i Båndene liggende Materiale for fremtidig Forskning, selv om de tilsvarende Bånd skulle forgå, dels at skaffe Materiale til Undervisning" (*Danske Dialektttekster 2 1973:5*).

I 1972 da udgivelsen af *Danske Dialektttekster* var godt i gang, fik Bent Jul Nielsen, lektor ved Institut for Dansk Dialektforskning, imidlertid en idé der var forholdsvis avanceret dengang: at indkode dialektteksterne på magnetbånd, og der blev nedsat et udvalg til at udarbejde en plan for arbejdet. I planen indgik også en slags opmærkning af ordene, så det blev muligt at foretage forskellige sproglige undersøgelser, og man lagde særlig vægt på at der skulle kunne foretages præcise ordsøgninger maskinelt – netop med henblik på ordbogsarbejdet (Pedersen 1974:31 og 1983:133). Projektet blev gennemført i midten af 1970'erne, men i de første år befandt teksterne sig på en computer uden for instituttet (på det daværende EDB-center ved Københavns Universitet), og det krævede i øvrigt en

særlig ekspertise at håndtere maskinen. Af og til – når redaktøren skønnede at det var ”strengt nødvendigt” – kunne man rekvirere en konkordans; i det daglige arbejde brugte man især det ordregister til serien Dialekttekster, der også var et afkast af indtastningen. Belæggene måtte skrives ud i hånden, men alene brugen af registret gav nye muligheder for at supplere grundmaterialet. Derudover excerperede man manuelt ”de allermest sjældne og interessante belæg”, dels i forbindelse med udskrivningen, dels i forbindelse med den gennemlytning der foretages når båndene katalogiseres og sagregistreres (Pedersen 1983:135).

Dialekttekster excerperes stadigvæk i hånden med henblik på udtaleformer og citater i fonetisk transskription, men nu udnytter vi især båndoptagelserne ved at søge på vores elektronisk lagrede korpus<sup>1</sup> af udskrifter i rigsmålsortografi. Og mens det tidligere kun var lejlighedsvis at udskrifterne blev udnyttet systematisk, har redaktørerne nu pligt til at søge de ord op i korpusset, som de redigerer, og til at tage hensyn til materialet herfra.

Selv om båndene er udskrevet i rigsmålsortografi, er udskrivningen en resurserkrævende proces, der har sat grænser for korpussets størrelse. Korpusset omfatter således kun 1,5 millioner ord, og kan altså ikke bruges til egentlige statistiske formål, men det har givet os let adgang til værdifuldt materiale fra båndene. Problemet med korpussets ringe størrelse opvejes også til en vis grad af at det ofte er de meget almindelige og frekvente betydninger og konstruktioner der er dårligt belagt i seddelsamlingen. Netop det meget almindelige og upåfaldende er som sagt det en optegner eller excerptist vil være tilbøjelig til at glemme at få med. Desuden må man også huske på at det var en meget omfattende indsamlingsopgave, og at der var meget andet man også skulle have indsamlet inden det eventuelt var for sent.

### 3.2 *Brugen af korpusset*

Båndoptagelserne har i særlig grad haft betydning for redaktionen af funktionsord o.l., men de har desuden gjort det muligt at bringe bedre og mere autentiske citater end vi ville have kunnet på grundlag af seddelsamlingen alene.

Ømålsordbogens grundmateriale er som sagt en seddelsamling på hen ved 3 millioner sedler. Nogle af sedlerne er udskrevet fra trykte kilder, fx bondedagbøger og dialektlitteratur eller fra utrykte ordsamlinger der stammer fra ældre indsamlinger, men grundstammen udgøres af det materiale som medarbejdere ved ordbogen har indsamlet. Det er meget forskelligartede optegnelser. Nogle indeholder lydformer, andre saglige beskrivelser af redskaber o.l., og andre igen giver eksempler, ofte lydskrevne, på ordenes brug. At eksemplerne ofte er i lydskrift slører den omstændighed at der ofte er tale om en slags redaktionelle eksempler, nemlig eksempler som en meddeler har givet spontant eller på opfordring til at illustrere et ords brug (jf. Hovmark 2004:4ff.).

1 En del af vores tekstkorpus er lagt ud på Internettet under navnet CorDiale (se <http://dialektforskning.ku.dk/tekstbaser/cordiale/>).

Ved indsamlingen er der i reglen brugt forskellige former for spørgelister. Spørgelisterne kan være besvaret enten ved indirekte metode, hvor meddeleren besvarer udsendte spørgsmål skriftligt, eller ved direkte metode, hvor en filologisk uddannet optegner udspørger meddeleren.

Følgende belæg er et eksempel på hvad man kan finde i seddelsamlingen (under verbet *banke*):

- (1) *der blev han ordentlig banket af* (Nordsjælland, Tune)

Det er en mundret replik som man sagtens kan tænke sig at nogen har sagt, men eksemplet er fremkommet som svar på et spørgsmål fra den såkaldte fraseologiske spørgeliste, udsendt omkring 1930 (Andersen 1932:186), hvor der spørges til kollokationen *banke af* ved at der anføres nogle betydninger som meddeleren skal tage stilling til. Spørgsmålet har nærmest karakter af en ordbogsartikel:

- (2) 1) gøre rent ved Bankning, fx banke Sengeklæder af, 2) gennemprygge, fx han bankede ham ordentlig af.

Den første betydning bekræftes, men det andet spørgsmål besvarer meddeleren kun indirekte, nemlig med sprogbrugseksemplet ovenfor. Som det ses, tager hun udgangspunkt i dialektologens forslag, men eksemplet modificeres, formentlig i mere mundtlig retning.

Spørgsmålene til meddelerne kan også være formuleret mere åbent. I en spørgeliste udsendt til de faste meddelere hedder det fx:

- (3) *dvæle* v. Mon dette ord kan bruges i ægte dialekt nogetsteds? Hvis De mener det, beder vi om eksempler, der viser brug og betydning.

Og en meddeler svarer:

- (4) Hører af og til (mest ældre) sige lige før de skal gå: *nå, vi sgate dvæle mer* (Thurø)

Her har meddeleren selv udformet eksemplet, og det gode i dette og mange andre svar er at eksemplet knyttes til en typisk situation. I dette tilfælde er vi altså langt tættere på et autentisk citat, men der er principielt tale om et redaktionelt eksempel.

Der er dog også i seddelsamlingen eksempler på ægte, autentiske replikker, replikker som en meddeler har sagt undervejs, og som optegneren har fået noteret ned i farten. Sådanne citater kan imidlertid være vanskelige at tolke, fordi de kan stå helt isoleret uden hverken sproglig eller situationel kontekst, det gælder fx:

- (5) *det ender nok med at det hele slår en kolbøtte* (Nordsjælland, Allindemagle)

Er *slå kolbøtte* brugt konkret eller overført, og hvad er *det hele*? Det står der ingen steder, så citatet må udgå, selv om det er interessant fordi *slå kolbøtte* er brugt med

upersonligt subjekt. Og det er desværre ikke helt sjældent at autentiske belæg fra seddelsamlingen ikke kan bruges i ordbogen, og citaterne fra båndoptagelserne har altså givet et godt tilskud til mange artikler.

Båndoptagelserne er også nyttige fordi de giver os et fyldigere materiale til beskrivelsen af det ordforråd der tilhører andelstiden og den begyndende mekanisering af landbruget (jf. Gudiksen 2006:155), men endnu vigtigere er det at båndoptagelserne har givet mulighed for at supplere materialet til beskrivelsen af funktionsord og andre frekvente ord, der som sagt ofte er underrepræsenteret i optegnelserne. Ordet *ja* er et godt eksempel. Artiklen fylder 16 spalter, og mere end 2/3 af materialet stammer fra båndoptagelser. Når det gælder *ja*'s kommunikative og samtalestrukturerende funktioner, så stammer praktisk talt alle belæg herfra, som fx det følgende citat der er et eksempel på at *ja* kan bruges som signal om at man indleder en tale (en fortælling, en forklaring osv.):

- (6) (hvordan gik det så at få ovnen varm)  
*ja, da lagde de jo noget halm ind*  
 (hvem gjorde det?)  
*ja, det gjorde manden jo gerne*  
 (Ømålsordbogen bind 9, *ja* 7.3)

Eller som her hvor *ja* bruges som signal om at man indleder en ny type information, her en korrektion eller præcisering:

- (7) (ølgæren) *den købte vi ved bryggeren, ja købte der, vi gav itte meget for den*  
 (Sydfyn, Horne)  
 (8) *og jeg har det jo allerbedste, for jeg er jo temmelig rask, ja det vil sige jeg har uro omkring, men det ..* (Østsjælland, Smerup)  
 (Ømålsordbogen bind 9, *ja* 8.4)

Det var svært at finde eksempler på *ja* brugt af tilhøreren til at vise at han følger og forstår hvad taleren siger, sådan som i eksemplet her:

- (9) *vi kommer jo så ud ved bækken eller der ude, og der var sådan noget småkrat – ja –*  
*og da .. fløj der en fasan op* (Østfyn, Revninge)  
 (Ømålsordbogen bind 9, *ja* 12)

Dvs. der var hundredvis af eksempler i korpusset – sagt af Poul Andersen eller andre kolleger. Heldigvis har vi dog enkelte båndoptagelser med lokale interviewere, så det lykkedes at finde belæg, men vi var for en gangs skyld nødt til at inddrage bånd der ikke indgår i korpusset.

### 3.3 Problemer ved udnyttelsen af båndoptagelserne

Vores problemer med *ja* afspejler at samtalerne på båndene er særlige samtaler. Den ene samtalepartner (der normalt er en rigsmålstalende dialektolog) lytter og

stiller spørgsmål, mens den anden samtalepartner fortæller (og forventes ikke at stille spørgsmål). Emnet for samtalen er for det meste livet på landet i meddelersens ungdom. I nogle passager fortæller meddeleren historier, dvs. beretter om tildragelser som han eller hun har oplevet eller fået fortalt af andre. Det er ofte historier som har en fast form, fordi indtaleren har fortalt dem mange gange før, og i de passager adskiller indtalerens bidrag sig ikke væsentligt fra hvad han kunne have sagt i en almindelig samtale. I andre passager gør indtaleren på opfordring rede for arbejdsprocesser og skikke. Disse passager rummer elementer der sjældent forekommer uden for optagelsessituationen, og man kan ofte diskutere hvor autentisk sproget egentlig er når det kommer til stykket. Det følgende belæg fra en båndoptagelse er bragt i artiklen *højtenbenet*:

- (10) “Efterhånden som de (dvs. skørterne) blev slidt og skulle have en frisk kant, der blev de jo kortere .. og så kunne vi se anklerne, og så blev de (dvs. pigerne) kaldt *højtenbenet*” (Ærø, Bregninge)

Det at der er brugt anførselstegn og ordinær i citatet undtagen i opslagsordet, følger Ømålsordbogens praksis. Det markerer at citatet er metasprogligt, og ikke et eksempel på hvordan ordet oprindeligt har været brugt i konkrete samtaler for 100 år siden. Samtalens ofte metasproglige karakter afspejler sig undertiden også i ordforrådet. *kalde* i den betydning som ses i citat 10 ovenfor: ‘bruge en vis betegnelse om nogen eller noget’, forekommer fx hyppigere på båndoptagelserne end normalt.

Man skal også være opmærksom på at interviewerens direkte kan påvirke meddelersens ordvalg, som i dette eksempel hvor det er usikkert om *kvinde* er det ord meddeleren normalt ville have valgt i betydningen ‘voksent menneske af hunkøn’, eller om han ville have brugt ordet *kone*:

- (11) (det var almindeligvis kvinder der malkede) *det var nærmest kvinder der malkede dengang* (Østsjælland, Nordrup)

Der er også tidsmæssige problemer ved udnyttelsen af båndoptagelserne. Ømålsordbogen beskriver som sagt de traditionelle dialekter. Ordbogen omfatter perioden fra 1750 – det tidspunkt hvor man overhovedet begynder at have optegnelser – til omkring 2. verdenskrig, med 1850-1920 som ordbogens kerneperiode. Men mange af båndoptagelserne er fra 1970’erne, i enkelte tilfælde senere. Vi bruger altså data fra slutningen af 1900-tallet som kilder til sproget i begyndelsen af århundredet eller før. Det tør vi bl.a. fordi indtalerne er gamle mennesker, der normalt også har en konservativ sprogbrug. Men vi må være på vagt over for sproglige fænomener der er indkommet i dialekterne efter krigen, og det vil næsten altid sige indslag fra standardsproget. At beskrive ord og betydninger fra 1945 eller senere i ordbogen vil jo være en direkte fejl, men det er også uheldigt hvis citaterne indeholder elementer der ikke hører hjemme inden for perioden, fx vil vi undgå citater som:

(12) *de folk de sidder og glør på det fjernsyn altid*

eller

(13) *der var itte noget med mejetærsker*

selv om begge citater i øvrigt er udmærkede sprogligt set. I enkelte tilfælde har vi dog medtaget et citat fra en båndoptagelse alligevel til trods for at det indeholder et for nyt ord. I artiklen *hvor* finder man fx dette citat under betydning 5 med den redaktionelle kommentar, ‘i relativ sætning med sb som korrelat i oversættelsen’:

(14) *jeg havde sådan et jern hvor jeg lavede nogen reder der til dem* (kartoffelreder, til gæsterne, som kogekone i nyere tid) (Lolland, Landet)

Der er kun få eksempler på at korrelatet, altså *jern*, betegner et redskab som man bearbejder noget med, så derfor er eksemplet blevet bragt trods alt.

Eksempler som indeholder betegnelser på tekniske fornyelser eller omtale af nye traditioner, er normalt lette at få øje på og undgå. Vanskeligere bliver det når der er tale om kollokationer, idiomatiske udtryk og overførte betydninger af etablerede ord, udtryk der kan virke helt upåfaldende for redaktøren.

Betydning 23 i artiklen *ind* udgøres af kollokationen *gå ind for* i betydningen ‘være gunstig stemt for noget, være tilbøjelig til noget’, der er belagt med et citat fra en båndoptagelse fra 1981:

(15) (læreren) *han gik meget ind for skrivning* (i skolen)

Tilsyneladende et uangribeligt eksempel, men i Ordbog over det danske Sprog (*gaa* 26.4, i bind 7 fra 1924) karakteriseres *gå ind for* som “l.br.” (lidet brugeligt) og som lån fra engelsk. Desuden tilføjer supplementet til Ordbog over det danske Sprog: “vist nyt o. 1920”. Det er derfor tvivlsomt om udtrykket har været etableret i dialekterne inden for ordbogens periode, og det burde ikke have været med.

Et andet eksempel der kan illustrere denne problemstilling, er *knald* i betydningen ‘fart, intensitet’ o.l. Det forekommer i nogle enkelte belæg fra båndoptagelser efter 1975 og fra en spørgelistebesvarelse fra samme tid, fx:

(16) *der var knald på om morgenen med at se og få dem i kurven* (Lolland, Vindeby)

Betydningen findes ikke i Ordbog over det danske Sprog, men er medtaget i supplementet. Det ældste belæg i supplementet er fra 1942, men eftersom forbindelser som *knald på* tilhører uformelt sprog og formentlig er sjældent i skrift, kan det sagtens være en del ældre. Og det kan heller ikke udelukkes at det er ældre i dialekterne end i rigsmålet. Den normale lånevej er jo fra standard til dialekt, men der forekommer også eksempler på det modsatte, fx *snalret* ‘fuld, beruset’ der oprindeligt stammer fra jyske dialekter. Endelig støttes det yngre

materiale til en vis grad af et belæg optegnet før krigen: *skafferen laste det samme som føringsaftenen med rigtig knald i*. På denne baggrund har vi med nogen tøven taget betydningen med.

Endelig udgør fremmedord også et problem. Det har altid været tanken at medtage de fremmedord der er indlånt og optaget i dialekterne, men fremmedord kan jo være mere eller mindre integrerede i dagligsproget. Oprindeligt skulle et fremmedord helst enten have udviklet en dialektal særbetydning eller gennemgået en lydlig særudvikling for at være interessant, og lån der ikke opfylder disse kriterier, især nyere lån, er normalt meget tilfældigt optegnet (Gudiksen 2006:158). Når et udtryk som *være i kontakt med nogen* kun er afhjemlet i båndoptagelser fra 1970'erne og 1980'erne, kan det derfor ikke udelukkes at det har været brugt også dialektalt inden for ordbogsperioden, og hvad angår et ord som *konkurrence*, kan man næsten ikke tænke sig andet end at det har været brugt i hvert fald i begyndelsen af 1900-tallet. Men det føles utrygt at tage *konkurrence* med, når materialet udelukkende består af et par belæg fra båndoptagelser optaget i 1980'erne.

#### 4. Afslutning

*Var det hårdt arbejdet i gamle dage*, spørger dialektologen på en båndoptagelse, og svaret lyder: *ja, det var sgu hårdt, det var hårdt alt hvad vi lavede, det var hårdt*. Ligesom alle andre kildegrupper giver båndoptagelserne problemer og spekulationer, men fordelene er klart større end ulemperne. Og eksempler fra båndoptagelsen giver ikke kun en bedre beskrivelse. Den gør også ordbogen mere interessant og sjovere at læse i.

#### Ordbøger:

*Ordbog over det danske Sprog*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal 1919–1956.

*Ordbog over det danske Sprog. Supplement*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal 1992–2005.

*Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer*. 1–. København: Institut for Dansk Dialektforskning og C.A. Reitzels Forlag 1992–.

*Ømålsordbogen. Tillægsbind*. København: Institut for Dansk Dialektforskning og C.A. Reitzels Forlag 1992.

#### Anden litteratur:

Andersen, Poul 1932: Beretning om Udvalg for Folkemaals Virksomhed 1909-32. I: *Fortid og nutid* 9, 161–210.

Andersen, Poul 1939: Udvalg for Folkemaals Virksomhed 1933-38 (med Udkast til Plan for den kommende Ordbog over de danske Ømaal). I: *Danske Folkemaal* 12, 17–64.

- Danske Dialektttekster* 1-5. Udgivet af Institut for Dansk Dialektforskning. København: Akademisk Forlag 1967–1977.
- Gudiksen, Asgerd 2006: Ømålsordbogen og mejetærskergenerationen. I: Henrik Lorentzen og Lars Trap-Jensen (red.): *Nordiske Studier i Leksikografi. Rapport fra konference om leksikografi i Norden, Sønderborg 24.–28. maj 2005*. København: Nordisk Forening for Leksikografi, 151–161.
- Hovmark, Henrik 2004: Jagten på det autentiske citat. I: *Danske Talesprog* 5, 3–22.
- Pedersen, Karen Margrethe 1974: Dialektttekster i rigsmålsnotation med becifring. Et EDB-projekt. I: *Danske Folkemaal* 20, 29–46.
- Pedersen, Karen Margrethe 1983: Om transskription af båndoptagelser og anvendelse af teksterne til leksikografiske og syntaktiske undersøgelser. Et EDB-projekt. *Folkmålsstudier* 28, 131–139.

Asgerd Gudiksen  
lektor, f. 1955  
Ømålsordbogen  
Nordisk Forskningsinstitut  
Afdeling for Dialektforskning  
Københavns Universitet  
Njalsgade 136, 5  
DK-2300 København S  
asgerd.gudiksen@hum.ku.dk

Henrik Hovmark  
adjunkt, f. 1966  
Ømålsordbogen  
Nordisk Forskningsinstitut  
Afdeling for Dialektforskning  
Københavns Universitet  
Njalsgade 136, 5  
DK-2300 København S  
hovmark@hum.ku.dk



## En ny tysk-islandsk ordbog

### Metode og problemer

**A new German-Icelandic dictionary. Methods and problems.** The paper describes a dictionary work in progress. PONS Dictionary Edition in Stuttgart provided the dictionary's source, besides the program and the technological knowledge. The source comprises approximately 40,000 headwords (main and sub-headwords including compounds) and 25,000 phrases (grammar constructions, collocations, idioms, proverbs and example sentences), covering the central lexicon of German. Acronyms, abbreviations, and short forms are listed. Specialized vocabulary from the fields of business, law and computing has been incorporated – and as a reflection of increasing specialization in everyday life, the macrostructure also includes specialized vocabulary that is vital in contemporary day-to-day communication at home, school, and in the business world. The emphasis is on modern, up-to-date language, and special attention has been paid to culture-specific vocabulary (foods, institutions, customs, etc.), as well as neologisms.

### 1. Indledning

Gode bilingvale ordbøger er mangelvare i Island, både med islandsk som målsprog og som kildesprog. Situationen er lidt forskellig med hensyn til enkelte sprog. Den største engelsk-islandske ordbog er f.eks. fra 1984 (Sörensson) og den fransk-islandske (Stefánsson og Hafsteinsdóttir) fra 1995, mens over 70 år er gået siden en tysk-islandsk ordbog (Ófeigsson 1935) udkom i Reykjavík, den første tysk-islandske ordbog overhovedet. Ordbogsredaktøren var en islandsk lærer i tysk ved gymnasiet. Bogen var omfattende, tog tillid til ordforrådet i forskellige discipliner og også til en god del af fremmedord i tysk som i sin tid ikke var en selvfølge. Den almene opfattelse var at fremmedord ikke skulle med i islandske ordbøger. Denne ordbog er blevet genudgivet gang på gang uden ændringer og ordforrådet er længst forældet selvom bogen endnu bruges i skolerne.

På markedet findes der også en islandsk-tysk, tysk-islandsk ordbog i serien Langenscheidts Universal-Wörterbücher (Brynjólfsson 1964) som er til hjælp hvad det nyere tyske ordforråd gælder, mens den ellers er mere tænkt til hjælp for turister som enten vil besøge Tyskland eller Island. Den er blevet genudgivet gang på gang. Der kan også nævnes en tysk-islandsk, islands-tysk lommeordbog (Eiðsdóttir og Böðvarsson 1991) og en tysk-islandsk ordbog til skolebrug (Mattíasson) som udkom 2004. Der arbejdes også i Greifswald in Tyskland på en

tysk-islandsk ordbog. Dette arbejde startede for omkring fem og tyve år siden og ordbogen er endnu ikke færdig selvom den hele tiden har været under bearbejdelse.

Nu arbejdes der på en ny ordbog hvor kildesproget er tysk og målsproget islandsk. Det drejer sig om et islansk-tyskt samarbejde mellem Vigdís Finnbogadóttir Institutet for Fremmedsprog, som er et forskningscenter ved Islands Universitet, og den leksikografiske afdeling indenfor Arni Magnussons institut for islandske studier på den ene side og Ernst Klett Spracen i Stuttgart, Bosch-Stiftung i Stuttgart og Würth-Stiftung i Künzelsau på den anden.

Samarbejdet foregår på den måde at Bosch og Würth-stiftelserne har skaffet penge som skal dække størstedelen af udgifterne. Resten kommer fra Island. Islands universitet har skaffet lokale og arbejdsverktøj og Guðrún Kvarans deltagelse i arbejdet bliver Arne Magnusson instituttets bidrag. Hun skal have tilsyn med værket både som rådgiver og kontrollør. Heimir Steinarsson blev ansat som hovedredaktør og han har allerede valgt sine medarbejdere. Redaktionsarbejdet skal være færdigt i 1. november 2008 og bogen skal være på markedet før jul samme år.

Der bliver lagt stor vægt på at bogen er til nytte såvel for almenheden og skolerne i Island som gerne vil slå op betydningen af tyske ord og ordforbindelser og for tyskere som vil prøve at finde frem til islandske ord og ordforbindelser. Den skal også være til nytte for oversættere som længe har klaget over mangel af dette vigtige hjælpemiddel. Indtil i dag er de fleste ordbøger i Island blevet lavet fra bunden, dvs opslagsordene er blevet samlet, nedskrevet, oversat og redigeret af en eller flere personer og siden sendt til et forlag. I den senere tid samarbejdes der oftest med et forlag lige fra begyndelsen.

Redaktøren af den store tysk-islandske ordbog, Jón Ófeigsson, brugte næsten halvdelen af sit liv på udarbejdelsen af sin bog. Vores opgave er imidlertid at skabe et tilsvarende opslagsværk af samme omfang og kvalitet på mindre end to år med et skrabet budget som ikke giver mulighed for store udsving. Så selvfølgelig havde vi i begyndelsen vore tvivl med hensyn til kvaliteten af sådan et værk men efterhånden er vi overbevist om at vi styrer imod en god tysk-islandsk ordbog.

## 2. Bogen som den tysk-islandske ordbog er baseret på: PONS Lommeordbog Engelsk-Tysk

### 2.1 Anvendt software

Det tyske forlag PONS har valgt at anvende XMLSpy programmet fra Altova til sine ordbøger. Programmet blev ikke udelukkende udviklet til leksikografisk anvendelse men til alle former for tekstbehandling baseret på XML standarden. Det har den fordel at det er meget fleksibelt og sætter brugeren istand til at arbejde med teksten på mange forskellige måder, som for eksempel at adskille ori-

ginal og måltekst, sammenligne nye opslag med den originale tekst osv. Det har imidlertid også den ulempe at det er temmelig omfattende, hvilket bevirker at en grundig kendskab til strukturen er en forudsætning og endvidere kendskab til programmering. Heldigvis ydede personalet fra PONS al den tekniske service vi havde brug for.

### *Arbejdsstruktur*

Den tysk-engelske udgave af *Kompaktwörterbuch* indeholder ca. 40.000 hovedord (lemmaer og sublemmaer, heriblandt sammensatte ord) og 25.000 fraser (grammatiske konstruktioner, kollokationer, idiomer, ordsprog og eksempler) som omfatter det centrale tyske leksikon. Akronymmer, forkortelser og korte former forefindes. Specielle ord og udtryk fra forretningsverdenen, jura og IT områderne er blevet inkorporeret og som et udtryk for den øgede specialisering i hverdagen omfatter makrostrukturen ligeledes specielle ord og udtryk som er af afgørende betydning i moderne daglig kommunikation; i hjemmet, i skolen og i forretningsverdenen. Vægten er lagt på moderne, nutidigt sprog og der er blevet lagt særlig vægt på kulturspecifikke gloser (mad, institutioner, traditioner m.m.) så vel som neologismer.

Opslagsordene reflekterer det sprog som der bruges idag; skriftligt, mundtligt, formel og uformel brug (herunder slang) og er opført med angivelse af stil og/eller anvendelighed i bestemte situationer. Følgende regionale varianter af tysk er ligeledes blevet inkluderet: nord, syd, øst, svejtsisk og østrigsk tysk.

Den tyske del af ordbogen blev til ved hjælp af en ensproget tysk sprogstruktur eller kildetekst. Den blev udarbejdet af indfødte tyskere med den hensigt at give maksimal grammatisk, semantisk og syntaktisk information for hvert opslag i en forudbestemt tydeligt mærkeret struktur og blev udarbejdet uden hensyntagen til målsproget.

Den samme tyske samling som vi arbejder med i øjeblikket og eksemplerne tages fra er ikke alene blevet oversat til engelsk men også til spansk, fransk og for nyligt til et mindre udbredt sprog som slovensk. Derfor var vores bekymringer over at gloserne blev udarbejdet udelukkende for en engelsksproget målgruppe tydeligvis ubegrundede. Alligevel besluttede vi at forbedre bogen med ca. 3.000 ord fra områder der er af stor betydning for Island såsom fiskeri, turisme osv.

### *2.3 Opslagsstruktur*

Information vedrørende grammatiske former forefindes, kollokationer er oversat eftersom disse udgør sproghederne der karakteriserer typisk idiomatisk brug og sammensatte ord er tilføjet på det rigtige sted på den alfabetiske liste over hovedord. Eksempler indeholder information vedrørende syntaktiske vanskeligheder, talesprog og kontekst som skaber specifikke oversættelsesproblemer.

Hver ordklasse (herunder brugen af transitive, intransitive og reflexive ud-sagnsord) bliver omtalt i en separat ROM sektion (figur 1–2):

**wäh-rend** ['vɛ:rənt] **I.** *präp +gen* during  
**II.** *konj* ① (*zur selben Zeit*) while ② (*wohin-gegen*) whereas

Figur 1: Grammatiske oplysninger

stich	während
STICH	wäh(%)rend
PHON	{'}w(E):r(e)nt
ROM	I
WORTK-PREP	PREP
OBJKAS	gen
ZIEL	during
ROM	II
WORTK	CONJ
ARAB	1
BED	zur selben Zeit
ZIEL	while
ARAB	2
BED	wohingegen
ZIEL	whereas

Figur 2: Grammatiske oplysninger i XMLspy

Forskellige betydninger af et ord bliver omtalt i separate arabiske kategorier (figur 3). Nogle former for skelnen af betydning skal følge ARAB (figur 4) og disse sondringer efterfølges altid af kendetegnende elementer som for eksempel en forklaring, en kollokation eller områdemærkater der viser området indenfor hvilken en bestemt brug er almindelig.

**Wer-bung** <-> *f kein pl* ÖKON ① (*Branche*) advertising ② (*Reklame*) advertisement;  
 ~ **für etw akk machen** to advertise sth  
 ③ (*Werbepot*) commercial; (*Werbeprospek-  
 te*) advertising literature ④ (*das Werben*) re-  
 cruitment; *von Kunden* attracting

Figur 3: Betydninger

stich	vWerbung
STICH	vWe{.}r{%}bung
infl-glo	
FLEX	-
WORTK-II	SUBST
GENIUS	f
ISYII-II	kein pl
prag-pos	
SACH	ökon
ARAB	1
BED	Branche
ZIEL	advertising
ARAB	2
BED	Reklame
ZIEL	advertisement
KOLLOK	
	Abc Text {TILDE} für etw
KASUS	akk
	Abc Text machen
Z-MWE	to advertise sth
ARAB	3
BED	vWerbespot
ZIEL	commercial
BED	vWerbeprospekte
ZIEL	advertising literature
ARAB	4
BED	das Werben
ZIEL	recruitment
KOLL	von Kunden
ZIEL	attracting

Figur 4: Betydning i XMLspy

Faste udtryk (idiomer) og ordsprog kan ikke sammenkædes med en bestemt mening af et hovedord. Derfor er de ikke opført som fraser i de enkelte ARAB-iske kategorier men i en separat blok i slutningen af opslaget eller i slutningen af hver ordklasse (figur 5–6).

Hele denne opslagsstruktur er ganske enkel at forstå og giver oversætterten af vores nye islandske projekt mulighed for at oversætte bogen med den hastighed der er nødvendig, såfremt vi ønsker at blive færdige til tiden.

Det har den fordel, at vi har mulighed for at opdatere samlingen (og rette evt. fejl om nødvendigt) næste gang vi beslutter at genoptrykke bogen, uden større besvær og uden at skulle udføre det elementære arbejde igen. Det betyder at det kun kræver én person til at tilføje rettelser og nye ord til den eksisterende database.

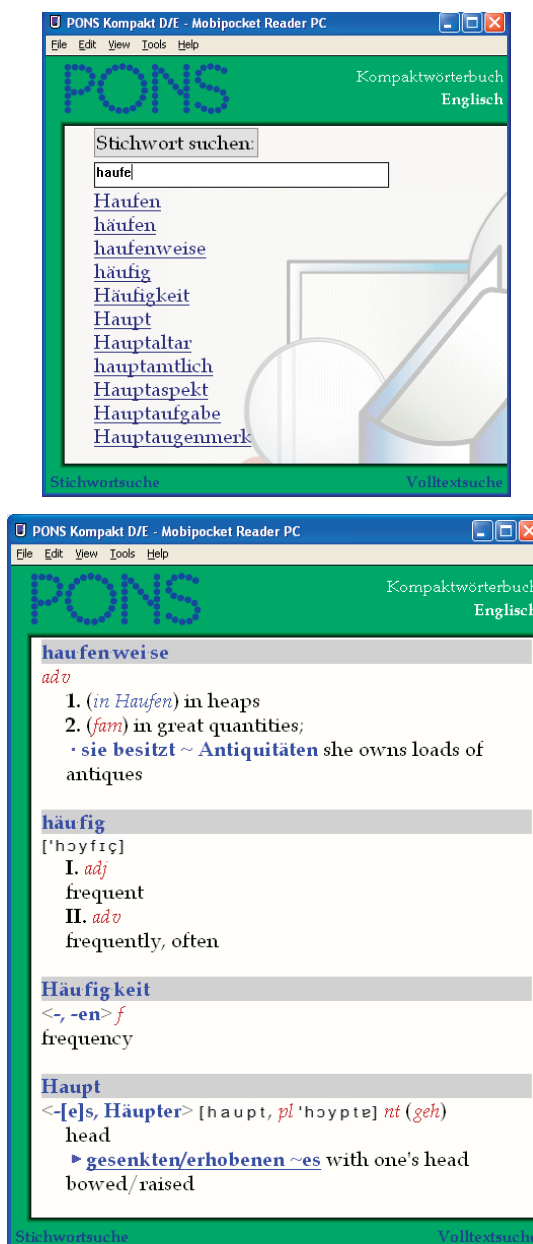
**Mund** <-[e]s, Múnder> ['munt, *pl* 'mʏndə]  
*m* mouth; **etw in den ~ nehmen** to put sth  
 in one's mouth; **mit vollem ~e** with one's  
 mouth full ▶ **den ~ [zu] voll nehmen** to talk  
 [too] big; **jdm über den ~ fahren** to cut sb  
 short; **[jd ist] nicht auf den ~ gefallen**  
 (*fam*) [sb is] never at a loss for words; **halt**  
**den ~!** shut up!; **jdm etw in den ~ legen** to  
 put [the] words into sb's mouth; **jdm nach**  
**dem ~[e] reden** to say what sb wants to he-  
 ar; **etw ist in aller ~e** sth is the talk of the  
 town; **wie aus einem ~e** with one voice

Figur 5: Faste udtryk

STICH	Mund
infl-glo	
FLEX	-[e]s, Múnder
PHON	
	Rbc Text (')m(U)nt,
ISYII	pl
	Rbc Text (')m(Y)nd(r)
WORTK-II	SUBST
GENUS	m
ARAB	VAKAT
ZIEL	mouth
KOLLOK	etw in den {TILDE} nehmen
Z-MWE	to put sth in one's mouth
KOLLOK	mit vollem {TILDE}e
Z-MWE	with one's mouth full
ARAB	2
WEND-ONLY	den {TILDE} [zu] *voll nehmen
Z-MWE	to talk [too] big
WEND	jdm über den {TILDE} *fahren
Z-MWE	to cut sb short
WEND	[jd ist] nicht auf den {TILDE} *gefallen
prag-mwe	
STIL	fam
Z-MWE	[sb is] never at a loss for words
WEND	*halt den {TILDE}!
Z-MWE	shut up!
WEND	jdm etw in den {TILDE} *legen
Z-MWE	to put [the] words into sb's mouth
WEND	jdm nach dem {TILDE}[e] *reden
Z-MWE	to say what sb wants to hear
WEND	etw ist in *aller {TILDE}e
Z-MWE	sth is the talk of the town
WEND	wie *aus einem {TILDE}e

Figur 6: Faste udtryk i XMLspy

Takket være XMLspy softwaren er det muligt at konvertere data til andre formater, såsom den digitale udgave som kan benyttes ved hjælp af et gratis program der hedder Mobipocket Reader (figur 7) og som er kompatibelt til almindelige computere, Blackberries, Pocket PC'er og mange andre tilsvarende enheder. Denne konvertering af data kan udføres uden store problemer eller udgifter.



Figur 7: Fra Mobipocket Reader

### 3. Fra tysk-engelsk til tysk-islandsk

Skønt engelsk og islandsk er meget forskellige sprog udføres en stor del af arbejdet med at erstatte det engelske målord. Enkle navneord, udsagnsord og tillægsord giver ikke anledning til problemer. Vi erstatter blot målordet og hvis dette er et navneord tilføjer vi det grammatiske køn (figur 8–9).

year out  
**Jahr·buch** *nt* yearbook  
**jah·re·lang** [*ˈja:rələŋ*] **I.** *adj attr* lasting for ye-

Figur 8: *Jahrbuch* i Kompaktwörterbuch

stich	Jahrbuch
STICH	Ja_hr{%}buch
WORTK-II	SUBST
GENIUS	nt
ZIEL	yearbook

stich	Jahrbuch
STICH	Ja_hr{%}buch
WORTK-II	SUBST
GENIUS	nt
ZIEL	árbók
Z-GENUS	f

Figur 9: *Jahrbuch* i XMLspy

I andre tilfælde, specielt når der ikke findes et tilsvarende ord på islandsk, har vi flere muligheder. Vi kan tilføje en kollokation, konstruktion, idiom eller et eksempel eller vi kan ændre en konstruktion til et eksempel:

**jäh·ren** [*ˈjɛ:rən*] *vr (geh)* ■ **sich** ~ to be the anniversary of

Figur 10: *jähren* i Kompaktwörterbuch

I tilfældet med ordet *jähren* (figur 10) findes der ikke noget tilsvarende ord på islandsk og derfor ændrer vi konstruktionen KONSTR i kildeteksten til et eksempel BSP hvor anvendelsen af ordet er vist i sammenhæng (figur 11).



stich	jáhren
STICH	jäh{(%)}ren
PHOH	{'}(E);r(e)n
WORTK-V	VERB
VERBK	vr
STIL	geh
KOHSTR-OHLY	sich {TILDE}
Z-MWE	to be the anniversary of

stich	jáhren
STICH	jäh{(%)}ren
PHOH	{'}(E);r(e)n
WORTK-V	VERB
VERBK	vr
STIL	geh
BSP-OHLY	das Ereignis jährt sich nun zum zehnten Mal
Z-MWE	nú eru tíu ár liðin frá því að atburðurðurinn gerðist

Figur 11: *jähren* i XMLspy

Et udbredt problem var i de tilfælde hvor engelsk har adskillige oversættelser for et ord på kildesproget.

**Jahr-gang** <-gänge> m ① (Personen eines Geburtsjahrs) age-group; (Gesamtheit der Schüler eines Schuljahres) [school] year  
 ② (Erntejahr) vintage; (Herstellungsjahr) year

Figur 12: *Jahrgang* i Kompaktwörterbuch

I eksemplet ovenover (figur 12) er der mange mulige engelske oversættelser af det tyske kildeord *Jahrgang*, afhængigt af sammenhængen. På islandsk oversættes ordet *Jahrgang* altid med det samme ord *árgangur* uanset sammenhængen. Det betød at vi måtte slette numrene og de forskellige meninger (figur 13).

stich	Jahrgang
STICH	Ja_hr{%}gang
infl-glo	
FLEX	-gänge
WORTK-II	SUBST
GENUS	m
ARAB	1
BED	Personen eines Geburtsjahrs
ZIEL	age-group
DIF	Gesamtheit der Schüler eines Schuljahres
ZIEL	[school] year
ARAB	2
BED	Erntejahr
ZIEL	vintage
BED	Herstellungsjahr
ZIEL	year

stich	Jahrgang
STICH	Ja_hr{%}gang
infl-glo	
FLEX	-gänge
WORTK-II	SUBST
GENUS	m
ZIEL	árgangur
Z-GENUS	m

Figur 13: Jahrgang i XMLspy

Det omvendte var tilfældet de steder hvor islandsk havde flere forskellige oversættelser af det samme kildeord.

**in-i-ti-ie-ren**\* [initsi'i:rən] vt (*geh*) to initiate  
**In-jek-ti-on** <-, -en> [Injek'tsjo:n] f injection  
**in-ji-zie-ren**\* [inji'tsi:rən] vt (*geh*) ■ [jdm]

Figur 14: Injektion i Kompaktwörterbuch

stich	Injektion
STICH	In{ % }jek{ % }ti{ % }on
infl-glo	
	FLEX - , -en
PHON	{ l }nj{ E }k{ ' }ts{ i _ }o : n
WORTK-II	SUBST
GENUS	f
ZIEL	injection
infl-glo	
	FLEX - , -en
PHON	{ l }nj{ E }k{ ' }ts{ i _ }o : n
WORTK-II	SUBST
GENUS	f
ARAB	1
SACH	tech
ZIEL	innspýting
Z-GEN	f
ARAB	2
SACH	med
ZIEL	stunga
Z-GEN	f

Figur 15: Injektion i XMLspy

Men her skal man være forsigtig: Det er let nok at undlade meninger i den engelske målttekst men det er straks vanskeligere at sikre at alle mulige betydninger er blevet behandlet i den nye islandske oversættelse.

For at sikre at enhver betydning er blevet omtalt på en korrekt måde må oversætteren lægge den bilingvale til side et øjeblik og istedet slå opslagsordene op i en monolingval ordbog som for eksempel *Duden Universal Wörterbuch* (figur 16) for at se om der findes mere end en betydning af opslagsordet.

**Injektion**, die; -, -en [lat. *injection*, eigtl. = das Hineinwerfen, zu: *inicare*, † *injizieren*]: **1.** (Med.) *das Injizieren; Spritze*: eine intravenöse, intramuskuläre I.; jmdm. eine I. geben, verabreichen; eine I. machen, vornehmen. **2.** (Med.) *starke Füllung u. damit Sichtbarwerden kleinster Blutgefäße im Auge bei Entzündungen*. **3.** (Bauw.) *Einspritzung von Verfestigungsmitteln, z. B. Zement, in unfesten Bauuntergrund*. **4.** (Geol.) *das Eindringen magmatischer Schmelze in Fugen u. Spalten eines Gesteins*. **5.** (Physik) *das Einbringen von [Elementar]teilchen in einen Halbleiterbereich von bestimmter elektrischer Leitfähigkeit bzw. in der Hochenergie- u. Kernphysik in einen Teilchenbeschleuniger*.

Figur 16: Eksempel fra *Duden Universal Wörterbuch*

#### 4. Konklusion – fordele og ulemper

Vi begyndte med at spørge os hvilke fordele og hvilke ulemper der var ved at samarbejde med et forlag og bruge en allerede færdig tysk-engelsk ordbog som værktøj. Som indledningsvis blev nævnt gjorde vi os skrupler over hvorvidt en ordbog som denne kunne blive lige så god og nyttig for islandske læsere som en ordbog der var blevet udarbejdet i Island helt fra begyndelsen. På den positive side bør nævnes at:

- kildeordforrådet blev leveret klar til brug uden at vi skulle udarbejde det og rette det, hvilket sparede os både tid og penge
- klar struktur sikrer en hurtig og præcis arbejdsgang
- god teknisk service kommer fra forlaget

På den negative side bør især nævnes det stramme budget vi arbejder med og tidspresset som naturligvis er højt såfremt vi ønsker at udgive bogen før slutningen af 2008.

Men set i lyset af de ældre forsøg så må man sige at effektiviteten af sådan et projekt som dette berettiger dets eksistens. Hvis det lykkes os at blive færdige med en god ordbog inden for tids- og budgetrammen ville det være et stort skridt i den rigtige retning og forhåbentlig en inspiration til fremtidige ordbogsprojekter.

#### 5. Fremtidige projekter – muligheder for yderligere samarbejde med et ordbogsforlag.

Udgiverne hos PONS har udarbejdet store samlinger ikke alene på tysk men også på engelsk, fransk og spansk. Alle disse sprog er udarbejdet af forfattere og

udgivere som taler disse sprog som modersmål. Dette ordforråd er tilgængeligt på den nødvendige software med fonetisk skrift og eksempler, kollokationer og grammatiske konstruktioner. Det bliver løbende opdateret med neologismer og foraldede ord bliver slettet.

Eftersom forlaget udgiver mange ordbøger hvert år har det den tekniske knowhow, forbindelserne til erfarne og økonomiske trykkere og det har en stor erfaring i udgivelse af ordbøger. Vi er overbevist om at igennem samarbejde med et forlag kan ordbøger som behandler små sprog som f.eks. det islandske blive produceret og opdateret jævnlige. Det er i det mindste vores indtryk at vi ikke behøver at vente i yderligere 70 år på den næste tysk-islandske ordbog.

## Litteratur

- Brynjólfsson, Ingvar 1964: *Íslenzk-þýzk [orðabók]*. I. hluti: Íslenzk-þýzk. II. hluti: Þýzk-íslenzk. Langenscheidts Universal-orðabók. Isländisch. Teil I: Isländisch-Deutsch. Teil II: Deutsch-Isländisch. Berlin-München: Langenscheidts Universal-Wörterbuch.
- Duden Universal Wörterbuch*. 2006. 6. überarb. u. erw. Auflage. Mannheim: Bibliographisches Institut.
- Eiðsdóttir, Eygló, og Árni Böðvarsson 1991: *Þýsk-íslensk, íslensk-þýsk vasaorðabók. Deutsch-isländisches, isländisch-deutsches Wörterbuch*. (2. udg. forøget 1997.) Reykjavík: Orðabókaútgáfan.
- Matthíasson, Steinar 2004: *Þýsk-íslensk íslensk-þýsk orðabók*. Reykjavík: Iðnú.
- Ófeiggsson, Jón 1935: *Þýzk-íslenzk orðabók. Deutsch-isländisches Wörterbuch*. Reykjavík: Ísafoldarprentsmiðja.
- Pons Kompaktwörterbuch Englisch*. 2006. Stuttgart: Klett; Auflage: Neubearb. (September 2006).
- Stefánsson, Þór (red.), og Dóra Hafsteinsdóttir 1995: *Frönsk-íslensk orðabók*. Reykjavík: Örn og Örlygur.
- Sörensson, Sören 1984: *Ensk-íslensk orðabók með alfræðilegu ívafi*. Jóhann S. Hannesson bjó til prentunar ásamt fleirum. Reykjavík: Örn og Örlygur.

Guðrún Kvaran  
 prófessor, f. 1943  
 Stofnun Árna Magnússonar í íslenskum fræðum – Háskóli Íslands  
 Neshagi 16  
 IS-107 Reykjavík  
 gkvaran@hi.is

Heimir Steinarsson  
 B.A. í þýsku, f. 1967  
 Verkefnastjóri þýsk-íslensku orðabókarinnar  
 Stofnun Vigdísar Finnbogadóttur í erlendum tungumálum – Háskóli Íslands  
 heimirs@hi.is



HALLDÓRA JÓNSDÓTTIR OG ÞÓRDÍS ÚLFARSDÓTTIR

## ISLEX: Islandsk-skandinavisk webordbog

ISLEX is a multilingual dictionary with modern Icelandic as source language and Danish, Norwegian and Swedish as target languages. The project is a collaboration of four institutes in the Nordic countries with a central base in Reykjavík, Iceland.

With 50.000 entry words, ISLEX is a medium sized dictionary. Set in a custom-made relational database, it is designed as an on-line dictionary where the various possibilities of the web are utilised. So ISLEX has pictures and animated images, and pronunciation given as sounds. There is a special emphasis on Icelandic phrases and fixed expressions, all of which are translated into the three target languages.

The work on the dictionary takes place simultaneously in four countries. The processing of the base language and coordination is based in Iceland, the Danish translation team is located in Denmark, the Norwegian team in Norway, and the Swedish team in Sweden. ISLEX is a six year project, begun in 2005. The aim is to complete the dictionary in mid-year 2011.

### 1. Indledning

ISLEX er et elektronisk ordbogsværk hvor kildesproget er islandsk og målsprogene er dansk, svensk og norsk (både bokmål og nynorsk). Ordbogen genspejler islandsk sprogbrug i samtiden og målet er at den bliver gratis tilgængelig for alle på webben. Ordbogsprojektet er udarbejdet af Árni Magnússon instituttet for islandske studier i samarbejde med universitetsinstitutioner i tre nordiske lande, Danmark, Norge og Sverige. Projektet er omfattende og ressourcerkrævende og man regner med at det vil strække sig over seks år. ISLEX-ordbogen skildrer islandsk nutidssprog, tæller omkring 50.000 opslagsord og svarer således til en mellemstor ordbog, foruden at der lægges stor vægt på fraser og sprogbrugseksempler. Ved redaktionen anvendes der en web-baseret database, der er specielt udformet til dette projekt og gør det muligt at redaktionsarbejdet kan foregå samtidigt i mange lande.

Efter en forberedende periode blev det redaktionelle arbejde påbegyndt i året 2005, først i Island og siden efterhånden i de øvrige samarbejdslande. Efter planen bliver ISLEX-ordbogen færdig i 2011, men der er stor interesse og vilje for at en del af materialet bliver tilgængeligt noget før. Ordbogens målgruppe er defineret meget bredt og den henvender sig således til alle der har brug for islandsk-nordiske ordbøger. Der lægges dog en særlig vægt på at imødekomme danske, norske og svenske brugeres behov, ikke mindst ved oversættelser fra islandsk.

I Island har man længe manglet tidssvarende ordbøger mellem islandsk og de

skandinaviske sprog, ordbøger der giver et godt overblik over islandsk nutidssprog og indeholder en klar beskrivelse af islandsk ordforråd. De eksisterende ordbøger mellem islandsk og de skandinaviske sprog er enten gamle eller utidssvarende, de fleste er små og indeholder utilstrækkelige oplysninger om islandsk sprogbrug. Man har endvidere peget på at de ældre ordbøger er en begrænset hjælp for udenlandske brugere, f.eks. oversættere og studenter i islandsk som fremmedsprog. Den store interesse i Island for at bevare den sproglige og kulturelle kontakt med de øvrige nordiske lande er endvidere afgørende i denne sammenhæng og man har anset et moderne ordbogsprojekt som et vigtigt led i at styrke forbindelsen mellem Island og de skandinaviske lande inden for kultur, undervisning, uddannelse, administration, økonomi, handel og erhvervsliv. Projektet bygger således på den fællesnordiske idé og vil være til fællesnordisk nytte.

Af førnævnte grunde indledtes der i året 2004 drøftelser mellem Islands Universitets leksikografiske institut, *Orðabók Háskólans* (nu *Árni Magnússon instituttet for islandske studier*), med repræsentanter fra *Instituttet for svenska språket* ved Göteborgs Universitet og *Nordisk Institut* (nu *Institut for lingvistiske, litterære og estetiske studier*) ved Universitetet i Bergen om et fælles ordbogsprojekt, der kunne råde bod på denne mangel. I 2005 sluttede *Det Danske Sprog- og Litteraturselskab* sig projektet, i forståelse med *Vigdís Finnbogadóttir Instituttet for fremmedsprog* ved Islands Universitet.

## 2. Samarbejdspartnere og finansiering

ISLEX-ordbogen er således et samarbejdsprojekt mellem fire institutioner i Norden. Den største redaktion sidder i Reykjavík og dens rolle er at beskrive kilde sproget og koordinere arbejdet mellem de fire lande, foruden at være ansvarlig for udformningen og udviklingen af databasen. De øvrige institutioner er *Det Danske Sprog- og litteraturselskab* i København, *Institutionen for svenska språket* ved Göteborgs Universitet og *Institut for lingvistiske, litterære og estetiske studier* ved Universitetet i Bergen. Hvert land er en selvstændig arbejdsenhed der ansætter medarbejdere til oversættelse og korrekturlæsning og står endvidere for finansiering af sin del af projektet. Regeringerne i alle de fire lande har bevilliget penge til ISLEX-projektet, som hovedsagelig er finansieret af offentlige midler. Projektet har endvidere modtaget støtte fra *Nordplus Sprog*, som især er gået til det forberedende arbejde i 2004, såvel som til samarbejds møder der afholdes regelmæssigt.

## 3. ISLEX-ordbogen

Den islandske del af ISLEX bygger hovedsagelig på en eksisterende ordbogsbase samt Árni Magnússon instituttets tekstkorpus. Arbejdet med ISLEX omfatter en leksikalsk analyse af dette materiale, bl.a. en semantisk opdeling af ordforrådet. Foruden at være en multisproglig ordbog tjener ISLEX-projektet også som et



vigtigt led i en systematisk web-baseret formidling af oplysninger om islandsk ordforråd for islandske brugere.

Instituttets tekstkorpus tæller allerede 60 millioner ord fra elektroniske tekster, hvor hovedparten er yngre end fra 1980. Foruden at være et uundværligt redskab ved ordbogsbeskrivelsen har tekstkorpuset desuden en selvstændig gyldighed som en informationsbase om islandsk nutidssprog og anvendes bl.a. i andre ordbogsprojekter, i sprogteknologiske opgaver og anden forskning i moderne islandsk. Forskningen inden for det islandske ordforråd kan nu formidles på en mere varieret og målrettet måde end før, og den leksikalske og leksikografiske analyse og opdeling tager nu hensyn til forskellige behov inden for forskningen, hvad enten det drejer sig om videnskabelig eller praktisk anvendelse.

ISLEX-ordbogen indeholder som før sagt 50.000 opslagsord, foruden et stort antal fraser, både ordforbindelser og faste udtryk. Der lægges en særlig vægt på at udnytte web-mediets fordele, bl.a. ved at tilføje de leksikografiske oplysninger billed- og lydmateriale. Databasen byder på mange slags søgefunktioner som redaktørerne gør nytte af i deres arbejde.

### 3.1 Lemmaselektionen

ISLEX-ordbogen omfatter det islandske nutidssprog og de udvalgte opslagsord genspejler dette. Ældre ordforråd medtages ikke medmindre det tilhører det sprog som studerende og oversættere forventes at støde på i klassiske litterære tekster.

Lemmaselektionen i ISLEX hviler iøvrigt på tre grundpiller; 1. Frekvens i det moderne skriftsprog. 2. Frekvens i talesproget, hvad enten ordene findes på tryk eller ej (tekster som stammer fra web-kommunikation regnes for talesprog). 3. Fagord, d.v.s. ordforråd der stammer fra bestemte fagområder, f.eks. geologi, jura og medicin. Det er vigtigt at medtage et bredt udvalg af fagord, bl.a. af den grund at der i Island er en lang tradition for at danne nye fagord i stedet for at bruge de internationale termer.

### 3.2 Ordbogsstrukturen

ISLEX er en meget struktureret ordbog hvor hvert element har sin faste plads og navn i databasen. I hver ordbogsartikel forekommer der bestemte obligatoriske elementer. Når lemmaet er et sammensat ord så er der et felt der viser orddelene. Derefter følger en forklaring på islandsk. Formålet med den er todelt. For det første hjælper den oversætterne i Skandinavien med at finde den rigtige ækvivalent på deres sprog. For det andet er ordforklaringerne grundlaget for en ny islandsk elektronisk ordbog, da en islandsk ordbog af mellemstørrelse over nutidssproget forefindes ikke. Det fremhæves dog her at de islandske forklaringer ikke skal være synlige for brugeren af ISLEX.

I databasen er der to kategorier for oversættelser til hvert målsprog. Den ene er til ækvivalenter og den anden er til en forklaring i de tilfælde hvor en ækviva-

lent ikke findes. Det afhænger oversætteren om han vælger en ækvivalent eller en forklaring i hvert tilfælde. Der er ikke altid overensstemmelse mellem dette i målsprogene; selv om der f.eks. findes en ækvivalent i svensk gælder det ikke nødvendigvis for dansk, og derfor må den danske oversætter forklare ordet i stedet for at bruge en ækvivalent. Feltet MSVIÐ indeholder lemmaets semantiske felt, evt. flere felter. Dette drøftes nærmere i 3.3.

Nederst i hver ordbogsartikel er der et felt der kaldes STIG (trin), der angiver hvor langt arbejdet med lemmaet er fremskredet, i kildesproget såvel som i hvert af målsprogene. Dette vises med tallene 0-3. Billede 1 viser de redaktionelle omgivelser i ISLEX, hvor der vises mange felter der forekommer i lemmaerne.



**ISLEX**  
STOFNUN ÁRNA MAGNÚSSONAR Í  
ÍSLENSKUM FRÆÐUM





hörpu-diskur n m

1	HLUTAR	hörpu-diskur	
2	SKÝRING	skeldýr (chlamys islandica)	
3	MYND		
		<input type="text" value="1 WWW"/> <a href="http://www.fauna.is">→ www.fauna.is</a>	
4	DA-jafn	kammusling	
5	NO-B-jafn	haneskjell	
6	NO-N-jafn	haneskjell	
7	SÆ-jafn	kammussla	
8	MSVIÐ	dýr	
9	STIG	(0-3:) Í-2 D-1 N-1 S-1	

[brotið](#) [strúktúr](#) [XML](#)

Leit að flettu:  ofl:  númer:  merki:  [Hætta](#)

með texta:   nákvæmt af teg:

OG  með texta:   nákvæmt af teg:

eldri leit:

Billede 1. Billedet viser et lemma der er blevet oversat til tre sprog. DA-jafn står for en dansk ækvivalent, NO-B-jafn for en ækvivalent på norsk bokmål, NO-N-jafn for en ækvivalent på nynorsk, og SÆ-jafn for en ækvivalent på svensk.

*Ordforbindelser og faste udtryk*

I ISLEX forekommer der flere typer fraser, d.v.s. kollokationer, faste udtryk og idiomer. Foruden Jón Hilmar Jónssons store ordbogsværk, *Stóra orðabókin um íslenska málnotkun* (2005), foreligger der ikke et oversigtsværk over fraser i islandsk og da heller ikke med oversættelser til andre sprog ligesom det er tilfældet i ISLEX. Kollokationerne medtages for at vise typiske eller obligatoriske syntaktiske mønstre omkring hvert enkelt lemma. Andre fraser skal fordybe brugerens forståelse og give en bedre adgang til et materiale som ofte er svært at finde frem til i traditionelle ordbøger, men er dog en vigtig bestanddel af kildesproget (Anna Hannesdóttir og Jón Hilmar Jónsson 2001). Antallet af fraser i ISLEX bliver sandsynligvis det samme som antallet af lemnaer, dvs. omkring 50.000. Målet er at gøre fraserne tilgængelige for brugerne på lige fod med lemnaerne (se 6.2).

*Eksempler*

Der findes mange sprogbrugseksempler i ISLEX-ordbogen. De er ikke autentiske eksempler, men valgt og tilpasset af den islandske redaktion. Eksemplerne viser lemnaerne i typiske omgivelser i sproget og er endvidere en tilføjelse til de semantiske oplysninger.

*Henvisninger*

Som man kender det fra andre online-ordbøger anvendes links når der henvises mellem lemmaer i ISLEX-ordbogen. Eksterne links er endvidere indbygget i systemet og de bruges til at forbinde ISLEX med andre websites. Den vigtigste er en database med bøjninger i islandsk, hvor ordenes bøjningsparadigme fremstår i sin helhed.

*Etiketter/Brugsbestemmelse*

Etiketter i ISLEX anvendes på en lignende måde som man kender fra de fleste traditionelle ordbøger. Der er to slags etiketter, for det første for at markere fagområder som *jura* og *fysik* og for det andet for at vise stiletiketter, f.eks. *gammelt*, *formelt* og *uformelt*. Både kildesproget og ækvivalenter i målsprogene gør brug af etiketter.

*Udtale og andre lyde*

Det blev tidligt besluttet ikke at vise ordenes udtale med fonetisk skrift, men at bruge lydeksempler i stedet. Denne fremgangsmåde er velkendt i elektroniske ordbøger. Andre lyde forekommer endvidere i ISLEX, bl.a. for at forklare lydord som f.eks. *plask*, *knirk* og *bums*.

*Illustrationer*

I ISLEX-ordbogen lægges der en stor vægt på at illustrere lemnaerne med billeder. Formålet med billedmaterialet er at hjælpe til ved bestemmelsen af ordets betydning. Desuden vil illustrationerne gøre brugen af ordbogen mere interes-

sant, især for de yngste brugere, der ellers måske ikke var interesseret i at slå op i ordbogen.

### 3.3 Semantiske felter

Hele ordforrådet i ISLEX er inddelt efter semantiske felter som hvert tæller 5 - 400 ord. De semantiske grupper er meget varierede og repræsenterer mange sider af ordforrådet. Således findes der semantiske felter som 'landskab', 'uvejr', 'skib', 'udseende', 'finanser', 'sundhed' og 'planter'. Hvert lemma har et afgrænset felt hvor den semantiske gruppe anføres, og hvert ord kan tilhøre flere semantiske grupper. Ordet *agúrka* (da. agurk) tilhører f.eks. tre grupper: *mad*, *grøntsager*, *planter*. Et andet eksempel er ordet *brúðkaup* (da. bryllup) der tilhører grupperne *forsamling*, *festlighed*, *giftermål*. Ved redaktionen udarbejdes hvert semantisk felt som helhed for at sikre at der er overensstemmelse mellem ord af lignende betydning. Efter at hver semantisk gruppe er klar fra den islandske redaktions side åbnes den til oversættelse. Oversætterne overtager således denne fremgangsmåde og erfaringen viser at de synes godt om arbejdsmetoden. Når én semantisk gruppe oversættes ad gangen bliver informationssøgningen mere koncentreret og målrettet, og det sparer både tid og besvær.

## 4. Databasen

I 2004 blev der udformet en database for ISLEX-projektet, som både skulle bevare projektets data og bruges som projektets hovedredskab. Systemet blev opbygget på den måde at det hører hjemme på det leksikografiske institut i Island (dengang Orðabók Háskólans), men skulle endvidere gøre det muligt at redaktionsarbejdet blev udført i flere lande samtidigt. Fordelen var at samarbejdsinstitutionerne i Norge, Sverige og Danmark ikke behøvede specielt udstyr til opgaven andet end en internetkobling, og en spredt placering af deltagerne ikke skulle være en hindring.

IT-specialisten Ragnar Hafstað blev ansat til at stå for design og programmering af systemet. Programmet blev taget i brug tidligt i året 2004 og det islandske grundmateriale blev importeret i databasen selv om systemet ganske vist ikke var helt færdigt på dette tidspunkt. Siden har Ragnar fortsat med at udvikle og udbygge databasen.

ISLEX-databasen bliver som før nævnt drevet som en webside. Der lægges en stor vægt på at udnytte mediets muligheder for at gøre projektet tidssvarende og de redaktionelle omgivelser så nemme og imødekommende for medarbejderne som muligt. Mange af de egenskaber vil på et senere tidspunkt givetvis gavnes ordbogens brugere.

### 4.1. Nærmere om systemet

For ISLEX-projektet bruges der open-source-systemet PostgreSQL for Debian Gnu/Linux operating system. Programmeringssproget er Perl.

Forskellige datatyper er defineret for ISLEX-databasen. For det første er der

tale om et **lemma** og dets **ordklasse** (og evt. nummer og en særskilt markering). Lemmaet kan være et ord, to ord eller en flerordsforbindelse. Lemmaet er systemets grundlæggende enhed og hvert lemma medtager et varieret antal andre indføringer af forskellige typer i ordbogaartiklen.


Der forekommer forskellige datatyper inde i selve ordbogsartiklerne. **Tekst** er den mest anvendte datatype, derunder falder ækvivalenter på målsprogene, to typer fraser, sprogbrugseksempler, etiketter, markering af semantiske felter o.fl. Visse teksttyper er i drop-down-menu f.eks. stiletiketter, hvor indeholdet er begrænset og forudbestemt.

**Billeder** og **animationer** er en særskilt datatype der tilføjes databasen med en mekanisme for uploading. **Lyd** er en datatype som hovedsagelig bruges til udtale af ord, men desuden for at beskrive lydord. Der er to typer **links**, den ene bruges til henvisninger indenfor ISLEX, og den anden forbinder ISLEX med andre websider.

Hjælpe materiale i form af **tekstfiler** er en del af databasen, men det er ikke synligt i overfladestrukturen. Man skal klikke på en link for at åbne et nyt vindue med disse filer, som hovedsagelig er linjer fra konkordanser, der kan anvendes som hjælp ved oversættelsesarbejdet. Tekstfilerne er uploadet i systemet.

#### 4.2. Struktur

I ISLEX er mikrostrukturens opbygning tredimensional, d.v.s. hierarkisk uden restriktioner. Dette betyder at inden for hvert lemma er det muligt at indordne dataene i forskellige trin, hvor hver indføring kan have en underordnet indføring, som igen kan have en anden indføring. Dette gør indholdets opbygning mere hensigtsmæssig end hvis man anvendte en flad struktur. Billede 2 illustrerer dette hierarki og viser lemmaet *kort* (da. kort).

kort n n			
1	HLUTAR	kort	+
2	LIÐUR	1	
		1 Z-MERKING landakort	+
		2 SKÝRING uppdráttur af borg eða landi	+
		3 MYND 	
		4 DA-jafn kort	+
		5 NO-B-jafn kart	+
		6 NO-N-jafn kart	+
		7 SÆ-jafn karta	+
		8 OSTÆÐA kort af <landinu>	
		1 DA-jafn kort over <landet>	+
		2 NO-B-jafn kart over <landet>	+
		3 NO-N-jafn kart over <landet>	+
		4 SÆ-jafn karta över <landet>	+
3	LIÐUR	2	+

Billede 2. Det er muligt at se tre trin i hierarkiet. Tallet 1 står i det øverste trin og viser den første af ordets tre betydninger. På andet trin vises ordets fire ækvivalenter på målsprogene samt feltet OSTÆÐA (orðastæða) hvor ordforbindelsen 'kort af <landinu>' fremstår. Det tredje trin rummer oversættelser af ordforbindelsen.

### 4.3. Grænseflade

Blandt andre af databasens bestanddele kan endvidere nævnes at der er udformet tre typer grænseflade ved siden af den vanlige, automatiske arbejdsgrænseflade.

1. **XML-snit.** Lemmaer og deres indhold overføres til XML-snit ved at klikke på en speciel link.
2. **Layout-form,** dvs. indholdet fremstår som en side i en ordbog og er tæt på det layout som vil møde ordbogens brugere. Man vil se indholdet af valgte indførelser i databasen (f.eks. tekst, billede), men ikke indførelsernes navne. Det er nemt at styre hvilke indførelser fremstår i den layoutede form og hvilke ikke. Bemærkninger fra redaktørerne bliver f.eks. ikke gjort synlige. Billede 3 ligner den grænseflade som møder brugeren i lemmaet *kort*, jf. billede 2 som viser den redaktionelle grænseflade.
3. **Struktureret grænseflade.** Omvendt den layoutede form ser vi på billede 4 kun navnene på indførelserne, men ikke deres indhold. Denne grænseflade er velegnet når man i store lemmaer vil se på materialets opbygning linje for linje.



Billede 3



Billede 4

## 5. Redaktion i fire lande

I hvert land er der 1–3 redaktører der tager sig af oversættelser. Tilknyttet til redaktionen er der desuden et fagudvalg som har tilsyn med projektet. Oversætterne deltager på mange måder aktivt i udformningen af værket og sikrer en slags kvalitetskontrol da de til enhver tid kan melde tilbage hvis de støder på fejl eller uoverensstemmelser. De kan også fremsætte ønsker om nye/ændrede sprogbrugseksempler, hvis de synes at dette behøves. Når den islandske redaktion har gennemgået en semantisk ordgruppe, sender hovedredaktøren en melding om dette på ISLEX-forumsiden (se 5.1). Der gælder ingen regler for hvilken redaktion oversætter først, og der kan være forskellige grunde til rækkefølgen. Eftersom alle målsprogene er beslægtede kan de ofte have støtte af hinanden ved oversættelsen, men dette kan dog i visse tilfælde være farligt da visse nordiske homonymer kan som bekendt have forskellig betydning i målsprogene.

### 5.1 Kommunikation

Fordi ISLEX-projektet opererer i fire lande kan der ikke tit afholdes egentlige redaktionsmøder. Derimod anvender redaktørerne tre forskellige kommunikationskanaler til den daglige kontakt. For det første er der oprettet en særskilt forumside i forbindelse med ISLEX-projektet. Denne side anvendes især for

vigtige meldinger henvendt til alle, men der finder man endvidere en håndbog om systemet, mødereferater, statistik, fotografier o.fl. En anden kommunikationskanal mellem enkelte medlemmer af redaktionen er e-mails, der hovedsagelig anvendes når det drejer sig om vigtige redaktionelle meldinger, f.eks. ændringer i enkelte lemmaer og når der opstår situationer hvor der skal reageres hurtigt. Den tredje kommunikationskanal er via selve databasen, hvor der i de enkelte ordbogsartikler er indbygget en kategori for bemærkninger. Der kan alle redaktionens medlemmer fremsætte kommentarer eller spørgsmål til hver enkelt lemma. Kommentarerne er synlige for hele redaktionen og fremkalder ofte feedback fra de øvrige redaktører.

### 5.2 Arbejds møder

Der afholdes arbejds møder for hele ISLEX-redaktionen to gange om året. Møderne strækker sig over tre dage og er meget vigtige for hele koordineringen af projektet. Selv om redaktionen har god kontakt via webben er disse møder af stor betydning for samarbejdet, hvor man mødes og lærer hinanden bedre at kende. Det ene årlige møde afholdes sædvanligvis i Island og det andet afholdes på skift i de øvrige nordiske samarbejdslande.

## 6. Brugergrænsefladen

### 6.1 Søgning

Mange webbaserede ordbøger er oprindelig trykte værker der er blevet overført til mediet. Af den grund ligner de trykte bøger i deres fremsætning af materialet, (f. eks. ordnet.dk/ods). ISLEX er derimod fra starten designet for webben og den har ikke de samme restriktioner som ældre ordbøger. Brugergrænsefladen for ISLEX er endnu ikke udformet, men der er ret klare ideer om hvordan den skal være. Som i andre webbaserede ordbøger vil brugeren kunne lede efter et islandsk opslagsord eller en orddel (søgen med wildcard) og man kan indstille hvilke mål-sprog man ønsker at se på skærmen.

Eftersom oversættelsesarbejdet skrider frem i Danmark, Norge og Sverige tilføjes der til databasen en stor mængde ækvivalenter på målsprogene. Dette giver en mulighed for søgning i dette materiale, d.v.s. at man søger efter et dansk ord og får den islandske ækvivalent. Man må dog huske på at ordbogen er udformet med islandsk som kildesprog og at det derfor er begrænset sikkerhed når man anvender ordbogen i modsat retning.

Der er mulighed for mange flere søgefunktioner. For eksempel kan det tænkes at der gives adgang til søgning i det materiale der findes i sprogbrugseksemplerne og oversættelser af disse, både med islandsk og målsprogene som udgangspunkt.

### 6.2 Adgang til fraser

En stor del af ISLEX-ordbogens indhold er ordforbindelser og faste udtryk og



dette materiale er specielt markeret indenfor databasen. Planen er at gøre fraserne nærmest uafhængige af resten af ordbogen og således at have en parallel fraseologisk ordbog med ISLEX-projektet. Søgning i dette materiale forudsætter et nøgleordssystem, som allerede er indbygget i systemet. For eksempel kan man lede efter frasen *have hjertet på rette sted* og så kan man enten slå 'hjerter' eller 'sted' ind i søgeboksen. Denne søgefunktion kender man f.eks. fra den webbaserede ordbog *The Free Dictionary*.

## 7. Betydning for leksikografien

ISLEX-ordbogen genspejler det islandske nutidssprog og skaber dermed grundlag for andre ordbogsprojekter med islandsk som kildesprog. Man kunne tænke sig en udvidelse af ISLEX-projektet, hvor der indlemmes flere målsprog. En anden mulighed er at ISLEX-projektet bliver offentliggjort i trykt form, enten som en tosproglig ordbog mellem islandsk og et af de andre nordiske sprog, eller mellem islandsk og alle målsprogene.

Foruden at være en multisproglig ordbog er ISLEX-projektet samtidig en udforskning af det elektroniske mediums muligheder indenfor leksikografien. I den forbindelse kan nævnes et fællesnordisk projekt der på svensk kaldes *Nordisk nätordbok*, og dets brugeradgang *Tvärslå* ([http://ordbok.nada.kth.se:8070/ordbog\\_module/ordbog](http://ordbok.nada.kth.se:8070/ordbog_module/ordbog)). *Tvärslå* er en samling af tosproglige ordlister over svensk, norsk (nynorsk og bokmål), dansk og islandsk, ligesom ISLEX, og desuden færøsk og engelsk.

Projektet beskrives bl.a. med følgende ord: „I delprojektet *Tvärslå* utvecklar vi en tvärspråklig lexikontjänst där både mänskliga och omänskliga användare ska kunna slå upp i alla lexikon som gjorts tillgängliga i projektet“ (se websiden [www.csc.kth.se/tcs/projects/netordbog](http://www.csc.kth.se/tcs/projects/netordbog)).

Projektet minder lidt om ISLEX, men adskiller sig dog på den måde at *Tvärslå* ikke er en egentlig ordbog og ikke har noget bestemt kildesprog.

ISLEX-projektet drejer sig om et meget omfattende ordbogsværk hvor der er indsamlet en stor mængde systematisk lingvistisk data. I ISLEX er ordforrådet specielt godt inddelt efter semantiske felter og dette kan blive et værdifuldt redskab for semantiske analyser, både indenfor hvert enkelt sprog og mellem de sprog der findes i ISLEX. Materialet byder også på muligheder for andre produkter f.eks. islandske lemmaer med forklaringer på islandsk og ordforbindelser knyttet til semantiske begreber. Desuden indsamles der en stor mængde sproglige oplysninger om enkelte ord, f.eks. om verbers valens og argumentstruktur.

Store dele af materialet er velegnet til sprogteknologiske projekter og i Göteborg er man allerede begyndt på et forskningsprojekt indenfor fraseologi der henter data fra ISLEX. Man kan tænke sig at der i fremtiden vil udspringe mange andre forskningsprojekter af ISLEX-materialet.

**Litteratur:**

Anna Hannesdóttir & Jón Hilmar Jónsson, 2001: Að hafa í sig og á – Ísländsk fraseologi í ett isländskt-svenskt perspektiv. *LexicoNordica* 8: 67–91.

Jón Hilmar Jónsson 2005: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.

*The Free Dictionary*. <http://idioms.thefreedictionary.com>

*Tvärslå*. [http://ordbok.nada.kth.se:8070/ordbog\\_module/ordbog](http://ordbok.nada.kth.se:8070/ordbog_module/ordbog)

*Nordisk nätordbok - Tvärslå och Tvärsök*. <http://www.csc.kth.se/tcs/projects/netordbog>

*Ordbog over det danske sprog*. <http://ordnet.dk/ods>

Halldóra Jónsdóttir  
projektleder, f. 1957  
halldo@hi.is

Þórdís Úlfarsdóttir  
hovedredaktør, f. 1961  
disa@hi.is

Stofnun Árna Magnússonar í íslenskum fræðum – Háskóli Íslands  
Neshagi 16  
IS-107 Reykjavík

## Jysk Ordbog – Rapport fra en digital ordbog

The Peter Skautrup Centre has been publishing the Dictionary of Jutland Dialects (Jysk Ordbog) on the internet since October 2000, and everyone with access to a computer connecting the internet has free access to all articles from A to Hæ-. The multilayered site contains links to dialect geographical maps, references to applied literature, a list of editorial abbreviations, an overview of phonetical signs, and furthermore links to information on the editorial background of the dictionary, i.e. geographical distribution and the assembling of the large and varied material. As the process of editing is very complex and, more important, in a way hazardous due to out of date software, plans are being made to update the technical platform of the dictionary. This new, far more coherent platform will consist of first of all an editorial (webbased) database, written in a XML standard. As the Centre has succeeded in finding some of the acquired financing funds, the final phase will start in securing and making more easily available the online dictionary.

### 1. Forhistorien

11. oktober 2000 blev der på Institut for Jysk Sprog- og Kulturforskning (som Peter Skautrup Centret på daværende tidspunkt hed) holdt en reception i forbindelse med lanceringen af Jysk Ordbog på internettet. Efter mange års forberedelser var det lykkedes, med teknisk hjælp fra IT-virksomheden UNI-C, at offentliggøre færdigredigerede afsnit af dialektordbogen for alle med adgang til computer og internetopkobling under adressen: [www.jyskordbog.dk](http://www.jyskordbog.dk).

I første omgang var det ordbogsafsnittene A–F brugerne kunne søge i, men i dag er afsnittene til og med Hæ- lagt på nettet (se figur 1) (sidste ord er **hævte**, verb. ‘true (med en kæp el. lign.)’).

Hjemmesiden opfylder i store træk redaktionens ønsker om præsentationen af Jysk Ordbog. På hjemmesidens frontside finder man oplysninger om redaktionens adresse, dens sammensætning gennem tiderne og praktiske oplysninger med hensyn til at downloade ordbogens særlige skrift. Linkene på sidens venstre halvdel fører til oplysninger om ordbogens redaktionelle baggrund med information om ordbogens emne (“Hvad dækker ordbogen?”), ordbogens forskellige korpusser (“Ordbogens samlinger”), artiklernes opbygning, ordbogens hjælpemidler (lydskrift, ordbogens kilder, dialektgeografiske kort mv.). Endelig er her også et link til ordbogens historie, der starter tilbage i 1932 da Peter Skautrup oprettede Aarhus Universitets første rigtige institut. Det centrale link på frontsideen fører naturligvis læseren til opslag i ordbogen.

**Jysk Ordbog**

**Hvad dækker ordbogen?**  
 Område  
 Periode  
 Ordforråd

**Ordbogens samlinger**

**Artiklernes opbygning**  
 Ordklasser  
 Artiklernes hoved  
 - Opslagsform  
 - Udtale  
 - Bøjning  
 Kantet parentes  
 Artiklernes krop  
 - Betydning  
 - Citater

**Hjælpe midler**  
 Lydskrift  
 Områder mv.  
 Forkortelser mv.  
 Kildeangivelser  
 Dialektkort

**Ordbogens historie**

## Velkommen til Jysk Ordbog

**Slå op i ordbogen her (A - Hæ)**

Resten af ordbogen vil blive lagt ud i takt med redaktionen. Har du spørgsmål til afsnit Hø - Å, kan du klikke [her](#), skrive spørgsmålet - og få svar fra os senere.

Ordbogen anvender en særlig lydskrift, som kan downloades [her](#)

- til PC
- til MAC
- til LINUX

**Peter Skautrup** Centret for Jysk Dialektforskning  
 Aarhus Universitet  
 Jens Chr. Skous Vej 4  
 Nobelparken, bygn. 1483  
 8000 Aarhus C.

Tlf.: 8942 1913  
 Fax: 8942 1912  
 E-mail: [jyskordbog@hum.au.dk](mailto:jyskordbog@hum.au.dk)

Jysk Ordbog er grundlagt af Peter Skautrup.

Indsamlet og prøveredigeret af Peter Skautrup, K. Flingsgaard, Niels Age Nielsen, Ove Rasmussen og Hans Veirup.

Redigeret af Ove Rasmussen, Viggo Sørensen, Torben Arboe, Inger Schoonderbeek Hansen og Nina K. Grøtthauge.

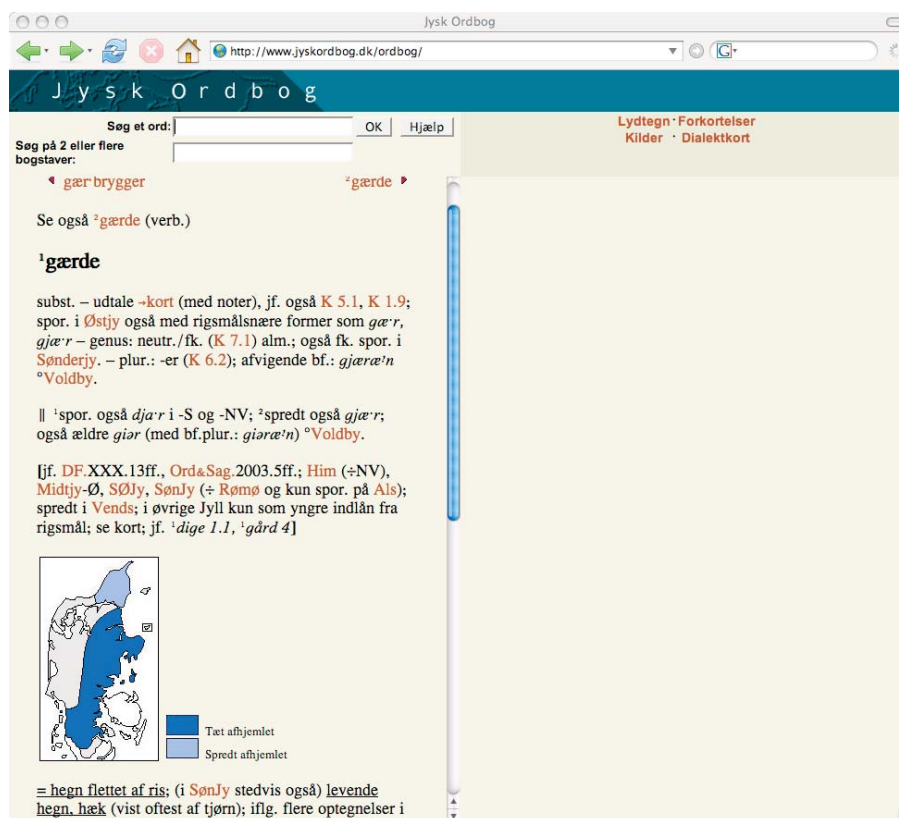
Omsat til skærmudgave af Nina K. Grøtthauge og Kirsten Lyshøj.

Implementeret på Internet af UNI-C ved Bengt Moss-Petersen.

Web-siderne er redigeret af Ove Rasmussen og Viggo Sørensen. Alle henvendelser bedes rettet til [web-editor Kirsten Lyshøj](#).

Figur 1. Jysk Ordbogs frontside

Ved hjælp af to søgefelter på opslagssiden kan man foretage opslag i ordbogen. I øverste søgefelt, "Søg et ord", kan man taste et helt ord ind. Og i søgefeltet nedenunder kan man, som nyt tiltag fra 2003–2004, søge ord i Jysk Ordbog ved at taste 2 eller flere startbogstaver af det ord man leder efter (se øverst i venstre spalte på skærbilledet fra figur 2). Øverst på opslagssidens højre halvdel fører linket "Lydtegn" brugeren hen til informationer om ordbogens lydskrift. Lydskriften er relativt grov, baseret på rigsmålets ortografi og Danias lydskrift. Via de øvrige link øverst på højre side kan brugeren finde informationer om de forkortelser der bruges i ordbogen, de kilder ordbogen anvender, og endelig via linket "Dialektkort" komme frem til en gruppe af dialektkort, vi hyppigt henviser til, vedrørende grovere eller finere inddelte dialektområder, forekomst (eller ej) af stød og længde i de jyske dialekter, jyske modsvarigheder til rigssproglige eller gammeldanske vokaler og konsonanter, særlige jyske bøjningsforhold mv.



Figur 2. Opslag <sup>1</sup>*gærde* ‘hegn flettet af ris ...’ i [www.jyskordbog.dk](http://www.jyskordbog.dk)

Alle link i ordbogen er markeret med rødt, også i de enkelte artikler (jf. figur 2). Opslag i Jysk Ordbog gør det således muligt at bladre frem og tilbage til foregående og efterfølgende artikel ved at klikke på henvisninger øverst i artiklen (se figur 2: “◀ gær·brygger” og “<sup>2</sup>gærde ▶”). Samme link er også placeret i bunden af den side artiklen optræder på (ses ikke på figur 2). I de tilfælde et opslag kunne føre til to eller flere artikler i Jysk Ordbog er der via et link mulighed for at komme frem til disse; det kan fx være homonymer (se figur 2: “Se også <sup>2</sup>gærde (verb.)”).

Den enkelte artikel er normalt opbygget på følgende måde (se figur 2):

1. Hoved: opslagsform; ordklassebetegnelse; oplysninger om ordets udtale og bøjning (med links til ordbogens dialektkort); evt. noter til udtaleangivelsen (eller til et helsides udtalekort).
2. Etymologisk parentes: evt. etymologi; ordets udbredelse (evt. med henvisning til et udbredelseskort); henvisning til evt. synonymer.
3. Kroppen: ordets oversættelse(r) til rigsdansk; efterfølgende citater, der illustrerer ordets brug.

I den enkelte artikel vil brugeren ofte via link blive ført til:

- a) Udtaleoplysninger, fx ved linket “udtale →kort” til udtalekort i skærmens højre side (evt. med noter); ved henvisning til standardkort i ordbogens atlasdel (dialektkortene) over accentforhold mv. ved link som “(K 5.1)”, “(K 1.9)” etc.
- b) Geografiske oplysninger ved link fra “se kort” til tilføjet udbredelseskort; ved link fra områdebetegnelser som Østjy, RougsøH; ved link fra ordbogens vigtigste kilder (de såkaldte punktoplysninger, fx “°Voldby”) til et kort, der åbnes i et separat vindue.
- c) Kildeoplysninger ved link fra ordbogens litteraturforkortelser til en liste over ordbogens kilder, der åbnes i højre halvdel af siden.

Sådan fungerer Jysk Ordbog i dag, meget kort fortalt.<sup>1</sup>

## 2. Nuværende teknisk platform

Bag Jysk Ordbogs meget brugervenlige grænseflade ligger en noget omstændelig og risikofyldt redaktions- og publikationsproces.

Selve redaktionen foregår ad følgende seks trin:

- I Redaktøren redigerer på baggrund af Jysk Ordbogs seddel- og spørgelistemateriale<sup>2</sup> sit ordbogsafsnit i et almindeligt tekstbehandlingsprogram (Word 5.1 for MacIntosh!); alle dialektgeografiske kort bliver tegnet i Adobe Illustrator (version 11.0.0).
- II Når et afsnit er færdigredigeret, gennemlæses det af en anden redaktør for til sidst at havne til endelig finpudsning, opsætning og omsætning til Word 97-fil, hvilket foregår på Centrets eneste pc.
- III Word 97-filen konverteres via en specielt designet makro til en ny fil i en slags HTML-format.
- IV Denne (pseudo) HTML-fil flyttes via en browser fra pc'en til en af UNI-C's servere ved hjælp af et filhåndteringssystem (WebDAV).
- V (Pseudo) HTML-en konverteres til en (rigtig) HTML via et af UNI-C specielt designet programmodul, et såkaldt CGI-script.
- VI Ny artikler (eller tilføjelser til eller ændringer i allerede publicerede artikler) kræver at et større omgivende afsnit af ordbogen opdateres sam-

1 Yderligere beskrivelse af [www.jyskordbog.dk](http://www.jyskordbog.dk): Torben Arboe (2003) “Jysk Ordbog på internettet” i *Nordiske Studier i Leksikografi* 6. Rapport fra Konferencen om leksikografi i Norden. Tórshavn 21.–25. august 2001. s. 31–41

2 Jysk Ordbogs seddelsamling består af mere end 3 millioner alfabetisk og topografisk ordnede papirsedler opbevaret i 14 kartoteksskabe med i alt knap 1300 skuffer. Derudover består spørgelistematerialet, dvs. besvarelser (fra gennemsnitligt 500 meddelere) af ca. 1200 spørgelister (af gennemsnitligt ca. 20 spørgsmål), der for 75%’s vedkommende er it-lagret, af oplysninger der svarer til godt to tredjedel af seddelsamlingen.

let. Ergo: Jo flere artikler der redigeres, jo længere tager det at opdatere ordbogen.

Den ovenfor beskrevne noget komplekse redaktions- og publikationsproces medfører en række risici:

- Ad I: Word 5.1 er et stærkt forældet tekstbehandlingsprogram.
- Ad II: Tyveri eller nedbrud af pc'en er en stadig fare.
- Ad III: Makroen er ensidigt kompatibel med Word 97, men ikke senere udgaver af Word; en autoriseret nyinstallation af den gamle software er ikke mulig da den ikke længere distribueres af Microsoft; det er desuden ikke sikkert at den specielt designede makro overhovedet kan fungere sammen med senere versioner af Word eller senere versioner af styringssystemet Windows.
- Ad V: Til trods for at systemet kun har haft ganske få nedbrud eller driftsproblemer i alle disse år, er det usandsynligt at UNI-C (endsige andre) på ret kort sigt råder over den fornødne ekspertise mht. de enkelte moduler da disse er baseret på 8-10 år gamle web-teknologier og applikationer (hvilket er *meget gammelt* i edb-verdenen).

Strengt taget er rekruttering af ny redaktører også et problem men det er en helt anden historie.

Processen er altså for kompliceret, for usikker, for langsom og gør brug af for gammelt programmel. Processen bør fremover baseres på åbne standardkomponenter, være mindre kompleks, være (mere) fremtidssikret; den bør simpelthen opdateres.

Kort sagt: Publikationen af Jysk Ordbog på nettet kræver en ny teknisk platform!

### 3. Fremtidig teknisk platform

Da det i sin tid var UNI-C der stod bag implementeringen af Jysk Ordbog på internet, og de lige siden har stået for hosting af Jysk Ordbog på en af deres servere, var det oplagt at UNI-C så det som sin opgave at komme med et forslag til en forbedret platform for [www.jyskordbog.dk](http://www.jyskordbog.dk).<sup>3</sup>

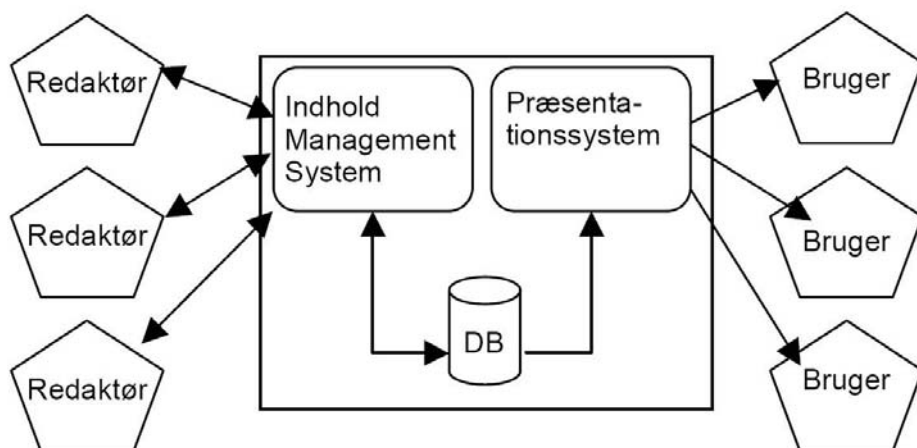
Den ny version af [www.jyskordbog.dk](http://www.jyskordbog.dk) tænkes opbygget som et samlet system centralt placeret på en hovedserver. Alt materiale, altså tekst, kort mv., skal ligge centralt på denne server, og der skal derfor ikke flyttes informationer mellem de enkelte elementer af systemet. Selve redaktionen vil foregå online via en browser idet redaktørerne kan arbejde uafhængigt af hinanden.

<sup>3</sup> I den forbindelse har UNI-C – efter aftale med Centret – udarbejdet to rapporter: “Forslag vedrørende teknisk platform for [www.jyskordbog.dk](http://www.jyskordbog.dk)”, november 2004, og “Tilbud om foranalyse til udvikling af ny platform for [www.jyskordbog.dk](http://www.jyskordbog.dk)”, oktober 2006.

I store træk vil den ny version se således ud:

- Det samlede system er mindre kompleks og mere fejlsikker end det gamle idet det bygger på en XML-standard.
- Datagrundlaget bliver mere struktureret, hvilket giver mulighed for senere at eksportere data (i XML-format) til en 3. part, fx via link fra andre netbaserede ordbøger mv.
- Der anvendes et let læseligt programmeringssprog, så systemet lettere kan vedligeholdes og modificeres fremover (evt. også i ny ordbogssammenhænge).

Grafisk fremstiller UNI-C Jysk Ordbogs ny version som det fremgår af figur 3.



Figur 3. Jysk Ordbogs ny teknisk platform efter UNI-C's forslag (jf. note 3)

Det samlede system indeholder en todelt database hvor den ene del indeholder data om ordbogens indhold, management og system mens den anden del af systemet udgør selve præsentationen. Redaktørerne redigerer fx via en browser direkte i systemet (indhold), og ad denne vej kan fx en webdesigner ændre ordbogens management og (til dels) system. Brugerne har naturligvis kun adgang til selve præsentationen af ordbogen.

I forhold til det tekstbehandlingsprogram (Word 5.1) vi bruger i dag, vil det ny, modulært opbyggede redigerings- og publiceringssystem betyde en væsentlig effektivisering af redaktionsprocessen i Jysk Ordbog. Særligt håber vi på at en del af Jysk Ordbogs store og differentierede henvisningsapparat i højere grad bliver automatiseret i den ny teknisk platform. For mange manuelle indtastninger på området medfører i dag en for høj fejlprocentage og en for stor andel i korrekturarbejdet.



#### 4. Jysk Ordbogs henvisningsapparat

Jysk Ordbogs mange forskellige typer henvisninger kan opdeles i *interne* og *eksterne* henvisninger. De interne henvisninger forbinder tekstenheder inden for ordbogen selv, de eksterne henvisninger peger ud til hvad vi kalder ordbogens fællesfaciliteter, dvs. dialektkort, kildefortegnelser mv. I figur 4 ses henvisningernes opdeling i interne og eksterne henvisninger (til dels på basis af artiklen fra figur 2). Der vises om henvisningen p.t. oprettes automatisk eller indtastes af redaktøren, og om de er opmærket eller ej. En firkantet klamme [ markerer at henvisningen er forberedt men endnu ikke implementeret, mens en tuborgklamme { markerer at henvisningen endnu ikke er forberedt men ønskes implementeret i fremtiden. De åbne pile i figur 4 viser i hvilken retning vores ønsker med hensyn til den ny platform går, dvs. fra mange manuelle til mange automatiske handlinger.

Interne henvisninger		Eksterne henvisninger	
Manuelt	Automatisk	Manuelt	Automatisk
til opslagsform – indtastet i database med side- og lydformer	til foregående og kommende artikel – i præsentationen markeret med ◀ ▶	til kildefortegnelse – opmærket foran og bag kildeforkortelsen ◊ ⇒	
[til anden artikel – opmærket @ foran fx synonym] ⇒		til atlasdel – opmærket fx “(K 5.1)” ⇒	
		til udtalekort – opmærket “→ kort” ⇒	
		til spaltekort – opmærket “se kort” ⇒	
		{til artikler i andre netordbøger etc.}	

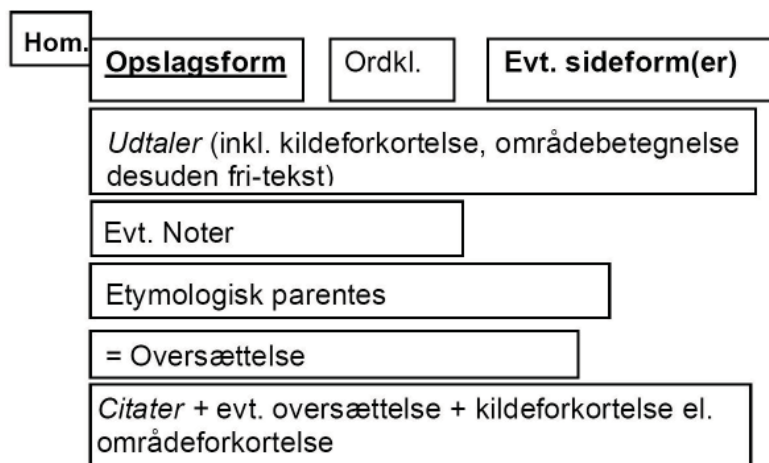
Figur 4. Jysk Ordbogs henvisningsapparat

Den ny platform vil kunne gavne og ikke mindst aflaste redaktionen i forhold til selve indtastningen af dataene, de enkelte artikler. Som sagt bruger vi i dag et almindeligt tekstbehandlingsprogram, og her er det redaktøren der har ansvar for evt. slåfejl, forglemmelser mv. Der bliver naturligvis læst grundig korrektur på hvert enkelt afsnit men der er utrolig mange detaljer at holde rede på: korrekte forkortelser (for områder og kilder), korrekte tegn som punktum, semikolon, komma – for ikke at tale om korrekt skift mellem fonte (fed, kursiv, understreget, antikva, særtegn), så det er svært at undgå småfejl i selve teksten. Dertil kommer opmærkningen til de forskellige eksisterende og planlagte interne og

eksterne link i ordbogen, fx de små hak foran og efter en kildeforkortelse, en bestemt pil foran en henvisningsform, et bestemt mellemrum foran og efter selve opslagsformen etc. Lige nu resulterer fx forveksling af et tegn eller forglemmelse af et punktum i en kildehenvisning i at der ikke oprettes link, og at brugeren altså ikke kan føres videre til kildefortegnelsen. Den ny platform skulle helst automatisk kunne kontrollere korrekt anvendelse af tegn, fonte og forkortelser; på andre punkter må den ikke hindre den enkelte redaktør i at give en artikel en vis form for individuelt, personligt præg).

## 5. Fremtidsplaner

Hvilke muligheder skal den ny platform til Jysk Ordbog have? Hvis man forestiller sig at databasen i den ny platform består af bokse der skal indeholde bestemt information, skal de enkelte bokse defineres ganske nøjagtigt så fejl eller forglemmelser udbedres automatisk eller eventuel fejlmeldes. For nogle af boksene ville fx gælde at skrifttypen i dem er defineret til at være fed og understreget mens redaktøren i andre bokse har større frihed, fx med hensyn til formuleringen i den etymologiske parentes (se figur 5).



Figur 5. Skitse af Jysk Ordbogs ny redigeringsdatabase

Databasen kunne være opbygget således at der i redaktionen først må tages stilling til om et opslagsord har homonymer, hvilket markeres ved små, hævede tal foran opslagsordet i fed og lidt hævet over linjen (jf. også figur 2). Næste boks indeholder opslagsformen som i præsentationen altid er fed og understreget. Dernæst indtastes ordklasse; boksen kunne fint fungere som rullevindue der indeholder de forskellige muligheder. Dernæst en boks der kan rumme eventuelle sideformer, der altid er fede. Så kommer en boks der skal give lidt større frihed, nemlig til at gøre rede for opslagsformens udtale. Udtalerne er i præsentationen i kursiv mens

al anden tekst er i antikva. De kilde- og områdeforkortelser der tastes ind her, må kunne tjekkes automatisk. Noteapparatet skal i præsentationen indeholde små hævede tal samt selve noteteksten, der refererer til udtaleoversigten eller/og et udtalekort. Noterne indeholder også kilde- og områdeforkortelser – der automatisk skal kunne kontrolleres. Den etymologiske parentes indeholder alle typer henvisninger, der automatisk skal kunne kontrolleres, mens oversættelsen er mere fri, dog optræder (hovedparten af) den altid som understreget tekst i præsentationen. Efterfølgende citater er altid i kursiv mens eventuelle oversættelser er i antikva; også her er der mange område- og kildehenvisninger.

I alle bokse skal der være indbygget en form for henvisning til en anden del af ordbogen. Fx skal homonymnummer i kombination med opslagsformen kunne linkes til artiklens udtalekort; boksen med eventuelle sideformer skal kunne linke til ordbogens søgefelt, så brugeren også kan finde disse afvigende former; område- og kildeforkortelser i boksen med udtaler skal forbindes med oversigten over områder hhv. kildefortegnelsen mv.

Redigeringsværktøjet skal kunne rumme flere forskellige layout så redaktøren fx også kan skrive sin artikel i en slags trappediagram der beskriver ordbogsstrukturen bag hver enkelt artikel. Desuden skal redaktøren frit kunne skifte mellem det ene og det andet layout og mellem de forskellige bokse eller elementer i strukturen, fx sådan at han/hun ikke er bundet til altid at skulle bringe en oversættelse (hvilket vi ikke gør ved ord der kun er afhjemlet fra en enkelt kilde).

Jysk Ordbog har et stærkt presserende behov for ny teknisk platform. I første omgang skal der laves en foranalyse der beskriver hver enkelt komponent i den ny platform, hvad det kræver af hard- og software, hvilke problemer man eventuelt måtte støde på mv. Centret skal her via en workshop bidrage med sit kendskab til ordbogens struktur og sine ønsker med hensyn til præsentationen. En del af dette arbejde er allerede gjort ved etableringen af vores nuværende netudgave mens andet skal revideres da den ny platform anvender ny værktøjer der med stor sikkerhed giver ny muligheder, fx med hensyn til intertekstuelle henvisninger.

Der er også planlagt en workshop om redigeringsværktøjet iLEX (udviklet af Jens Erlandsen), så dets fordele og mangler i forhold til Jysk Ordbog kan afdækkes. iLEX er efterhånden det tool der hyppigst anvendes i ordbogsbranchen i Danmark, fx hos DSL, Gyldendal, og det åbner klart nogle perspektiver. Det vil være en fordel, men ikke en nødvendighed, for det videre forløb hvis eksisterende delkomponenter til en ny platform kunne anvendes.

Endelig er det UNI·C's plan på basis af de to workshoper at udarbejde en rapport som skal indeholde en anbefaling om det videre arbejde med den ny platform. Rapporten vil give et bud på hvilken XML-standard der skal arbejdes med, hvilken editor der skal bruges (altså det redigeringsværktøj redaktørerne skal arbejde med til daglig), og hvordan det videre forløb skal se ud frem til det endelige system.

Siden konferencen i Akureyri maj 2007 har projektet omkring fremtidssikringen af [www.jyskordbog.dk](http://www.jyskordbog.dk) været igennem en turbulent periode. Først opsagde UNI-C ensidigt deres samarbejde med os, og hermed faldt alt det planlagte på jorden. Ikke længe efter så fremtiden dog lysere ud da Jysk Ordbog blev medspiller i det langt større projekt DK-CLARIN, hvis mål er lettere digital adgang for såvel forskere som offentligheden til humanistisk og socialvidenskabeligt kildemateriale. Udviklingsarbejdet bliver nu forestået af ekspertise fra Det Humanistiske Fakultets egen IT-afdeling i tæt samarbejde med partnerne i DK-CLARIN-projektet. En stor del af finansieringen til foranalyse og udvikling af en ny platform er dermed kommet på plads, og samtidig vil der blive mulighed for gensidig linking til CLARIN's øvrige partnere.

Inger Schoonderbeek Hansen  
f. 1970  
Peter Skautrup Centret for Jysk Dialektforskning  
Nordisk Institut, Aarhus Universitet  
Jens Chr. Skous Vej 4, bygn. 1483, 6. etage  
8000 Århus C  
[jysis@hum.au.dk](mailto:jysis@hum.au.dk)

## Om noen praktiske spørsmål i tospråklig leksikografi

The article discusses the relation between conventional textbooks of grammar, dictionary grammars and grammatical information in the lemma list section (the dictionary proper) in a bilingual dictionary. The treatment of synonym groups is also commented upon.

Particular attention is given to grammatical categories that exist only in one of the two languages involved, such as the definite article in Nordic languages versus Russian and other Slavic languages, the aspect category in the latter which is absent in Nordic languages. It is important that these differences are clarified and explained as meticulously as possible by concrete indications and examples of use. Instances of this are shown, e.g. the problem of translating examples containing the only Russian past tense into Germanic languages which possess a number of past tenses, e.g. the choice between "aoristic" and "perfectivic" meaning of the Russian perfective past tense.

Furthermore it is of paramount importance not to overlook differences in the meaning structure and extension (range of meanings) of the source language lemma and its primary equivalent in the target language. A comparison of the solution of such problems in three different dictionaries is presented.

The use of a certain category, existing in both source and target language, may reveal subtle differences not to be ignored. An instance of this is a specific use of the grammaticalized definite article in Norwegian and Icelandic, an idiosyncrasy that is not heeded in textbooks. Needless to say, it is practically impossible to cover all the above-mentioned differences by examples in every concrete case, and *ad hoc* solution must be applied to a great extent.

### 1. Grammatikkens kår i en tospråklig ordbok

Et av de spørsmål som alltid står på dagsorden når en tospråklig ordbok skal lages, er hvilke grammatiske (morfologiske og syntaktiske) opplysninger ordboken skal formidle til begge målgrupper og, ikke minst, hvordan. Dertil kommer informasjon om stil, distribusjon, diverse konnotasjoner o.m.a.

Jo mindre en tospråklig ordbok er, desto mer må den stole på at brukeren er tilstrekkelig bevandret i grammatikk til å klare seg uten eksplisitt informasjon av nevnte type; i høyden angis lemmaenes ordklasse. I praksis utgjør dette et paradoks, fordi (biskopale) lommeordbøker gjerne brukes av nybegynnere. Slike brukere, spesielt i de tilfeller der ordboken brukes til tekstproduksjon på fremmedspråket, er da ikke i stand til å aktivisere bokens leksikalske opplysninger på en adekvat måte.

I mellomstore verk av denne typen er spekteret av valgte løsninger enormt. Man får inntrykk av at det er vanskelig å finne to ordbøker med identisk løsning

på nevnte spørsmål. Dessuten dikterer mål- og kildespråkets karakter en spesiell ad hoc tilnærming i hvert enkelt tilfelle. Jo større typologisk forskjell mellom de to språk, desto mer kompliserte løsninger. Disse bøker omfatter som regel ikke en ordboksgrammatikk eller bare en meget primitiv sådan. Også her forutsettes tilstrekkelig opplæring hos brukerne.

Store tospråklige ordbøker som f. eks. V. Berkovs *Russisk-norsk ordbok* (RNO, 1987 og 1994) og *Stor norsk-russisk ordbok* (SNRO, 2003 og 2005) tilstreber maksimal informasjon på et minimum av plass, noe som innebærer et kodesystem med henvisninger til en skreddersydd ordboksgrammatikk. Det er nettopp denne ordbokstypen som de mest avanserte brukerne benytter seg av.

Fremtidens nettbaserte ordbøker behøver ikke å ta hensyn til plass. Dessuten har slike bøker tekniske muligheter som bare i beskjeden grad er tilgjengelige for papirbaserte ordbøker (f.eks. bruk av diverse farger, mangfold av lenkekjeder o.m.a.).

## 2. Hvordan skal grammatiske opplysninger formidles?

Til og med i store ordbøker av den type som ble nevnt i 1 ovenfor (RNO og SNRO), er spørsmålet om forholdet mellom den grammatikk brukeren forutsettes (ideelt) å ha tilegnet seg og ordbokens grammatiske og kombinatoriske opplysninger, brennaktuelt.

Det er tre måter å gi denne informasjonen på. Den kan gis eksplisitt, herunder vha. vanlige standardforkortelser, eller via kodesystem, som f.eks. med bøyningskoder som viser til ordboksgrammatikk, som i RNO, SNRO og *Russisk-islandsk ordbok* (RIO, 1996).

Den tredje måten er den vanskeligste, dvs. informasjon innebygget i brukseksempler med hovedvekt på det kontrastive aspekt. I vellykket og gjennomtenkt form er denne metoden brukervennlig og informativ på et enkelt og direkte vis og bør følgelig tilstrebes i den grad plasshensyn tillater. Et konkret eksempel fra russisk der brukseksempler kan støtte opp om tilegnelsen av fremmedartede konstruksjoner, er upersonlige setninger og passivekvivalerende konstruksjoner versus vanlig passiv i nordiske språk.

I en papirbasert ordbok er det viktig å lage mest mulig ”tettpakkede” brukseksempler, dvs. å mestre kunsten å ”slå flere fluer i én smekk”.

Brukseksempler er en av de viktigste komponentene i en adekvat tospråklig ordbok. De skal gi begge målgrupper verdifull informasjon, gi de kodede grammatiske og syntaktiske opplysningene ”kjøtt på beina”, demonstrere kombinabilitet, grammatiske egenskaper, karakteristiske ordforbindelser og kollokasjoner. I denne sammenheng er det viktig å understreke forskjellen mellom brukseksempler i tospråklige ordbøker og enspråklige ordbøker for morsmålsbrukere. I de sistnevnte går man ut fra at brukeren er godt kjent med det typiske, vanlige, og at dette ikke trenger noen utdypning. Tvert imot representerer brukseksempler

i en enspråklig ordbok gjerne perifere betydningsnyanser, spesielle, idiomatiske uttrykk som en vanlig bruker kan tenkes å trenge forklaring av. Det er fare for at en uerfaren ordboksforfatter ukritisk vil kopiere slike eksempler selv om de har lite å bestille i en tospråklig ordbok og i verste fall kan forvrengte resepsjonsbrukerens oppfatning av leksemet og dets funksjon. Dette var et aktuelt problem for oss redaktører under utarbeidelsen av "Stor norsk-russisk ordbok" der manusforfatterne, de fleste uerfarne, i hovedsak støttet seg til de enspråklige kildene *Norsk illustrert ordbok* og *Norsk bokmålsordbok*.

I en tospråklig ordbok er det derimot de typiske kombinatoriske sider ved både lemma og ekvivalent som må belyses, valens i vid forstand, herunder kasus- og preposisjonsbruk.

I ordbøker mellom beslektede og typologisk ensartede språk er problemet med grammatisk og kombinatorisk informasjon forholdsvis uproblematisk. Germanske (og til dels romanske) språk har verbale fortidsformer som i stor grad tilsvarende, de har alle grammatikalisert bestemthet. Likevel er brukseksempler viktige her nettopp for å belyse forskjeller i bruk av dette fellesgods, f.eks. bruk av bestemte former i svensk og engelsk, norsk og islandsk (jf. avsnitt 8). Hvor og hvordan skal kunnskap om slike fenomener formidles? I en (kortfattet) grammatisk skisse, i eksempler ved de mest aktuelle ordene eller (helst) begge deler? Til sammenligning kan det nevnes at selv om tysk gjengivelse av nordisk inkluderende perfektum og pluskvamperfektum er grundig behandlet i lærebøker, vil en oppfølging av dette i form av sporadiske brukseksempler være pedagogisk tjenlig.

Mange praktiske spørsmål er knyttet til bruk av eksempler: Skal man - mest av plasshensyn - prioritere pronomener som aktører i brukseksemplene? Hvis ja, hvilke? Hvilke verbformer er å foretrekke i brukseksempler, hvilke tidsformer?

I ordbøker mellom beslektede språk med lik grammatisk (og til dels lik syntaktisk struktur) er disse spørsmål mindre aktuelle. Likevel kan brukseksempler her spille en viktig rolle i klargjøring av eventuelle forskjeller i f.eks. bruk av bestemte former, av preteritum versus perfektum, o.m.a. Mer om dette senere.

### 3. Grammatiske kategorier

Det er mellom språk i ulike familier og/eller med ulik grammatisk og typologisk struktur at det stilles størst krav til forholdet mellom forutsatt grammatikkunnskap og formidling i selve ordboken. Her er det særdeles viktig at de forskjeller som det er vanskeligst for språkstudenten å beherske i mest mulig grad belyses vha. brukseksempler i ordboksartiklene, selv om de utgjør et selvskrevent innslag i de angjeldende lærebøker og grammatikker.

For de fleste slaver vil bestemthet og verbets mange fortidsformer i nordiske språk representere slike kategorier. Når det gjelder syntaks viser det seg f. eks.

at det er vanskelig for folk med slavisk morsmål å tilegne seg visse ufravikelige regler om omvendt nexus i skandinaviske språk.

For personer med nordiske språk som morsmål representerer til gjengjeld den slaviske verbalaspekten, retningskategorien (ofte upresist kalt for ”bevegelsesverb”) og uttrykk for bestemthet i språk der den ikke er grammatikalisert, et evig hodebry.

Mangeårig pedagogisk virksomhet har overbevist meg om at selv grundig teoretisk grammatisk opplæring ofte svikter når det kommer til praktisk anvendelse i tekstproduksjon. Derfor skal en avansert ordbok mellom slike språk hyppigst mulig minne brukerne om de vanskeligste forskjellene språkene imellom og demonstrere riktige løsninger. Men her er det mange skjær i sjøen.

#### 4. Synonymer

Ett av de problemer som gjør krav på spesiell oppmerksomhet, er behandlingen av synonymrekker.

Altfor ofte ser vi at et lemma A oversettes med en rekke synonymer på målspråket uten skikkelig spesifisering. Denne samme ramsen gjentas så som ekvivalenter til flere, kanskje alle av As synonymer i kildepråket.

For det første er 100% synonymer en sjeldenhet. Forskjellene kan gjelde betydningsnyanser, distribusjon, stilart o.m.a. Et ord (i en rekke av synonymer) i språk A som betydningmessig svarer nokså nøyaktig til en ekvivalent i språk B, kan ha vidt forskjellig hyppighet, stilverdi, o.m.a.

Adekvat behandling av synonymer i tospråklige ordbøker forutsetter tilgang til avanserte synonymordbøker i begge språk, der betydning, stilart, distribusjon osv. redegjøres for på en tilfredsstillende måte. Synonymordbøker der alle disse hensyn er tilstrekkelig ivaretatt i alle henseende, er dessverre en sjeldenhet. Her må vi konstatere at russisk leksikografi er blant dem som er kommet lengst på dette området (se forøvrig Haraldsson 1995).

#### 5. Slaviske versus nordiske verbformer i brukseksempler

Tilbake til morfologi. Som eksempel på problematikken kan vi ta følgende russiske setning med perfektivt preteritum av verbet ”komme opp, spire”:

(1) Огурцы взoшли (Ogurcý vzošl<sup>1</sup>)

Avhengig av sammenhengen kan dette oversettes med:

- a) Agurkene kom opp (aoristisk betydning)
- b) Agurkene er *eller* har kommet opp (perfektivisk nåtidsbetydning)
- c) Agurkene var *eller* hadde kommet opp (perfektivisk fortidsbetydning)

1 Ordtrykk markeres i transkripsjonen, som gis i parentes



- d) Agurkene hadde kommet opp (*f. eks. tidligere enn vanlig*) (pluskvamperfektum).

Et annet eksempel med aspektparet одеваться (odevát'sja, imperfektiv) og одеться (odét'sja, perfektiv) 'kle på seg':

- (2) она оделась тепло (oná odélas' tepló) (perf. preteritum)
- 'hun tok på seg varme klær' (aoristisk preteritum, handlingsbeskrivelse, ordrett: "hun kledde seg varmt (teplo) ")
  - 'hun har på seg {har tatt på seg} varme klær' (perfektivisk preteritum, tilstandsbeskrivelse)
  - 'hun hadde på seg {hadde tatt på seg} varme klær' (perfektivisk pluskvamperfektum)
- (3) она одевалась тепло (oná odeválas' tepló) (imperf. preteritum)
- 'hun pleide å ha {ta} på seg varme klær', dvs. 'hun gikk (vanligvis) varmt påkledd' (iterativt preteritum)
  - 'hun var i ferd med {holdt på å ta på seg varme klær' (durativt preteritum)

Det er innlysende at en oversettelse av disse og liknende brukseksempler aldri kan bli uttømmende.

I Valerij Berkovs prisbelønte *Russisk-norske ordbok* (RNO) har vi f.eks. under delbetydning 8 til lemmaet спустить (spustít') 'slippe (ut) o.m.a.' eksemplet: шина спустила (šína spustíla) 'ringen har gått ned {mistet lufta, er blitt flat}'. Her velger Berkov m.a.o. perfektivisk perfektum som ekvivalent, selv om betydningen like godt kunne vært 'ringen hadde gått ned {mistet lufta, var blitt flat}', dvs. perfektivt pluskvamperfektum, eller 'ringen gikk ned {mistet lufta, ble flat}', dvs. aoristisk preteritum.

Hvilke retningslinjer skal man så følge ved oversettelse av brukseksempler i preteritum, især perfektivt? Alternativene er:

- (4)
- å alltid bruke samme konstruksjon som ekvivalent
  - å velge subjektivt en ad hoc-konstruksjon i hvert enkelt tilfelle

I visse tilfeller peker ett alternativ seg ut, f.eks. vil betydningen av живого места на нём не осталось (živógo mésta na něm ne ostálos') (intransitivt verb) som oftest være perfektivisk 'han var mørbanket', mens живого места на нём не оставили (živógo mésta na něm ne ostávili) (transitivt verb) i de aller fleste tilfeller være aoristisk 'han ble mørbanket, de slo ham helseløs'.

## 6. Generelle retningslinjer for brukseksempler

Mange med forkjærlighet for klare retningslinjer vil foretrekke alternativ (4) a)

ovenfor som det metodisk mest akseptable. Jeg mener det er et pedagogisk tvilsomt standpunkt. Ordboksbrukeren bør i eksemplene få demonstrert hele spekteret av en konstruksjons betydninger, selvfølgelig ikke alle i hvert enkelt tilfelle, men valget av konstruksjon bør variere. Hyppige kongruensfeil hos nybegynnere i russisk av typen hankjønnsform av verb i preteritum når subjektet er hunkjønn, kan skyldes klar dominans av maskulinum i eksempler både i de lærebøker og ordbøker studentene er avhengige av.

Ordboksforlaget Russkij jazyk i Moskva hadde f.eks. som regel å prioritere det maskuline personpronomenet он (on) 'han' i brukseksempler. Dette var nærmest et påbud til ordboksforfattere, og mens undertegnede hadde kontrakt med dette forlaget under utarbeidelsen av Russisk-islandsk ordbok, fikk jeg reprimande for ikke å følge dette påbudet. Dette valgte jeg å ignorere ikke bare ut fra ønsket om mest mulig variasjon, men også av følgende grunn: Et vesentlig argument i favør av hunkjønnspronomen er at nominativ og akkusativ av pronomenet *hann* faller sammen i islandsk. *Hann langar* 'han vil, han ønsker' kan (og høyst sannsynlig vil) av en utlending oppfattes som vanlig nexus: subjekt (*nom.*) – verbal, mens *hana langar* 'hun vil, hun ønsker' røper med all ønskelig tydelighet at verbet er upersonlig og at det logiske subjektet står i akkusativ. Slike eksempler er det mange av i islandsk.

## 7. Medium. Passiv. Adjektiver

Mange russiske verb med postfikset -ся (-sja) kan ha flere betydninger. De kan evt. svare til

- s-form/medium av nordiske verb i passiv
- og/eller resiprok betydning
- refleksive verb

Man må være på vakt for å få med begge (evt. alle) betydninger, jf. verbet мазаться (mázat'sja) i RIO:

мазаться *нсв* 1. (*пачкаться*) *lita*<sup>3</sup>, *káma*<sup>3</sup>} frá sér 2. *св из-* (*пачкаться*) (*чем-л. - í e-u*) *atast*<sup>3</sup>, *makast*<sup>3</sup>, *kámast*<sup>3</sup>} út, *skitna*<sup>3</sup>, *verða*<sup>Δ</sup> *kámugur*<sup>5c</sup>; *ata*, *klína*, *sóða*, *maKa*<sup>3</sup>} sig út (*себя*) 3. *св на-, по-* (*T*) *bera*<sup>Δb</sup>, *smyrja*<sup>1b</sup>} á sig *A*; *mála*<sup>3</sup> sig (*краситься*).

I RNO kommer dette ikke klart til syne. I RED nøyer man seg med å oversette de to refleksive betydningene mens 3. delbetydning henviser til det transitive verbet:

3. *спрд.* к мазать 1, 2, 3,

dvs. *passiv* til delbetydning 1, 2 og 3.

Liknende sammenfall mellom personlig og inanimat referent i russisk preger visse adjektiver. I RNO trakteres adjektivet **neumélyj** slik:

неумелый *prl* ùðvd, ùerfaren (*неопытый*); klòsset(e), kèitet(e), klønet(e) (*неуклюжий*).

Av dette fremgår klart at de to første ekvivalentene refererer til personer, de øvrige kan også referere til inanimat referenter, både på russisk og på norsk. På islandsk derimot må man forklare forskjellige bruksfelt eksplisitt, jf. RIO:

неумелый *prl* klaufskur<sup>5</sup>; óreyndur<sup>5</sup> (*неопытный*); klaufalegur<sup>5c</sup> (*о движениях и т.п. 'om bevegelser o.l.'*).

Også i dette tilfelle oppviser RED en krevende korthet:

неумелый clumsy; (*о человеке - т.ж. 'om personer - også'*) unskilful.

Dvs. *clumsy* kan brukes både om personer og andre referenter.

Under arbeidet med RIO viste det seg at i tallrike tilfeller brukte islandsk et adjektiv på *-legur* om inanimat referenter, men et adjektiv (av samme rot) uten *-legur* om personer (og evt. andre animata), noe som krevde en permanent årvåkenhet i og med at russisk bruker samme adjektiv(er) i begge typer tilfeller.

## 8. En særegenhet ved bestemthet i islandsk

Jeg vil nå drøfte en særegenhet ved bestemthet i islandsk som etter min mening fortjener grundig behandling både i lærebøker i islandsk for utlendinger og i tospråklige ordbøker der islandsk er involvert.

Bestemthet i islandsk uttrykkes bl.a. vha. følgende konstruksjon:

- a) Tilstedsadverb + bevegelsesverb + preposisjon + stedssubstantiv uten artikkel.
- b) Påstedsadverb + handlings- eller eksistensialverb + preposisjon + stedssubstantiv uten artikkel.

Substantivene det her er tale om refererer til bestanddeler av nærmiljøet til en bestemt gruppe eller (mikro)samfunn. Når en slik gruppe mennesker omtaler sitt felles nærmiljø og alt dette miljøet består av, har fellesnavnet for hver av disse enhetene bare én bestemt referent, og i konstruksjoner av nevnte type brukes ubestemt form på linje med egnavn. Den adverbiale bestemmelsen fungerer som bestemthetsmarkør.

Når familiemedlemmer snakker om sitt eget kjøkken, heter det f.eks.: *hún er frammi í eldhúsi* 'hun er på kjøkkenet'; *hún fór fram í eldhús* 'hun gikk ut på kjøkkenet'; videre: *pabbi fór inn í stofu* 'far gikk inn i stua'; *Einar fór niður {er niðri} í*

*kjallara* 'Einar er gått ned {er nede} i kjelleren'; *bíllinn stendur úti á hlaði* 'bilen står på gårdsplassen'; *fundurinn verður uppi í samkomuhúsi* 'møtet blir holdt i samfunnshuset'; *það situr hrafn uppi á þaki* 'det sitter en ravn på taket (vårt)'; *krakkarnir eru að leika sér niðri í fjöru* 'ungene leker nede i fjæra'; *rollurnar stukku upp í hlið* 'sauene sprang opp i skråningen'; *þeir eru að vinna inni í frystiklefa {frammi í móttöku}* 'de driver og jobber inne i fryserommet {i (fiske)mottaket}'. I alle disse eksemplene er det tale om "vårt kjøkken", "vår fjære", "vårt samfunnshus" osv. Dette vil også gjelde i beretninger der narratøren betrakter et slikt miljø fra innsiden, f. eks. som gjest: *hún stóð upp og fór fram í eldhús* 'hun reiste seg og gikk ut på kjøkkenet'.

Uten adverbial bestemmelse kommer bestemt form til anvendelse: *við ætlum að mála eldhúsið* 'vi har tenkt å male kjøkkenet', *það þarf að gera við þakið hjá okkur* 'taket vårt må repareres' osv.

Denne gruppen substantiver omfatter det meste av et hus eller en leilighets interiør, alle rom og deres deler: gulv, vegg, vindu, vinduskarm osv; møbler som f.eks. bord, skap, kjøleskap o.l.; "individualiserte" deler av nærmeste omgivelser og arbeidsplass: tun, gate, eng, port, kolle, skråning, elv, bekk; fjøs, stall, låve; (kjøkken)hage; fjell, skog (sic!), o.m.a. Denne foreteelsen er på ingen måte marginal, den er hyppig og karakteristisk for dagligtalen og bør følgelig ikke ignoreres i lærebøker og helst ikke i ordbøker heller.

Det er vanskelig å anslå antallet slike substantiver, men det er i hvert fall flere titall av dem og gruppen er ikke klart avgrenset. Det ville være meget plasskrevende i en ordbok der islandsk er mål- eller kildespråk å gi eksempler på på denne uttrykksmåten for bestemthet ved hvert enkelt av disse ord. En mulighet er å velge et egnet ord som representant for denne gruppen substantiver etter prinsippet "en for alle" (jf. Svensén 1987 : 152-153) med henvisning til dette fra de øvrige. Dette prinsippet er et grep som bl.a. RIO, RNO og SNRO gjør utstrakt bruk av. Også denne måten å formidle denne islandske særegenheten på er problematisk pga. hvor uensartede de angjeldende substantiver er og hvor mange verb og adverbialer det kan være tale om.

Så vidt jeg har kunnet erfare, er denne bestemthetstypen i islandsk lite påaktet i lærebøker, enn si i ordbøker. Den nevnes i Einarsson (1945 : 115): "Unlike English, Icelandic often has no article with generic nouns, with place names that borders on proper nouns, with names of parts of the human body, parts of a house, parts of a ship, and with certain collective words." Blant Einarssons eksempler er *Frammi í stofu* '(out) in the parlor' men uten at sammenhengen med stedsadverbialet understrekes. Einarssons øvrige eksempler er mer idiompregede og isolerte.

## 9. Defekte verb i ordbøker

Så vidt jeg vet, gir ikke eksisterende islandske ordbøker (hverken enspråklige eller tospråklige) noen opplysninger om defekte islandske verb, skjønt dette er

av ikke ubetydelig praktisk betydning. At et ord har defekt bøyning, har lakuner i sitt paradigma, er en relevant opplysning både i enspråklige og tospråklige ordbøker.

Som eksempler på defekte islandske verb kan vi velge de tre homonyme verbene: *valda I*, *valda II*, *valda III*. Det første er et helt regelmessig verb og vi ser bort fra det. *Valda II* har betydningen 'forårsake' og bøyes sterkt, pres. 1. pers. *veld*, pret. *olli*. *Valda III* betyr 'greie, klare; løfte'. Det bøyes som *valda II* men mangler preteritumsformer i moderne islandsk. Man kan si *hún veldur (ekki) verkefninu* 'hun klarer (ikke) oppgaven', men det går etter mitt skjønn ikke an å si (i preteritum) *\*hún olli ekki verkefninu*. Et annet eksempel er verbet *látast* som i betydningen 'gi seg ut for; late som' er vanlig sterkt verb. *látast* i betydningen 'dø' savner derimod presensformer. Til gjengjeld mangler *ólátast* 'bråke' (om barn) preteritum og perfektum partisipp. Man tyr som regel til durativ: *börnin voru að ólátast úti í garði* 'barna boltret seg høylydt i hagen'.

Det er innlysende at opplysninger om slike forhold i høyeste grad er relevante for ordboksbrukere.

## 10. Konklusjon

Konklusjonen på ovenstående er følgende: Gjengivelse av og opplysninger om divergerende grammatiske forhold mellom språk i en tospråklig ordbok avhenger av en rekke faktorer og må løses ad hoc i hvert enkelt tilfelle. Slike faktorer er f.eks. ordbokens størrelse, språkernes slektskap og typologiske likheter, og – ikke minst – om boken er monodireksjonal eller bidireksjonal. Det er innlysende at en nettbasert ordbok har muligheter som aldri kan bli aktuelle i sin fulle bredde for en papirutgave med hensyn til løsninger på denne typen spørsmål.

## Anvendt og sitert litteratur

- Einarsson, Stefán 1945: *Icelandic. Grammar. Texts. Glossary*. Baltimore: The John Hopkins Press.
- Haraldsson, Helgi 1995: Plasshensyn i tospråklige ordbøker. I: *Nordiske studier i leksikografi III*, 185–194.
- Kress, Bruno 1982: *Isländische Grammatik*. Leipzig : Verlag Enzyklopädie.
- Новый объяснительный словарь синонимов русского языка* (ред. Ю. Д. Апресян). «Языки русской культуры». Moskva. Første hefte 1999, Annet hefte 2000.
- RED = *Russian-English Dictionary*. Moscow: «Soviet Encyclopaedia» Publishing House 1965.
- RIO = Haraldsson, Helgi 1996: *Rússnesk-íslensk orðabók*. Reykjavík: Nesútgáfan..
- RNO = Berkov, Valerij 1994: *Russisk-norsk ordbok*. Oslo: Universitetsforlaget.
- SNRO = Berkov, Valerij et al. 2003: *Stor norsk-russisk ordbok*. Oslo: Kunnskapsforlaget.
- Svensén, Bo 1987: *Handbok i lexikografi*. Esselte Studium.TNC. Stockholm.

228 HELGI HARALDSSON

Helgi Haraldsson  
Prof. Emeritus, f. 1938  
Otto Ruges vei 84 B  
NO-1361 Østerås  
helgi.haraldsson@ilos.uio.no

HENRIK HOVMARK

## Konventionaliserede forbindelser med danske retningsadverbier

### Leksikografisk repræsentation og funktion<sup>1</sup>

**Conventionalized expressions: lexicographic representation and function.** The semantics of Danish directional adverbs (DDA) (for instance *op* 'up', *ned* 'down', *ud* 'out', *ind* 'in') are closely connected to the semantics of other words, primarily verbs and prepositions (VB + DDA + PP). In this article, I explore the specific representation of conventionalized expressions with DDA, i.e. expressions which encode specific conceptualizations and semantic patterns, but which are not idiomatic or fixed. The representational solutions in two major monolingual Danish dictionaries, The Danish Dictionary and The Dictionary of the Insular Dialects, are described. Special attention is given to the possible lexicographic use potential of conventionalized expressions as linguistic items highly suitable for conveying information about semantic fields and typical use situations.

#### 1. Indledning

Danske retningsadverbier (RA) (*op/ned/ud/ind/hen/over/om/hjem/tilbage/..*) forekommer i tre typer: en dynamisk, en statisk og en processuel:

- |                 |                      |   |
|-----------------|----------------------|---|
| (1) dynamisk:   | <i>ud/-Ø</i>         | ( <i>han kørte <u>ud</u> til lufthavnen</i> ) |
| (2) statisk:    | <i>ud/-e</i>         | ( <i>mappen ligger <u>ude</u> på bordet</i> ) |
| (3) processuel: | <i>ud/-ad/-efter</i> | ( <i>vandet løb <u>udad</u> mod gaden</i> )   |

Og de kan i lighed med præpositioner betegnes som meget kontekstafhængige ord. Deres præcise betydning opstår i høj grad i et samspil med andre ord eller leksikalske enheder. RA indgår fx i skabelonen VB + RA + PP (fx *hun løb over til bageren* eller *kufferten stod ned på gulvet*) og udgør et vigtigt element i sprogbrugens beskrivelse af hvilke rumlige karakteristika der gør sig gældende i forholdet mellem to størrelser, nogen eller noget, i dansk. Med udgangspunkt i kognitiv lingvistik kan et sådant forhold beskrives som konceptualiseringen af relationen

<sup>1</sup> Denne artikel bygger på min ph.d.-afhandling: *Danske retningsadverbier og rumlig orientering* (Hovmark 2007). Afhandlingen er udarbejdet ved Ømålsordbogen, Afdeling for Dialektforskning, og rummer et leksikografisk tillæg, der dannede udgangspunkt for mit foredrag ved NFL-konferencen i Akureyri 2007. Denne artikel er en let tilpasset version af dette tillæg.

(Path) mellem en Figur (fx *hun* eller *kufferten*) og en Grund (fx *bageren* eller *gulvet*) (jf. Talmy 2000).

Betydningen og brugen af de tre typer af RA er fx knyttet til bestemte typer af VB. Fx kan modalverber ikke kombineres med statiske RA (fx *jeg skal op til Island* vs. *\*jeg skal oppe til Island*). Til gengæld kan den statiske type, i modsætning til den dynamiske og den processuelle type, forekomme som frit lokativ i sætninger hvor prædikatet både angiver en overgang (fx *hun gik ind i butikken henne i Nørregade*), en tilstand (fx *hun stod ude i haven*) og en aktivitet (fx *hun løb rundt ude i haven*) (jf. Harder et al. 1996:187).

RA indgår altså på den ene side i en række semantiske og syntaktiske forbindelser med VB og PP som kan beskrives nøjere. På den anden side er der ikke tale om egentlige faste forbindelser (selvom disse også forekommer i rigt mål). Hvordan kan eller skal RA i den slags forekomster behandles leksikografisk? Ordbøger, i hvert fald i deres klassiske papirudgave, arbejder normalt kun med to indgange: 1) lemma (i de allerfleste tilfælde et enkelt ord) og 2) sublemma (fast, ofte idiomatiseret forbindelse). De omtalte forbindelser med RA falder ofte et sted midt mellem. Hvornår skal de pågældende forbindelser – eller mønstre – oprettes som sublemma? I hvor høj grad er man nødt til at beskrive betydningerne under alle relevante elementer (fx *komme ud til NGN/NGT* under både VB *komme*, RA *ud* og P *til*)? I hvor høj grad og hvordan kan/skal der henvises mellem artikler (betydninger eller sublemmaer) i disse tilfælde (jf. Viehweger 1989:890)?

Jeg vil i det følgende afsnit kigge nærmere på den leksikografiske repræsentation af konventionaliserede forbindelser som *ud til fyret* eller *ned til stranden*. En konventionaliseret forbindelse vil være en kombination som ikke kun er frekvent, men som også er typisk, dvs. etableret som genkendelig og mundret. I mere tekniske termer kan man sige at den skal være “indlejret” (*entrenched*), og at både type- og token-frekvens spiller en rolle (jf. Croft & Cruse 2004:309). Konventionaliserede forbindelser som fx *ud til fyret* eller *ned til stranden* kan analyseres som repræsentative for en bestemt konceptualisering af forholdet mellem en rumlig formation og et (basalt) billedskema. Når en dansk sprogbruger skal angive relationen til en strand, vil han eller hun ofte vælge at sige *ned til stranden*. Der er imidlertid ikke tale om en idiomatiseret fast forbindelse, for valget vil altid være situationsbestemt, og afhængigt af situationen kan sprogbrugeren også vælge at sige fx *ud til stranden* eller *ind til stranden*.

Brugen af de forskellige RA er knyttet til bestemte situationstyper, dvs. til bestemte typer konceptualiseringer af rumlige ytringssituationer. Fx bruges *ud* særdeles hyppigt ved en rumlig situationstype der kan beskrives som ‘bevægelse mellem rum’ (fx *ud af stuen*, *ud i køkkenet*, *ud i gården*), dvs. hvor der overskrides en grænse mellem noget lukket til noget åbent i bevægelsen fra ét sted til et andet. De konventionaliserede forbindelser med RA der konceptualiserer ‘bevægelse mellem rum’, kan siges også samtidig at ekspliciterer de typiske rumlige ytrings-



situationer som den konceptuelle skematisering af 'bevægelse mellem rum' er knyttet til. Forbindelserne *ud af stuen*, *ud i køkkenet* og *ud i gården* kan siges ikke kun at konceptualisere 'rum' på det meget abstrakte plan ('BEHOLDER', 'LUKKET'-'ÅBEN'), men også de mere konkrete socio-kulturelt indlejrede situationstyper og ytringssituationer som den meget abstrakte konceptualisering er knyttet til (konkrete 'boliger', 'opholdssteder', 'pladser', 'indhegninger' som fx *stuer*, *køkkener* og *gårde*).

Jeg vil i det følgende prøve at se på de konventionaliserede forbindelser med RA i leksikografien, ikke kun som et problem, men også som en resurse: Hvad kan den slags forbindelser bruges til? Jeg vil tage udgangspunkt i eksempler hentet primært fra *Ømålsordbogen* (ØMO) og *Den Danske Ordbog* (DDO), dvs. store beskrivende videnskabelige ordbøger. Jeg vil begrænse mig til repræsentationen af RA i konkret rumlig betydning. Jeg vil altså ikke komme ind på den leksikografiske behandling af relationen mellem konkret rumlig betydning og de forskellige overførte betydninger (fx tid eller hierarki). Jeg vil til gengæld særligt interessere mig for hvordan en leksikografisk repræsentation kan tydeliggøre og udnytte den sammenhæng mellem abstrakte billedskemaer og typiske ytringssituationer som kan siges at komme til udtryk i de konventionaliserede forbindelser med RA.

## 2. Leksikografisk repræsentation af konventionaliserede, ikke-faste forbindelser

Viehweger (1989) nævner at ensproglige ordbøger traditionelt har underbelyst de semantiske relationer mellem leksikalske enheder som ikke kan siges at være faste forbindelser (og derfor oprettet som sublemma) – eller at de har behandlet dem usystematisk.

Angaben zum Kombinierbarkeit von Wortschatzelementen sind demgegenüber unterrepräsentiert und beschränken sich in der Mehrzahl der Fälle auf die Aufzählung mehr oder minder aussagekräftiger Beispiele oder Beispielgruppen bzw. authentischer Belege; durch sie sollen charakteristische Verwendungsweisen des jeweiligen Wortschatzelements deutlich gemacht werden. (Viehweger 1989:889)

Det er det der fx kan ses i artiklen *ud* fra en af de nyeste udgaver af *Nudansk Ordbog* (2001). Her følges en betydningsdefinition af syv mere eller mindre typiske brugseksempler. Den meget generelle definition svarer nogenlunde til en basal billedskemabetydning, nemlig billedskemaet BEHOLDER, dvs. konceptualiseringen af et inklusivt lukke og et omgivende større område – samt tilstedeværelsen og overskridelsen af en grænse mellem det mindre lukkede og det større åbne (jf. fx Croft & Cruse 2004:44f.).

- (4) **ud** [..] **1.** i retning fra det indre af noget mod noget mere åbent el. noget udenfor  
 [..]  
 □ *gå ud af stuen • gå ud ad døren • gå ud i køkkenet • vinduerne vender ud mod gaden • grenene strittede ud fra træets stamme • han hænger ud over gelænderet • han så ud over byen*  
 [..] (Nudansk Ordbog (2001), *ud*)

Betydningsdefinitionen baserer sig altså på en basal, kernesemantisk billedskemabetydning i det pågældende RAs rod ('ud'), og ud fra den betragtning er der ikke noget "forkert" i at nøjes med en leksikografisk repræsentation der består i en kernesemantisk definition og tilhørende eksempelmateriale under en eller anden form der illustrerer den kernesemantiske betydnings brugspotentiale. Hvor omfattende en ordbogsartikel skal være, er naturligvis også et spørgsmål om både tid, plads og brugergruppe, og det er klart at DDO, ØMO og ODS (Ordbog over det danske Sprog) som videnskabeligt beskrivende ordbøger generelt giver semantisk og grammatisk komplekse ord som fx RA en langt mere udførlig behandling end Nudansk Ordbog. Spørgsmålet er imidlertid om disse oplysninger kunne gives på en mere struktureret og oplysende måde end det er tilfældet i Nudansk Ordbog?

I leksikografisk sammenhæng kan man prøve at udskille det typiske og konventionaliserede på forskellig måde. Jeg vil i det følgende se på praksis i dels DDO, dels ØMO.

### 3. DDO: konventionaliserede forbindelser som særskilt oplysningstype

Under artiklen *ud* i DDO ses følgende betydningsbeskrivelse.

- (5) **ud** [..]  
**1** bort fra et nuværende el. hidtidigt opholdssted og hen til en position udenfor; bort fra det indre el. midten af noget · især om bevægelse bort fra et hus el. et rum; [..] □ *~ af rummet, ~ på gaden, ~ af stuen, ~ af huset* □ *"Nej, Bjarne, du kan ikke gå ud i det vejr!" "Sneen flyger, og man kan ikke se en hånd for sig!"* [..] (DDO, *ud*)

Som det fremgår, udskilles konventionaliserede forbindelser i DDO som en særskilt oplysningstype fra citater. Ved den pågældende betydning er anført fire typiske samforekomster (*~ af rummet* osv.). Disse konventionaliserede forbindelser er fundet på basis af statistisk søgning i det korpus der ligger til grund for udarbejdelsen af ordbogen, og derefter indplaceret under de relevante betydninger. Citaterne i ordbogen er derimod valgt af redaktøren – og her kan mange forskellige hensyn spille ind (jf. Lorentzen 2001). Den ovennævnte betydningsbeskrivelse er imidlertid særlig interessant fordi en del af beskrivelsen består i

en angivelse af den typiske brugssituation ved den pågældende betydning. Det der angives, kan som sagt beskrives som en typisk rumlig situationstype (*især om bevægelse bort fra et hus el. et rum*). Angivelsen af en bestemt ytringskontekst der er karakteristisk for brugen af et ord i en bestemt betydning, og som afspejles i de typiske samforekomster, kan komme til udtryk på forskellige måder og være mere eller mindre ekspliciteret i DDO. En af underbetydningerne til den nævnte betydning 1 af *ud* i DDO lyder fx sådan her.

- (6) **ud** [..]  
 1 [..]  
 • hen til et fremmed el. nyt tilholdssted, ofte for en længere periode; **JF**  
 væk □ *drage* ~, *rejse* ~, ~ *i verden* □ *Nordboerne opholdt sig ikke kun ved deres gårde i de varme fjordområder, de sejlede også ud på opdagelse og fangst*  
 (DDO, *ud*)

I underbetydningerne i DDO rykker angivelsen af ytringskontekst ofte op i selve betydningsdefinitionen, hvor den kan være integreret i selve definitionen eller markeret med et *fx, især, som regel* e.l. Det kan være særlig relevant netop ved kontekstafhængige ord som RA, hvis betydning hele tiden hænger snævert sammen med semantikken i det VB og PP de forudsætter og normalt er omgivet af. Det fremgår i eksemplet netop af de tre typiske samforekomster, hvoraf to er med VB og en med PP (*drage* ~, *rejse* ~, ~ *i verden*). På den måde fungerer den særskilte angivelse af typiske samforekomster som en vigtig eksplicitering af hvordan en brug af opslagsordet er bundet til en bestemt ytringskontekst og til bestemte sproglige udtryk i form af konventionaliserede forbindelser.

Konventionaliserede forbindelser må medtages og angives på basis af en vurdering af deres typicitet ved det enkelte ord eller den enkelte betydning. Det betyder at en meget prægnant samforekomst kan optræde enslydende ved flere forskellige opslagsord. Et klassisk eksempel må være *røde roser*. Associeringerne af den røde farve med roser og roser med den røde farve er begge stærkt konventionaliserede, og forbindelsen optræder da også under både *rød* og *rose* i DDO. Til gengæld giver det ikke mening at samle forbindelsen ét sted. Typiske samforekomster er netop ikke faste forbindelser (med eventuel idiomatisk eller overført betydning), men må anføres begge steder hvis kombinationen er typisk ved begge opslagsord.

Selvom den syntaktiske konstruktion ved RA i alle tilfælde er den samme, nemlig VB + RA + PP, kan det være forskelligt hvilke oplysninger om konventionaliseringer og typiske ytringssituationer der er medtaget under henholdsvis VB og RA. Ved VB *drage* genfinder man fx den typiske samforekomst *drage ud i verden* (jf. eksempel (6)).

(7) **drage**<sup>2</sup> vb.

[..]

2 rejse et sted hen; begive sig et sted hen · ofte længere væk el. for længere tid; **JF** tage; [NGN drager +RETNING, NGN drager af\_sted/bort/videre/...,hjælpevb.: være] □ ~ udenlands/sydpå, ~ på turné, ~ ud i verden, ~ fra sted til sted, ~ i krig □ Til august drager vi alle fire til Bangkok. Vi bliver der et år [..] (DDO, drage II)

Betydningsbeskrivelsen er den samme som før, denne gang dog med særskilt markering af ytringskonteksten (*ofte længere væk el. for længere tid*). Men man finder også andre typiske samforekomster eftersom opslagsordet nu er et andet. Forbindelsen mellem *drage* og *ud* er ikke en leksikaliseret fast forbindelse, men en typisk samforekomst som er relateret til en bestemt ytringskontekst ('rejse til et sted, langt væk og/eller i længere tid'). Ved *drage* gives derudover en konstruktionsoplysning – i modsætning til *ud* er den syntaktiske skabelon (VB + RA + PP) her antydet (NGN drager +RETNING).

Som det fremgår af eksemplet, tager DDO på flere forskellige måder højde for og inddrager den bagvedliggende skabelon VB + RA + PP, både i betydningsdefinitioner, valensbeskrivelser og typiske samforekomster. Det fremgår måske endnu mere tydeligt hvis man tager en forbindelse hvor der ikke er fundet typiske samforekomster i korpus. Det er fx tilfældet ved forbindelsen NGT flyder ud over NGT – her dukker skabelonen VB + RA + PP op på forskellig måde, helt eller delvis.

(8) **flyde** hovedbetydning:

1. brede sig og glide forholdsvis langsomt hen over et område  
[NGT flyder + ADVL]

**flyde** sublemma:

flyde ud

**ud** sublemma:

ud over

Noget af forbindelsen NGT flyder ud over NGT (VB + RA + PP) bliver beskrevet som særskilte sublemmaer, noget som del af hovedbetydningen under *flyde*.

1. *flyde ud* er udskilt, dog formodentlig ikke kun fordi der er tale om en særlig brug af hovedbetydningen (uden PP), men måske mere for at kunne gøre opmærksom på dannelsen *udflydende* der er med som opslagsord i DDO, og for at kunne angive en overført betydning (*flyde ud (i ét/hinanden)* 'gå i ét; ikke længere kunne adskilles af ens sanser; blive uskarp at se på').
2. *ud over* er udskilt som en særlig karakteristisk forbindelse, uden tvivl fordi forbindelsen forekommer i flere funktioner og betydninger – og fordi forbindelsen også findes sammenskrevet (*udover*, medtaget som

opslagsord i DDO). Foruden den rumlige betydning ('hen over en overflade, et areal e.l.'; 'hen over en forhindring, kant el. grænse (og ned)'), findes betydninger som 'som noget ekstra el. yderligere i forhold til noget allerede eksisterende el. forekommende' (fx *ud over mælk købte jeg brød, pålæg, kødvarer, lys og servietter*) og 'bortset fra' (fx *Det [: at blive kendt] har ikke forandret noget for mig ud over, at det er blevet nemmere at hæve i banken uden at vise ID-kort*). Det interessante i denne sammenhæng er betydningsdefinitionerne under de rumlige betydninger. De medtager nemlig det følgende led i skabelonen, nemlig PP (henholdsvis 'en overflade, et areal e.l.' og 'en forhindring, kant el. grænse'). I en meget streng definition af betydningsdefinitioner som værende ideelt set udskiftelige (jf. Svensén 1987:111 og 121 eller Bergenholtz et al. 1997:92), burde disse elementer ikke være med. På den anden side er de en næsten nødvendig hjælp til brugeren, som ville være på næsten bar bund med hensyn til at forstå og skelne mellem de to betydninger hvis kun den udskiftelige del var anført ('hen over' vs. 'hen over (og ned)').

3. Det samme er tilfældet under verbets betydningsdefinition hvor både RA og PP medtages ('brede sig og glide forholdsvis langsomt hen over et område'). Hertil kommer valensmønstret ([NGT flyder + ADVL]), der også viser skabelonen VB + RA + PP, de to sidste dele dog kun indirekte, som del af overkategorien ADVL.

Samlet kan man sige at konventionaliserede forbindelser i form af typiske samforekomster faktisk fremhæves i DDO, og at der i hvert fald indirekte knyttes en forbindelse mellem angivelsen af ytringssituation i betydningsdefinitionen og de typiske samforekomster der rent faktisk koder den ytringsrelaterede betydning sprogligt på ytringsniveau. Men man må samtidig sige at markeringen er temmelig diskret. Man kunne nemt have markeret oplysningstypen med et TYPISK e.l. Mere interessant er det om det ville være muligt at markere angivelsen af ytringskontekst eller situationstype mere eksplicit i betydningsdefinitionen? En mulighed var at begrænse sig til én markør (fx *især*) og konsekvent anbringe oplysningstypen efter den lille sorte prik.

Det er klart at markeringen af ytringskontekster også kan ske på andre måder. I Brugervejledningen til DDO (I:13–52) findes bl.a. en oversigt over forskellige "brugsbestemmelser" (I:27–29 og 32–35). Disse omfatter kodificeringer af fx stilleje (fx *spøgende, forskønnende, ironisk*) og teksttype (fx *især i sportsjournalistik, især i juridiske tekster*). En anden oplysningstype er særlig interessant, nemlig "kommunikationssituation" (I:34) der fx omfatter markeringerne *bruges ofte som skældsord* og *især i tiltale*. I eksemplet er der imidlertid anført en noget længere beskrivelse.

(9) **levvel** sb. itk.

[..]

afskedshilsen hvor man ønsker det bedste for en person idet man siger ordene: *lev vel (ofte i forb. med at nogle skilles for en længere periode el. for altid)* **JF** farvel (DDO.I:34)

Denne særlige oplysningstype er blevet udnyttet meget lidt i DDO, bl.a. fordi oplysningstypen retter sig mere snævert mod angivelse af en bestemt diskursiv samtalsituation. Men det er alligevel tankevækkende at DDO i øvrigt er spækket med oplysninger om ytringssituationer i bredere forstand. De står bare oftest i tilknytning til selve betydningsdefinitionen. Det er fx tilfældet ved underbetydningen af *ud*.

(10) **ud** [..]

1 [..]

• hen til et fremmed el. nyt tilholdssted, *ofte for en længere periode*; **JF** væk □ *drage ~, rejse ~, ~ i verden* □ *Nordboerne opholdt sig ikke kun ved deres gårde i de varme fjordområder, de sejlede også ud på opdagelse og fangst* (DDO, *ud*)

Man kan vel godt tillade sig at sige at både *ud* i den pågældende betydning og *levvel* tilhører samme semantiske felt og ytringssituation. En person der *drager ud i verden* kan ytre et *levvel* inden afrejsen. Man kan hævde at DDO er uafklaret i sin praksis med hensyn til hvordan en eventuel markering af ytringskontekst skal tage sig ud, og hvor den skal stå, men eksemplet viser også at det kan være vanskeligt præcist at indkredse og definere en specifik ytringssituation.

#### 4. ØMO: konventionaliserede forbindelser som del af definitionen

Konventionaliserede forbindelser kan fremhæves i ØMO på flere forskellige måder. Jeg skal i det følgende skitsere en typisk praksis ved dels kombinationer af RA + PP, dels VB + RA. Især den sidste model er af interesse i forhold til eksemplerne fra DDO.

ØMO baserer sig primært på en seddelsamling (udskrevne båndoptagelser inddrages dog i det omfang det er muligt, jf. Gudiksen & Hovmark, denne publikation). En korpusbaseret statistisk fremhævelse af typiske samforekomster er derfor ikke mulig. ØMO er snarere tilbøjelig til at fokusere på en semantisk afgrænsning af det typiske ved en betydning eller underbetydning. I lighed med eksemplet *ud* fra DDO, angives der ganske ofte typer af konceptualiseringer og ytringssituationer, fx i artiklen *ind*.

(11) **ind** adv [..]

–**1.1**) i forb i n d e i n g t (i et rum, en bygning, en bebyggelse, en beplantning, et landskab, et farvand mm; mht i en “beholder” jf bet 3)

Ø(alm): *ellers så havde vi dem* (duerne) *inde i loen* svS(SlagLd), *gården lå inde i byen og al jorden den lå ude omkring* nS(Karl) (ØMO, ind)

Til gengæld vil typiske samforekomster normalt ikke være fremhævet, men "kun" forekomme i de autentiske citater (fx *inde i loen* og *inde i byen*). Men som i DDOs tilfælde kan citater være udvalgt af mange grunde, og der er derfor ikke nogen sikkerhed for at typiske samforekomster systematisk fremhæves i ØMO på samme måde som i DDO. Man skal i den forbindelse også være opmærksom på at fastere konventionaliserede forbindelser (kollokationer) i ØMO heller ikke nødvendigvis er fremhævet som selvstændigt opslag (dvs. anført spatieret og indledt med *i forb*, som fx *i forb hist og her* under opslagsordet *hist*). Det kommer helt an på frekvens og på redaktørens vurdering af om det ud fra materialet er muligt at sige noget sikkert om samforekomstens status (hos den enkelte meddeler og i den enkelte dialekt). Citaterne i ØMO kan altså godt rumme både typiske samforekomster og kollokationer.

Typiske samforekomster kan imidlertid godt være fremhævet i ØMO. Det ses fx i artiklen *historie*.

(12) **historie** [..]

–1) fortælling om en virkelig ell opdigtet begivenhed; *ofte* i forb *fortælle en historie* (ell *historier*) [..]

–2) forløb, begivenhed, affære (især om ubehageligt ell besværligt forløb), *ofte* i forb *som en dyr, værre, kedelig historie* [..]

–3) om virksomhed, foretagende, bedrift, arbejde olgn, *især* i forb *en stor (større) historie* [..] (ØMO, *historie*)

Som det fremgår, har man i ØMO mulighed for at anføre de sproglige udtryk – de konventionaliserede forbindelser i form af typiske samforekomster og/eller eventuelle kollokationer – som en del af betydningsdefinitionen, og der er derfor en mere direkte sammenhæng mellem konventionaliseringer og betydningsbeskrivelse i den leksikografiske repræsentation. Modellen er i virkeligheden udviklet for at kunne gengive valensoplysninger (i lighed med DDO), og ses derfor nok oftere ved forbindelser med VB, men som det fremgår, kan der også indsættes typiske samforekomster eller kollokationer på denne plads i artikelstrukturen. Jeg vil i det følgende se på nogle eksempler med RA. Man skal i det følgende være opmærksom på at *forb* i ØMO bruges om alle fraserer, både konventionaliserede forbindelser, kollokationer og idiomer.

(13) **hen** adv [..]

–3.3) hvor målet anses for irrelevant; *især* i udtryk for at skille sig af med ngt, lægge ngt til side (til opbevaring)

|| *i forb som forpagte, fæste, give, love (ngt, én, sig) hen*  
dvs ud af ens besiddelse ell rådighed [..]

|| *i forb som gemme, hænge, lægge, sætte ngt hen* dvs til

side, af vejen [..]

|| i forb som k o g e, s a l t e, s t e g e n g t h e n koge, salte, stege ngt hen koge, salte, stege ngt (så det kan holde sig) og gemme det hen (til opbevaring) [..]

|| jf u de resp vb. (ØMO, *hen*)

I det anførte eksempel indledes der med en overordnet betydningsdefinition som både rummer kernesemantiske træk (*skille sig af med, lægge til side*) og ytringsrelaterede træk (*til opbevaring*). Ligesom i DDO er der ikke tale om en streng udskiftelig definition, snarere om en form for parafrase der ekspliciterer en praksis som er forbundet med den pågældende betydning og brug af *hen*. Betydningsdefinitionen inddrager med andre ord semantikken i alle elementerne i skabelonen VB + RA + PP (hvoraf det sidste element, PP, netop ikke ekspliciteres i de pågældende betydninger). Efter definitionen følger de forbindelser, dvs. de sproglige kontekster, hvor den overordnede betydning kommer til udtryk. De fordeler sig i tre klumper der fremstår som tre mere specifikke anvendelsesområder hvortil der knytter sig et mindre antal konventionaliserede forbindelser. Disse forbindelser er netop anført som en del af selve underopslaget, der på den måde kommer til at fremtræde som en slags "klynge-sublemma" med samme betydning og fælles, lignende ytringssituation (fx *ud af ens besiddelse ell rådighed*).

Som nævnt er ØMOs behandling af forskellige typer samforekomster undertiden vaklende, og nogle af de anførte forbindelser under *hen* 3.3 må nok betegnes som kollokationer snarere end typiske samforekomster (fx *koge/salte/stege NGT hen*). Jeg vil i det følgende imidlertid ikke gå nærmere ind i denne problematik, men i stedet se på ØMOs praksis med at anbringe konventionaliserede forbindelser direkte i betydningsdefinitionen som en måde at knytte en forbindelse mellem kernesemantik og ytringskontekst i den leksikografiske repræsentation.

Normalt beskrives de nævnte forbindelser under både VB og RA i ØMO. Som eksempel kan nævnes den første underbetydning under *hen* 3.3. Denne betydning er fremhævet på tilsvarende måde under de nævnte VB, dvs. *forpagte, fæste* og *give* (ØMO er endnu ikke nået til *love*). Her bruges opslagstypen til at anføre valensoplysninger (*ngt/én/sig*) og til at angive den leksikalske variation på RAs plads (*bort/hen/ud/væk*).

- (14) **forpagte** v [..] i forb som f o r p a g t e n g t b o r t, h e n, u d leje ngt bort, hen, ud (til ngn) (ØMO, *forpagte*)
- (15) **fæste** v [..] –2.1 [..] i forb som f æ s t e é n (ell s i g) b o r t, v æ k, h e n få én i tjeneste, tage tjeneste hos fremmede (ØMO, *fæste*)
- (16) **give** v [..] –2 overdrage, tildele, forære [..]  
|| i forb g i v e n g t b o r t, h e n, v æ k (ØMO, *give*)

Der sker altså i høj grad en dobbeltbehandling af en række temmelig enslydende og stabile sproglige forbindelser. Kunne de i virkeligheden behandles samlet ét sted, med henvisning fra andre relevante steder? En sådan model følges faktisk undertiden i ØMO, fx under en anden underbetydning til *hen* 3.



- (17) **hen** adv [...] –3.2) hvor man vægrer sig ved at nævne målet || i forb *bære, forsyne, følge, koste* (én) *hen* (til graven) Ø(vist alm, jf *u bære* 11.4, *forsyne* 6, *følge* 1.5, *koste*): *jeg var med til at bære ham hen* Lgl(Tul), *vi skal lige have vor mor forsynet hen først* Fa(NVed) [...] (ØMO, *hen*)
- (18) **følge** III v  
 –1) bevæge sig i samme retning som ell sammen med (en anden ell andre)  
 –1.5) i spec anv: være ledsager ved barnedåb, kirkegang, bryllup, begravelse [...] || ledsage en afdøds kiste til graven og overvære begravelsen [...]; i forb *følge lig* [...]; i forb *følge hen* [...]; i forb *følge til jorde* [...]; i forb *følge til jorden* [...]; i forb *følge til grave* [...]; i forb *følge til graven* [...] (ØMO, *følge* III)

En henvisningsløsning er imidlertid ikke altid den bedste. Der er ganske enkelt for store og mange variationsmuligheder. Det kan være meget svært at finde ét enkelt sprogligt udtryk der kan siges at være den faste kerne. I forbindelserne med *hen* er der også variation på RA-pladsen (*hen, bort, væk, ud*, jf. eksempel (14)–(16)) – og semantisk kan RA-pladsen også erstattes af andre adverbialer, fx *af vejen, til side, afvæk*. Samtidig er der behov for at kunne gengive netop de semantiske træk og konventionaliserede variationsmuligheder der knytter sig til det specifikke opslagsord, i dette tilfælde fx brugen af *følge* ved flere forskellige slags kirkelige ceremonier (ikke kun begravelse) og variationen i de konventionaliserede forbindelser med *følge* i den pågældende betydning (*følge lig, følge hen, følge til jorde* osv.) (jf. samme forhold mellem *drage* og *ud* i DDO).

Henvisninger er dog ikke ualmindelige og ses også ved typiske samforekomster. Det kan ske i form af en generel henvisning (i det følgende eksempel til de nævnte VB i de forskellige betydninger).

- (19) **hen** adv [...] –3.3) hvor målet anses for irrelevant; især i udtryk for at skille sig af med ngt, lægge ngt til side (til opbevaring)  
 || i forb som *forpakte, fæste, give, love* (ngt, én, sig) *hen* dvs ud af ens besiddelse ell rådighed [...]  
 || i forb som *gemme, hænge, lægge, sætte* ngt *hen* dvs til side, af vejen [...]  
 || i forb som *koge, salte, stege* ngt *hen* koge, salte, stege ngt (så det kan holde sig) og gemme det hen (til opbevaring) [...]  
 || jf *u de resp vb.* (ØMO, *hen*)

Eller det kan ske i form af en mere direkte henvisning.

- (20) **ind** adv [...] –2.1) i udtryk for bevægelse udefra til et sted indenfor i en bygning, et rum i en bygning ell lign [...]  
 i forb *binde* (II.3.4), *lede, lukke, slå, tage* *kørne* (kreaturet olgn) *ind* (for natten ell for vinteren) *se u de resp vb* og

jf *indbinding, indstalde* (1), *indvintre*; [..]  
 i forb *k ø r e* (k o r n, h ø o l g n) i n d *se u k ø r e*;  
 i forb *f l y t t e, f ø r e, t a g e* → *h æ s* (I.1) ell → *s t a k i n d* [..] (ØMO,  
*ind*)

Det er klart at jo mere fast en konventionaliseret forbindelse er, jo lettere og mere oplagt er det at vælge en ren henvisningsløsning. Men man skal samtidig være opmærksom på et særligt hensyn som ØMO er nødt til at tage. ØMO beskriver en lang række dialekter og er forpligtet til at angive udbredelsen af de ord og betydninger der medtages i ordbogen. Fx har forbindelses-klyngen *binde, lede, lukke, slå, tage kørerne (kreaturet) ind* i eksempel (20) ovenfor forskellig udbredelse, og derfor er det under alle omstændigheder nødvendigt at lave en leksikografisk løsning hvor de enkelte forbindelsers udbredelse kan angives (dvs. under de respektive verber). Uanset om en redaktør måtte foretrække en dobbeltbeskrivelse uden henvisning eller en gennemført henvisningsløsning, vil hensynet til udbredelsesangivelsen kunne spille ind og indebære at en henvisning under alle omstændigheder er nødvendig.

Dobbeltbeskrivelser af konventionaliserede forbindelser er ikke nødvendigvis af det onde. En vigtig funktion i en almensproglig videnskabelig beskrivende ordbog er at vise det enkelte ords semantiske kombinationsmuligheder og tilknytningen til forskellige ytringskontekster. Især den sidste funktion er måske værd at hæfte sig ved, ikke mindst hvis man har et funktionalistisk sprogsyn, for så er koblingen mellem 1) kernesemantiske betydninger og betydningsdefinitioner og 2) kommunikationssammenhænge og ytringssituationer vigtig. Alle underbetydningerne under betydning 3 af *hen* knytter fx en forbindelse mellem 1) en særlig kernesemantisk betydning og brug af *hen* hvor målet er i baggrunden konceptuelt og ofte ikke ekspliciteres (undertiden af pragmatiske grunde, jf. *vægrer sig ved at nævne ell finder det irrelevant at nævne*), og 2) forskellige klynger af konventionaliserede forbindelser, dvs. sproglige udtryk.

- (21) **hen** adv [..] –3) uden ang af mål; i tilfælde hvor målet fremgår af kontekst ell situation, ell hvor man vægrer sig ved at nævne det ell finder det irrelevant at nævne, ell hvor målet ikke er kendt ell ikke er nærværende for tanken (ØMO, *hen*)

Det eventuelle konstante element i de anførte eksempler på brug af *hen* ligger i denne meget generelle betydning som er parallel med fx *bort, væk* og *ud*, og som består i et konceptuelt fokus på at noget eller nogen ikke (længere) er (aktivt) til stede, men (nu) befinder sig et andet sted. Denne konceptualisering knytter sig til forskellige ytringskontekster med dertil hørende konventionaliserede forbindelser.

Man har med ØMOs model mulighed for at knytte en mere direkte forbindelse mellem sproglige udtryk og ytringssituationer end det er tilfældet i DDO.

Og angivelser af ytringskontekster, pragmatisk-kommunikative såvel som socio-kulturelle, er ofte udførlige (det sidste hænger sammen med den vægt ordbogen lægger på encyklopædiske oplysninger, jf. Pedersen 1994). Til gengæld er ØMO upræcis eller uafklaret i sin skelnen mellem og markering af forskellige typer frasemer og i det hele taget i sin markering af forskellige oplysningstyper, og det skyldes bl.a. at ordbogens redaktionsprincipper blev formuleret før redigering i databasebaserede programmer blev almindeligt inden for den praktiske leksikografi – en sådan redigeringspraksis kræver at man foretager en mere stringent definition af bestemte oplysningstyper. Datamaterialets karakter (seddel-samling, excerpering, optegnelse osv.) giver som sagt heller ikke mulighed for at lave statistisk baserede frekvens- og typicitetsundersøgelser.

## 5. Afslutning

Jeg har i denne artikel kigget nærmere på brugen af konventionaliserede forbindelser i leksikografien, med forbindelser med danske RA som eksempel, og jeg har set på hvordan disse forbindelser repræsenteres i henholdsvis DDO og ØMO. Konventionaliserede forbindelser forekommer ofte i form af typiske eksempler i ordbøger, dvs. som en hjælp til brugeren. Men også i det redaktionelle arbejde kan disse forbindelser måske være nyttige fordi de peger på centrale semantiske felter i et ords betydninger. Moderne leksikografi udnytter i stadig stigende grad store tekstkorpora i det redaktionelle arbejde, og statistiske værktøjer der tager højde for både type- og token-frekvens, giver netop mulighed for at finde konventionaliserede forbindelser der kan udnyttes på forskellige måder (jf. fx Stubbs 1996, Gries & Stefanowitsch 2004), ikke kun til at finde de mest typiske eksempler på et ords brug, men også til semantisk og leksikografisk strukturering af et ords betydninger.

## Ordbøger

*Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal 2003–2005.

*Nudansk Ordbog*. København: Politikens Forlag 2001.

*Ordbog over det danske Sprog*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal 1919–1956.

*Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer*. 1–. København: Institut for Dansk Dialektforskning og C.A. Reitzels Forlag 1992–.

## Anden litteratur

Bergenholtz, Henning et al. 1997: *Nordisk leksikografisk ordbok* (Skrifter utgitt av Nordisk forening for leksikografi 4). Oslo: Universitetsforlaget.

- Croft, William & Cruse, D. Alan 2004: *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Gries, Stefan Th. & Stefanowitsch, Anatol 2004: Covarying Collexemes in the *into-causative*. I: Achard, Michel & Kemmer, Suzanne (udg.): *Language, Culture and Mind*. Stanford: CSLI Publications, 225–236.
- Gudiksen, Asgerd & Hovmark, Henrik, denne publikation: Båndoptagelser som kilde til Ømålsordbogen.
- Harder, Peter & Heltoft, Lars & Nedergaard Thomsen, Ole 1996: Danish directional adverbs. Content syntax and complex predicates: A case for host and co-predicates. I: Engberg-Pedersen, Elisabeth et al. (udg.): *Content, Expression and Structure. Studies in Danish functional grammar*. Amsterdam & Philadelphia: John Benjamins, 159–198.
- Hovmark, Henrik 2007: *Danske retningsadverbier og rumlig orientering* (ph.d.-afhandling). København: Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet.
- Lorentzen, Henrik 2001: Jagten på det gode citat. I: Gellerstam, Martin et al. (udg.): *Nordiske studier i leksikografi 5* (Meijerbergs arkiv för svensk ordforskning 27). Göteborg: Nordiska föreningen för leksikografi & Meijerbergs institut för svensk etymologisk forskning, 202–216.
- Pedersen, Karen Margrethe 1994: Ordbog og encyklopædi. I: *Nordiske studier i leksikografi 2* (Skrifter udgivet af Nordisk Forening for Leksikografi 2). København: Foreningen af Leksikografer i Danmark, 248–256.
- Stubbs, Michael 1996: *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture* (Language in Society 23). Oxford: Blackwell.
- Svensén, Bo 1987: *Handbok i leksikografi. Principper och metoder i ordboksarbetet*. Stockholm: Esselte Studium & Tekniska nomenklaturcentralen.
- Talmy, Leonard 2000: *Toward a Cognitive Semantics I–II*. Cambridge, Massachusetts & London: MIT Press.
- Viehweger, Dieter 1989: Probleme der Beschreibung semantischer Vereinbarkeitsrelationen im allgemeinen einsprachigen Wörterbuch. I: Hausmann, Franz Josef et al. (udg.): *Wörterbücher. Ein Internationales Handbuch zur Lexicographie* (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1). Berlin & New York: Walter de Gruyter, 888–893.

Henrik Hovmark  
 adjunkt, f. 1966  
 Ømålsordbogen  
 Nordisk Forskningsinstitut  
 Afdeling for Dialektforskning  
 Københavns Universitet  
 Njalsgade 136, 5  
 DK-2300 København S  
 hovmark@hum.ku.dk

ANN-KRISTIN HULT

## Från ord till handling

### En studie i ordboksanvändning på nätet

The article presents the results of a log file analysis of the searches in *Lexin Svenska ord*, a free Swedish monolingual Internet dictionary. It focuses on the non-successful searches, called not-found. The results show, firstly, that a considerable amount of the not-found are correctly spelled Swedish words, but they are not included in the lemma selection of the dictionary. Users also often search for two (or more) word expressions. Secondly, the dictionary's programmed response in the case of not-found gives far too many suggestions for it to be user-friendly; sometimes the user is given as many as forty suggestions on words that he or she could have been searching for. Thirdly, users tend to repeat their searches in order to obtain a successful search, e.g. by trying another spelling. It is shown that log files can be a helpful tool in the lexicographers' work improving electronic as well as paperback dictionaries.

#### 1. Inledning

Som en följd av att ordböcker har blivit tillgängliga i elektroniskt format har möjligheten uppstått att registrera alla sökningar som utförs i så kallade loggfiler. Via de automatiskt registrerade loggfilerna är det möjligt att få tillgång till stora material, i detta fall användarnas sökningar i en Internet-ordbok. Vid traditionella användarstudier av pappersordböcker är informanterna färre och brukarsituationen (oftast) konstruerad. Via loggfilerna kan man få reda på vilka ord användarna faktiskt slår upp och, då mer avancerade sökningar är möjliga, vilken typ av information de är ute efter. Det är således dels ett relativt nytt medium som undersöks, Internetordböcker, dels ett nytt verktyg som används för att studera ordboksanvändning, loggfiler. Befintliga studier visar också att loggfiler kan fungera som ett effektivt hjälpmedel i arbetet med att förbättra såväl pappersordböcker som elektroniska ordböcker.

I denna artikel presenteras resultaten av en studie av loggfilerna från den svensk-svenska delen av *Lexin*-seriens ordböcker, *Svenska ord*, som ligger fritt tillgänglig på Internet. Syftet med studien har varit att ta reda på vilka typer av ord som användarna söker efter och hur de gör då de inte får någon träff på det ord som efterfrågas. Detta undersöks med hjälp av de uppemot 17 000 registrerade sökningarna i loggfilerna från den första oktober 2006. Andelen icke framgångsrika sökningar visar sig vara mycket hög, hela 41 %. Det är därför

intressant att undersöka vad användarna vill men inte kan få reda på, samt vilka strategier användarna verkar ta till då de inte får någon träff. Studien avgränsas således till de icke framgångsrika sökningarna, hädanefter kallade icke-träffar. Målet med undersökningen har varit att få kunskap om användarnas faktiska sökbeteende och därigenom också vad de förväntar sig att (e-)ordboken ska innehålla.

Artikeln är disponerad på följande sätt: i kapitel 2 redogör jag för *Lexin*-serien och i kapitel 3 ger jag en inblick i forskningen kring ordboksanvändning där jag också behandlar tidigare loggfilsstudier. I kapitel 4 presenteras materialet. Resultaten redovisas sedan i tre avsnitt under kapitel 5. Till sist sammanfattas resultaten i kapitel 6 där jag också blickar framåt mot vidare användarforskning.

## 2. *Lexin*-serien

*Lexin*-serien består för det första av en enspråkig svensk ordbok, *Svenska ord*. Den har tagits fram av Myndigheten för skolutveckling tillsammans med Språkdata under ledning av Martin Gellerstam (se vidare Gellerstam 1999). För det andra ingår ett trettiotal tvåspråkiga lexikon på stora minoritetsspråk i Sverige. *Svenska ord* har fungerat som underlag vid framställning av de tvåspråkiga delarna av ordboksserien.

*Svenska ord* har 28 500 uppslagsord. Ordurvalet är gjort med tanke på invandrare som ska lära sig svenska och få tillgång till vanliga ord samt samhällsord som är viktiga för att man ska kunna orientera sig i det svenska samhället. I förordet påpekas det dock att *Svenska ord* också kan användas som skolordbok. I övrigt innehåller ordboksartiklarna språklig information kännetecknande för en definitionsordbok (t.ex. definitioner, uttalsangivelser, böjningsformer och språkexempel), men också uppgifter av mer grammatisk karaktär (t.ex. satskonstruktion). I den tredje upplagan (1999) har samhällsorden uppdaterats och ord som har med datorer och informationsteknologi att göra har lagts till (*Svenska ord* 1999:V).

Internetversionen av *Lexin*-serien har funnits sedan 1994 och distribuerades fram till och med 2006 av Myndigheten för skolutveckling i samarbete med KTH (Kungliga Tekniska högskolan i Stockholm). Från och med den första januari 2007 tillhandahålls tjänsten i det nya Språkrådets regi. I dagsläget finns femton lexikon fritt tillgängliga via webbsidan. Bild 1 visar webbsidans utseende efter en sökning på ordet *nyckel* i *Svenska ord*.

The screenshot shows the Lexin online dictionary interface. The header includes 'Språkrådet' and 'Institutet för språk och folkminnen'. The main content area is titled 'Svenska ord' and 'Översättning av nyckel'. The entry for 'nyckel' includes its phonetic transcription [n'yk:el], a pronunciation icon, and a definition: 'nyckeln nycklar subst. redskap att öppna el. stänga lås med (bildligt "verksamt medel, lösning"; i sms även bildligt "viktig, central")'. It also provides a link to 'bildtema 10'. Below the definition, there are sections for 'Exempel' and 'Sammansättningar/avledningar', each with a bulleted list of examples. The left sidebar contains navigation options like 'Svenska till svenska', 'Rättstava', and a list of other Lexin dictionaries.

Bild 1. Exempel på uppslagsordet *nyckel*, hämtat 2007-08-28.

Artikeln ger först information om uttal, som man även kan lyssna till, böjningsformer och ordklassstillhörighet. Sedan följer en betydelsebeskrivning och som synes har *nyckel* både en bokstavlig och bildlig betydelse. Det är även möjligt att se en bild på uppslagsordet via länken "bildtema 10", där man får se i vilka sammanhang en nyckel används. Vidare ges syntaktiska och morfologiska exempel samt sammansättningar och avledningar. I nedre delen av spalten till vänster om uppslagsordet finns länkar till de övriga lexikonerna i *Lexin*-serien.

### 3. Ordboksanvändning

#### 3.1. Forskning om ordboksanvändare och ordboksanvändning

Hartmann (1989:102–103) ställer upp tolv hypoteser om ordboksanvändaren. En av dem handlar om att lexikografer alltid har en viss uppfattning om vem användaren är och vilken utformning av ordboken som är mest användarvänlig. Ett annat antagande slår fast att det inte finns något ett-till-ett-förhållande mellan det sätt på vilket ordboken är avsedd att användas och det sätt som ordboken sedan faktiskt brukas på. Användarna missförstår informationen, till exempel vad en huvudkommentar motsvarande "vard." innebär (Bejoint 1989:208). De har också en övertro på ordbokens definitioner och ordurval (jfr SAOL 13:V).

Det råder dessutom en okunskap om det varierande utbud av ordböcker som finns och deras olika syften (Bejoint 1989:208). Ordboksanvändaren kanske till exempel letar efter definitioner i en synonymordbok. Beträffande e-ordböcker tillkommer svårigheten att välja rätt sökalternativ bland den mängd av avancerade sökmöjligheter som finns.

Vad gäller användningen av e-ordböcker befinner sig forskningen i en initial fas och än så länge är det svårt att veta vad utvecklingen av e-ordböcker har betytt för ordboksanvändningen. Enligt Svensén (2004:553f) har användarforskningen tidigare i huvudsak inriktat sig på inlärningsordböcker och tvåspråkiga ordböcker. Däremot vet man lite om ordboksanvändning av svenska, enspråkiga allmänordböcker och de elektroniska varianter som finns. De flesta empiriska undersökningar har dessutom genomförts med enbart studenter som informanter. Hur olika yrkesgrupper, framför allt de som yrkesmässigt arbetar med språk, eller människor i allmänhet använder ordböcker har man mindre kunskap om. Genom att studera loggfiler, där användaren är anonym, kommer man åt (de verkliga) användarna utan att veta deras ålder, kön eller sysselsättning. Det finns dock goda möjligheter att få reda på mer information om användarna om de registrerar vilka de är.

### 3.2. Loggfilstudier

Tidigare studier av loggfiler visar att de är ett utomordentligt hjälpmedel i arbetet med att göra ordböcker bättre. Genom att analysera loggfiler kan den typen av hypoteser som Hartmann ställer upp prövas eftersom man får reda på vad användarna faktiskt efterfrågar. Dessutom kan antaganden om den typiska användaren undersökas genom att man "bjuder in" brukarna till att kommentera och svara på frågor on-line om sitt ordboksanvändande.

Bergenholtz & Johnsen har undersökt loggfilerna från *Den Danske Netordbog*, <[www.netordbog.dk](http://www.netordbog.dk)>, en enspråkig ordbok med 109 000 lemman, och konstaterar inledningsvis följande:

Analyses of log files reveal exactly which lemmas and which types of information have been requested, and, perhaps more significantly, which lemmas and which types of information have been requested but were not found in the dictionary. (Bergenholtz & Johnsen 2005:117)

Med hjälp av loggfilerna kunde författarna ta reda vilka typer av problem användarna stötte på när de sökte på ett ord och inte fick träff. Loggfilerna visade till exempel att användarna sökte på passiv- och imperativformer av verb utan att få träff. Det fanns alltså ett behov av att kunna utföra den typen av sökningar och därför bör man fundera över om inte dessa former ska vara sökbara.

Studien visade också att nästan alla uppslagningar i topp 100-listan av sökningar som misslyckades hade sin förklaring i felstavningar. Det finns olika orsaker till att man stavar fel och de vanligaste anledningarna torde vara slarv, okunskap och



dyslexi. En stor del av felstavningarna bland de undersökta loggfilerna berodde enligt författarna på en uttalsmässig stavning. En annan orsak till att man inte fick träff var att man inte visste om ett uttryck skrevs som ett eller två ord. Icke-träffarna bestod också av ord som inte finns, vilket exempelvis kunde bero på att ett ord skrevs med suffixet *-ning* (*beskrivning*) i stället för det i danskan korrekta *-else* (*beskrivelse*). Användarna verkade också vara osäkra på när de skulle använda *foge-s*.

Ett av de största användningsområdena för loggfilerna är dock upptäckten av lemmaluckor genom att man ser vilka sökningar som inte fått träff. Många av de ord som användarna inte fann i ordboken var datarelaterade, finansiella, juridiska och medicinska termer och begrepp. De vanligaste begreppen inom respektive område, de som redan används frekvent i allmänspråket, borde enligt författarna ingå i ordboken (Bergenholtz & Johnsen 2005:137–138).

Vidare har de Schryver & Joffe (2004) undersökt loggfiler i en tvåspråkig Internetordbok mellan ett sydafrikanskt bantuspråk, Sesotho sa Leboa, och engelska. Även här bestod en överraskande hög andel av sökningarna av felstavade ord. Författarna redogör också för hur loggfilerna regelbundet utnyttjades under projektets gång. Ett sätt var att frekventa icke-träffar, som visade sig vara lemmaluckor i ordboken, lades till när ordboken uppdaterades.

I studien redogör de Schryver & Joffe även mer detaljerat för hur de kunde följa varje specifik användares sökningar och därmed undersöka dennes sökstrategier. De lade därtill upp ett formulär on-line där användarna kunde ge feedback på ordboken. Författarna påpekar till sist hur kontinuerlig analys av loggfiler tillsammans med användarnas feedback kan utnyttjas till att anpassa innehåll och utformning efter brukarnas behov (jfr Bergenholtz & Johnsen 2007).

#### 4. Material

Den studie som presenteras här grundar sig alltså på en analys av loggfilerna från Lexin *Svenska ord* från den första oktober 2006, ett slumpmässigt valt datum, och utgörs av 16 903 sökningar.<sup>1</sup> Av dessa sökningar är 41 % (6 941) felaktiga uppslagningar, vilka jag har valt att fokusera min studie på. Felaktiga uppslagningar är de sökningar som inte leder till träff i ordboken, alltså icke-träffar. Av tabell 1 framgår hur siffrorna ser ut i *Svenska ord* under delar av 2006 och 2007.

---

1 Tack till Viggo Kann, professor i datalogi vid KTH i Stockholm, som möjliggjort denna studie genom att tillhandahålla loggfilerna och svara på frågor om loggfiler och statistik.

Månad	År 2006			År 2007		
	Totalt antal uppslagningar	Felaktiga uppslagningar	Procent felaktiga ord	Totalt antal uppslagningar	Felaktiga uppslagningar	Procent felaktiga ord
Maj	410 154	173 978	42	627 593	247 945	40
Juni	221 865	89 891	40	300 567	116 328	39
Juli	200 048	69 670	34	231 654	91 496	39

Tabell 1. Antal uppslagningar i *Svenska ord* under maj-juli 2006 respektive maj-juli 2007.

I de två tidigare loggfilsstudierna jag refererar till, (se avsnitt 3.2), är både antalet sökningar och andelen icke-träffar lägre. I den danska nätordboken utfördes i snitt drygt 2 200 per dag mellan januari 2003 och augusti 2004 och andelen misslyckade sökningar var 19,5 % under den period som sökningarna loggfördes. I den afrikansk-engelska ordboken var snittet per dag drygt 3 500. Studien genomfördes från det att Internetordboken blev tillgänglig, april 2003, och sex månader framåt. Två uppdateringar genomfördes under de undersökta månaderna och efter den andra uppdateringen sänktes felprocenten från 33 % till 25 %. I båda ordböckerna ökade antal besökare snabbt med tiden.

Antalet uppslagningar är alltså mycket högre i *Svenska ord* än i de två ovan nämnda ordböckerna. Det beror troligtvis på att *Lexin*-seriens ordböcker har funnits tillgängliga via Internet uppemot tio år längre. De har dessutom distribuerats av Myndigheten för skolutveckling med fokus på andraspråksinläring och som skolordbok. Den har därmed haft en tydligt avgränsad målgrupp och myndigheten har på olika sätt aktivt spritt kunskap om e-ordbokstjänsten.

Efter en första genomgång av de uppemot 7 000 icke-träffarna visade det sig att uppemot 45 % av dem bestod av felstavningar. Det som dock fångade mig var, för det första vilka övriga typer av sökningar som resulterade i icke-träffar, och för det andra användarnas benägenhet att upprepa sina sökningar. Jag valde således att inte fokusera på felstavningarna vilka, tillsammans med en grupp oidentifierbara uppslagningar, endast kort berörs i avsnitt 5.2. Det material som redovisas i det följande utgör ungefär en tredjedel av loggfilerna.

## 5. Resultat

Resultaten presenteras under tre avsnitt där jag först tar upp vilka typer av sökningar som resulterar i icke-träff. Sedan beskrivs hur ordboken är programmerad att svara när sökordet inte finns i ordbokens lemmauppsättning. Till sist redogör jag för de sökstrategier som användarna tar till då de inte lyckas i sin första sökning.

### 5.1. Sökningar som resulterar i icke-träff

De uppslagningar som inte ger träff i *Svenska ord*, efter ovan nämnda avgränsning, är av olika slag och delas in i olika typer så som tabell 2 visar.

Typ av sökning	Antal	Procent
Enkla ord, avledningar och sammansättningar	800	58
Utländska ord (90% engelska)	160	12
Encyklopediska och Google-liknande sökningar	140	10
Fasta konstruktioner: fraser, idiom, kollokationer	90	6,5
Tillfälliga konstruktioner	65	5
Årtal, sifferkombinationer	50	3,5
Förkortningar	30	2,5
Felaktigt särskrivna uttryck	25	2
En/ett+substantiv och att+verb	10	<1
Totalt	1370	100

Tabell 2. Typer av sökningar som resulterar i icke-träff ordnade efter frekvens.

Det är i en del fall svårt att avgöra i vilken grupp ett ord ska placeras, t.ex. om en fras är en fast eller mer tillfällig konstruktion eller om en encyklopedisk sökning i stället tillhör gruppen med enkla ord, avledningar och sammansättningar. En del ord och uttryck har därför placerats i två grupper. Siffrorna och procenttalen har avrundats men torde ändå vara representativa för hur vanliga de olika typerna av sökningarna är i förhållande till varandra.

#### 5.1.1. Sökning på enkla och sammansatta ord, avledningar samt utländska ord

Mest frekventa är sökningarna som utgörs av enkla ord, avledningar och sammansatta ord. Användarna söker på såväl grundform som pluralformer och bestämda former. Nedan anförs exempel på de tre kategorierna:

- Enkla ord: *egal, mejla, hölster*
- Avledningar: *integration, finansiering, modernisering*
- Sammansättningar: *friskvård, ansiktsmimik, fängelsevistelse, hissfobi*

Det kan först konstateras att användarna har höga förväntningar på ordboken vad gäller dess urval av uppslagsord. Därför är det inte enbart negativt att många av de undersökta icke-träffarna utgörs av söksträngar som är rättstavade befintliga ord i svenska språket. Att man inte får träff beror dels på det begränsade lemmaurvalet (28 500 lemmen) i *Svenska ord*. Urvalet är gjort med tanke på invandrarens behov och *Lexin* verkar attrahera en mycket mer heterogen grupp än så.

Med tanke på den stora mängd olika ord användarna söker på verkar de inte veta vilken typ av ordbok *Lexin* är och vilken målgrupp den har. Precis som med tryckta ordböcker, där förord och inledning ofta inte läses, är det säkert så att webbsidans besökare inte tar reda på vilken typ av ordbok de har framför sig på skärmen. Eller så vet de det, men hittar inget annat alternativ och prövar med *Lexin*. Orden i denna kategori utgör alltså lemmaluckor i ordboken och visar hur loggfiler kan utnyttjas till att uppdatera och utöka lemmauppsättningen. Det skulle tillfredsställa en större och bredare grupp av användare utan att för den skull missgynna den primära målgruppen, andraspråksinlärare.

Näst efter den nyss nämnda gruppen kommer sökningar på utländska ord, ca 160 stycken, varav närapå 90 % engelska, till exempel *education*, *throw* och *influential*. Det är något förvånande att utländska ord återfinns bland sökningarna i *Svenska ord* av det skälet att man via *Lexins* startsida aktivt måste välja vilken av ordböckerna man vill använda.

### 5.1.2. Flerordssökning

Tvåordslemman som partikelverb och mer eller mindre fasta konstruktioner som t.ex. idiom och kollokationer samt tillfälliga konstruktioner är också en av de vanligaste typerna av sökning som leder till icke-träff:

- Reflexiva verb: *föresätta sig*, *anstränga sig*
- Partikelverb: *känna igen*, *fälla ihop*
- Idiom: *kalla fötter*, *lägga lök på laxen*
- Kollokationer: *psykisk ohälsa*, *civil olydnad*
- Tillfälliga konstruktioner: *hierarkiskt samhälle*, *säljande image*

Med sökfunktionen fritextsökning och generösa hänvisningar skulle användaren kunna hitta fram till idiom, kollokationer, vanliga sammansättningar samt annan information som ligger gömd i ordboksartiklarna. I nuvarande skick får användaren inte träff på ord som faktiskt finns i ordboken, t.ex. det reflexiva verbet *föresätta sig* eller partikel verbet *känna igen*, p.g.a. att verbets uppslagsform står i presens (*föresätter sig* respektive *känner igen*). Användaren får således informationen att orden inte finns i ordboken vilket är rent felaktigt. Ett enkelt verb rättas dock automatiskt till uppslagsformen presens i de fall man söker på en annan tempusform.

### 5.1.3. Encyklopediska och Google-liknande sökningar

En beaktansvärd andel uppslagningar består av vad jag kallar encyklopediska och Google-liknande sökningar. I den förstnämnda gruppen ingår bland annat sökning efter geografiska platser, namn på skådespelare och medicin- och historierelaterade begrepp:

- Geografiska platser: *Asien*, *Tanzania*, *Brunei*
- Skådespelare o.d.: *Aretha Franklin*, *Bengt Andersson* (skådespelare)
- Medicin: *alzheimer*, *leukemi*, *rohypnol*
- Historia: *Konstantin I*, *hellenism*, *Versaillesfreden*, *postmodernism*
- Övriga: *SAOL*, *Dawsons 's Creek*, *Vattuman*, *Burma* (kattras)

De Google-liknande sökningarna utgörs av flerordssökningar som liknar informationssökning, t.ex. *VM i innebandy 2006* och *kategori: orter i Bayern*, där man kan anta att användaren är ute efter information om något snarare än en definition av den sökta strängen. Andra exempel är *de fyra grundlagarna*, *borgerliga partier*, *bolibompa* och *Karolinska institutet*. Dessa typer av sökningar visar att användarna är vana att söka på det sätt som Google och andra sökmotorer tillåter.

Det verifierar också mitt tidigare påstående att användarna inte verkar vara medvetna om vilken typ av ordbok *Lexin* är och vilken målgrupp den har.

#### 5.1.4. Övriga sökningar

Andra typer av uppslagningar som är värda att kommentera är för det första en grupp övriga sökningar bestående av årtal, nummer, förkortningar och egennamn:

- Årtal och nummer: *1170-talet, 1099, 1395, 1449*
- Förkortningar: *mg/l, dna, bnp, sms, ACSE, JWST, EAPR*
- Egennamn: *Daniel, Jas*

De i sammanhanget frekventa fyrsiffriga talen framstår som en aning kryptiska. Är de årtal, koder eller något annat, och i så fall vad? Förkortningarna är av två slag, igenkännbara som *mg/l* och *dna* samt vid första anblicken icke igenkännbara bokstavskombinationer. Exempelen ovan har senare identifierats via Google.<sup>2</sup> Ett telefonnummer och tre datum återfanns också i materialet.

För det andra finns det en samling felaktigt särskrivna uttryck bland loggfilerna, men dessa är färre än vad jag förväntade mig. I de flesta fall har användaren utan framgång först sökt på uttrycket i sin hopskrivna, korrekta form, och sedan valt att särskrivna det. Detta har jag tolkat som en sökstrategi och det är en av de sju strategier som jag beskriver i avsnitt 5.3.

#### 5.2. Ordbokens programmerade svar vid icke-träff

När man söker på ett ord som inte finns i ordboken, och rutan "Rättstava okända ord" är ikryssad, får man i många fall förslag på ord som man kan ha menat. Dessa förslag varierar i antal mellan ett och upp till ett fyrtiotal. En genomgång av en tredjedel av materialet ger vid handen att en inte obetydlig mängd av icke-träffarna, ca 130 stycken, får över tio ordförslag. Nedan följer exempel från de olika typer av sökningar som resulterar i höga förslagsantal:

- Existerande ord: *blogg, lakej, ninja*
- Oidentifierbara och icke-existerande ord: *frola, aaaa*
- Felstavade ord: *betuta* (jfr *betuttad* 'förtjust i')
- Engelska ord: *event, exam*
- Övrigt: *Daniel, bla*

Exempelorden i sammanställningen får ett förslagsantal mellan 30 och 40. Att ge så många förslag på vad användaren kan ha menat är näppeligen användarvänligt. Hur det ser ut om man söker på ordet *blogg* framgår av bild 2:

<sup>2</sup> ACSE står för Association Control Service Element, JWST för James Webb Space Telescope och EAPR för Euroatlantiska Partnerskapsrådet.



Bild 2. Sökning på ordet *blogg*, hämtat 2007-08-28.

Alla förslag som ges ryms inte på bilden, men scollar man neråt finner man att en sökning på *blogg* ger 40 ordförslag. Många av förslagen är dessutom, både i detta och i de flesta andra fall, ologiska. Det beror sannolikt på att rättstavningsverktyget endast är baserat på editeringsavstånd. Det innebär omkastning av två intilliggande bokstäver eller insättning, borttagning eller utbyte av en bokstav (Viggo Kann, personlig kommunikation).

### 5.3. Användarnas strategier vid icke-träff

Användarna tar till olika strategier när de inte får träff i sitt första försök. Några lämnar ordet och går inte vidare med sin sökning, och andra fortsätter att söka. Gemensamt för dem som fortsätter att söka är att de tenderar att upprepa sina sökningar, men de gör det på olika sätt. Jag har funnit sju sökstrategier, nämligen följande:

- 1) Upprepad sökning med identisk stavning
- 2) Sökning på del av ord
- 3) Sökning med annan stavning
- 4) Borttagning eller tillägg av böjningsformer/ändelser
- 5) Särskrivning
- 6) Utbyggd sökning
- 7) Kombination av flera strategier

De sju strategierna presenteras i det följande i nummerordning.

Strategi 1 är helt enkelt att upprepa sökningen med samma stavning:

- *tvärhand* → *tvärhand* → *tvärhand* → *tvärhand* → *tvärhand*
- *inskriftion* → *inskriftion* → *inskriftion*

Användaren söker i det första exemplet efter ordet *tvärhand*, ett korrekt stavat svenskt ord som dock inte ingår i *Svenska ord*. Sökningen upprepas fyra gånger. Såväl i detta fall som i exemplet *inskriftion* får användarna beskedet att ordet inte finns i ordboken men de väljer ändå att upprepa sin sökning.

Strategi 2 är att upprepa sökningen och söka på en del av ett ord:

- *enhetssträvandena* → *strävandena*
- *biosfär* → *bio* → *sfär* → *biosfär*

I det första exemplet söker användaren på ordet *enhetssträvandena*, utan resultat, och sedan på *strävandena*. Det ger en träff på ordet *strävan*. Vad gäller det andra exemplet får användaren träff på både *bio* och *sfär*, men inte på *biosfär*. Att användaren återgår till den första söksträngen, som han eller hon gör i det här fallet, är också vanligt.

Strategi 3 är att pröva en annan stavning. Först har vi de som söker på rätt stavning i sin första sökning och inte får träff. Ordet finns alltså inte i *Svenska ord*. Då prövar de en annan stavning och de börjar följaktligen stava fel:

- *desorienterad* → *desorigenterad*
- *habituell* → *hatibuell* → *habutiell*
- *klassificering* → *kalsificering* → *klassificering* → *klassificiering* → *klaceficering* → *klassifisering* → *klassifisera*

I första exemplet verkar användaren pröva en mer uttalsenlig stavning, i det andra kastas bokstäverna *b*, *t*, *i* och *u* om. Det tredje exemplet är ett av många belägg på den envishet som användarna visar prov på. Användaren får här till slut träff på ordet *klassificerar* men borde, enligt mig, kunna få träff tidigare med ett vassare rättstavningsverktyg.

I det andra fallet stavar användaren fel i sin första sökning och börjar sedan, som ovan, att pröva med att stava ordet på ett annat sätt:

- *geslong* → *geslång* → *jeslong* → *cheslomg* → *sheslong*
- *kyvär* → *kuvär* → *kyver* → *kuver* → *kuvär* → *kuvett*
- *shinettzel* → *shinettzel* → *skinettzel* → *skinetsel* → *schinettzel* → *schinetsel* → *schitsel*

Gemensamt för huvuddelen av sökningarna i den här gruppen är att orden användarna efterfrågar är svårstavade, som till exempel *schäslong*, som inte finns med i *Lexin*, och *kuvert*, som finns med och som användaren till slut får träff på. Det tredje fallet är ytterligare ett exempel på användarnas envisa beteende.

Strategi 4 innebär att man tar bort eller lägger till böjningsformer respektive

ändelser. I det andra exemplet nedan söker användaren till exempel först på particip- och preteritumformer och till sist på verbets grundform:

- *transferering* → *transfereringar*
- *internaliserade* → *internaliserad* → *internalisera*

Även här består de sökningar som inledningsvis felstavas av svårstavad ord, men den här gruppen innehåller få belägg.

De två näst sista strategierna, 5 och 6, innehåller också få exempel. För det första har vi användare som söker på ett sammansatt ord utan resultat och som då väljer att särskrivna ordet:

- *språkkunskap* → *språk kunskap*
- *innanhav* → *innan hav*

För det andra finns det de som bygger ut sin sökning:

- *tillkorta* → *tillkortakommanden*
- *kongruens* → *kongruensböjning*

Den sista strategin är att kombinera två eller flera av ovan nämnda taktiker. Det kan vara att man i sin andra sökning ändrar stavning och vid tredje försöket lägger till en pluraländelse.

I genomsnitt upprepades varje sökning tre till fyra gånger. I tabell 3 ställs strategierna upp ordnade efter frekvens och även här har procentsiffrorna avrundats. Strategi 3, 4 och 7 är uppdelade i två grupper – sökningar som är rättstavade respektive felstavade vid första uppslagningen – så som de presenterats ovan.

Strategi	Antal	Procent
Sökning med annan stavning – felstavat vid första sökningen	52	19
Borttagning/tillägg av ändelse – rättstavat vid första sökningen	50	18,5
Kombination av strategier – felstavat vid första sökningen	38	14
Sökning på del av ord	33	12
Sökning med annan stavning – rättstavat vid första sökningen	30	11,5
Kombination av strategier – rättstavat vid första sökningen	28	10
Identisk stavning	13	5
Borttagning/tillägg av ändelse – felstavat vid första sökningen	9	>3
Särskrivning	9	>3
Utbyggd sökning	9	>3
Totalt	271	100

Tabell 3. De sju strategierna ordnade efter frekvens.

De två vanligaste strategierna är för det första att pröva annan stavning i de fall man stavar fel i första sökningen och inte får träff, och för det andra att ta bort eller lägga till en ändelse i de fall man stavar rätt i första sökningen men inte får träff.



## 6. Sammanfattning och diskussion

I artikeln redogör jag för en analys av de icke framgångsrika sökningarna gjorda i Internetordboken *Lexin on-line Svenska ord*. De icke-träffar som jag ytterligare avgränsar min studie till utgör ca en tredjedel av materialet.

För de första delar jag upp icke-träffarna i olika typer av sökningar. Det visar sig att en stor mängd icke-träffar utgörs av rättstavade svenska ord, men de ingår inte i *Svenska ords* lemmauppsättning. Vidare är sökningar på fraser av olika slag och encyklopediska sökningar frekventa. Dessa resultat visar inte bara att loggfiler kan utnyttjas till att upptäcka lemmaluckor utan också ett behov av att kunna utföra fritextsökningar så att man hittar fram till sammansättningar, idiom m.m. Det visar sig också att många utländska ord återfinns i *Svenska ords* loggfiler. Det finns redan en lista på länkar till andra webblexikon via *Lexin* men dessa bör lyftas fram mer och kompletteras med fler. Det bör också finnas en lättillgänglig länk till Google och andra sökmotorer. Detta för att hjälpa användaren att hitta fram till rätt källa.

Det andra avsnittet behandlar det höga antal ordförslag som många gånger presenteras för användaren vid en icke framgångsrik sökning. Jag efterlyser här ett mer avancerat verktyg, baserat på såväl editeringsavstånd som uttal, för att kunna uppnå ett mer logiskt urval av förslag på ord. Det skulle resultera i färre antal icke-träffar och därmed fler nöjda användare.

Slutligen redovisas de sju sökstrategier användarna tar till när deras sökning misslyckas. Användarna tenderar att upprepa sina sökningar på olika sätt. De två vanligaste strategierna är att pröva annan stavning eller att använda en kombination av flera taktiker. Benägenheten att upprepa sina sökningar skulle, i värsta fall, kunna leda till att användarna börjar ifrågasätta sina språkkunskaper. Det vore mycket olyckligt i de fall brukarna faktiskt stavar rätt i sin första sökning och utgör ytterligare ett skäl till att ha ett kraftfullare rättstavningsverktyg.

Materialet består av sökningar från en dag och får anses vara relativt litet, trots att de uppgår till nästan 7 000. Jag har heller inte haft möjligheten att identifiera varje specifik användare och kan således inte med säkerhet knyta de upprepade sökningarna till en och samma användare. Men det är uppenbart att användarna har höga förväntningar på ordbokens innehåll och att de är envisa med tanke på hur de tenderar att upprepa sina sökningar. Det faktum att många sökningar på vanliga, rättstavade ord inte ger träff kan tolkas på ett positivt sätt, nämligen att användarna har en mycket stor tilltro till ordboken vad beträffar dess innehåll. Att *Lexin* fyller en viktig funktion för många användare står också klart i och med de jämförelsevis mycket höga antal uppslagningar per månad i förhållande till den danska respektive afrikansk-engelska ordboken (se kapitel 4 ovan). Det är dock ett problem att användarna så ofta inte lyckas nå den önskade informationen genom sina sökningar. Det bör dock här poängteras att det inte har varit min mening att kritisera *Lexin*. Ordboken lever upp till de mål som låg inom ramen för *Lexin*-projektets uppdrag.

Det står klart att loggfiler är ett mycket användbart redskap för att upptäcka lemmaluckor. Jag har i denna artikel visat att det finns anledning att vidare undersöka loggfilernas funktion som ett effektivt verktyg inom lexikografi och efterlyser fler studier av och med hjälp av loggfiler. Man kan diskutera hur stort inflytande användarna ska ha i det lexikografiska arbetet, men jag menar att man inte ska vare sig underskatta eller överskatta deras möjliga insatser. Användarna bör stimuleras till att ge feedback i form av nyordsförslag, definitioner m.m. och även uppmantras att kommentera hur olika sökfunktioner uppfattas och om gränssnittet är tillfredsställande. Det vore även intressant att förhöra sig om i vilken situation behovet av en ordbok uppstår och inte minst vad användarna gör med den information de får genom sina sökningar. På så sätt skulle man kunna kartlägga användarnas sökbeteende mer ingående och identifiera den verkliga målgruppen. Allt i syfte att förbättra såväl elektroniska ordböcker som pappersordböcker.

## Litteraturförteckning

### Ordböcker

Den Danske Netordbog: <<http://www.netordbog.asb.dk>>

LEXIN = *Lexikon för invandrare, Svenska ord* <<http://lexin.nada.kth.se/svenskaord>>

SAOL 13 = *Svenska Akademiens ordlista* 2006. (13:e uppl.). Stockholm: Norstedts Akademiska Förlag.

*Svenska ord – med uttal och förklaringar* 1999 (3 uppl.) Stockholm: Norstedts.

### Övrig litteratur

Béjoint, Henri 1989: The Teaching of Dictionary Use: Present State and Future Tasks.

I: *Wörterbücher, Dictionaries, Dictionnaires. First Volume*. Berlin/New York: Walter de Gruyter, 208–215.

Bergenholtz, Henning & Mia Johnsen 2005: Log files as a Tool for Improving Internet Dictionaries. I: *Hermes – Journal of Linguistics* 34, 117–141.

Bergenholtz, Henning & Mia Johnsen 2007: Log Files Can and Should Be Prepared for a Functionalistic Approach. I: *Lexikos* 17, 1–21.

Gellerstam, Martin 1999: LEXIN – lexikon för invandrare. I: *LexicoNordica* 6, 3 – 18.

Hartmann, Reinhard R. K. 1989: Sociology of the Dictionary User: Hypotheses and Empirical Studies. I: *Wörterbücher, Dictionaries, Dictionnaires. First Volume*. Berlin/New York: Walter de Gruyter, 102–111.

de Schryver, Gilles & David Joffe 2004: On How Electronic Dictionaries are Really Used. I: Geoffrey, W & S. Vessier (eds.), *Proceedings of the eleventh EURALEX International Congress. Lorient, France, July 6–10, 2004*, Lorient: UBS, 187–196.

Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. (2 uppl.) Stockholm: Norstedts Akademiska Förlag.

Anki Hult  
Fil. mag., f. 1977  
Institutionen för svenska språket  
Box 200  
SE-405 30 GÖTEBORG  
ann-kristin.hult@svenska.gu.se



## På strejftog indenfor Retskrivningsordbogens territorie

### Retskrivningsordbogen 2011

The Danish Language Council (Dansk Sprognævn) is preparing a new edition of the official dictionary of Danish orthography (Retskrivningsordbogen). In this paper we discuss some orthographical changes. The changes concern the spelling of words which end in *-ium* (e.g. *akvarium*, *gymnasium*), the spelling of hyphenated compounds like *e-mail-adresse*, *S-togs-station*, and the spelling of multi-word prepositions like *bag ved*, *inden for*. In the new edition of the dictionary, the entry words will contain word-division points. The proposed end-of-line divisions are not the only acceptable ones, and we briefly discuss different principles of word-division (syllabic versus morphemic).

#### 1. Introduktion

Vi er i Dansk Sprognævn gået i gang med at forberede en ny udgave af Retskrivningsordbogen, som efter planen skal udkomme i 2011. Der er tale om en helt ny ordbog med et fuldstændig ajourført ordforråd og med væsentlige forbedringer hvad angår bøjningsoplysninger og oplysninger om ordenes sammensætningsformer. Der vil også være korte ordidentificerende betydningsangivelser ved langt flere opslagsord end i den nuværende udgave, og der vil være oplysning om orddeling på alfabetisk plads, sådan at man straks ved opslaget kan se et forslag til hvordan ordet kan deles.

I denne artikel vil vi først gennemgå det orddelingsprincip vi har valgt for opslagsordene, og dernæst redegøre for tre retskrivningsændringer i den kommende ordbog. Den første ændring drejer sig om forbindelser af adverbium og præposition, fx *inden for/indenfor*, den anden om brugen af bindestreger i sammensætninger som *e-post-adresse* og *S-togs-station* og den sidste om *-ium*-former i ord som *territorium* og *ministerium*.

#### 2. Orddelingsangivelser på alfabetisk plads

I de officielle danske retskrivningsordbøger har man hidtil været henvist til at slå op i afsnittet med retskrivningsregler hvis man vil have oplysning om hvordan man deler ordene ved linjeskift. I den kommende udgave af retskrivningsordbogen vil man få serveret orddelingsoplysningerne på alfabetisk plads – i form af diskrete

prikker eller lodrette streger i selve opslagsordet. Den form for orddelingsinformation får man i en del andre mere eller mindre officielle ordbøger, bl.a. Svenska Akademiens ordlista (fra og med 12. udgave, 1998) og Duden – Die deutsche Rechtschreibung:

**har|mon|is|era** (SAOL)

**har|mo|ni|sie|ren** (Duden)

Man kan også vælge at vise orddelingsoplysningerne i et særskilt ord. Det har man gjort i fx Politikens Betydnings- og Retskrivningsordbog:

**harmonisere**

(..)

DELES: har-mo-ni-se-re

Fordelen ved at gøre det på den sidstnævnte måde er at man undgår at forstyrre skriftbilledet i opslagsordet. Vi har imidlertid valgt at gøre som i SAOL og Duden; det sparer plads, og det er jo vigtigt i en ordbog der skal kunne ligge i skoleelevernes tasker.

### 2.1 Orddelingsprincipper

I dansk kan ordene deles efter to forskellige principper, nemlig efter det morfologiske princip og efter det mere mekaniske stavelsesprincip. Af hensyn til pladsen nævner vi her kun principperne for deling ved en eller flere konsonanter mellem vokaler.

Efter det morfologiske princip deles sammensatte, afledte og bøjede ord. Der deles således mellem sammensætningsled, fx *halv-abe*, *spejl-æg*, efter præfiks, fx *for-stå*, *mis-tanke*, foran suffiks, fx *køns-lig*, *dyr-isk*, og foran bøjningsendelse, fx *mand-en*, *klar-est*.

Ord der ud fra en synkron dansk synsvinkel ikke kan inddeles i flere betydningsbærende elementer, må nødvendigvis deles efter et andet princip. Her bruges stavelsesprincippet, der kan beskrives som en blanding af det gamle enkonsonantprincip og et lidt vagt formuleret fonotaktisk/grafotaktisk princip: Der deles enten sådan at der går én konsonant til den efterfølgende vokal, eller "foran så mange konsonanter som almindeligt kan stå først i et ord" (RO 2001, § 17.2.b), fx

*gis-pe* eller *gi-spe*

*tørs-te* eller *tør-ste*

*æb-le* eller *æ-ble*

Stavelsesprincippet kan også bruges ved deling af komplekse ord, forudsat at man ikke skiller et enkeltbogstav eller en konsonantgruppe fra det ordelement som bogstavet eller konsonantgruppen hører til (fx *syvar-met*, *mistan-ke*, ikke "sy-varmet", "mi-stanke").

Som man kan se, er de danske orddelingsregler beslægtet med de norske regler og meget nært beslægtet med de traditionelle svenske regler. I Svenska Akademiens ordlista har man valgt at gennemføre det morfologiske princip: Ordene inddeles her i synkront identificerbare elementer. Flerstavelsesord som ikke lader sig inddele i flere identificerbare elementer, har ingen delingsangivelser i SAOL. Indførelsen af et rent morfologisk princip har som bekendt givet anledning til en del debat, dels fordi det er svært at gennemføre, dels fordi det i mange tilfælde resulterer i orddelinger som kan virke kontraintuitive eller direkte afstikkende.

De danske orddelingsregler er meget fleksible, og der er derfor ofte mange forskellige korrekte måder at dele et ord på. Fx kan et tastavelsesord som *svensker* deles på hele tre måder:

*sven-sker, svens-ker, svensk-er*

Fleksibiliteten er praktisk for den skrivende, men det er ikke hensigtsmæssigt at angive så mange delepunkter i et opslagsord. Mange ordbogsbrugere ville nok føle sig ladt i stikken hvis vi gjorde sådan her:

**sven|s|k|er**

Brugerne vil gerne have klarere besked. Vi har derfor besluttet at der i opslagsordet skal gives ét og kun ét orddelingspunkt pr. stavelse. Vi følger altså stavelsesprincippet men respekterer naturligvis de begrænsninger der gælder for sammensatte og afledte ord.

## 2.2. Orddelingsprincippet i Retskrivningsordbogen 2011

I vores variant af stavelsesprincippet følger vi enkonsonantprincippet foran ikke-fuldvokal. Det vil sige at der som hovedregel går én konsonant til en efterfølgende ikke-fuldvokal:

<VCV> *a-be, ma-le, no-te, si-ger, vi-ser*  
 <VCCV> *an-ker, bur-de, hen-te, læn-ge, tør-te, vin-der*  
 <VCCCV> *ank-re, jung-le, mest-re, slent-re, stolp-re*  
 <VCCCCV> *blomst-re, fængs-le, venst-re*

Foran fuldvokal deler vi så langt til venstre som muligt:

<VCV> *ba-sar, bi-list, du-sin, e-po, me-tyl, sta-tion, vi-sum*  
 <VCCV> *del-ta, di-plomat, fak-tisk, ma-krel, tek-nik, æ-dru*  
 <VCCCV> *ast-ma, a-stral, in-tro, man-dril, trans-port*  
 <VCCCCV> *ek-stra, eks-trovert*

Det at orddelingsangivelserne på alfabetisk plads følger stavelsesprincippet, indebærer ikke en indskrænkning af de eksisterende orddelingsregler. Det vil fortsat være muligt og lige så korrekt at dele ordene efter det morfologiske princip.

Når vi har valgt at følge stavelsesprincippet, skyldes det først og fremmest at det er enkelt at anvende, og at det i de fleste tilfælde fører til delinger der er hensigtsmæssige ud fra et lydligt synspunkt.

### 3. Adverbium + præposition

I dansk retskrivning har vi en regel der først og fremmest udmærker sig ved at udløse meget stort antal stavfejl. Reglen siger at forbindelser af adverbier og præpositioner skrives i to ord når præpositionen har en styrelse, ellers i ét ord:

Forbindelser af adverbier og præpositioner skrives i to ord når præpositionen har en styrelse, ellers i ét ord:

*Vejen gik stejlt op ad bakke. (Men: Vejen gik stejlt opad).*

*Neden under gulvet vrimlede det med mus. (Men: Nedenunder vrimlede det med mus).*

I et eksempel som *Neden under gulvet vrimlede det med mus* skrives *neden under* i to ord, men i *Nedenunder vrimlede det med mus* skrives *nedenunder* i ét ord.

Det er en ældgammel retskrivningsregel. Vi har haft den siden i hvert fald 1889, og siden da har generation efter generation af skoleelever måttet kæmpe for at lære den, eller rettere: for at lære at bruge den. For reglen i sig selv er jo egentlig såre simpel. Der er bare det ved den at den forudsætter en ganske stor grammatisk indsigt hvis man skal følge den konsekvent. Man skal fx kunne indse at relativordet *som* er styrelse til *over for* i et eksempel som dette:

*Vi har et problem som ledelsen desværre er helt uforstående over for*

Man skal også vide at et ord som er udeladt, fx et elliptisk *som*, også kan tælle som en styrelse, og at ordforbindelsen derfor skal skrives i to ord:

*Vi har et problem [...] ledelsen desværre er helt uforstående over for*

Det kan vistnok kaldes subtilt. Man ville muligvis kunne forsvare at bevare reglen hvis den kunne bruges til at adskille betydninger med. Men det kan den ikke. Den information man meddeler ved at følge den, er 100 % redundant.

Reglen medfører som nævnt et meget stort antal stavfejl, også hos personer som i øvrigt er gode og sikre stavere. Vi har derfor besluttet at gøre det valgfrit at skrive i ét eller to ord når forbindelsen af adverbium og præposition har en styrelse, dvs. når den fungerer som præposition. Når forbindelsen ikke har en styrelse, dvs. fungerer som et adverbium, skal der ligesom i dag altid skrives i ét ord. Reglen kommer altså til at se nogenlunde sådan ud:

Mange almindeligt brugte forbindelser af adverbium + præposition kan



valgfrit skrives i ét eller to ord når præpositionen har en styrelse. Når præpositionen ikke har en styrelse, skrives der i ét ord:

*Vejen gik stejlt op ad* el. *opad bakke.* (Men: *Vejen gik stejlt opad*).

*Neden under* el. *nedenunder gulvet vrímlede det med mus.* (Men: *Nedenunder vrímlede det med mus*).

På denne måde får vi elimineret næsten alle de ligegyldige fejl som reglen er skyld i. Men reglen lever stadig og kan fortsat praktiseres af elitestavere der vil skille sig ud fra mængden.

Måske burde vi i stedet tage skridtet fuldt ud og gøre som i norsk og svensk, hvor almindelige ordforbindelser af denne type skrives i ét ord uanset deres grammatiske funktion:

#### Norsk

*en løser ikke problemene ved å gå utenom dem* (præposition)

*det er ikke til å komme utenom* (adverbium)

#### Svensk

*utanför dörren* (præposition)

*stå utanför* (adverbium)

I ordbøgerne står de opført som både adverbium og præposition:

#### SAOL

**inn|an|för** prep. o. adv.

**ut|an|på** prep. o. adv.

#### Tanum

**utenom** prep. og adv.

**henover** adv. og prep.

I den kommende danske retskrivningsordbog kunne ordene behandles på samme måde:

**indenfor** *præp., adv.*; kom indenfor!; indenfor døren.

**udenpå** *præp., adv.*; udenpå tøjjet; bogen har et billede udenpå.

På den måde slipper vi af med en overflødig retskrivningsregel og med et stort antal lige så overflødige stavfejl.<sup>1</sup>

## 4. Bindestreger i sammensætninger som e-post-adresse, S-togs-station og 1900-tals-digter

I den nuværende Retskrivningsordbog (§ 57.7) står der at der sættes bindestreg

<sup>1</sup> Dansk Sprognævn har sidenhen vedtaget at gøre som i norsk og svensk, nemlig at skrive almindelige forbindelser af adverbium + præposition i ét ord uanset om præpositionen har en styrelse eller ej.

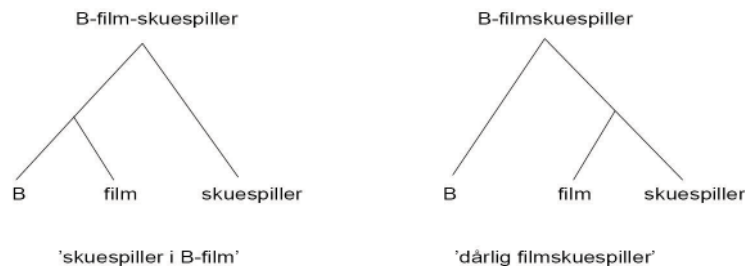
mellem en sammensætnings dele hvis første- eller sidsteleddet selv indeholder en bindestreg. Ord som fx *e-post* og *S-tog* indeholder en bindestreg fordi det er sammensætninger med en forkortelse, og hvis sådanne ord sættes sammen med henholdsvis *adresse* og *station*, skriver man *e-post-adresse* og *S-togs-station*, med to bindestreger. Man skriver ligeledes *1900-tals-digter* med to bindestreger – den første bindestreg bliver man nødt til at have fordi det er efter et taltegn, og når der er bindestreg i sammensætningens første led, skal man også sætte en bindestreg mellem førsteleddet og sidsteleddet.

Vi der er redaktører af ordbogen, finder denne regel logisk og systematisk. Den sidste bindestreg tydeliggør hvilke dele sammensætningen består af. Den adskiller det sidste led i sammensætningen fra det første led. Hvis denne bindestreg udelades og kun den første bindestreg sættes, ser det ud som om at leddene i sammensætningen består af de to dele der vises på hver side af bindestregen. Dette fremgår af figur 1.



Figur 1.

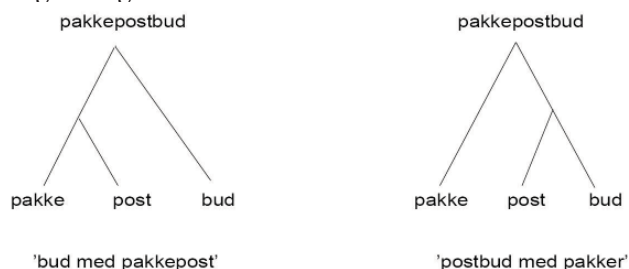
På denne måde er også bindestregerne med til at markere betydningsforskelle som *B-film-skuespiller* 'skuespiller i en B-film' og *B-filmskuespiller* 'en dårlig filmskuespiller' (jf. figur 2).



Figur 2.

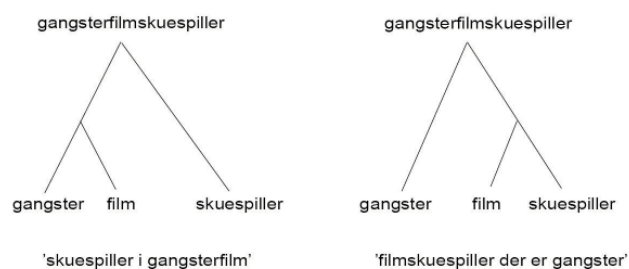
Men det er tilsyneladende kun nørder som os der synes denne regel er logisk og smuk. Sprogbrugerne følger den ikke – de udelader den sidste bindestreg. Og man kan da også hævde at det ikke er rimeligt at skulle lave en analyse af sammensætningens dele for at kunne skrive sådanne sammensætninger korrekt. Det gør man jo ikke ved "almindelige" sammensætninger som fx *pakkepostbud* og

*gangsterfilmskuespiller*. Sådanne sammensætninger er strukturelt tvetydige, som det fremgår af figur 3 og 4.



Figur 3.

Figur 3 og 4 viser at sammensætninger som *pakkepostbud* og *gangsterfilmskuespiller* kan analyseres på to måder da der ikke er nogen markering af hvilke dele sammensætningen består af. Vi klarer os fint uden denne markering i ”almindelige” sammensætninger så hvorfor skal vi have den i sammensætninger med forkortelser (*e-post-adresse*)?



Figur 4.

Det er ofte også vanskeligt at afgøre hvor grænsen mellem første og andet led går, og det er nok også ligegyldigt for betydningen. I et ord som fx *OL-guldvinder* er det svært at afgøre om der skal én eller to bindestreger i ordet. Er det en vinder af OL-guld (*OL-guld-vinder*), eller er det en guldvinder ved OL (*OL-guldvinder*)?

Der er endvidere problemer med en sammensætning som fx *C-vitamin-pille* som er korrekt med to bindestreger da det er ’en pille med C-vitaminer’. Men to bindestreger i denne sammensætning ser forkert ud fordi vi er vant til ordet *vitaminpille*, som er uden bindestreg. Derfor skriver langt de fleste sprogbrugere *C-vitaminpille* i stedet for *C-vitamin-pille*.

På baggrund af ovenstående betragtninger er vi nået frem til at vi i den næste Retskrivningsordbog vil afskaffe den sidste bindestreg i de nævnte sammensætninger således at de kun skrives med én bindestreg. Det betyder altså at man ikke længere har mulighed for at skelne mellem *B-filmskuespiller* (med én bindestreg) og *B-film-skuespiller* (med to bindestreger). Der vil fra 2011 kun skulle være én bindestreg i en sådan sammensætning uanset betydningen (jf. tabel 1).

Retskrivningsordbogen 2001	Retskrivningsordbogen 2011
e-post-adresse	e-postadresse
1900-tals-digter	1900-talsdigter
OL-guldvinder/OL-guld-vinder	OL-guldvinder
B-filmskuespiller/B-film-skuespiller	B-filmskuespiller
C-vitamin-pille	C-vitaminpille

Tabel 1.

### 5. Valgfrihed mellem *-ium* og *-ie*: *gymnasium*, *studium*, *kollegium* mv.

Få ord (dvs. 10 ord) i Retskrivningsordbogen har valgfrihed mellem *-ium* og *-ie*, fx *ovarium/ovarie* (valgfri siden RO 1955), *kranium/kranie* (valgfri siden RO 1986), *podium/podie* (valgfri siden RO 2001), *stadium/stadie* (valgfri siden RO 2001). I den nye Retskrivningsordbog vil vi give valgfrihed mellem *-ie* og *-ium* ved yderligere 35-40 opslagsord.

De ord det drejer sig om, er de almindeligere ord, dvs. de ord som de fleste af os kender og bruger, fx *akvarium*, *gymnasium*, *kollegium*, *kompedium*, *kriterium*, *laboratorium*, *mysterium*, *solarium*, *studium*, *territorium*.

Formerne på *-ie* er almindelige i talesproget ved disse ord. Studerende siger ”jeg bor på kollegie”, ministre siger ”sådan er det ikke i mit ministerie”, og mange af os siger fx *kriterie*, *territorie* og *mysterie* i stedet for *kriterium*, *territorium* og *mysterium*. Men i øjeblikket er det altså sådan at det ifølge Retskrivningsordbogen kun er formerne på *-ium* der er korrekte i skriftsproget. Når man nu siger *-ie*, er det mærkeligt at være tvunget til at skrive *-ium* for at stave korrekt – her kan man da tale om en stor uoverensstemmelse mellem lyd og skrift.

Formerne på *-ie* er naturlige og velkendte fordi de i forvejen findes i bøjningsformer, fx *kriteriet*, *territorier*, *mysterierne*, og det er også *-ie*-formerne der bruges i sammensætninger når ordene indgår som første led i sammensætninger, fx *kollegieværelse*, *ministeriebygning*, *territorieplanlægning*. For mange mennesker er der desuden en stilforskel mellem *-ie*- og *-ium*-formerne. Nogle finder det højtideligt eller krukke at bruge former som *mysterium*, *studium* og *territorium* til forskel fra *mysterie*, *studie* og *territorie*. Der er altså gode argumenter for at indføre valgfrihed mellem de to former ved endnu flere ord i den nye ordbog.

Men det er som sagt kun ved de almindeligere ord at vi vil indføre valgfrihed mellem *-ie* og *-ium*.

Fx skal de mere lærde ord fortsat kun kunne skrives med *-ium*, fx *amfibium*, *beneficium*, *herbarium*, *offertorium*, *vivarium*. En *-ie*-form er simpelthen utænkelig ved disse ord.

Andre ord som heller ikke vil få en valgfri form på *-ie*, er grundstofferne, fx *aluminium*, *helium*, *kalium*, *magnesium*, *plutonium*, *titanium*. En *-ie*-form er helt

udelukket ved disse ord, og ved bøjning og sammensætning ændrer disse *-ium-* ord sig da heller ikke til *-ie*. Det hedder *aluminiummet*, *heliummet*, *kaliummet* osv. i bestemt form singularis, og det hedder fx *aluminium(s)gryde* og *heliumballon*; *-ie-*formerne er altså slet ikke inde i billedet ved disse ord.

## Litteraturliste

### Ordbøger

- Duden = *Duden. Die deutsche Rechtschreibung*. 24. udgave. Dudenverlag: Mannheim, Leipzig, Wien, Zürich 2006.
- Politikens Betydnings- og Retskrivningsordbog*. København 2001.
- RO 55 = *Retskrivningsordbog*. Dansk Sprognævn. Gyldendal: København 1955.
- RO 1986 = *Retskrivningsordbogen*. Dansk Sprognævn. Gyldendal: København 1986.
- RO 1996 = *Retskrivningsordbogen*. 2. udgave. Aschehoug Dansk Forlag A/S: København 1996.
- RO 2001 = *Retskrivningsordbogen*. 3. udgave. Dansk Sprognævn. Alinea A/S – Aschehoug Dansk Forlag A/S: København 2001.
- SAOL = *Svenska Akademiens ordlista över svenska språket*. 12. udgave. 1998.
- Tanum: *Tanums store rettskrivningsordbok*. 9. udgave. Kunnskapsforlaget: Oslo 2005.

Anita Ågerup Jervelund  
videnskabelig medarbejder, f. 1969  
Dansk Sprognævn, Njalsgade 136  
DK-2300 København S  
aagerup@dsn.dk

Jørgen Schack  
seniorforsker, f. 1961  
Dansk Sprognævn, Njalsgade 136  
DK-2300 København S  
schack@dsn.dk



JÓN HILMAR JÓNSSON

## Lemmatisering og tilgangsstruktur Refleksjoner på bakgrunn av *Íslenskt orðanet*

This article discusses how an electronic presentation of a dictionary description provides the means of a wider, but simultaneously more specific, lemmatisation than previously practised in printed dictionaries. This especially applies to multiword lemmas. At the same time users are offered a more varied access to the lexicographic information contained within the dictionary. Here, these possibilities are being evaluated with respect to an ongoing Icelandic lexicographic project named *Íslenskt orðanet* (Icelandic Wordnet). The material of the wordnet is mainly drawn from the contents of dictionaries of Icelandic phrases, and the data is used to identify semantic relations within the Icelandic vocabulary, for single words as well as for phrases. In the lemmatisation of the material the primary consideration is semantic disambiguation. Furthermore, multiword units, particularly various verbal phrases, are given a lemma status equal to that of single words. That way, the dictionary description is made accessible from many different directions, both towards particular semantic fields and related lemmas within those fields, and towards formally related lemmas.

### 1. Innledning

Vi som har deltatt i det nordiske leksikografiske samarbeidet helt fra starten i 1991 da Nordisk forening for leksikografi ble grunnlagt i Oslo på vår første leksikografiske konferanse, har opplevd ganske store endringer i vårt daglige ordboksarbeid. De mest iøynefallende og gjennomgripende endringene knytter seg i større eller mindre grad til tekniske fremskritt og nyheter, som har gjort det lettere å foreta innsamling og utarbeidelse av leksikografiske data, muliggjort en mer variert og målrettet søking etter leksikografisk informasjon og skapt forutsetninger for å få til en klarere og mer brukervennlig ordbokstekst. Men selv om teknologien har vært en avgjørende faktor i denne utviklingen, finnes det også andre drivkrefter bak disse endringene; de vitner nemlig også om leksikografenes interesse, vilje og evne til å definere og ta fatt på sine oppgaver på en ny måte, sprengte de rammer som lenge har begrenset formatet og dimensjonen av leksikografiske produkter, og komme fram til en strukturering og presentasjon som på best mulig måte tar hensyn til det materialet ordboken bygger på, og den leksikografiske beskrivelse den sikter mot. Mye av det som nå er blitt realistisk og aktuelt å utføre, har sikkert lenge streift leksikografenes tanker i en eller annen form, men kunne ikke bli til virkelighet mens de ytre forutsetninger sto i veien for gjennomføringen.

Når jeg ser tilbake i denne sammenheng, vil jeg gjerne framheve at for ti år siden, våren 1997, holdt NFL sin fjerde konferanse i Helsingfors. Disse vakre maidagene i Finland feiret vi bl.a. utgivelsen av Nordisk leksikografisk ordbok. NLO er altså bare ti år gammel og beholder forhåpentlig fortsatt sin verdi som hjelpemiddel og oppslagsverk for nordiske leksikografer. Uansett er det klart at visse grunnleggende forutsetninger for definisjonen og behandlingen av viktige leksikografiske begreper nå må ses i et nytt lys. Jeg vil i det følgende ta opp noen av de spørsmål det her dreier seg om, og da først og fremst på bakgrunn av et ordboksprosjekt jeg har vært opptatt av de siste årene.

## 2. De trykte ordbøkernes tradisjon

### 2.1 En fastlåst lemmaliste

Tradisjonelle ordbøker i trykt form er forbundet med mange og viktige begrensninger som i stor grad også har preget forutsetningene for definisjonen og oppdelingen av mange leksikografiske begreper. For det første så har en trykt ordbok uunnngåelig bare plass for et begrenset antall oppslagsord eller lemmaer, og denne lemmabestanden kan ikke så lett revideres etter at den er blitt konkretisert i form av en trykt ordbokstekst. Jo større og innholdsrikere ordboken er, desto vanskeligere blir det å foreta revidering og oppdatering av ordboksteksten. Og det er ikke lett å komme fram til klare kriterier for avgrensingen av det ordforrådet som får plass som oppslagsord i ordboken overfor det vokabularet som ikke blir tatt med. Dette problemet, lemmaseleksjonen og den konkrete lemmatiseringen, får sjelden noen omtale når ordboken blir presentert overfor brukerne, og de leksikografer som har sittet med dette problemet, er tilsynelatende ikke noe særlig interessert i å redegjøre for hvordan de har arbeidet med det. I ordboksreklamen blir oppmerksomheten også gjerne ledet i en annen retning ved at man framhever hvor mange lemmaer ordboken inneholder eller man utpeker enkelte lemmaer som skal være karakteristisk for lemmavalget.

### 2.2 Uttrykket fremfor innholdet – den allmenne ordbokens karakter

Ordet med henvisning til sine uttrykksmessige egenskaper har lenge vært en grunnleggende enhet i allmenn leksikografisk beskrivelse, der det framtrer som lemmatisk enhet overfor brukerne samtidig som det utgjør en overskrift for en avgrenset ordboksartikkel som inneholder en selvstendig beskrivelse av vedkommende lemma. Som regel blir alle tilgjengelige opplysninger om lemmaet plassert her uansett om eller i hvilken grad ordet forekommer i beskrivelsen av andre lemmaer eller ikke. Innenfor artikkelen blir hovedvekten lagt på betydningsdifferensiering ved at det skilles mellom delbetydninger i form av nummererte ledd osv. Denne oppbyggingen er ikke minst karakteristisk for den ordbokstype som skal være mest dekkende og minst spesialisert, dvs. allmenne definisjonsordbøker.



Man anser også gjerne at denne typen for lengst har etablert seg som en autonom og grunnleggende ordbokstype som ikke har mye å hente fra de mer spesialiserte ordbøkene. Men når man ser nærmere på saken, er forholdet ikke så enkelt. Allmenne definisjonsordbøker i sin endelige form, uansett hvor omfattende og innholdsrike de er, utgjør f.eks. et ganske begrenset og ufullstendig grunnlag for de ordbøker som har til oppgave å redegjøre for ordenes semantiske relasjoner, og det gjelder ikke i mindre grad for de ordbøker som skal beskrive ordenes syntagmatiske relasjoner og fraseologi.

### *2.3 Problemer med struktureringen av ordboksteksten*

Når det gjelder allmenne ordbøker, er det naturlig at leksikografen retter oppmerksomheten mot helhetsbildet av de enkelte lemmaer, ser etter at beskrivelsen blir strukturert på en regelrett måte, at hvert element i ordboksteksten er på sin plass osv. Det ligger i sakens natur at brukerne ikke i samme grad vil være opptatt av dette, fra deres synsvinkel er et slikt helhetsbilde sjelden av særlig interesse utover det at det viser hvor i artikkelen den aktuelle informasjon er å finne. Deres oppmerksomhet og interesse gjelder først og fremst en bestemt angivelse lemmaet gir adgang til, men som kan være adressert til et underordnet tekstelement. Her vil det selvfølgelig være en fordel for brukeren å ha en klar innsikt i struktureringen av ordboksteksten for å kunne lokalisere den informasjon han eller hun er ute etter. Det er i dette berøringsfeltet mellom leksikografens oppgave å vise et helhetlig bilde av lemmaenes egenskaper og brukerens behov for en problemfri tilgang til en bestemt informasjon ordbokens mikrostruktur får sin utforming, med sin ofte utstrakte og kompliserte oppdeling og hierarki som brukeren kan ha problemer med å orientere seg i og beherske.

Hvor mikrostrukturen er innfløkt og komplisert er makrostrukturen som regel tilsvarende enkel, slik at den først og fremst fungerer som en slags portal til mikrostrukturen som brukeren må kunne åpne, men som i seg selv ikke byr på noen vesentlig informasjon. Men dessuten viser den på sin ufullkomne måte det eneste sammenhengende bildet av ordbokens lemmabestand brukeren kommer i kontakt med, en bestand der alfabetet er den styrende faktor som også gir et glimt av den morfologiske sammenhengen. Ut fra lemmalisten kan brukeren til en viss grad danne seg et bilde av hva slags ordforråd ordboken dekker, og hvordan det blir avgrenset. Men sammenhengende deler av ordforrådet når det gjelder semantiske egenskaper og begrepsfelt, står utenfor brukerens synsfelt, og det blir sjelden nevnt om og i tilfelle i hvilken grad hensynet til slike relasjoner har påvirket lemmaseleksjonen og leksikografens avgrensning av lemmabestanden.

### *2.4 De semantiske og syntagmatiske relasjonene står i skyggen*

De semantiske relasjonene innenfor ordforrådet står tradisjonelt utenfor den allmenne definisjonsordbokens område. Dette aspektet hører til spesialiserte ordbøker av ulike slag, først og fremst synonymordbøker og begrepsordbøker. Det

ordforrådet som disse ordbøkene inneholder, blir dermed vanligvis framstilt uten noen tilknytning til den behandling tilsvarende ord eller lemmer får i en allmenn ordboksbeskrivelse, f.eks. med hensyn til oppdeling i delbetydninger og homonymseparering.

I allmenne ordbøker blir redegjørelsen for ordenes syntagmatiske relasjoner, kollokasjoner, konstruksjoner og faste ordforbindelser i hovedsak underordnet den semantiske betraktningssmåten. En selvstendig redegjørelse for slike relasjoner som finnes i konstruksjonsordbøker, kollokasjonsordbøker og andre spesialiserte ordbøker, er heller ikke knyttet til den behandling tilsvarende ord eller lemmer får i en allmenn ordboksbeskrivelse.

Denne strenge typologiske oppdelingen har satt sterke grenser for utfoldelsen av allmenn ordboksbeskrivelse. Det gjelder ikke bare den allmenne ordbokens autonomi, de spesialiserte ordbokstypene jeg har nevnt, har tradisjonelt heller ikke vært i kontakt med hverandre, og det er ganske uvanlig å se en sammenkobling av en begrepsorientert beskrivelse på den ene side og en syntagmatisk eller fraseologisk orientert beskrivelse på den andre. Her gjelder den langvarige tradisjonen som går ut på at det semasiologiske og det onomasiologiske aspektet skal være adskilt, og at det semasiologiske skal ha prioritet som en slags grunntype.

### 3. Framveksten av den elektroniske ordboken

Framveksten av elektronisk databehandling og den elektroniske ordboken har endret dette bildet og skapt forutsetninger for revidering av den tradisjonelle typologien. I og med at man nå i mange tilfeller med en enkel søkeoperasjon kan gå direkte til den ønskete informasjon, blir det mindre aktuelt for brukeren å støtte seg til en fast makro- eller mikrostruktur. Avgrensingen av det ordforrådet ordboken dekker, er også usynlig, det kan gjennomgå en kontinuerlig revidering og oppdatering der brukerne til og med kan opptre som aktive rådgivere. Som følge av dette får vi nå en sterkere fokusering på ordforrådet som helhet og dets interne relasjoner istedenfor å være opptatt av en selvstendig beskrivelse av et stort antall uavhengige og isolerte lemmer. Da spør det bl.a. om den allmenne ordboksbeskrivelsen kan få en utvidet funksjon ved å få en direkte tilknytning til en omfattende og detaljert redegjørelse for de semantiske og syntagmatiske relasjoner ordene opptre i.

### 4. *Íslenskt orðanet*

#### 4.1. *Semantisk ordnett på grunnlag av syntagmatiske og morfologiske relasjoner*

Disse refleksjonene har sin konkrete bakgrunn i et ordboksprosjekt jeg har arbeidet med de siste to årene sammen med Þórdís Úlfarsdóttir. Dette prosjektet, som har fått navnet *Íslenskt orðanet*, går i hovedsak ut på å utnytte ordenes syntagmatiske og fraseologiske relasjoner med utgangspunkt i konkrete leksikografiske

produkter for å avdekke ordenes semantiske relasjoner og bygge opp et sammenhengende semantisk ordnett der også den syntagmatiske og fraseologiske sammenheng blir ivaretatt.<sup>1</sup>

Oppbyggingen av dette ordnettet foregår i mange etapper, og mye arbeid gjenstår før prosjektet er ferdig. Et grunnleggende spørsmål i den innledende fasen besto i å vurdere og bestemme hvilke leksikalske enheter skulle fungere som grunnelementer i ordnettet og dermed inngå i de semantiske relasjoner ordnettet ville vise, og samtidig være tilgjengelige som søkeelementer for brukerne. Oppfatningen av dette har bl.a. en avgjørende betydning for hvor omfattende ordnettet vil være, om det først og fremst skal knytte sammen enkeltord, eller om det også skal kunne dekke flerordsenheter.

#### 4.2 Ordnettets kildemateriale

Det konkrete materialet bak ordnettet er ganske omfattende og mangfoldig, men kjernen består av ordforbindelser, kollokasjoner og fraser av ulike slag som er hentet fra tre forskjellige leksikografiske kilder: a) to fraseologiske ordbøker, *Orðastaður* (Jón Hilmar Jónsson 2001) og *Orðaheimur* (Jón Hilmar Jónsson 2002), senere slått sammen i *Stóra orðabókin um íslenska málnotkun* (STOB, Jón Hilmar Jónsson 2005), og b) et register over ordforbindelser og fraser i seddelarkivet til *Orðabók Háskólans* (Íslands universitets leksikografiske institutt; jf. *Orðasambandaskrá Orðabókar Háskólans*), til sammen over 200.000 ordforbindelser av ulike slag. Dertil kommer omtrent 100.000 sammensetninger fra *Orðastaður* som i ordboken er semantisk klassifisert og morfologisk knyttet til sine orddeler.

I tillegg til dette materialet inneholder ordnettets database leksikografiske opplysninger av semantisk karakter som er hentet fra to ulike kilder. For det første dreier det seg om en samling med islandske ekvivalenter hentet fra to tospråklige ordbøker med islandsk som målspråk, *Ensk-íslenska orðabókin* og *Dönsk-íslensk orðabók*. Disse ekvivalentene er valgt slik at de i ordboksteksten står sammen med en eller flere ekvivalenter slik at man kan gå ut fra at det (i de fleste tilfeller) dreier seg om en synonymirelasjon. For det andre inneholder databasen et ordtilfang fra en eldre arkivsamling med opplysninger fra informanter om betydningen av ord og ordforbindelser, som i samlingens database får en markering som viser at de står i en synonymi- eller hyponymirelasjon til hverandre.

#### 4.3 Oppbyggingen av ordnettet

4.3.1 En liste med nøkkelord dekker samtlige ordforbindelser og sammensetninger. Hele fraseologien i ordnettets kildemateriale blir knyttet sammen i et register over samtlige nøkkelord (fortrinnsvis innholdsord) i ordforbindelsene. Den ene delen av dette registeret ble til under utarbeidelsen av STOB, den andre er hentet

1 For nærmere informasjon om prosjektet se: <http://www.ordanet.is/>

fra det store fraserregisteret til seddelarkivet. Dette nøkkelordregisteret, som også dekker alle sammensetninger fra *Orðastaður* og de orddeler de består av, utgjør kjernen i en lemmaliste som i samsvar med de regler som gjelder for lemmatiseringen, etter hvert blir bygget opp. Lemmatiseringen er dermed ikke gitt på forhånd, men inngår som en del av bearbeidelsen av det leksikografiske materialet. Foruten å dekke de konkrete ordforbindelser og sammensetninger lemmaet forekommer i (i alfabetisk rekkefølge), viser de enkelte artikler også en særskilt liste med de lemmaer (nøkkelord) disse ordforbindelsene og sammensetningene knytter seg til, sortert etter ordklasse og alfabetisk innenfor hver ordklasse for seg (jf. figur 1), med umiddelbar tilgang til disse lemmaene. Dermed får lemmaene en utvidet funksjon som gjenspeiler hvordan de knytter sammen formelt og betydningsmessig sammenhengende vokabular.

#### 4.3.2 Typemarkering av ordforbindelsene

Den videre bearbeidelsen av grunnmaterialet består bl.a. i å markere skilnaden mellom de enkelte typer ordforbindelser det dreier seg om. Det blir gjort for det første ved å markere idiommer som en særskilt type som da skiller seg fra de mer allmenne ordforbindelsene. For det andre blir ordpar (med sideordnete ledd) definert som en egen type. Denne typemarkeringen blir synliggjort blant nøkkelordene i de enkelte artikler i form av ikoner (jf. figur 1), som da viser hvilket forhold de enkelte nøkkelordene har til det aktuelle lemmaet. Som det framgår av figur 1 dekker denne ikoniseringen også de nøkkelord som er relatert til lemmaet gjennom en allmenn ordforbindelse eller en sammensetning.

5 dýrlegur lo	<input type="checkbox"/> PAR: hof og hallir	hof no hvk
6 glæsilegur lo	<input type="checkbox"/> PAR: skrautbyggingar og hallir	skrautbygging no kvk
7 glæstur lo	<input type="checkbox"/> PAR: stórhýsi og hallir	stórhýsi no hvk
8 gulllegur lo	<input type="checkbox"/> OSAMB: byggja höll	byggja <hús; verksmiðju>
9 háreistur lo	<input type="checkbox"/> OSAMB: dýrlegar hallir	dýrlegur lo
10 hátimbraður lo	<input type="checkbox"/> OSAMB: endurbyggja höllina	endurbyggja <húsið> so
11 stór lo	<input type="checkbox"/> OSAMB: glæsileg höll	glæsilegur lo
12 veglegur lo	<input type="checkbox"/> OSAMB: glæstar hallir	glæstur lo
13 aðsetur no hvk	<input type="checkbox"/> OSAMB: gullleg höll	gulllegur lo
14 aðseturshöll no kvk	<input type="checkbox"/> OSAMB: gæta hallarinnar	gæta <hallarinnar> so
15 álfahöll no kvk	<input type="checkbox"/> OSAMB: háreist höll	háreistur lo
16 álfur no kk	<input type="checkbox"/> OSAMB: hátimbruð höll	hátimbraður lo
17 bíó no hvk	<input type="checkbox"/> OSAMB: höll konungs	konungur no kk
18 bíóhöll no kvk	<input type="checkbox"/> OSAMB: höllin rís <uppi á hæðinni, í miðjum garðinum>	<húsið> rís so3bn
126 stórbygging no kvk	<input type="checkbox"/> OSAMB: höllin stendur <uppi á hæðinni, í miðjum garðinum>	<húsið> stendur <við götuna> so3bn
127 stórhýsi no hvk	<input type="checkbox"/> OSAMB: loftsvallir hallarinnar	loftsvallir no kvk flt
128 sumar no hvk	<input type="checkbox"/> OSAMB: reisa höll	reisa <hús; skóla> so
129 sumarhöll no kvk	<input type="checkbox"/> OSAMB: vegleg höll	veglegur lo
130 sund no hvk		

Figur 1: Íslenskt orðanet. Artikkelen **höll** i forkortet form. Den venstre spalten inneholder nøkkelordene til lemmaet, sortert etter ordklasse. Spalten i midten inneholder de enkelte ordforbindelser, sortert etter ordforbindelsestype og alfabetisk innenfor hver type.

## 4.3.3 Markering av semantiske relasjoner blant lemmaene

Gjennomgåingen av fraseologien til de enkelte lemmaer sikter mot det formål å knytte sammen semantisk sammenhørende ord og andre leksikalske enheter som i de enkelte artikler opptrer som nøkkelord. Her synliggjør adjektivartiklene sammenhørende substantiver og omvendt, substantivartiklene gir dessuten adgang til sammenhørende verb, verbartikler viser sammenhørende substantiver og adverb, og adverbiale artikler knytter sammen beslektede verb. I tillegg kommer de sideordnete kollokasjoner der oppslagsordet er knyttet til semantisk nærstående ord, typisk et nærsynonym eller antonym.

Hvis vi tar utgangspunkt i et innholdsrikt lemma som substantivet *saga* 'historie', kan vi forvente at materialet der vil avdekke relevante semantiske relasjoner mellom de adjektiver ordet står sammen med, og at det fører sammen semantisk beslektede verb og verbale konstruksjoner. Det får vi også et konkret bevis på:

eftirminnileg ('minneverdig') saga  
minnisstæð ('uforglemmelig') saga

þýða ('oversette') söguna  
snara ('oversette') sögunni

hlusta á ('lytte til') söguna  
hlýða á ('lytte til') söguna

Men substantiviske konstruksjoner viser seg også å ha denne funksjon:

meginefni ('hovedinnhold') sögunnar  
efniskjarni ('kjerneinnhold') sögunnar  
meginþema ('hovedtema') sögunnar

Og blant sammensetningene finnes det gjerne nært beslektede ord:

kjaftasaga 'sladderhistorie'  
slúðursaga 'sladderhistorie'  
sögupersóna 'romanperson'  
söguhetja 'romanhelt'

Det er et aktuelt spørsmål i denne sammenheng hvor gjennomgripende og presis denne semantiske sammenføyingen kan bli. Svaret vil først og fremst være avhengig av hvordan de lemmatiske enhetene blir utformet og presentert, om man kan nøyes med ettordsenheter i samsvar med den leksikografiske tradisjonen, eller om det er behov for flerordslemmaer. Et annet viktig spørsmål gjelder homonymi og polysemi: Kan man operere med flertydige lemmaer, eller stilles det krav om entydiggjøring av lemmaene?

## 4.3.4 Lemmatisering av flerordsforbindelser

Ved utarbeidelsen av ordnettet har det vært av avgjørende betydning at det fraseologiske grunnmaterialet allerede er framstilt på en standardisert måte i de leksikografiske kilder det er hentet fra. Det kommer sterkest til uttrykk når det dreier seg om verb, verbale konstruksjoner og verbforbindelser av ulike slag, både kollokasjoner og mer faste forbindelser som idiomer. Det er nettopp disse ordbokselementene som lenge har vært mest påtrengende og plasskrevende i ordbøkernes mikrostruktur, noe som har betydd at de ikke i tilstrekkelig grad er synlige i tekstbildet, og at de ikke kan ses i en aktiv sammenheng med betydningsmessig eller formelt beslektede enheter. I *Orðabheimur* og senere i STOB blir det tatt et skritt mot å gjøre slike forbindelser til selvstendige leksikografiske enheter ved å vise den semantiske sammenhengen under begrepsoverskrifter og tilføye en særskilt registerkomponent der samtlige ordforbindelser blir angitt i alfabetisk rekkefølge under vedkommende verb (jf. figur 2). I ordnettet blir skrittet tatt fullt ut, og verbforbindelsene blir gjort til regelrette lemmaer. Dermed kan de fritt inngå i semantiske relasjoner på linje med ettordslemmaene, samtidig som de byr på sortering og klassifisering med hensyn til deres syntagmatiske egenskaper.

láta ekki sitt eftir liggja → DUGNADUR, FRAMLAG,  
FRAMTAKSSEMI, liggja, ÓSERHLÍFNI/FORNFÝSI  
láta ekki sjá sig → FJARVERA, sjá  
láta ekki svo lítið að <heilsa gestunum>  
→ DRAMB/HROKI  
láta ekki teyma sig á asnaeyrunum → STADFESTA  
láta ekki teyma sig á eyrunum → STADFESTA  
láta ekki troða á sér → DUGUR, VIDNÁM  
láta ekki vaða ofan í sig → DUGUR, <sup>2</sup>vaða  
láta ekki vaða ofan í sig á skitugum skónum  
→ DUGUR  
láta ekki þar við lenda → ÁFRAMHALD  
láta ekki þar við sitja → ÁFRAMHALD  
láta ekki <tækifærið> ganga sér úr greipum  
→ FRAMTAKSSEMI  
láta ekki <tækifærið> hjá líða → FRAMTAKSSEMI  
láta engan bilbug á sér finna → ÁKVEDNI, bilbug-  
ur, DUGUR, KIARKUR  
láta fallast ofan í stól → stóll  
láta fallast með <honum> → ÁSTARSAMBAND  
láta fangann lausan → fangi  
láta fara fram athugun á <málinu> → athugun  
láta fara illa með sig → DUGLEYSI  
láta fara sem fara vill → KÆRULEYSI  
láta fara vel um sig → fara, MAKINDI, VELLÍÐAN  
láta fá sig til að <taka á sig sökina> → VEIKLYNDI  
láta fé af hendi rakna → fé  
láta félagsmál til sín taka → félagsmál  
láta fjálmár skarast → fjöl  
láta fjúka í kveðlingum → kveðlingur  
láta fjúka í kviðlingum → kviðlingur  
láta fjölinna fljóta → KÆRULEYSI  
láta flakka lausu → KÆRULEYSI  
láta fótinn fæða sig → ÁGENGNI, FRAMFÆRSLA  
láta frá landi → BROTTFOR

Figur 2: *Stóra orðabókin um íslenska málnotkun*. Et avsnitt av artikkelen *láta* i registerdelen, der de enkelte ordforbindelser er angitt i alfabetisk rekkefølge med henvisninger til vedkommende lemma(er) i ordbokens hoveddel.

Denne lemmatiseringen fører til en voldsom økning av verblemmaene ut over det som er vanlig i leksikografisk sammenheng. I alt dreier det seg om bortimot 70.000 verbale lemmer mot omtrent 125.000 substantivlemmer.

Motivasjonen bak lemmatisering av flerordsforbindelser er ikke begrenset til verbene selv om den kan sies å ha størst virkning i den sammenheng. Adverbiale flerordsforbindelser fortjener også klart lemmastatus, slik at semantisk beslektede adverbialer kan knyttes sammen og bygge opp et større semantisk felt. Substantiviske flerordsforbindelser bør også kunne lemmatiseres, særlig når det dreier seg om klare begreper som kan være semantisk relatert til andre substantiver. Blant adjektivene er forholdene litt mer komplisert, som det framgår av redegjørelsen i neste avsnitt.

#### 4.3.5 Entydiggjøring av lemmer

En utstrakt lemmatisering av flerordsforbindelser innebærer i grunn at mikrostrukturen blir oppløst samtidig som flerordslemmerne blir gjort til makrostrukturelle enheter på linje med ettordslemmerne. For verbene har denne operasjonen en avgjørende betydning for såvidt at hvert enkelt verblemma i hvert fall langt på vei blir entydiggjort, foruten at selve lemmerformen i mange tilfeller gjenspeiler betydningen og viser et syntagmatisk mønster som gjerne er felles for flere semantisk beslektede lemmer.

Behovet for entydiggjøring gjelder i prinsippet for hele lemmabestanden, med hensyn til at de semantiske relasjoner kollokasjonene avdekker, i alle tilfeller gjelder en bestemt delbetydning av flertydige ord. Spørsmålet er hvordan man best kan foreta en klar entydiggjøring av flertydige ettordslemmer der selve lemmerformen ikke uten videre kan avsløre den aktuelle betydningen.

Generelt kan flertydighet blant substantivene oppheves ved å tildele de enkelte delbetydningene ulike indeksnumre, på liknende måte som homonymer blir adskilt med homonymindeks i en tradisjonell lemmaliste. En slik separering vil være tilstrekkelig for å få til en feilfri sammenkobling av beslektede ord, men overfor brukeren må den kompletteres med en markering som identifiserer den aktuelle betydningen.

Blant adjektivene er skillelinjene mellom de enkelte delbetydningene mye mer uklare. Men her gjelder det også å forhindre at flertydighet fører sammen ord med ulik betydning, i dette tilfellet substantiviske kontekstpartnere. Den konkrete løsningen i ordnettet består delvis i å utvide lemmerformen med et veiledende substantiv som identifiserer den aktuelle betydningen, f.eks. av **hlýr** 'varm': **hlýr** <**vindur**> ('vind'), **hlýr** <**sumar, vetur**> ('sommer, vinter'), **hlýr** <**flík**> ('klesplagg'), **hlýr** <**bros**> ('smil'). Men dessuten blir den bruken av adjektiver som er relatert til personer, realisert i form av en predikativ konstruksjon: **vera hlýr** ('være hyggelig'), **vera hlýlegur** ('være blid'). Denne fremstillingen blir også overført på karakteriserende substantiver med en liknende funksjon: **vera dugnaðarforkur** 'være en dyktig person', **vera letingi** 'være lat'.

#### 4.4 Ordnettets tilgangsstruktur

Her er det ikke anledning til å gi en utførlig beskrivelse av ordnettets tilgangsstruktur, den er heller ikke fastlagt i detaljer. Følgende kommentarer er for det meste begrenset til noen allmenne forutsetninger som etter min mening preger utformingen av brukerens tilgang til den leksikografiske informasjon han eller hun er ute etter.

##### 4.4.1 Verbale konstruksjonsmønstre

Lemmatiseringen av flerordsforbindelser og kravet om mest mulig entydiggjøring av flertydige lemmaer er bestemmende faktorer for hvordan ordnettets tilgangsstruktur kan utformes. Før vi går inn på den semantiske komponenten i ordnettet, er det verdt å fremheve betydningen av å kunne foreta klassifisering og sortering av de verbale flerordslemmaene ut fra deres formelle egenskaper. Grammatisk tagging av lemmaformene, som allerede er blitt foretatt, har vist seg å være hensiktsmessig både når det gjelder oversikt over konstruksjonsmønstre for enkelte verb, og for å knytte sammen lemmaer som viser samme konstruksjonsmønster (se Þórdís Úlfarsdóttir 2006). Den medfører også at en semantisk sortering av lemmaene kan vurderes og klassifiseres i forhold til lemmaenes formelle egenskaper og omvendt.

##### 4.4.2 Søking i lemmalisten

Man kan gå ut fra at en direkte søking i lemmalisten vil være den mest naturlige søkeoperasjon for de fleste brukere, i hvert fall i første omgang. Lemmatisering av flerordsforbindelser åpner her for nye søkemuligheter, f.eks. slik at en bestemt del av lemmaformen blir valgt som søkeelement. Dette vil f.eks. være aktuelt overfor mange idiomer med en stor formell variasjon der et konstant ledd kan knytte sammen de enkelte varianter. Man kan også regne med at brukeren vil ha behov for en samlet oversikt over de verbale lemmaer av ulik syntagmatisk karakter som inneholder et bestemt verb, for dermed å kunne foreta en målrettet søking. En slik liste vil kunne fremstilles på ulike måter, bl.a. sortert etter konstruksjonsmønstre. I en større sammenheng kan en slik fremstilling by på en interessant sammenlikning mellom enkelte verb. Også i andre tilfeller vil det være aktuelt å fokusere på en formelt sammenhengende del av lemmalista, enten det gjelder et avgrenset ordforråd innenfor en ordklasse, et bestemt konstruksjonsmønster eller forekomsten av et bestemt ord eller en orddel.

##### 4.4.3 Valens og tilknytning til nøkkelordindeksen

De verbale flerordslemmaene gir en eksplisitt informasjon om verbets syntaktiske og semantiske valens. Valensangivelsen er avgrenset innenfor lemmaformen med hakeparentes der de konkrete ord og ordkombinasjoner fungerer som kategorisatorer for leksikalske elementer som syntagmatisk og semantisk kan stå i den aktuelle omgivelsen: *lesa* <*bókina; söguna; textann*> 'lese <boken; historien; teksten>', <*fyrrtækið*> *skilar* <*gróða, hagnaði*> 'bedriften gir <overskudd>'.



For å få innsikt i verbenes syntaktiske valens vil det være mest hensiktsmessig å utnytte den grammatiske markeringen av lemmaformene. Når det gjelder de elementer som uttrykker verbenes semantiske valens, så er de tilgjengelige for direkte søking som i mange tilfeller vil føre sammen semantisk og/eller syntagmatisk beslektede lemmaer. Men her bør man tenke på at de konkrete kontekstpartnerne blir valgt som representanter for betydningslike ord som kan stå i den aktuelle omgivelsen. For å kunne aktivisere hele det ordfeltet og dermed gjøre søkingen uavhengig av et subjektivt valg av kontekstpartner, kreves det at søkefeltet blir utvidet fra å være begrenset til lemmaene til også å dekke nøkkelordindeksen eller i hvert fall deler av den. En slik utvidelse vil være nyttig i mange sammenhenger, men det spørs allikevel om den ikke vil komplisere tilgangsstrukturen i for høy grad.

#### 4.4.4 Søking på semantiske relasjoner

Det er naturlig å tenke seg at de semantiske relasjonene innenfor ordforrådet står i sentrum for brukernes oppmerksomhet, og det overordnede målet for leksikografen består også i å gi brukeren tilgang til en klar og systematisk oversikt over semantiske felt og relasjonstyper, enten det gjelder synonymer og antonymer eller større ordfelt. Her spiller kollokasjonenes vitnesbyrd en hovedrolle ved å føre sammen semantisk beslektet vokabular. I den forbindelse kan man tenke seg ulike forutsetninger for samkjøring av sammenhengende vokabular som da vil gi ulike resultater når det gjelder omfanget og innholdet av et semantisk felt. Slike alternative grupperinger av semantisk beslektede ord kan være interessant i mange brukssammenhenger. Men brukeren vil også gå ut fra at det foreligger en bevisst semantisk sammenkobling og gruppering av lemmabestanden ut fra leksikografens vurdering, enten det gjelder synonymi, antonymi eller andre semantiske relasjoner. Brukerens tilgang til denne informasjon kan utføres på ulike måter. Fra leksikografens synsvinkel spørs det om det er hensiktsmessig å operere med konkrete og synlige overskrifter for de enkelte felt slik man kjenner fra trykte ordbøker (f.eks. *Orðaheimur*), eller om det er mer fleksibelt og brukervennlig å tenke seg at brukeren i alle tilfeller går ut fra de formelle egenskapene (et ord, en frase eller et konstruksjonsmønster), og at veien til de semantiske relasjonene og ordfeltene alltid starter som en utvidet informasjon til et bestemt lemma. I det tilfellet kan konkrete overskrifter være overflødige og til og med ha forstyrrende effekt på den indre dynamikken i begrepsrelasjonene. Her ville man eventuelt kunne nøye seg med veiledende elementer i form av sentrale ord eller fraser innenfor et semantisk felt for å markere det semantiske innholdet i et mindre (underordnet) felt, mens tilgangen til et utvidet felt ikke ville kreve en særskilt markering. Men uansett må leksikografen foreta en omfattende og presiserende etikettering av lemmaene for å gi brukeren tilgang til semantisk sammenhengende enheter.

#### 4.5 Noen sammenliknende kommentarer

Jeg har her redegjort for hvordan ordnettets genuine formål bestemmer prinsippene for lemmatisering av de leksikalske enhetene det inneholder, og hvordan en utstrakt lemmatisering av flerordsforbindelser samt et krav om entydige lemmaer preger utformingen av brukerens tilgang til den leksikografiske beskrivelse. Hvis vi sammenlikner med den leksikografiske tradisjon vi kjenner fra den allmenne trykte definisjonsordboken, består ordnettets særpreg i følgende: a) Kjernen i lemmabestanden er i utgangspunktet nøkkelord til de ordforbindelser og sammensetninger ordnettet inneholder; b) Oppbyggingen av lemmalisten pågår kontinuerlig i takt med utarbeidelsen av det leksikografiske materialet; c) Lemmaene er entydige; d) Lemmaene kan klassifiseres og sorteres utfra ulike kriterier i samsvar med brukerens søkemotiv; e) Den leksikografiske beskrivelsen knytter sammen det semasiologiske og onomasiologiske aspektet. I tillegg kommer at bak lemmabestanden ligger en selvstendig fraseologisk beskrivelse ordnettet kan knyttes til.

## 5 Sluttord

Hvilken posisjon har så en ordboksbeskrivelse som er utformet på denne måte? Er den begrenset av den ramme som den her er blitt presentert innenfor, eller er det tenkelig å bygge den ut med flere komponenter? Innenfor ordnettet blir det ikke regnet med at enkelte lemmaer, enten de består av ett eller flere ord, vil få betydningsforklaringer eller definisjoner slik man kjenner det fra allmenne definisjonsordbøker. Men det er faktisk logisk at slike forklaringer og definisjoner blir tilføyet når det semantiske ordnettet er blitt ferdiggjort. Det henger ikke minst sammen med verdien av å kunne arbeide med definisjonene ut fra de opplysninger ordnettet kan formidle om forskjellige leksikografiske egenskaper til ordene, og om betydningsgrupperinger innenfor ordforrådet. Ordnettets vil også kunne fungere som veiledende verktøy ved organiseringen og framdriften av definisjonsarbeidet, med hensyn til hvilket ordtilfang skal prioriteres, hvilke ord eller lemmaer som ikke krever noen betydningsforklaringer, og kunne nøye seg med den kombinatoriske og klassifiserende informasjon som foreligger i ordnettet. Det ville gjenspeile den realistiske innstilling at lemmaene har ulikt behov for betydningsforklaringer, og at det finnes lemmaer som godt kan settes i en leksikografisk sammenheng uten å bli definert på en tradisjonell måte. En slik innstilling bidrar til å oppheve de strenge restriksjoner lemmabestanden gjerne blir utsatt for, og den medfører at oppmerksomheten i større grad får gjelde ordforrådet i sin fulle bredde og mangfoldighet. Ordnettets kan også veilede om en innbyrdes vurdering av de delbetydninger ordene opptrer i, f.eks. ut fra antall kollokasjoner og andre forbindelser vedkommende lemma er knyttet til.

Det er verdt å framheve at det ordforrådet som ligger til grunn for ordnettet, er blitt satt sammen ut fra helt andre vurderinger enn de som karakteriserer

lemmavalget i en allmenn ordbok. Ordnettet er i grunnen uavgrenset og åpent for tilføyelser og nye lemmaer, som dels vil oppstå under utarbeidelsen av materialet. Og det gir for såvidt et mer virkelighetstro bilde av ordforrådet og ordbruken at de mest sentrale og aktive ordene får størst vekt, de trekker til seg de fleste og mest varierende forbindelsene og de krever mest leksikografisk analyse. Ordnettet inneholder også viktige opplysninger som kan utnyttes ved vurdering og en eventuell markering av ordenes verdi innenfor ordforrådet i ulike sammenhenger. Her kunne f.eks. antall nøkkelord være en aktuell målestokk, og det kunne også være interessant å se hvilke ord som er mest karakteristiske for idiomer, hvilke konstruksjonstyper som kjennetegner bruken av enkelte verb osv.

Men overfor de muligheter leksikografen har til å bygge opp en helhetlig ordboksbeskrivelse med en fleksibel tilgangsstruktur og mangfoldige søke- og sorteringsrutiner for aktive og målbevisste brukere, står de ytre omstendigheter leksikografien og utgivelsen av ordbøker er avhengig av, der markedet bestemmer, og nye ordbøker må kunne appellere til de gjennomsnittlige ordboksbrukerne og oppfylle deres forventninger og behov. Og utgivelsen må forrente seg. Det er ikke gitt på forhånd, aller minst i våre små språksamfunn.

## Litteratur

- Dönsk-íslensk orðabók*. 1992. Ritstj. Hrefna Arnalds, Ingibjörg Johannesen. Reykjavík: Ísafoldarprentsmiðja.
- Ensk-íslenska orðabókin*. 2006. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2001. *Orðastaður. Orðabók um íslenska málnotkun*. Önnur útgáfa, aukin og endurskoðuð. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2002. *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2005. *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- Orðasambandaskrá Orðabókar Háskólans*: [http://www.arnastofnun.is/page/arnastofnun\\_ord\\_ordasambond](http://www.arnastofnun.is/page/arnastofnun_ord_ordasambond), 8. april 2008.
- STOB = Jón Hilmar Jónsson 2005.
- Þórdís Úlfarsdóttir. 2006. Málfræðileg mörkun orðasambanda. *Orð og tunga* 8:117–144.

Jón Hilmar Jónsson  
 forskningsprofessor, f. 1947  
 Stofnun Árna Magnússonar í íslenskum fræðum – Háskóli Íslands  
 Neshaga 16  
 IS-107 Reykjavík  
 jhj@hi.is



KJELD KRISTENSEN

## Svensk-Dansk Ordbog – tradition og fornyelse

At present, the Society for Danish Language and Literature is preparing a new large-scaled Swedish-Danish dictionary comprising about 50,000 entries. This paper explores the functionalities of the dictionary, the lemma selection and the structure of the entries from the perspective of lexicographic tradition and innovation. First and foremost, the dictionary is to be a L2-L1 reading comprehension dictionary for Danish users, either for consultation on specific comprehension or translation problems, or for general information on Swedish language. The lemma selection is made on the basis of occurrences in various monolingual Swedish dictionaries and bilingual Swedish-Danish dictionaries as well as electronically accessible text banks. Finally, three example entries are presented to demonstrate how the dictionary will provide information about Swedish language on a broader basis than traditionally done by internordic bilingual dictionaries.

Arbejdet med den nye, store svensk-danske ordbog har været i gang siden 2004, og det må være på tide at præsentere projektet i et nordisk leksikografisk forum. Det er ét af to internordiske ordbogsprojekter ved Det Danske Sprog- og Litteraturselskab; det andet er ISLEX, som også præsenteres ved denne konference. Med den svensk-danske ordbog bliver der taget et væsentligt skridt i retning af en bedre dækning med interskandinaviske ordbøger, sådan som den nordiske sprogdeklaration lægger op til. Projektet finansieres da også delvis af nordiske midler, fra Nordplus sprog og Nordisk Kulturfond, men især fra privat side takket være A.P. Møllers almenfond. Der er udsigt til at den svensk-danske ordbog kan udkomme om et par år, i første omgang som papirordbog. Ordbogen skal rumme mellem 45.000 og 50.000 opslagsord på ca. 1000 sider.

I foredraget omtales ordbogens koncept og nogle udvalgte emner, nemlig ordbogens funktionalitet, lemmaselektionen og ordbogsartiklernes struktur, set i forhold til leksikografisk tradition og fornyelse.

### 1. Ordbogens funktionalitet.

Begrebet funktionalitet er i stigende grad kommet i fokus inden for leksikografien og metaleksikografien i de senere år, og enhver leksikograf bør da også gøre sig overvejelser over hvem den ordbog man vil lave, henvender sig til, og hvilke sproglige aktiviteter den skal hjælpe de påtænkte brugere med. På Det Danske Sprog- og Litteraturselskabs hjemmeside står der bl.a. følgende i præsentationen af Svensk-Dansk Ordbog:

Ordbogen skal indeholde ca. 50.000 opslagsord, såvel farlige ord, lumske ligheder og andre svære ord som almindelige svenske ord der ikke eller kun ubetydeligt afviger fra de tilsvarende danske. Målgruppen skal være bred og vil i sagens natur især omfatte danskere med behov for at få oplysninger om svenske ord og udtryk; det er, foruden en stor almenhed, skoleelever, studerende, oversættere og ansatte i offentlige og private virksomheder med svensk-danske interesser. Det er tanken at ordbogen skal oplyse bredt om svensk sprog sådan at den ikke blot kan løse brugerens specifikke problem med dette eller hint bestemte svenske ord eller udtryk, men også give den interesserede bruger yderligere oplysninger [om svensk sprog].

Ordbogen er altså først og fremmest beregnet for danskere; det er den primære målgruppe. Hertil kommer islændinge, færinger, grønlandere og andre der har dansk som sproglig indgangsportal til det skandinaviske sprogområde. Ordbogens kildesprog er svensk, målsproget er dansk, og metasproget, det redaktionelle sprog, er dansk; hermed er ordbogen en L2-L1-ordbog beregnet for danskere. Funktionaliteten for ordbogen beskrives i præsentationen som det at oplyse om et givet problem inden for den del af sproget som man kalder leksikon, ordforråd. Det er en traditionel funktion for en L2-L1-ordbog; det gør ordbogen til en receptionsordbog, i første omgang en passivordbog. Men ligesom andre L2-L1-ordbøger kan Svensk-Dansk Ordbog have en aktiv funktion i forlængelse af den passive, nemlig ved oversættelse af svenske tekster til dansk. Skal man også tilgodese oversætterne, må man i endnu højere grad end ellers sørge for at de danske ækvivalenter virkelig er ækvivalente i flere dimensioner, både denotativt, konnotativt, stilistisk og pragmatisk. Det betyder at man for hvert opslagsord eller hver betydning skal spare på ækvivalenternes antal, så vidt muligt undgå definitioner og i stedet sigte mod det ene og rigtige ord eller udtryk der slår hovedet på sømmet. Vi vil altså gerne undgå den opremsning af flere mere eller mindre synonyme ækvivalenter der præger denne ordbogs forgænger.

I sidste punktum i citatet fra hjemmesiden står der at ordbogen skal oplyse bredt om svensk sprog og ikke blot hjælpe brugeren med dette eller hint specielle sproglige problem. I Svenséns "Handbok i lexikografi" (2004:36) er jeg stødt på en opdeling af ordbøger i *konsultationsordbøger* og *læseordbøger*. Og vel skal Svensk-Dansk Ordbog først og fremmest være et opslagsværk som brugeren konsulterer for at få løst et specifikt problem i forbindelse med sin beskæftigelse med svensk sprog, men vi vil altså gerne give ordbogens brugere noget mere end det han eller hun i første omgang var interesseret i. Det kan være oplysninger om synonymer, beslægtede ord, falske venner, afledninger og sammensætninger, og det kan være et bredere udvalg af eksempler end man er vant til i skandinaviske tosprogsordbøger, alt sammen for at tilfredsstille den bruger der er mere interesseret i svensk sprog end som så, fx studerende eller danske indvandrere i Sverige. Måske får brugeren lyst til at orientere sig bredere end i første omgang og lære sig mere om

svensk. Ordbogen bliver vel ikke en læseordbog af det, men alligevel håber vi at der her er tale om en frugtbar fornyelse af genren skandinaviske tosprogsordbøger. Når jeg senere viser et par ordbogsartikler, vil man se nogle eksempler på hvordan det bredere sigte med ordbogen kommer til udtryk.

Kan Svensk-Dansk Ordbog også bruges af svenskere, nemlig som L1-L2-ordbog til produktion af dansk, fx ved oversættelse fra svensk til dansk? Svenskeren vil finde ordbogen delvis utilstrækkelig til dette formål, for ordbogen indeholder ikke oplysninger om de danske ækvivalenter bøjning og om de danske substantivers køn, men disse oplysninger kan man dog forholdsvis let få over internettet, hvor den danske retskrivningsordbog og andre danske ordbøger er eller bliver tilgængelige. Der er heller ikke i Svensk-Dansk Ordbog oplysninger om de danske ords konstruktionsmuligheder bortset fra det der fremgår af de danske oversættelser til de faste svenske udtryk og svenske eksempler der er med i artiklerne. Svenskeres brug af ordbogen er en sekundær funktion i første omgang. Men hvis man senere som led i en elektronisk publicering kan linke de danske ækvivalenter til de danske opslagsord i Det Danske Sprog- og Litteraturselskabs ordnet-projekt, åbner der sig videre perspektiver for svenskeres aktive brug af den svensk-danske ordbog til produktion af dansk.

## 2. Lemmaselektion.

Hvordan finder vi frem til de 45.000-50.000 opslagsord som vi vil medtage i ordbogen? Vi kender ikke til nogen ordbase, elektronisk eller ikke-elektronisk, som er egnet til vores særlige formål, heller ikke det ordforråd der findes i den forholdsvis nye, store ordbog over moderne svensk der hedder Nationalencyklopedins ordbok (1995-1996) med over 60.000 opslagsord. Den har altså flere opslagsord med end vi kan bringe, og mere væsentligt: Den er overvejende konciperet ud fra kriteriet hyppighed i svensk. Det er udmærket, men vi må have et yderligere udvælgelseskriterium i spil, nemlig den kontrastive faktor, forskellen mellem svensk og dansk. Vi må sørge for at få de almindelige og måske lidt mindre almindelige svenske ord med i ordbogen som volder eller kan volde en dansker problemer. Så kunne man gå til eksisterende svensk-danske ordbøger og nøjes med ordforrådet i dem, men enten er de for gamle eller ganske vist nye, men for små. Der er ingen patentløsning på problemet med at tilvejebringe ordudvalget til den svensk-danske ordbog og af tekniske grunde slet ikke nogen elektroniske muligheder.

De to forskellige kriterier for lemmaselektion: hyppighed i svensk og forskel mellem svensk og dansk, kalder på to grupper ordbøger som leksikografisk grundlag for udvælgelsen af opslagsord: svenske modersmålsordbøger og svensk-danske ordbøger. Vi udnytter disse to grupper materiale ved manuel simulering af en automatisk metode som vi brugte på Den Danske Ordbog til udpegning af lemmakandidater. Som udgangspunkt bruger vi fotokopierede sider fra Natio-

nalencyklopedins ordbok, hvor vi så med forskellige tegn og farver markerer ordenes forekomst i de forskellige ordbøger. Mange markeringer tyder selvfølgelig på stærk kandidatur som opslagsord i Svensk-Dansk Ordbog og få eller ingen markeringer det modsatte. Der er yderligere nogle kilder som vi tager i betragtning ved afvejningen, først og fremmest forekomst på internettet undersøgt vha. Google indstillet på svenske tekster, og forekomst i Språkbanken, som er en meget omfattende svensk elektronisk tekstsamling bestående af avisartikler og fiktionsprosa.

Vi har nogle supplerende kilder, bl.a. Svenska Akademiens ordlistas seneste udgave og et par nyordsordbøger plus Svenska språknämndens årlige nyordslister. Vi sørger desuden for at få mange finlandssvenske ord og udtryk med.

Man kan vel sige at vores lemmaselektion sker ved at kombinere en traditionel metode der tager udgangspunkt i eksisterende ordbøger, med en mere moderne elektronisk sprogstatistisk metode, tilsat en god portion personligt skøn. Endelig har vi vores svenske gennemlæsere som bagstopperer hvis der skulle mangle nogle vigtige opslagsord.

### 3. Ordbogsartiklernes struktur.

Inden vi ser på et par ordbogsartikler, har jeg en enkelt vigtig oplysning om ordbogens makrostruktur. Som det vil være bekendt, alfabetiserer svenskerne Å før Ä og Ö, mens danskerne alfabetiserer Å efter Æ og Ø. Da Svensk-Dansk Ordbog primært er beregnet for danskere, bryder vi med traditionen og følger reglerne for dansk alfabetisering; dermed mener vi faktisk at vi servicerer den primære målgruppe bedst. Desuden følger vi den nye udgave af Svenska Akademiens ordlista i at regne W for et selvstændigt bogstav og ikke en variant af V.

**krage** *sb.* -n, kragar, kragarna

1 krave; **JF** *kråka* krage; *kragknapp*

2 flip; **SYN** *löskrage*

*kappa och krage* kjole og krave (: stilling som præst)

*ta sig i kragen el. ta sig själva i kragen* tage sig (selv) i nakken;

tage sig sammen;  $\diamond$  *nu måste du ta dig i kragen och söka ett nytt jobb.*

Nu til artiklerne, først den forholdsvis korte artikel **krage**, der dog har to betydninger og to faste udtryk eller idiomer. I artiklens første linje står opslagsord, ordklasse og bøjning; der er ingen udtale her da den følger de regler for omsætning fra svensk skrift til udtale der er kendte af danskere. Der er to betydninger med hver sin ækvivalent. **JF** markerer en struktur der viser at opslagsordet er en falsk ven; en dansker kan nemlig tro at opslagsordet betegner den fugl, *krage*, der på svensk hedder en *kråke*. Men den måde vi i øjeblikket angiver dette på, er en nødløsning og ikke noget der skal stå sådan i den trykte ordbog. Vi forestiller os i stedet at oplysningen om den falske ven kan gives således:



Det svenske ord *krage* betyder 'krave'.

Krage hedder *kråka* på svensk.

Denne boks kan bringes i forbindelse med artiklen: før, inde i eller efter – eller måske i en særlig sprogpædagogisk afdeling i ordbogen sammen med mange andre tilsvarende.

Under betydning 1 har vi førsteledssammensætningen *kragknapp* i fed kursiv; bemærk i øvrigt at vi tilstræber at svenske ord, udtryk og eksempler står med kursiv skrift, danske med ordinær skrift. Men vi medtager altså som noget vistnok nyt i internordiske tosprogsordbøger førsteledssammensætninger og i øvrigt også afledninger. Disse orddannelser fungerer som et supplement til ordbogens egentlige opslagsord; ofte står de på alfabetisk plads, og da de står med fed skrift, er de lette for brugeren at finde. I en elektronisk udgave kan de slås op direkte, og under alle omstændigheder kan man ved at opøge førsteledets danske ækvivalent i artiklen let finde ud af hvad de betyder, for andetledet er nemlig altid transparent, sommetider helt sammenfaldende med det tilsvarende danske ord. Endelig kan man sige at førsteledssammensætningerne og afledningerne i visse tilfælde understøtter forståelsen af opslagsordets betydning ud over hvad ækvivalenten siger.

Ved betydning 2 har vi et synonym, som vi altså gerne vil oplyse brugeren om.

Bemærk forklaringen i parentes og efter kolon ved det første faste udtryks danske ækvivalent. Her har vi et eksempel på at vi gerne vil hjælpe de brugere der ikke er fortrolige med en dansk ækvivalents betydning. Vi bruger også sådan en betydningsangivelse ved fremmedord og andre svære ord hvis betydning måske ikke er særlig kendt, fx af yngre ordbogsbrugere. Måske har brugeren slået ordet op i den tro at det er et særligt svensk ord, får så at vide at det på dansk hedder det samme eller noget andet som han eller hun ikke forstår, og er derefter lige klog. Her giver vi igen lidt mere.

**kaka** *sb.* -n, kakor, kakorna

1 kage (ofte mindre, uden fyld); **JF** *bakelse, tårta; kaktång*

2 fladt, rundt brød, evt. med hul i midten; **JF** *brödkaka, rågkaka*

● (*ovfö.*) kage (: samlet økonomisk udbytte til deling); ◇ *dela på kakan* dele kagen · *kräva sin del av kakan*

3 (*edb*) cookie (: datafil); **SYN** *cookie, kakburk, kakfil*

*baka/grädda en kaka* bage en kage

*fylld kaka* (*finl.sv.*) lagkage; **SYN** (*sv.*) *tårta*

*hela kakan* hele molevitten

*kaka choklad* (*tobak, ..*) plade chokolade (tobak, ..)

*kaka söker maka* krage søger mage

*statens kaka är liten men säker* (*ofte iron.*) en statsansættelse er tryk, men giver en ringe løn

*man kan inte både äta kakan och ha den kvar* man kan ikke både blæse og have mel i munden  
*mjuk kaka* skærekage  
*plocka russinen ur kakan* (ovfö.) skumme fløden (: vælge det bedste)  
*sju sorters kakor* overdådigt kaffebord  
*torr kaka* (især finl.sv.) skærekage; **SYN** (sv.) *mjuk kaka*;  
**JF** *småkaka utan fyllning* tørkage.

Næste artikel, **kaka**, har tre betydninger, betydning 2 har en overført underbetydning, og der er adskillige faste udtryk. Ved betydning 1 bør man som dansker ikke være tilfreds med at få at vide at *kaka* betyder 'kage'. Der er brug for en præcisering, anført i parentes, og ved at følge de indirekte henvisninger efter **JF** får brugeren mulighed for at få noget mere at vide om hvilke svenske ord der dækker spektret af det man på dansk kalder kager. Også her er der en førsteledssammensætning: *kaktång*. Ved betydning 2 må vi have en definition; der findes ikke nogen dansk ækvivalent.

Også her er der indirekte henvisninger til betydningsmæssigt beslægtede ord. Under betydning 2 er der en overført betydning hvor ækvivalenten er forklaret. To korte eksempler viser først en særlig svensk konstruktion, med dansk oversættelse, og derefter en brug der svarer til den danske. Ved betydning 3 ser vi at vi prøver at give brugeren lidt begreb om hvad en cookie er. Blandt de faste udtryk observerer vi to finlandssvenske udtryk: *fyld kaka* og *torr kaka*, begge med oplysning om hvad det hedder på standardsvensk, måske lidt overflødig. Under *torr kaka* er der oplysning om falsk ven.

**kant** sb. -en, -er, -erna

- 1 kant; ◇ *skarpa kanter · fyll upp till glasets kant; kantsten*
- 2 margin; **SYN** *marginal(1)*; ◇ *göra anteckningar i kanten* skrive notater i marginen
- 3 del (af en by); egn; retning; **JF** håll; ◇ *i östra kanten av staden*
- 4 skorpe; **SYN** *skorpa*; ◇ *kanten av brödet · kanten av osten*  
*fin i kanten* fin på den; kræsen  
*bålla sig på sin kant* være reserveret; passe sig selv  
*kant i kant* kant mod kant; ◇ *numera sätts nästan alla tapeter kant i kant* efterhånden sættes næsten alle tapeter kant mod kant  
*kant i kant med ngt* i flugt med ngt  
*nagga ngt i kanten* (ovfö.) slå skår i ngt; ◇ *hans rykte har naggats i kanten efter skandalen* der er slået skår i hans rygte efter skandalen  
*plus i kanten* (dagl.) stor ros; stor fordel  
*på alla håll och kanter* overalt  
*ställa på kant* stille på højkant  
*vara med på en kant* være lidt med (: deltage, men lidt forsigtigt).

Sidste artikel, **kant**, blev ikke omtalt i foredraget. Her er der fire betydninger og adskillige faste udtryk. Under betydning 1 er der to korte eksempler samt en førsteledssammensætning, *kantsten*, der kan forekomme banal; men så stor som ordbogen bliver, kan den ikke kun vise hvad der er forskelligt i de to sprog, den må også pege på de ligheder der er, i dette tilfælde at *kantsten* faktisk betyder det som man som dansker skulle tro. Ved betydning 2 er der angivet et synonym, ved betydning 3 et betydningsbeslægtet ord, *håll*, og ved betydning 4 et synonym. Bemærk at et par af eksemplerne er forsynet med dansk oversættelse; det gør vi fx hvis der er en interessant konstruktionsforskel i de to sprog, en forskel vi gerne vil gøre opmærksom på, eller hvis der er ét eller flere svære ord i det svenske eksempel. Det giver igen brugeren mulighed for at lære sig lidt mere om svensk end det han eller hun umiddelbart søgte.

### Litteratur

*Nationalencyklopedins ordbok 1-3*. Höganäs: Bokförlaget Bra Böcker 1995-1996.  
Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Kjeld Kristensen  
Cand.mag., ledende redaktør, f. 1943  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
kk@dsl.dk



PATRICK LEROYER

## Maultasche og Novillada? En teori for turistordbøger

In addition to tourist guides and books, millions of information hungry tourists all around the world use dictionaries referred to as ‘travel dictionaries’, including phrase books containing glossaries and lexicographically structured information. Also, software companies propose multilingual, downloadable dictionary solutions that can be printed on demand or consulted on a PDA or a WAP phone. In brief, the lexicographic commercial supply and demand are enormous. However, no specific lexicographic theory has ever been developed in this field of lexicography which remains scientifically neglected. The purpose of this article is to develop a specific theory for tourist dictionaries and thereby lay the foundation of tourist lexicography as a new field of dictionary research and practice. In order to demonstrate its potential, the theory will be applied to a prototypical tourist dictionary concept. The core of the theory is to bring targeted, useful cognitive data types in tourist dictionaries, in particular cultural words such as ‘Maultasche’ (huge ravioli served in Swabia, Germany) and ‘Novillada’ (corrida for young bulls in Spain and Southern France).

### 1. Rejseordbøger – fra definition til illusion

#### 1.1. Leksikografisk definition

Rejseordbøgerne<sup>1</sup> er givetvis blandt de mest udbredte ordbøger i verden. Det leksikografiske udbud er enormt, den leksikografiske efterspørgsel lige så. For folk, der rejser, er fundamentalt set nysgerrige. De vil gerne i tæt kontakt med de destinationer, de besøger. De vil gerne opleve, nedbryde sprogbarrierer, lære om fremmede kulturer og deres befolkning, og har derved tilsyneladende ubegrænsede, interkulturelle og interlingvale informationsbehov.

I teorien er rejseordbøger og rejseparlører beregnet til at yde fremmedsproglig assistance til brugere, der rejser over landegrænser og har brug for hjælp til at løse praktiske kommunikationsproblemer i det fremmede land. Assistanzen består som regel af en oversættelse til og fra modersmål af de specielle turistrelevante ord og sætninger, der kan indgå i en konversation. I Nordisk Leksikografisk

---

1 I denne artikel vil der blive fokuseret på de ‘fuldgyldige’ leksikografiske produkter – rejseordbøger og parlører – og der vil blive set bort fra ‘hybridprodukter’ såsom rejsebøger, -guides og -håndbøger, der ofte indeholder en beskedent ordbogs- eller parlørdel, som fx rejsebøgerne fra Politikens forlag i den velkendte serie “Turen går til...”.

Ordbok (1997) er rejseordbøger og parlører defineret ud fra deres form og substans, deres sigte og deres målgruppe. Målgruppen anføres først og defineres som brugere, der ingen eller kun meget begrænset viden har om fremmedsproget i forvejen.

#### reiseordbok

- Reiseordboken er tospråklig (eller flerspråklig) og beregnet på personer som har ingen eller liten kunnskap om det fremmede språket. Det er en typisk konsultasjonsordbok, og hovedhensikten er å løse praktiske kommunikasjonsproblemer.
- Reiseordboken må ikke forveksles med en kortfattet tospråklig allmennordbok; dels er reiseordbokens utvalg og presentasjon av leksikalske enheter klart innsiktet på det å reise, dels er materialet ofte disponert på en særskilt måte [...].

#### Eksempel 1. *Artiklen 'reiseordbok' i NLO*

Samme teoretisk udgangspunkt finder man hos Abend (1991: 2903), som anfører: "Das Reisewörterbuch ist als lexikographisches Hilfsmittel für Reisende gedacht, die nur geringe oder keine Kenntnisse der Sprache ihres Urlaubslandes besitzen." Det særligt interessante er dog, at Abend ikke henviser til fremmedsproget som sådan, men til det sprog, der tales i ferielandet. Dette åbner op for ny teoretisk tænkning, da associationen af sprog og ferieland peger i retningen af en mere nøjagtig beskrivelse af de potentielle ordbogsbrugeres behov, herunder sproglige behov:

- De potentielle brugere er ferierejsende, der udøver en turistaktivitet
- De potentielle brugere besøger et bestemt ferieland eller en landsdel, hvortil der kan associeres et eller flere sprog

#### 1.2 *Den leksikografiske illusion i skrift og tale*

Rejseordbøgerne med parlørdel og parlører med ordbogsdel er beslægtet med sætningsordbøgerne til skrift og tale, som Hausmann (1989) sporer tilbage til sekstenhundrede tallet. Den leksikografiske grundtanke er at hjælpe brugeren med sætningsdele eller færdige sætninger, der kan stykkes sammen til at skrive med (som i STK 1901 eller DSS 1956) i bestemte situationer, specielt handelskorrespondance, eller kan bruges direkte til at tale med i specifikke turistsituationer, fx på hotel og restaurant. Skønt besnærende virker denne pragmatiske, kommunikative tankegang dog sproglig set temmelig naivt. Hovmark (2005) har ret i sin anmeldelse af tre franske ordbøger til rejsebrug, når han gør opmærksom på, at "forestillingen om at en parlør(del) kan/skal bruges af brugere med absolut minimal kompetence, holder næppe i praksis". Men Hovmark fører desværre ikke tanken til ende: Det turde være indlysende for enhver, at ingen rejseordbog eller parlør i verden nogensinde vil kunne hjælpe en turist, som ingen eller kun meget lille kendskab har til det fremmedsprog, der tales i ferielandet.

Britiske Monty Python (Chapman [Her med to –n-, i litteraturlisten med ét] et al: 1970) var de første, der påviste den leksikografiske illusion i deres berømte

sketch – ‘den ungarske parlør’ – hvori en ungarsk turist i London kommer i håndgemæng i en butik på grund af en sjofel parlør. Under det retslige efterspil er det udgiveren, der sidder på anklagebænken:

- Bailiff: That’s fine. On the 28th of May, you published this phrase book.
- Publisher: I did.
- Bailiff: I quote an example. The Hungarian phrase meaning ‘Can you direct me to the station?’ is translated by the English phrase ‘Please fondle my bum.’
- Publisher: I wish to plead incompetence [...].

Det, som Monty Python illustrerer *ex absurdo*, er, at parlører faktisk er temmelig ubrugelige. Inkompetencen er af metaleksikografisk art og skyldes et teoretisk underskud inden for området rejseordbøger, som Abend (1991: 2908) også erkender – “Die Wörterbuchforschung hat dem RWB (Reisewörterbuch) bisher nur wenig Aufmerksamkeit geschenkt” – og foreslår løst på følgende måde: “Eine stärker an der Praxis und am Wörterbuchbenutzer orientierte Wörterbuchforschung wird das Reisewörterbuch in Zukunft stärker berücksichtigen müssen”. At undersøge brugen af rejseordbøger i praksis og/eller spørge brugerne vil dog ikke kunne føre til nævneværdig teoretisk udvikling. En sådan udvikling må nødvendigvis komme fra en klarlæggelse af ordbøgernes formål og af deres funktioner (jævnfør Bergenholtz/Tarp 2005 og specielt Tarp 2006). Men først skal status quo ganske kort undersøges.

## 2. Den leksikografiske status quo – ‘innsiktet på det å reise’?

### 2.1. Udvalgte lemmata

Rejseordbøgernes erklærede formål er at løse praktiske kommunikationsproblemer og er derfor pragmatisk opbygget i vante temaer. De tager udgangspunkt i de konventionelle rejsesituationer og dermed associerede kommunikationsbehov, der typisk vil kunne opstå under turen: overnatning på hotel (bestille et værelse), restaurantbesøg (bestille mad og drikke), sightseeing (bede om et kort eller program på det lokale turistbureau) osv. Problemet er dog, at listen over rejsesituationer er ubegrænset, da alle situationer og dermed alle kommunikationsproblemer i princippet kan forekomme under rejsen. Velkomstsiden på online Dansk Parlør (DP 2007) illustrerer afgrænsningsproblemet, med en lemmabestand, der har karakter af en leksikografisk rodebutik:

**Strand, Svinekød, Bil, Grøn, Butik, Blå, Opad, Nord, Grænse, Gå, Pas, Til venstre, Hus, Spiseseddel, Hotel, Kobling, Vindue, Udsigt, Køre, Flyvemaskine, Lufthavn, Station, Høne, Pris, Tjener, Søge, Hospital, Hjælpe, God, Hund, Edderkop, Seng, Danmark, Syg, Vand, Vej, Værdifuld, Kone, Hjem, Penge, Brød, Vandfald**

Rejser man i bil, og er koblingen defekt (eller gået i stykke), kan det jo være rart at kunne slå op og sige 'kobling', når bilen er på det udenlandske værksted. Man kan muligvis også klage over, at der er 'edderkopper' på hotelværelset... Faktisk kan man antage, at netop det at rejse udvider antallet af lemmata, der er 'innsikket'... Da antallet af rejsesituationer er ubegrænset, synes lemmaantallet derved også at være ubegrænset.

Det er dog ikke tilfældet i andre rejseordbøger, der synes at anvende en reduceret udgave af den almindelige lemmaliste fra de større bilingvale ordbøger. Et tilfældigt opslag i ordbogsdelen af Fransk-Dansk/Dansk-Fransk Ordbog i handy format (FDDFO) afslører et begrænset antal rejserelevante lemmata; her findes kun **restaurant, resto, restoroute**, der umiddelbart kan associeres til en rejseaktivitet – alle de andre er der ikke specifikt behov for i en rejsesituation.

**resquilleur, ressac, ressaisir, ressemblance, ressemblant, ressembler, ressemeler, ressentiment, ressentir, resserrer, ressort, ressortir, ressortissant, ressource, ressusciter, restant, restaurant, restaurateur, restauration, restaurer, reste, rester, restituer, restitution, resto, restoroute, restreindre, restreint, restrictif, restriction, restructuration, résultat, résulter, résumé, résumer, résurrection, retable**

Da bogens handy format begrænser den plads, der er til rådighed, er den udvalgte information, man i øvrigt kan finde under de relevante opslagsord, temmelig spartansk og har temmelig begrænset nytteværdi:

**restaurant** m. restaurant; □~  
*d'entreprise* kantine

## 2.2. Udvalgte sætninger

Samme udvælgelsesproblematik finder man i parlørerne og i ordbøgernes parlørudel, som udgør den 'særskilte måde' at præsentere de leksikografiske datatyper på. Disse datatyper er som regel konversationssætninger, der er omgrupperet i typisk forekommende situationer, hvor sætningerne kan komme til deres ret. Mange af disse sætninger er i øvrigt udformet som spørgsmål og anmodninger, som her i FPBD fra Berlitz, der er beregnet til at skulle hjælpe engelske brugere:

Can you suggest?	<b>Pouvez-vous nous conseiller?</b>
	<i>Poovay voo noo kawngsayay</i>
things to do	<b>quelque chose à faire?</b>
	kelker shoaz a fehr?

På denne baggrund rejser den leksikografiske "parlørtradition" to spørgsmål:

1. Hvad er kriterierne for udvælgelse af sætningerne? Er alle situationer – og alle sætninger – da ikke tænkelige?<sup>2</sup>
2. De franske Guides de conversation Routard (GCR 2007) anfører, at alle de situatio-



2. Hvordan kan turistbrugeren blive forstået? Og hvis turistbrugeren bliver forstået, hvordan vil han eller hun kunne forstå det svar, der uvilkårligt vil blive givet?

Det manglende leksikografiske informationsbehov synes at gælde for mange af de selekterede sætninger. Dette gælder fx i FDDFO, hvor der næppe er behov for at vide, hvordan man spørger, om den franske taxa er fri, da taxaen normalt skilter med det, eller holder på en taxaholderplads. At assertere, at man er sulten eller tørstig, kan heller ikke bruges til ret meget (og under ingen omstændigheder til at bestille mad eller drikke):

■ **Taxa**

**Er denne taxa fri?**

Vous êtes libre?

■ **ALMINDELIGT VELBEFINDENDE**

**J'ai faim/soif**

Jeg er sulten/tørstig

Eksempel uddrag fra FDDFO

### 2.3. Udtaleangivelser

Lyd udgør i sagens natur et særligt problem i rejseordbøgerne, der som hovedfunktion skal hjælpe brugerne med at løse specifikke problemer i forbindelse med produktion og reception af mundtlige tekster. Ordbøgerne kan vælge én af følgende fire muligheder: ingen angivelser, IPA-notation, tilpassede udtaleangivelser eller lydfiler.

Skulle en dansk kvinde komme i den ubehagelige situation at blive plaget af en sexchikanerende franskmand, kunne hun forsøge at løse problemet med at slå op og fremsige følgende sætning fra PFR (2005: 113) – problemet her er dog, at der mangler udtaleangivelser, og man må blot håbe, at vedkommende kan klare sig uden:

Min mand kommer lige om lidt **Mon mari va venir tout à l'heure**

I GSFP (2005:46) findes lignende sætninger, der kan bruges – med og uden eftertryk! – og med IPA-notation. Her synes brugeren at være bedre hjulpet:

**Lad mig være i fred!**

Laissez-moi tranquille, je vous en prie [lese mwa trākil zvuz\_ā pri]

(med eftertryk) Foutez-moi la paix ! [fute mwa la pɛ]

ner, der kan opleves under rejsen, er blevet undersøgt og sorteret, og at nyttige sætninger er nøje udvalgt: "Nous avons passé au crible toutes les situations vécues au cours d'un voyage. Les phrases utiles ont été minutieusement choisies..." De egentlige kriterier for bestemmelse af nytteværdien forbliver dog i det uvisse.

Problemet med IPA-notationen er, at brugeren skal være lingvist for at kunne udnytte angivelserne (selv lingvisterne vil dog kunne komme til kort og mangle artikulationsbeskrivelserne af de særlige fonemer, som er helt specielle for det pågældende fremmedsprog). Problemet er også, at brugeren helst skal kunne anvende elementær intonation, prosodi samt en lang række andre fonologiske fænomener, hvis brugeren skal gøre sig sikkert håb om at blive fuldstændig forstået og opnå den tilsigtede virkning.

De tilpassede udtaleangivelser virker umiddelbart mere brugervenlige, men kan næppe garantere opnåelse af fælles forståelse, som fx i den allerede citerede sætning fra Berlitz parlør, hvor den engelske bruger skulle forsøge sig med en "Poovay voo noo kawngsayay kelker shoaz a fehr".

Den fjerde og sidste mulighed er lydfiler. I fremtiden vil disse sikkert blive meget mere udbredte, og kunne hjælpe turistbrugerne i en vis udstrækning. Brugeren behøver nemlig ikke at efterligne lyden, men kan blot nøjes med afspille ord og sætninger direkte på sin PDA eller WAP-telefon i en kommunikations-situation. Lydfiler i sig selv er dog ingen garanti for, at de praktiske kommunikationsproblemer hermed er løst. Den mere eller mindre forudsætningsløse bruger vil stadig være på bart bund og ude af stand i at forstå, hvad der bliver sagt eller svaret i konversationen. Samtalen forbliver en utopi.

### 3. Turisternes leksikografisk relevante informationsbehov

#### 3.1. *Det genuine formål*

Brugen af termen 'rejseordbog' i leksikografien er beklageligt, for det skygger for den virkelige leksikografiske situation. Folk rejser ikke blot for at rejse. De rejser for at leve og opleve, de rejser, fordi de er turister. De fleste turister i dag er veluddannede og er oprigtig nysgerrige over for de destinationer, de besøger. Grundlæggende er turisme en social aktivitet<sup>3</sup>, der er motiveret af den menneskelige nysgerrighed og af lysten til at lære om de destinationer, der besøges, både før og under turen. Turisterne søger på en eller anden måde at finde frem til destinationens sande identitet, de er nysgerrige og vil gerne vide noget om selve destinationen (geografi, historie, erhverv osv.), om det eller de sprog, der tales, om befolkningen og kulturen. Turisterne vil gerne kommunikere med de folk, de besøger i det fremmede, selvom kommunikationen på fremmedsproget i langt de fleste tilfælde forbliver en utopi, da det netop kræver et godt forhåndskendskab til det pågældende sprog. Turisterne vil også gerne opleve destinationen og har brug for opdaterede informationer om lokale begivenheder, lokal politik osv.

3 Det skal bemærkes at gængse definitioner af turisme – faglige eller lingvistiske (se fx World Tourism Organization 2002 eller den franske almensproglig ordbog LNPR 1993) meget påfaldende har fokus på selve rejseaktiviteten. Dette kan være forklaring på, hvorfor leksikografien har fokus på rejse og rejsende.

Det er indlysende, at turistordbøger umuligt kan tilfredsstillende alle turistrelevante informationsbehov, og der bør derfor skelnes skarpt mellem leksikografisk relevante behov og leksikografiske irrelevante behov. Eksempelvis skal turistordbøger ikke formidle de turistinformationer, der har en litterær eller en reklamemæssig dimension: disse formidles bedst af brochurer, kataloger, rejsebøger og rejsebeskrivelser, turistguides, samt turistkontorerne hjemmesider.

Dette fører til opstillingen af tre nye teser samt en definition af turistordbøgernes genuine formål. Definitionen af det genuine formål er en tilpasning af Tarps generelle definition (2006: 107):

#### **Turistordbogens genuine formål**

Det genuine formål af en turistordbog er tilfredsstillende af de typer leksikografisk relevante behov, som kan opstå hos en eller flere typer af potentielle turistbrugere i en eller flere typer turistsituationer.

### *3.2. Tre nye teser*

Den første tese er af begrebsmæssig karakter og tager udgangspunkt i selve ordbogsbenævnelsen. Termen 'rejse' er upræcis og giver ikke megen mening i leksikografisk sammenhæng. Med andre ord er rejseordbog et uklart begreb. Da ordbøgerne er værktøjer, der skal bruges af bestemte brugere, bør rejseordbøgerne omdøbes til hvad de i virkeligheden er: turistordbøger, dvs. værktøjer der skal hjælpe turisterne, fordi at være turist netop er relateret til en meget specifik social og kulturel aktivitet, hvor ordbogen kan bringe megen relevant hjælp. Hermed er det brugeren og brugerens behov<sup>4</sup>, der kommer i centrum:

#### **Tese 1**

##### **Rejseordbøger bør omdøbes til turistordbøger**

Det leksikografiske begreb 'rejseordbog' er videnskabeligt skadeligt og bør erstattes af begrebet 'turistordbog', der klart rummer essensen af denne type ordbogs genuine formål.

Den anden tese tager udgangspunkt i de påtænkte ordbogsfunktioner. De traditionelle rejseordbøger og parlører kan ikke løse turisternes praktiske kommunikationsproblemer, men kan kun i bedste fald yde kommunikativ nødhjælp i kommunikative nødsituationer.

#### **Tese 2**

##### **Turistordbøgerne er ikke velegnet til at opfylde kommunikative funktioner**

Turistordbøger, der er beregnet for turistbrugere med ingen eller ringe kendskab til det fremmedsprog, der tales i turistlandet, kan ikke løse turisternes praktiske kommunikationsproblemer i forbindelse med mundtlig

<sup>4</sup> På samme måde refererer man til lærerordbøger, fordi fokus ligger på lærere og deres specifikke behov (se Tarp 2004).

tekstproduktion og tekstreception, men kan kun i yderst begrænset omfang yde kommunikativ assistance.

Den tredje tese flytter fokus fra kommunikativ assistance til kognitiv assistance:

### Tese 3

#### **Turistordbøgerne er velegnet til at opfylde kognitive funktioner**

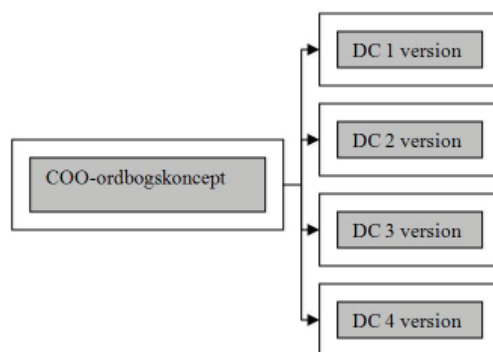
Turistordbøgerne kan yde værdifuld assistance til turistbrugerne i alle de specifikke turistsituationer, hvor turisternes leksikografisk relevante vidensbehov kan dækkes ved at bringe kontrastive, turistvidensrelaterede datatyper: dette gælder først og fremmest sproglig og kulturel viden.

## 4. Perspektivskift – fra ordbogsstandardisering til ordbogslokalisering

Bestemmelsen af turistordbøgernes genuine formål og af deres ordbogsfunktioners arbejdsfordeling medfører et perspektivskift, der ændrer præmisserne for udarbejdelsen af nye ordbogskoncepter. Perspektivskiftet er forankret i en erkendelse af, at turisternes leksikografiske behov bedst kan identificeres og analyseres i et kontrastivt, lokaliseret perspektiv. Der skelnes i det følgende mellem det land, turisterne kommer fra – med den internationale kommunikationsbetegnelse ‘Country of origin’ (=COO) – og det land, turisterne besøger, ‘Destination Country’ (=DC).

### 4.1 Ordbogsstandardisering

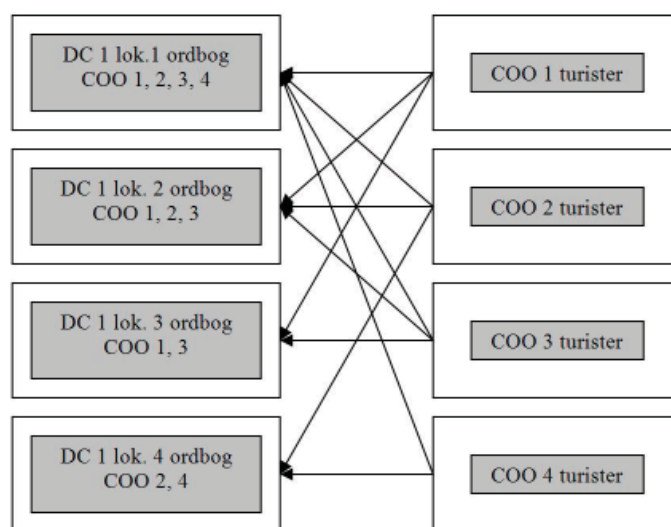
Konventionelle turistordbøger udarbejdes normalt i serier som forskellige sproglige versioner af én og samme ‘turen går til’ ordbogskoncept. Figur 1 viser, hvordan en COO-ordbogskoncept afføder en række DC-versioner af samme koncept. Trods tilpasningen til de forskellige sproglige versioner er der tale om en ordbogsstandardisering, der vanskeliggør indarbejdelse af de væsentlige kulturforskelle, som er så centrale for turisterne, og som Hovmark (2005) netop efterlyser.



Figur 1. Ordbogsstandardisering

#### 4.2 Ordbogslokalisering

I det nye ordbogsperspektiv er ‘turisterne går til’ byttet med ‘turisterne kommer fra’. Det er den destination, der modtager turister, der bedst kan udarbejde de forskellige ordbøger, afhængigt af hvor turisterne kommer fra og af turisternes leksikografisk relevante behov, når de befinder sig på destinationen. Perspektivet åbner op for en nøjagtig lokalisering i henhold til turisternes behov, og for den præcise, kontrastive indarbejdelse af kulturinformationer. Figur 2 illustrerer, hvordan der i samme destinationsland (DC1), men i 4 forskellige lokaliteter (lok. 1, 2, 3, 4) kan udarbejdes forskellige ordbøger, der hver især inddrager forskellige COO-turistkombinationer.



Figur 2. Ordbogslokalisering

### 5. Sprogkombinationer og ordbogstyper

Som tidligere anført åbner Abends (1991) begrebsassociation ‘sprog’ og ‘ferieland’ op for en mere nøjagtig og nuanceret beskrivelse af de potentielle ordbogsbrugeres sproglige behov. Disse behov kan nu indsættes i et COO-/DC-perspektiv. Tabel 1 viser den leksikografiske relevans af turisternes sproglige behov afhængigt af sprogkombinationer i de enkelte lande eller landsdele, og anviser den ordbogstype, der bedst kan opfylde disse behov.

I første kolonne – COO sprogbrugere – er opstillet turisternes sprog – både modersmål (L1), fremmedsprog (LX), og lingua franca (LF), som oftest er engelsk.

I anden kolonne – DC-sprog – er opstillet status af det eller de sprog, der tales på turistdestinationen.

I tredje kolonne er opstillet den leksikografiske relevans af de kommunikative datatyper. I flere tilfælde er relevansen afhængig af den aktuelle sprogpolitik i de forskellige DC-lande eller landsdele.

I fjerde og sidste kolonne er anvist de ordbogstyper, der bedst modsvarer de forskellige brugersituationer og sprogkombinationer.

COO sprog	DC sprog	Leksikografisk relevans af de kommunikative datatyper	Ordbogstype
(1) L1	L1	Ingen	(Monolingval)
(2) L1 andet	L1	Yderst begrænset	(Monolingval)
(3) L1	L1+L1 lokal variation	Begrænset	Varietal
(4) L1	L1+L1 tæt regional	Afhængig af sprogpolitik	Varietal
(5) L1	L1+LX regional	Afhængig af sprogpolitik	Bilingvaliseret
(6) L1	L1+LX truet	Afhængig af sprogpolitik	Bilingvaliseret
(7) L1	L1 tæt national	Afhængig af sprogpolitik	Varietal
(8) L1	L1 tæt national + LXa, LXb, ..., LXn	Afhængig af sprogpolitik	Varietal + multilingvaliseret
(9) L1	LX	Stærkt begrænset	Bilingvaliseret
(10) L1 +LX lørn	LX	Ingen	Lørner
(11) L1 + LF	LX + LF	Afhængig af status for LF	Bilingvaliseret

*Tabel 1.* Sprog, leksikografisk relevans og ordbogstyper

Da pladsen ikke tillader en fyldestgørende gennemgang af de mange sproglige, sprogpolitiske og leksikografiske parametre, skal de enkelte kombinationer blot ganske kort introduceres.

I situation (1) er turistbrugerens ferieland et L1 land, hvorfor behovet for sproglig hjælp er irrelevant.

I situation (2) har turistbrugeren L1 som andet sprog, og ferieland er også et L1 land – behovet for sproglig hjælp er yderst begrænset.

I situation (3) besøger turistbrugeren en landsdel, hvor der også tales dialekt – behovet for hjælp er begrænset, og kan klares af en ordbog, der har fokus på sprogvariation.

I situationer (4)-(8) er ferielandet eller landsdelen et multilingvalt samfund<sup>5</sup>.

<sup>5</sup> Eksempler på sådanne er Spanien og catalansk (4), Spanien og baskisk (5), Bretagne

Relevansen af hjælpen afhænger i hvert enkelt tilfælde af de aktuelle sprogpoltikker i de enkelte lande eller landsdele.

I (9) har vi en situation, hvor en L1 turistbruger besøger en LX destination. Dette er faktisk grundlaget for langt de fleste rejseordbøger og parlører. Som allerede anført er den leksikografiske relevans af de kommunikative datatyper stærk begrænset. Den type ordbog, der passer her, er ikke en konventionel, bilingval ordbog, men en L1-ordbog, der er bilingvaliseret (dvs. en L1 ordbog med en beskeden LX dimension), se i øvrigt Tarp 2005 for en diskussion af begrebet 'bilingval ordbog' og dets glidende overgange.

I (10) har vi en situation, hvor en L1 turistbruger besøger en LX destination, men hvor turistbrugeren også er lærer af LX. Turistbrugeren har derfor ikke brug for en bilingval turistordbog, idet ordbogsbehovet her bør dækkes af en lærerordbog.

I (11) har vi endelig en situation, hvor en L1 turistbruger behersker et lingua franca (LF): der åbnes derved mulighed for udarbejdelse af turistordbøger, der, afhængigt af det aktuelle LF og dets status i destinationslandet, i en vis udstrækning kan udarbejdes på det pågældende LF (fx tysk i Kroatien).

## 6. Kognitive behov og videnstyper

Tabel 2 viser turisternes leksikografiske vidensbehov, afhængigt af de forskellige typer af viden og af, hvornår ordbogen bruges – før eller under ferieturen.

Videnstyper	Leksikografisk behov	Leksikografisk relevans	Tidsfaser
(1) Encyclopædisk	-	-	Før
(2) Promotionel	-	-	Før og under
(3) Sproglig	(+)	+	Før og under
(4) Praktisk	+	+	Under
(5) Faktuel	++	++	Under
(6) Kulturel	++	++	Under

Tabel 2. Videnstyper, leksikografisk behov og relevans, tidsfaser

Turistviden kan opdeles i følgende videnstyper:

1. **Encyklopædisk** – omfatter den geografiske, økonomiske, historiske osv. viden om turistdestinationslandet, som er genstand for den konventionelle encyklopædiske dataselektion og datastrukturering. Denne type viden er uegnet til præsentation i turistordbøger, og konsulteres normalt før afrejsen til turistlandet.

og bretonsk (6), Syd og Central Amerika og spansk (6), Québec og fransk (7), Sydafrika og engelsk, afrikaans, bantu sprogene (8).

2. **Promotionel** – omfatter den viden om de reklamebudskaber, som man finder i turistbrochurerne og turistfoldere, rejsekataloger, turisthjemmesider mm.
3. **Sproglig** – omfatter viden om selve sproget.
4. **Praktisk** – omfatter her og nu viden om praktiske forhold, såsom lægekonsultationers adresse og åbningstid, apotekernes adresse og åbningstider, osv.
5. **Aktuel** – omfatter viden om aktuelle events såsom koncerter, udstillinger mm.
6. **Kulturel** – omfatter viden om destinationens skik og brug, traditioner, moderne selvopfattelse, mm.

Det er kun de sidste 3 typer viden, hvor informationsbehov og -relevans er leksikografisk i top i turistordbøgerne.

- den praktiske
- den aktuelle
- den kulturelle

## 7. Online turistordbog – datastrukturering og dataindhold

Det ligger i turistordbogens essens, at ordbogen skal lokaliseres. På landsplan, på regionsplan eller på lokalt plan.

Til illustration af teorien i praksis er valgt den sydfranske by Nîmes. Byen, der ligger i Languedoc Roussillon region, er her taget som eksempel på en by, hvis turistkontoret med fordel kunne tilbyde en online ordbog, der henvender sig til sine turistgæster, der ikke kan fransk.

En sådan ordbog vil nødvendigvis skulle bestå af en række monolingvale ordbøger, på turisternes modersmål (L1), og med en minimalistisk bilingval dimension (L1-L2)<sup>6</sup>. En sådan ordbog vil nødvendigvis indeholde både kommunikative og kognitive datatyper.

De kommunikative datatyper kan underopdeles i to kategorier: de datatyper, der altid er brug for, og de datatyper, der kun er brug for i nødsituationer.

I begge tilfælde kan den leksikografiske strukturering opbygges som følger:

- L1 lemma
- L2 ækvivalent
- L2 udtale og intonation på lydfil
- Evt. pragmatisk note på L1

<sup>6</sup> Ordbogen må nødvendigvis også indeholde en funktionel brugervejledning/online hjælp (se fx Nielsen 2005), da brugerne ikke er fortrolige med det nye koncept.



### 7.1. Kommunikative datatyper – altid brug for i alle situationer

De kommunikative datatyper, som der altid er brug for, er de franske sætninger, som fungerer som markører og signaler for høflig taltale, som fx i følgende:

**au revoir (monsieur, madame)**  
**bonjour (monsieur, madame)**  
**merci**  
**merci beaucoup**  
**non merci**  
**pardon**  
**s'il vous plaît**  
**s'il vous plaît!**  
**excusez-moi, je ne comprends pas le français.**  
**excusez-moi, je ne parle pas le français.**  
**parlez-vous l'allemand, l'anglais, l'espagnol, ...?**

### 7.2. Kommunikative datatyper – kun brug for i nødsituationer

De kommunikative datatyper, som der kun er brug for i nødsituationer, bør omfatte data, der kan hjælpe turisterne i nødsituationer, fx under sygdom og lignende. Disse datatyper kan kun anvendes som nødkommunikation, da det ikke er mening, de skal bruges i en samtale. Dataindholdet vil fx kunne omfatte følgende:

**brûlure**  
**coup de soleil**  
**diarrhée**  
**grippe**  
**migraine**  
**piqûre d'insecte**  
**rage de dent**  
**rhume**

### 7.3. Kognitive datatyper – lemmatisering af 'kulturonymer'

De kognitive datatyper skal hjælpe turistbrugerne med at bringe løsninger til de leksikografisk relevante vidensbehov, der kan opstå i mødet med den fremmede kultur. Der er derfor tale om interkulturel viden. Den leksikografiske strukturering kan opbygges som følger:

- L2 lemma
- Kulturnote redigeret af lokal ekspert på L1<sup>7</sup>
- Illustration eller mediaclip

7 Et godt eksempel på, hvordan kulturnoter kan opbygges kan ses i den fransk-engelsk/engelsk-fransk elektronisk ordbog LGRCE, der netop anvender disse som særskilte lister og søgeoptioner.

- Udvalgte eksterne links på L1, der opfylder en klar vidensfunktion<sup>8</sup>

Her skal alene bringes nogle få eksempler, der illustrerer selektion og lemmatisering af de data, der understøtter turistrelevant vidensbehov om byens kultur. De ord og fraser, der selekteres og lemmatiseres, kunne passende kaldes for 'kulturonymer', dvs. ord og fraser der ikke alene er bærere af en kulturviden<sup>9</sup>, men også rummer en fortælling om byens historiske og moderne identitet. Subsidiært tjener lemmatisering af kulturonymer også en kommunikativ funktion – tekst-reception – idet mange kulturonymer optræder flittigt i skriftlige tekster, der er særdeles synlige i bybilledet (plakater, informationstavle, skilte, osv.). Kulturonymer kan opdeles i forskellige kategorier.

### 7.3.1 Patronymer og 'gastronymer'

Der selekteres og lemmatiseres navne på kendte personer og kendte specialiteter, der så at sige 'tegner' byens ansigt udad til, som fx:

**Bigot** (byens barn, occitansk digter og forsvarer af occitansk)

**Brandade** (byens specialitet, fiskeret tilberedt af afsaltet klipfisk)

**Costières de Nîmes** (områdets AOC-vin)

**Daudet** (byens barn og forfatter til de berømte "Breve fra min Mølle")

**Foster** (verdensberømt britisk arkitekt og ophavsmand til 'Carré d'Art', byens kunstmuseum og bibliotek – og til en heftig arkitektonisk polemik blandt byens borgere)

**Nicot** (byens barn, der indførte tobakken i Frankrig)

### 7.3.2 Toponymer

Toponymer omfatter navne på steder, som normalt ikke kommer med i turist-brochurer, men som er vigtige for forståelsen af byens identitet, som fx:

**Ecusson** (refererer til Nîmes' gamle bydel, der på bykortet har form som et våbenskjold eller 'écusson' på fransk)

**Coupole des halles** (byens overdækket marked og moderne indkøbscenter)

**Garrigues** (refererer til det særlige landskab, der omringer byen)

### 7.3.3 Modernymer

'Modernymer' er navne og fraser på moderne informationsfænomener, som er centrale for forståelse af byens aktualitet, af dens aktuelle politik, af dens fremtidsplaner, kommunikationsbudskaber m.m. Den type information selekteres normalt aldrig som egentlig turistinformation, men bør ikke desto mindre opfattes som værende helt central for den gensidige kulturforståelse.

<sup>8</sup> Se Leroyer 2006 for en gennemgang af principperne for opbygning af funktionelle, eksterne links.

<sup>9</sup> Ord, der er i sig selv er bærere af en kulturviden og fortjener en særskilt leksikografisk behandling, kaldes hos Galisson (1991), Pruvost (2000) og Lino/Pruvost 2003 'lexiculture'.

**AEF** (meget omfattende anlægsarbejde, der har sat sit præg på byens centrum, og har til formål at skabe et nyt, sammenhængende bymidte)

**la ville avec un accent** (byens slogan, der refererer til den dobbelte betydning af ordet accent: et grafisk tegn ‘’ i byens grafi – Nîmes – og til en fransk by, der godt nok taler fransk, men har sin helt egen accent, sin egen identitet)

**ville active** (moderne bydel med en stor koncentration af butikker, indkøbscentre m.m.)

#### 7.3.4 Events

Der selekteres navne på aktuelle events, der markerer eller fejrer byens identitet, herunder navne på festivals o.l., og der selekteres ligeledes ord og fraser, der repræsenterer den viden, der er i centrum af disse festivals og events. Denne type information bør selvfølgelig løbende opdateres.

**féria de la Pentecôte** (den årlige byfest i Pinsen, med tre dages corrida, og en lang række andre festlige indslag)

**manade** (bestemt opdræt af tyre til tyrefægtning fra udvalgte opdrættere)

**novillada** (spansk corrida udelukkende med unge tyre)

### 8. Konklusion – “samtale fremmer forståelsen”

Det bør afslutningsvis medgives, at opstillingen af en ny teori for turistordbøger ikke udgør nogen garanti for, at den kommercielle leksikografi vil ændre praksis fremover. Rejseordbøger og parlører er populære, og vil trods deres begrænsede brugbarhed i kommunikative situationer fortsat blive udgivet som hidtil. Nogle brugere vil givetvis kunne drage nytte af dem i begrænset omfang, men vil dog langt fra få den hjælp til at løse deres praktiske kommunikationsproblemer, som det fejlagtigt er blevet stillet dem i udsigt. Hjælpen vil nødvendigvis altid være yderst rudimentær.

Der er derfor behov for en helt ny generation turistordbøger, ikke alene som papirordbøger, men i særdeleshed som elektroniske ordbøger med avancerede søgefaciliteter. Den nye teori åbner for lokaliserede online turistordbøger, der kan udarbejdes af de nationale, regionale og lokale turistorganisationer i de enkelte destinationer og kan være til store gavn for deres turistgæster.

Grundet den altoverskyggende dominans af den sproglige tilgang i rejseleksikografien er det turisternes kommunikationsbehov, der har været i fokus hidtil, og det sproglige fokus har skygget for den videnskabelige klarlægning af turisternes øvrige leksikografisk relevante behov. Den nye teori bygger på en erkendelse af, at turistordbøger er uegnet som kommunikationsværktøjer; omvendt anviser teorien, hvordan turistordbøger kan blive effektive vidensværktøjer og hjælpe turisterne med værdifuld, leksikografisk relevant turistviden.

For få år siden bragte udbyderen af mobiltelefoni ‘Orange’ i Danmark en række Tv-reklamer for deres telefonløsninger, hvori der anvendtes følgende slogan:

“Samtale fremmer forståelse”. Sloganet blev fremsagt af selskabets daværende, kvindelige direktør, som var franskmand, og som på ironisk vis og med megen charme gjorde særlig meget ud af at anvende en stærk, fransk klingende udtale og betoning! Til forskel fra telefonien er turistordbøgerne uegnet til at fremme forståelsen gennem samtalen – til gengæld er de velegnet til at fremme forståelsen gennem interkulturel viden.

## Litteratur

### Ordbøger:

- DP = *Dansk Parlør*. <<http://www.parlor.dk/>> 2007.
- DSS = Lohman, C.: *Dansk-spansk sætningsbog*. Hirschsprungs sætningsbøger for forretningsfolk. København: H. Hirschsprungs Forlag 1956.
- FDDFO = Juul Hansen, Else: *Fransk/Dansk/Dansk-Fransk Ordbog* (Gyldendals Mini-ordbøger). 1. udgave, 2. oplag. København: Gyldendal 2004.
- FPBD = *French Phrase Book & Dictionary*. Berlitz Publishing. <[http://www.berlitzpublishing.com/berlitz/tourist\\_phrase.asp?TAG=&CID](http://www.berlitzpublishing.com/berlitz/tourist_phrase.asp?TAG=&CID)> 2007.
- GSFP = *Gyldendals Store Fransk parlør*. Redigeret af Gyldendals ordbogsredaktion, 1. udgave, 1. oplag. København: Gyldendals Boghandel, Nordisk Forlag A/S 2005.
- GCR = *Guides de conversation Routard*. Paris: Editions Hachette, en coopération avec Larousse. <<http://www.routard.com/boutique.asp>> 2007.
- LGRCE = *Le Grand Robert-Collins Electronique/The Unabridged Collins-Robert Electronic French Dictionary*. (Version 1.0.). Paris: Dictionnaires Le Robert/VUEF, Development: Bureau Van Dijk, Brussels, <[www.bvdep.com](http://www.bvdep.com)>.
- LNPR = Rey-Debove, Josette/Rey, Alain: *Le Nouveau Petit Robert, Dictionnaire alphabétique et analogique de la langue française* (Première édition 1967). Paris: Dictionnaires Le Robert 1993.
- NLO = Bergenholtz, Henning/Cantell, Ilse/Fjeld, Ruth Vatvedt/Gundersen, Dag/Jónsson, Jón Hilmar/Svensén, Bo: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget 1997.
- PFR = *Politikkens Fransk Rejseparlør*. København: Politikkens Forlag 2005.
- STK = Johannson, A/Bruinier, J. W/Morén, G. G.: *Svensk-tyskt konstruktionslexikon for temaskrivning och korrespondens*. Stockholm: Fr. Skoglund's Förlag 1901.

### Anden litteratur:

- Abend, Heike 1991: *Das Reisewörterbuch? I: Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand and Ladislav Zgusta (Hrsg.): Wörterbücher/Dictionnaires/ Dictionnaires, Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie, Third volume*. Berlin, New York: Walter de Gruyter, 2903 – 2908. (= *HSK, Handbücher zur Sprach- und Kommunikationswissenschaft, band 5.3.*).
- Bergenholtz, Henning/Tarp, Sven 2005: *Wörterbuchfunktionen. I: Irmild Barz/Henning Bergenholtz/Jarmo Korhonen (red.): Schreiben, Verstehen, Übersetzen, Lernen. Zu*

- ein- und zweisprachigen Wörterbüchern mit Deutsch*. Frankfurt a.M./Bern/Bruxelles/New York/Oxford/Wien: Peter Lang, 11-25.
- Chapman Graham/Gilliam, Terry/Cleese, John/Idle, Eric/Jones, Terry/Palin, Michael 1970: Dirty Hungarian Phrase Book. I: *Monty Python's Flying Circus*, The Second Series, Episode twenty-five, Recorded 1970/06/25, Aired 1970/12/15, Prod. #60129. London: BBC Comedy, <<http://www.bbc.co.uk/dna/h2g2/brunel/A687945>>. (Retrieved 2007/01/03).
- Galisson, Robert 1991: *De la Langue à la culture par les mots*. Paris: Clé International.
- Hausmann, Franz Josef 1989: Das Satzwörterbuch. I: Franz Josef Hausmann/Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Hrsg.), *Wörterbücher/Dictionaries/Dictionnaires, Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography/ Encyclopédie internationale de lexicographie, Third volume*. Berlin, New York: Walter de Gruyter, 1050 – 1053. (= *HSK, Handbücher zur Sprach- und Kommunikationswissenschaft, band 5.1.*).
- Hovmark, Henrik 2005: Tre franske miniordbøger til rejsebrug og daglig tale. I: *LexicoNordica* 12, 213-231.
- Leroyer, Patrick 2006: Ordbogseksterne henvisninger – fra belæg til funktioner. I: Henrik Lorentzen/Lars Trap-Jensen (red.): *Nordiske Studier i Leksikografi 8. Rapport fra Konferencen om Leksikografi i Norden*, Sønderborg 24.-28. maj 2005. København: Nordisk Forening for Leksikografi i samarbejde med Språkrådet i Norge og Det Danske Sprog og Litteraturselskab, 239-251.
- Lino, Marie Thérèse/Pruvost, Jean 2003: *Mots et lexiculture. Hommage à Robert Galisson*. Paris: Honoré Champion.
- Nielsen, Sandro 2005: User Guides. I: Irmhild Barz/Henning Bergenholtz/Jarmo Korhonen (red.): *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Frankfurt a.M./Berlin/Bern/New York/Oxford/Wien: Peter Lang: 137-146.
- Pruvost, Jean 2000: *Dictionnaires et nouvelles technologies*. Paris: Presses Universitaires de France.
- Tarp, Sven 2004: Basic Problems of Learner's Lexicography. I: *Lexikos* 14: 222-252.
- Tarp, Sven 2005. The concept of a bilingual dictionary. I: Irmhild Barz/Henning Bergenholtz/Jarmo Korhonen (Hrsg.): *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Bern/Berlin/Bruxelles/ Frankfurt a.M./New York/Oxford/Wien: Peter Lang, 27-41.
- Tarp, Sven 2006. *Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi*. Doktorafhandling. Århus: Handelshøjskolen i Århus.
- World Tourism Organization, UNSD, Eurostat-Unité du tourisme 2002. *Tourism Satellite Account: Recommended Methodological Framework*. <http://pub.world-tourism.org:81/WebRoot/Store/Shops/Infoshop/Products/1194/1194-1.pdf>

Patrick Leroyer  
 lektor, ph.d., f. 1955  
 Center for leksikografi, Handelshøjskolen  
 Aarhus Universitet  
 pl@asb.dk



NINA MARTOLA

## Angivelse av verbkonstruktioner i svenska ordböcker

In this study I analyze how verb valency is treated in five Swedish monolingual dictionaries with the preposition *åt* ('to; for; at') as example. I also look at the presentation of phrasal verbs with *åt*. Initially I examine how prepositional particles are classified as lemmas, and I can establish that in most dictionaries they are looked upon as adverbs.

As to the valency information the differences between the dictionaries are considerable, but in general one can say that the larger dictionaries, not unexpectedly, cover verb valencies better than the shorter ones. The larger dictionaries also give more explicit information while the shorter dictionaries give valency information only in the form of (causal) examples.

The most problematic items are the phrasal verbs. In two of the analyzed dictionaries the treatment is good, though. The phrasal verbs are given as sublemmas after the verb entry with their own descriptions. The other three dictionaries subsume phrasal verbs in the verb article, and it is not always clear whether *åt* is a valency preposition or a particle in the given examples. One of the dictionaries even mix prepositions and particles in the formal constructional information.

### 1 Inledning

Ordet *verbkonstruktion* har en tämligen vag innebörd. Att jag ändå låtit det ingå i rubriken beror på två olika omständigheter. För det första är undersökningen bred i den meningen att den innefattar såväl mer traditionella valens- eller konstruktionsuppgifter i ordboksartiklar som ordböckernas behandling av partikelförbindelser. För det andra är incitamentet till denna undersökning en omfattande studie av verbfraser med *åt* där jag anlägger ett konstruktionsgrammatiskt perspektiv (Martola 2007).

Jag kommer i huvudsak att begränsa undersökningen till konstruktioner med funktionsordet *åt*, eftersom jag därvidlag alltså har en omfattande, korpusbaserad studie av *åt* att relatera till (Martola 2007). Dessutom kan en näranalys av en representant för ett funktionsord (preposition/partikel) antas ge en förhållandevis god bild av hur behandlingen av ordtypen ser ut generellt i ordböckerna, även om det naturligtvis kan förekomma en viss variation i behandlingen.

Denna studie består av tre delar. Jag inleder med att se på kategoriseringen av lemmat *åt* och andra prepositioner/partiklar. Därefter undersöker jag hur valensuppgifterna vid verb är utformade och i hur hög grad konstruktionsuppgifter finns för olika verb i ordböckerna. I den sista delen ser jag på behandlingen av partikelförbindelser med *åt*.

Innan jag går in på själva undersökningen ska jag ge en översikt över vilka ordböcker jag behandlar. Jag presenterar också kort den studie av funktionsordet *åt* som jag kommer att hänvisa till i analyserna samt tar upp några termfrågor.

## 2 Bakgrund

### 2.1 De undersökta ordböckerna

Föremålet för min undersökning är fem allmänspråkliga svenska ordböcker.

- *Nationalencyklopedins ordbok* (NEO), 1995–1996
  - Definitionsordbok, ca 137 000 lemman, 1 959 sidor.
  - För sökningar efter vissa kombinationer har jag utnyttjat webbutgåvan (ne.se).
- *Svenskt språkbruk* (SSB), 2003
  - Ordbok för konstruktioner och fraser, ca 19 600 lemman och sublemman, 1 433 sidor.
- (Norstedts) *Svensk ordbok* (NSO), 1999
  - Definitionsordbok, ca 60 000 lemman, 1 492 sidor.
- *Natur och kulturs ordbok* (NKO), 2001
  - Definitionsordbok för utlänningar, ca 32 000 lemman och sublemman, vissa illustrationer, 1 226 sidor.
- *Bonniers svenska ordbok* (BSO), 2002
  - Definitionsordbok med vissa illustrationer, ca 65 000 lemman, 704 sidor.

NEO och NSO är i mångt och mycket två olika omfattande versioner av samma ordbok. De baserar sig på samma lexikala databas. Ursprungligen kom NEO till som en utbyggd version av *Svensk ordbok*, men *Svensk ordbok* har fortsatt att komma ut i nya egna upplagor. Artiklarna i NSO är färre och åtminstone de mer omfattande artiklarna är något förkortade jämfört med NEO.

Både NSO och NKO finns i senare upplagor än de jag undersökt, den förra från 2004 och den senare från 2006, men dem har jag tyvärr inte haft tillgång till för näranalyser<sup>1</sup>.

I den första delen av undersökningen, om kategoriseringen av verbpartiklar, har jag dessutom kontrollerat *Svenska Akademiens ordbok* (SAOB), *Svensk handordbok* (SHO) och *Svenska Akademiens ordlista* (SAOL).

### 2.2 En jämförande studie av *åt*

Under åren 2003–2006 arbetade jag som forskare inom projektet *Svenskan i Fin-*

<sup>1</sup> Varken NSO eller NKO återfinns i sin nyaste upplaga på biblioteken i Helsingforsregionen. I en bokhandel har jag varit i tillfälle att bläddra i böckerna och kunnat konstatera att de inte tycks ha genomgått någon revidering i fråga om de aspekter jag undersökt.



*land* – syntaktiska drag i ett jämförande perspektiv finansierat av Svenska litteratursällskapet i Finland (jfr <http://www.sls.fi/media/pdf/svenskanifinland.pdf>). I mitt eget delprojekt undersökte jag hur den finlandssvenska användningen av prepositionen (och partikeln) *åt* skiljer sig från den sverigesvenska användningen. Undersökningen är korpusbaserad och utgör en jämförelse av ett finlandssvenskt och ett sverigesvenskt textmaterial. Bägge materialen finns tillgängliga hos CSC (Center for Scientific Computing) i Finland (<http://www.csc.fi/kielipankki/index.phtml>). De är uppmärka med samma automatiska analysprogram och är därmed sökbara med samma sökverktyg. Den sverigesvenska korpusen hos CSC är innehållsligt identisk med den svenska Parolekorpusen och omfattar drygt 19 miljoner ord (<http://spraakbanken.gu.se/parole/>).

Eftersom ordböcker för svenska bygger på sverigesvenskt språkbruk kommer jag här inte att gå in på skillnaderna mellan sverigesvenskan och finlandssvenskan, och vid jämförelser mellan ordböckerna och korpusmaterialet redovisar jag endast siffror för den sverigesvenska korpusen.

### 2.3 Några termer

Ovan nämnde jag att jag använder termen *verbkonstruktion* som sammanfattande benämning på olika valensförbindelser (t.ex. *frysas [åt ngt]*) och partikelförbindelser (*komma åt ngt*). Det finns en mängd andra verbkonstruktioner men de faller utanför denna undersökning.

För kombinationer som *komma åt ngt*, *skiljas åt* och *kapa åt sig ngt* kommer jag att använda termen *partikelförbindelse* (jfr SAG 1: 207, 3: 417). Den mer hävdvunna benämningen är *partikelverb*, men i lexikografiska sammanhang är det en poäng med den förra termen, eftersom den betonar att det handlar om kombinationer av ord.

Ordet *åt* i *frysas åt ngt* är entydigt en preposition. Hur *åt* ska etiketteras i kombinationen *komma åt ngt* är inte givet. I det första analysavsnittet (avsnitt 3) undersöker jag hur ordböckerna tacklar problemet. Som en sammanfattande benämning på prepositioner och partiklar använder jag *funktionsord*. Det finns andra typer av funktionsord men de aktualiseras inte i denna undersökning.

## 3 Kategoriseringen av funktionsorden i ordböckerna

Det som är påfallande när man ser på själva artiklarna för *åt* i de olika ordböckerna är att kategoriseringen varierar och att relationen mellan *åt* som preposition och *åt* som partikel verkar uppfattas väldigt olika. Jag har därför vidgat perspektivet något och kontrollerat hur ett antal olika prepositioner/partiklar behandlas i de undersökta ordböckerna jämte SAOL, SAOB och SHO. Nedan ger jag en kort redogörelse för skillnaderna och sammanfattar sedan mina iakttagelser i tabellform (tabell 1).

NEO har två separata artiklar för *åt* betecknade med *prep.* och *adv.* Det samma

gäller orden *av*, *för*, *i*, *med*, *på* och *till*, men för *från* finns det bara prepositionsartikel.

SSB ansätter bara en artikel för *åt* betecknad ”partikel vid verb”. I artikeln ges några exempel på partikelförbindelser följda av uppmaningen ”se resp. verb”. På samma sätt behandlas *i*, *med* och *på*. För *av*, *för* och *till* finns förutom partikelartikeln en separat adverbartikel där verblösa fraser tas upp (t.ex. *vid sidan av*, *för mycket*, *lite till*). För *från* finns utöver partikelartikeln en prepositionsartikel.

NKO anför två separata artiklar för *åt* betecknade med *prep.* resp. *verbpartikel* med den senare markeringen följt av angivelsen ”verb + ’åt’”, där *åt* alltså har betoningssmarkering. *Åt*-artikeln är dock ett undantag. De andra partiklarna har den dubbla etiketteringen *adverb/verbpartikel* följt av samma verb+partikel-angivelse som för *åt*. Ordet *från* anføres bara som preposition, liksom i NEO. Det beror sannolikt på att partikelfunktionen mestadels uppbärs av varianten *ifrån*.

Den ordbok som i minst utsträckning skiljer ut partiklarna är BSO. I den är *åt*-artikeln semantiskt momentindelad och ordklassangivelse saknas. Partikelfallen finns anförda i moment 2. På samma sätt är *från* behandlat, utom att artikeln inte är momentindelad. För *av* och *på* finns dubbel ordklassangivelse (*prep. o. adv.*). Samma gäller ordet *med*, för vilket funktionsordet behandlas som homonym 1 och substantivet som homonym 2. För *i* finns bara beteckningen *prep.* och inga partikelförbindelser tas upp. Artikeln för *till* slutligen är momentindelad med beteckningen *prep.* för moment 1 och angivelsen *betonat med vissa verb* för moment 3 (i moment 2 anføres *en gång till* utan närmare kategoriangivelse).

I fråga om kategoriangivelser har jag också kontrollerat SAOL 13. Där ansätts två homonyma lemman för *åt* betecknade med *prep.* resp. *adv.*, vilket är en ändring från föregående upplaga där preposition och adverb anføres i en gemensam artikel. För de andra prepositionerna jag har tittat på har inga ändringar gjorts mellan upplagorna: *av* och *på* tas upp i homonyma artiklar medan *från*, *i*, *med* och *till* anføres i en artikel med angivelsen *prep. o. adv.*

I tabell 1 återfinns en översikt över hur de kontrollerade prepositionerna/partiklarna är ansatta som lemman. Siffrorna 1 och 2 osv. utdragna till vänsterkanten i kolumnerna står för två separata artiklar, alltså homonymer, medan siffra inom rad anger polysemi. Siffrorna följer respektive ordboks numrering men endast kategoriseringar som är relevanta för denna framställning anføres i tabellen och inte t.ex. homonyma substantiv och konjunktioner. I BSO skiljs homonymi och polysemi inte riktigt klart åt, men uppenbarligen betraktas prepositioner och likalydande partiklar som fall av polysemi, eftersom betydelsebeskrivningarna är gemensamma.

Lemma	NEO/ NSO	SSB	NOK	BSO	SAOL 13
av	1 prep. 2 adv.	1 adv. 2 partikel vid verb	1 prep. 2 adverb/verb- partikel	(prep. o. adverb)	1 prep. 2 adv.
från	prep.	1 prep. 2 partikel vid verb	prep.	[ang. saknas]	prep. o. adv.
för	1 prep. 2 adv.	2 adv. 3 partikel vid verb	1 prep. 2 adv. 3 verbpartikel	2 prep. o. adv.	1 prep. 2 adv.
i	1 prep. 2 adv.	2 partikel vid verb	1 prep. 2 adverb/verb- partikel	2 prep.	prep. o. adv.
med	1 prep. 2 adv.	partikel vid verb	2 prep. 3 adverb/verb- partikel	1 prep. o. adv.	prep. o. adv.
på	1 prep. 2 adv.	partikel vid verb	1 prep. 2 adverb/verb- partikel	(prep. o. adv.)	1 prep. 2 adv.
till	1 prep. 2 adv.	1 adv. 2 partikel vid verb	1 prep. 2 adverb/verb- partikel	1 (prep), 3 (be- tonat med vissa verb)	prep. o. adv.
åt	1 prep. 2 adv.	partikel vid verb	1 prep. 2 verbpartikel	[ang. saknas]	1 prep. 2 adv.

Tabell 1. Översikt över hanteringen av funktionsord i ordböckerna

Behandlingen av funktionsorden är alltså rätt brokig och skiljer sig inte bara mellan ordböckerna utan också (och delvis omotiverat) inom en och samma ordbok. Alla ordböcker utom SSB betecknar partiklarna i partikelförbindelser som adverb. NOK menar möjligen att verbpartikeln är en typ av adverb. NEO, NSO, SSB och NOK anför preposition och motsvarande verbpartikel i separata artiklar (homonymi) medan BSO konsekvent anför dem i samma (polysemi). SAOL vacklar i sin syn.

Det tycks med andra ord vara vanligt inom svensk lexikografi att beteckna prepositioner i partikelfunktion som adverb. Så görs också i SAOB. Särskilt i de artiklar som tillkommit senare kompletterar man dock försiktigtvis beteckningen *adverb* med andra uppgifter. Detta framgår av tabell 2.

<i>av</i>	I) prep.		II) adv.
<i>från</i>	I) prep. betecknande ngt ss. [...]		II) adv. (se vidare under de verb med vilka från förekommer i särsk. förb.)
<i>för</i>	I) prep.		II) adv. (l. i absolut anv.)
<i>i</i>	I) prep.		II) adv.
<i>med</i>	I) ss. prep. o. ss. adv. i särsk. förb.		
<i>på</i>	I) ss. prep. med det styrda ordet utsatt.	II) ss. prep. utan utsatt styrt ord l. ss. adv.	III) ss. prep. l. adv. i vissa verbförbindelser (s. k. särskilda förbindelser),
<i>till</i>	I) ss. prep. med utsatt rektion,		II) ss. prep. i absolut anv. l. ss. adv. ; i sht ss. partikel i förb. med verb;

SAOB-förkortningar: ss. = såsom, särsk. förb. = särskild förbindelse, l. = eller, anv. = användning, s.k. = så kallad(e), i sht = i synnerhet, förb. = förbindelse

Tabell 2. Kategoriseringen av funktionsord i SAOB

SAOB ansätter alltså en gemensam artikel för vart funktionsord och delar in artikeln i huvudmoment genom romerska siffror. Prepositionsbeskrivningen står efter den romerska siffran I och adverbbeskrivningen efter II. Under den senare redovisas partikelförbindelser. I artikeln *med* innehåller moment I en gemensam beskrivning för preposition och adverb medan moment II ägnas betydelsen 'också' (*hon blev glad – jag med*). I artikeln *på* innehåller moment II verblösa förbindelser medan partikelförbindelserna får ett eget moment numrerat III. Trots att huvudindelningen i princip sker enligt kategorierna preposition och adverb, markerar SAOB:s makrostruktur att preposition och motsvarande partikel är fall av polysemi och inte av homonymi. Vanligen brukar olika ordklassstillhörighet för likalydande ord tjäna som ett kriterium för homonymi inom lexikografen. Också i SAOB ansätts orden *med* och *för* som egna lemman när de är substantiv (och det senare också adjektiv). Att preposition och partikel sammanförs i en artikel speglar uppenbarligen det faktum att det är betydelsebeskrivningen som står i förgrunden.

Samma praxis som i SAOB följer *Svensk handordbok* (SHO). För *åt* har moment II beteckningen "adverbiellt i förb. med verb" medan moment II för artiklarna *från*, *i*, *med* och *på* har kategoriseringen *adv.* För *till* finns en ännu djupare hierarkisk indelning med *prep.* och *adv.* som A och B och med romerska siffror som huvudmoment under dem. För *av* slutligen finns det två separata artiklar med kategoriseringarna *prep.* respektive *adv.*

Medan flera ordböcker alltså klassificerar prepositionella verbpartiklar som adverb är den stora svenska grammatiken SAG (3: 417 ff.) entydigt av den meningen att prepositioner i partikelfunktion har kvar sin ordklassstillhörighet. Det konstateras att de flesta osammansatta prepositionerna kan användas absolut som partikeladverbial (id., 419). SAG använder inte termen (*verb*)partikel utan talar om *partikeladverbial*. Inte heller den förhållandevis hävdvunna benäm-

ningen *partikelverb* används, utan man introducerar *partikelförbindelse* som benämning.

#### 4 Valensangivelser – verb bestämda av prepositionsfras med *åt*

Sättet att ange valens varierar kraftigt i de undersökta ordböckerna. De mest explicita ordböckerna är NEO och SSB. NEO ger valensen i slutet av varje artikel eller i slutet av betydelsemoment (siffrumrerat huvudmoment) vid momentindelade artiklar. Uppgiften markeras med förkortningen *KONSTR.* Verbet anges med tildetecken (~) och optionella valensbestämningar står inom parentes (se illustration 1). (Diskussionen om bestämningars bindingsgrad lämnar jag därhän i detta sammanhang och jag tar heller inte ställning till om angivelserna för optionalitet är korrekta eller inte.)

SSB inleder genomgående verbartiklarna med regelrätta konstruktionsuppgifter. Därefter följer, föregångna av en romb, lexikala fraser med fler positioner fyllda. Som sista grupp, markerad med stjärna/asterisk, står regelrätta idiom och pragmatiska fraser. För den som läst användaranvisningarna eller som själv kommit underfund med systemet är de regelrätta konstruktionsuppgifterna därmed lätta att känna igen. Att en bestämning är optionell framgår av att det finns en konstruktionsuppgift med och en konstruktionsuppgift utan bestämningen i fråga (illustration 1).

<p><b>NEO</b>  <b>fnysa</b> verb <i>fnös</i> el. <i>fnyste</i>, <i>fnyst</i>, pres. <i>fnys-er</i>  <small>ORDLED: <i>fnys-er</i></small>  <small>SUBST.: <b>fnysande</b>, <b>fnysning</b></small>      • häftigt och ljudligt andas ut genom näsan som tecken på missnöje el. förakt: <i>han fnös indignerat</i>  <small>BET.NYANSEK: a) överfört visa förakt eller missnöje: <i>kultureliten fnyser åt alla TV-serier</i> b) om djur: <i>tjuren fnös ilsket</i></small>  <small>KONSTR.: ~ (åt ngn el. ngt), ~ till</small>  <small>HIST.: sedan ca 1600; sv. dial. <i>fnysa</i>; gemens. germ. ord, urspr. ljudhärmande</small></p>	<p><b>SSB</b>  <b>fnysa</b> verb <i>ngn fnyser</i> om att andas ut ljudligt när man är missnöjd el. föraktfull <i>tjuren fnös ilsket</i>; ofta som tecken på missnöje el. föraktfullhet hos människor <i>han bara fnyser (frustar) när fällskärmsavtal kommer på tal</i>; <i>fnysa av ngt</i> en tallrik med groddar fick många män att ~ av förakt; <i>fnysa åt ngn/ngt</i> äv. <i>fnysa över ngn/ngt</i> vuxna fnyser alltid åt (är nedlåtande mot) ungdomarnas musiksmak; det är lätt att ~ över dåligt pålästa medborgare</p>
<p><b>NSO</b>  <b>fnysja</b> verb <i>-te</i> el. <i>fnös</i>, <i>-t</i> • häftigt och ljudligt andas ut genom näsan som tecken på missnöje el. förakt: <i>han fnös indignerat</i> □ äv. överfört visa missnöje på ett föraktfullt sätt: <i>på kultursidorna -er man åt såpoperorna</i> □ äv. om djur: <i>tjuren fnös ilsket</i></p>	<p><b>NKO</b>  <b>fnysa</b><sup>v</sup> verb (fnyser, fnös eller fnyste, har fnyst; fnys!)  <b>1</b> andas 'ut luft genom näsan så att det hörs  <b>2</b> visa att man inte tycker något är bra • <i>Hon föreslog att de skulle fråga någon om vägen, men han bara fnös.</i> • <i>Hon fnös åt förslaget.</i></p>
<p><b>BSO</b>  <b>fnysa</b> häftigt andas ut genom näsborrarna; <i>f. åt rynka på näsan åt</i>, se ned på    <i>fnyser, fnyste</i> el. <i>fnös, fnyst</i> <b>fnysning</b>    <i>-en; -ar</i></p>	

Illustration 1. Artikeln *fnysa* i de olika ordböckerna

De tre andra undersökta ordböckerna (NSO, NKO, BSO) är betydligt mindre explicita i fråga om valensuppgifter. Valensen framgår – om den framgår – av

språkexempel (illustration 1). Möjligen är tanken att ovana ordboksanvändare har svårt att tolka mer schematiska angivelser av den typ NEO använder, och det ligger det kanske något i, men å andra sidan har man åtminstone som van användare svårt att veta vilken produktiv kraft man ska tillmäta ett språkexempel. Det syns t.ex. ingen skillnad på mer eller mindre fritt utbytbara element (markerade med *ngn* och *ngt* i NEO:s och SSB:s konstruktionsuppgifter) och element som är begränsat utbytbara.

NEO:s valensangivelser är formellt explicita. Däremot är de problematiska genom att de inte skiljer på valensbestämningar och partiklar. I artikeln *fnysa* ovan t.ex. ges partikelförbindelsen *fnysa till* som en konstruktionsuppgift. Jag återkommer till frågan nedan.

Efter att ha sett på hur valensuppgifterna formellt är utförda ska jag härnäst undersöka i vilken utsträckning valens ges i de olika ordböckerna. Jag använder verb som konstrueras med *åt* som exempel. Jag har i första hand gått igenom artiklarna för kommunikations- och reaktionsverb, eftersom *åt* är vanlig som valenspreposition för dessa i Parolekorpusen. Jag ser därtill på de verb som allra mest frekvent förekommer tillsammans med *åt* i Parolematerialet. Antalet belägg i Parolekorpusen anges i tabellerna 3 och 4, och noteras bör att siffrorna avser förekomster på verbet med en *åt*-fras som bestämning på noll till fyra elements avstånd, inte det totala antalet belägg på varje verb.

I fråga om kommunikations- och reaktionsverben (tabell 3) ser man att de ordböcker som har de explicitaste konstruktionsuppgifterna, NEO och SSB, också i högst utsträckning noterar vilken preposition som används. För NEO har jag också angett om prepositionsangivelsen förekommer i exempel, eftersom alla konstruktionsuppgifter inte illustreras. Det finns därtill fall där möjligheten att konstruera ett verb med *åt*-fras framgår av exempel men inte av konstruktionsuppgiften. Sådana är t.ex. artiklarna *gapskratta*, *morra* och *vifta*. SSB ger genomgående ett eller ett par autentiska exempel för varje konstruktionsangivelse, så jag redovisar därför ingen exempelkolumn. Vad gäller *åt* är verbet *skrika* i alla fall ett undantag; *skrika till ngn* och *skrika åt ngn* anges som likvärdiga alternativ (angivelsen *el.* 'eller') men autentiskt exempel ges bara på det förra alternativet.

Grå markering i en ruta i tabellen anger att verbet eller förbindelsen inte finns ansatt som lemma i ordboken i fråga. Förbindelsen *anförtro sig* förekommer alltså som uppslagsord bara i SSB, däremot förekommer det enkla *anförtro* i alla ordböckerna, och det är i den artikeln som konstruktionsmöjligheten *anförtro sig åt ngn* framkommer i NEO och NSO.

I NEO saknas uppgift om att verbet *ropa* kan uppträda med *åt*-fras, medan SSB inte noterar att *ryta* kan göra det. För *ryta* är konstruktionsuppgiften i NEO inexact: ~ (*ngt el. att+SATS*) (*åt ngn*). De två parenteserna bör rimligen kunna tolkas så att bägge elementen kan vara närvarande samtidigt, och vid *att*-satser ger det fel utslag: *?han röt att jag skulle sticka åt mig*. I fråga om *ryta* kan som kuriositet

dessutom nämnas att NEO anför som ett exempel *han röt åt barnen* medan NSO har samma exempel med prepositionen *till*, *han röt till barnen*.

I BSO ges *fnysa åt* som exempel med en egen betydelseangivelse. Eftersom objekt saknas kommer kombinationen att likna partikelförbindelser (t.ex. *huta åt*, *klämna åt*, *komma åt*), om man inte håller i minnet att BSO (i regel) betoningsmarkerar partiklar.

Förekomsten av angivelse av *åt* som valenspreposition för ett antal kommunikations- och reaktionsverb i de olika ordböckerna finns sammanfattad i tabell 3.

Verb	Parole	NEO		SSB	BSO	NKO	NSO
		Konstruktion	Exempel	Konstruktion	Exempel	Exempel	Exempel
skratta	269	1	1	1		1	1
skrika	92	1		1			
nicka	74	1		1			
le	62	1	1	1			1
fnysa	46	1	1	1	1	1	1
vinka	40	1	1	1		1	1
anförtro sig	20		1	1			1
teckna	20	1	1	1	1		1
blinka	19	1	1	1		1	1
ryta	18	1	1			1	
jubla	17	1		1			
ropa	15			1		1	
svära	15	1		1			
fnissa	14	1		1		1	

Tabell 3. Valensangivelserna vid kommunikations- och reaktionsverb

Ser man på de verb som förekommer allra mest frekvent med *åt*-fras i mitt material (tabell 4) kan man notera en viss brokighet i fråga om angivelserna i ordböckerna. När det gäller det vanligaste verbet, *ägna* och kombinationen *ägna sig*, är konstruktion med *åt* genomgående noterad. För simplex verbet *ägna* framgår typen av objekt (t.ex. *liv*, *tid*, *uppmärksamhet*) förhållandevis bra, om än enbart implicit, av exemplen. I SSB anges objekttypen på sätt och vis explicit i och med att ingen konstruktionsuppgift med objektsangivelsen *ngt* förekommer. I BSO framgår kombinationen *ägna sig åt ngt* av artikeln för det enkla *ägna* medan NKO rentav ansätter två olika sublemman för de olika objekten: *ägna sig åt något* och *ägna sig åt någon*.

För verbet *ge* är det bara SSB som noterar konstruktionsmöjligheten *ge ngt åt ngn* som en produktiv konstruktion, både för betydelse 1 'erbjuda, skänka' och 2 'lämna, räcka'. NEO anför bara alternativet *ge ngn ngt* och *ge ngt till ngn*. I NKO

framgår bara bitransitiv konstruktion (*ge ngn ngt*) och mottagarlös konstruktion av exemplen, inte konstruktion med *till*. I BSO ges exempel enbart på mottagarlös konstruktion i *ge*-artikeln. Däremot noteras *ge ngt åt ngn* i artikeln för *åt*.

Det som inte framgår av någon av ordböckerna är att *ge* visserligen konstrueras med *åt*-fras, men att objektet då är av en viss typ, nämligen abstrakt. I SSB finns ett autentiskt exempel med *möjlighet*. I Parolematerialet är de vanligaste objekten *plats, stöd, namn, uttryck* och *liv*. Denna omständighet nämns av SAG. Man konstaterar att *åt* förekommer i ”lexikaliserade vändningar” och nämner *ge liv åt* som exempel (SAG 3: 466).

Verben *glädja sig* och *glädjas* anförs i NKO i ett gemensamt sublemma under *glädja*: ”*glädja sig över (åt) något* eller *glädjas över (åt)*”. Alternativa sublemman avviker från ordbokens normala praxis. För *lämna* förekommer frasen *lämna ngt åt framtiden* i NEO och NSO. SSB tar upp fraserna *inte lämna ngt åt slumpen* och *lämna ngn/ngt åt sitt öde* i form av hänvisningar till andra artiklar.

Att verbet *göra* inte i någon högre grad annoteras för valens med *åt*-fras i ordböckerna är naturligt. Det höga förekomsttalet i Parolekorpussen beror på att *göra* används som ersättningsverb för andra verb (t.ex. *Vi måste få gräsmattan klippt. – Jag kan göra det åt er.*). Därtill förekommer *göra* i vissa funktionsverbsfraser (t.ex. *göra rättvisa åt ngt, göra i ordning ngt [åt ngn]*). *Åt*-fras ingår alltså inte i verbet *göra*:s egen valens. Den fras som däremot noteras under simplex verbet *göra* i NEO, SSB och NSO är *göra någonting åt ngt* i olika varianter. Den tas dock också upp som partikelförbindelse och jag återkommer därför till den i avsnitt 5. Belägen på frasen ingår inte i siffran 108 i tabell 4.

Verb	Parole Antal belägg	NEO			SSB		BSO	NKO		NSO
		Konstr.	Exempel	Idiom	Konstr.	Fras	Exempel	Sublemma	Exempel	Exempel
ägna sig	1037	1	1		1		1	1	1	1
ge	821				1					
ägna	569	1	1		1		1		1	1
glädja sig	169	1	1		1			1		1
glädjas	141	1					1	1		
lämna	123	1	1			1				1
göra	108			1		1				1

Tabell 4. Valensangivelser för de verb som frekventast konstrueras med *åt*-fras i Parolekorpussen

I tabell 4 återfinns en översikt över ordboksbehandlingen av de vanligaste verben med *åt*-fras. Rutorna markerade med mörkare grått anger att verbet/förbindelsen saknas som lemma i vederbörande ordbok. Den ljusare grå markeringen i NKO-kolumnen anger att de båda verben är sammanslagna till ett sublemma.



Efter att ha gått igenom hur ordböckerna hanterar valensuppgifterna ska jag i det sista avsnittet av studien undersöka hur partikelförbindelser med *åt* behandlas.

## 5 Partikelförbindelser med *åt*

Också partikelförbindelserna anföras på varierande sätt i ordböckerna. NEO tar upp vissa som lemman och andra som konstruktionsuppgifter och några bara i exempel. Jag återkommer strax till NEO.

<p><b>NEO</b>  <b>ka'pa</b> verb <i>-de -t</i>  ORDLED: <i>kap-ar</i>  SUBST.: <b>kapande, kapning; kaperi</b> (till 2)  [...]</p> <p><b>2</b> med (hot om) våld överta kontrollen av farkost e.d.; för direkt vinning el. i (pol.) utpressningssyfte {→<b>uppbringa 2</b>}: <i>planet ~des över Medelhavet; ~ ett fartyg</i>  BET.NYANSER: <b>a</b>) spec. förr som ett led i sjökrigföringen {se <b>kapare 3, kaperi</b>} <b>b</b>) allmänare (otillbörligt) lägga beslag på (vanl. med partikel): <i>~ åt sig den största biten</i>  KONSTR.: <i>~ (åt sig) ngt</i>  HIST.: sedan 1749; av nederl. <i>kapen</i> med samma bet., till frisiska <i>kap</i> 'köp'</p>	<p><b>SSB</b></p> <p><b>2kapa</b> verb (ta över med våld) <i>kapa ngt en terrorist ~de ett flygplan i morse</i>  <b>kapa åt sig</b> äv. <b>kapa till sig</b>: vard.; <i>kapa åt sig ngt</i> företaget har lyckats ~ åt (tillskansa) sig över 10 procent av marknaden; han menade att staten hade ~t (roffat) åt sig oskäligt mycket av den enskildes egendom</p>
<p><b>NSO</b>  <b>kapja</b> verb <i>-ade -at 1 korta av (ngt)</i>  [...]</p> <p><i>kostnadskrävande åtagandena 2</i> med (hot om) våld överta kontrollen av farkost e.d.; för direkt vinning el. i (pol.) utpressningssyfte: <i>planet -ades mitt över Medelhavet; ~ ett fartyg</i> □ spec. förr som ett led i sjökrigföringen (jfr <b>kapare, kaperi</b>) □ äv. allmänare orättfärdigt lägga beslag på (vanl. med partikel): <i>~ åt sig den största biten</i></p>	<p><b>NKO</b></p> <p><b>kapa</b><sup>v</sup> verb (kapar, kapade, har kapat; vara kapad, kapat)</p> <p><b>1</b> göra något kortare, t.ex. genom att hugga eller såga • <i>kapa en gren</i>  <b>2</b> ta kontrollen över t.ex. ett flygplan och tvinga piloten att flyga någonstans • <i>Planet kapades på väg till Tokyo.</i>  <b>kapa 'åt sig något</b> vardagligt ta något före alla andra • <i>Han har kapat 'åt sig den bästa platsen.</i></p>
<p><b>BSO</b>  <b>kap 2</b> fångst, fynd: <i>göra ett gott k.    -er; pl. =, best. pl. -en</i>    <b>kapa 1</b> erövra (fartyg, flygplan) med vapenhot el. våld; <i>k. åt sig riva till sig, ruffa åt sig    -de</i></p>	

Illustration 2. Partikelförbindelsen *kapa åt sig* i ordböckerna

SSB och NKO ansätter genomgående partikelförbindelserna som sublemman i artikeln för det enkla verbet, vilket är tydligt och bra (illustration 2). NKO anför även andra fraser som sublemman, men eftersom betoning anges är partikelförbindelserna åtminstone i princip lätta att känna igen. En viktig orsak till att NKO ger betoning är säkert att den riktar sig till icke-svenskspråkiga, men det kunde vara praxis i fler ordböcker. Visserligen är betoningsfördelningen i löpande tal inte alltid så klar, men principiellt vore det bra att få partiklar/partikelförbindelser markerade på ett tydligt vis, eftersom de hör till de mer komplicerade fenomenen i svenskan. Partikelförbindelser har dels klart lexikala drag, dels syntaktiska drag (Norén 2005, Booij 2002).

BSO anför partikelförbindelserna som exempel under de enkla verben och anger i princip betoningen med en punkt under vokalen (illustration 2). Tyvärr är punkteringen inte konsekvent genomförd; för exempelvis *roffa åt sig* under *roffa* saknas betoningsangivande punkt och likaså för *ta åt sig*. Den senare förbindelsen ingår i ett betydelsemoment med angivelsen ”med betonad partikel” i *ta*-artikeln. Avsaknaden av betoningsangivande punkt under *åt* blir extra inkonsekvent i och med att partikeln i de flesta andra partikelförbindelserna i momentet är punktmarkerade.

Att ge partikelförbindelser som sedvanliga exempel i artikeln för motsvarande simplexverb är enligt min mening ingen lyckad lösning, även om partiklarna betoningsmarkeras. Både syntaktiskt och lexikalt vore det bättre att gruppera dem separat.

NEO:s hantering av partikelförbindelser är den mest brokiga och därmed också den mest problematiska. Vissa förbindelser ansätts som lemman medan andra anges som konstruktionsuppgifter och vissa förekommer enbart som exempel. Sammanlagt är det elva partikelförbindelser med *åt* som finns med som lemman, varav två reflexiva, *bära sig åt* och *ta åt sig*. När de prepositionella partiklarna ingår i konstruktionsuppgiften går de inte att skilja från regelrätta obetonade prepositioner. I artikeln *kapa* (illustration 2) anges det visserligen i betydelsemoment b) att verbet vanligen står med partikel i den allmännare betydelsen, men konstruktionsuppgifterna hänför sig ju till hela momenten och formellt kommer det att verka som om *åt sig* var en optionell bestämning både för huvudbetydelsen och för de två betydelsenyanserna. I t.ex. artikeln *klämma* (illustration 3) förekommer ingen angivelse om att verbet förekommer med partikel i betydelsenyans d) där exemplet *~ åt en falskdeklarant* återfinns. Därmed ser exemplet likadana ut som t.ex. *frysas åt TV-serier* i artikeln *frysa* (jfr illustration 1). Osäkra användare kan inte veta att det ena exemplet illustrerar en valenskombination och det andra en partikelförbindelse. En tydligare gruppering hade varit önskvärd eller åtminstone en betoningsmarkering av något slag för partiklar.

<sup>1</sup>**klämm`a** verb *klämde klämt*, pres. *klämmer*  
 • (ofta med partikel) trycka samman från två eller flera sidor på en mindre yta: [...]  
 BET.NYANSER: [...]  
**d)** överfört: *klämd mellan föräldrarnas ambition och egna önskemål*; *~ åt en falskdeklarant*; *~ ur sig svaret*; *orkestern klämde i med en marsch*; *~ till med ett citat*; *~ fram en ursäkt*  
 [...]  
 KONSTR.: *~ (fast el. ihop el. in el. ut) ngt*, *~ på ngt*, *~ till med ngt*

### Illustration 3. Artikeln klämma i NEO

Av tabell 5 framgår vilka av de vanligaste partikelförbindelserna med *åt* som finns noterade i de olika ordböckerna samt i vilken form de tas upp. Av de 24 vanligaste

partikelförbindelserna med *åt* i Parolematerialet är det bara ett, *skilja sig åt*, som inte noteras i en enda av ordböckerna. Förbindelserna *rycka åt sig*, *slita åt sig* och *suga åt sig* finns noterade i bara en ordbok, resten i fler. Kombinationen *göra åt* är gråmarkerad i tabellen, eftersom det är diskutabelt om den ska anses utgöra en partikelförbindelse, trots att den behandlas som en sådan i några av ordböckerna. Jag återkommer strax till den problematiken.

Förbindelse	Parole	NEO			SSB	BSO	NKO	NSO	
		Lm	Kon.	Ex.	Sblm	Ex.	Sblm	Lm	Ex.
1–24	Belägg								
<i>göra åt</i>	527	1			1		1	1	
<i>komma åt</i>	415	1			1	1	1	1	
<i>gå åt</i>	329	1			1	1	1	1	
<i>bära sig åt</i>	221	1			1	1	1	1	
<i>säga åt</i>	215	1			1		1	1	
<i>skilja åt</i>	166		1	1	1		1		1
<i>hjälpas åt</i>	125	1			1	1	1	1	
<i>ta åt sig</i>	84	1			1	1	1	1	
<i>skiljas åt</i>	71			1			1		1
<i>skilja sig åt</i>	60								
<i>dra åt</i>	59				1	1	1		
<i>följas åt</i>	40	1			1		1	1	
<i>vilja åt</i>	39				1	1	1		
<i>strama åt</i>	38		1	1		1	1		1
<i>rycka åt sig</i>	32			1					1
<i>ta åt</i>	30	1			1		1	1	
<i>kapa åt sig</i>	28		1	1	1	1	1		1
<i>roffa åt sig</i>	24		1	1	1	1	1		1
<i>dra åt sig</i>	23			1	1				1
<i>klämma åt</i>	22			1	1	1	1		1
<i>plocka åt sig</i>	21			1	1		1		1
<i>slita åt sig</i>	18				1				
<i>suga åt sig</i>	17						1		
<i>sätta åt</i>	15				1		1		

Tabell 5. Noteringen av partikelförbindelser med *åt* i ordböckerna

För NEO kan man notera att det överlag är de mest frekventa förbindelserna som är ansatta som lemmen. Av de reflexiva förbindelserna med ordningsföljden V *åt* SIG är det bara *ta åt sig* som finns med som lemma. Två finns noterade som konstruktionsuppgifter och några bara som exempel medan vissa inte noteras alls. För *dra åt sig* är det enda exemplet idiomat *dra öronen åt sig*. Möjligen har förbindelsetypen ansetts stå nära prepositionella uttryck, vilket har ett visst fog för sig, eftersom det reflexiva SIG kunde tolkas som, men ordföljden visar att *åt*-frasen inte är en prepositionell bestämning; man säger normalt *han plockade åt sig några broschyrer* och inte *?han plockade några broschyrer åt sig*. Ordföljden framgår också

av konstruktionsuppgifterna (jfr *kapa* i illustration 2). En alternativ förklaring till den bristande noteringen är att typen räknats som ett produktivt mönster, men den är knappast ett fullgott skäl. Produktiviteten är dessutom begränsad (Martola 2007), och typen borde i så fall åtminstone ha noterats i adverbartikeln *åt*, vilket inte är fallet. Man kan slutligen notera att förbindelserna *dra åt*, *vilja åt* och *sätta åt*, som inte tas upp vare sig i artiklarna för respektive simplexverb eller som lemmen, finns noterade i adverbartikeln för *åt* i NEO.

NEO:s behandling av partikelförbindelser är såsom ovan framgått rätt brokig. Av de 24 vanligaste partikelförbindelserna med *åt* i Parolekorpusen finns dock sammanlagt 17 noterade i någon form. SSB noterar 19 och NKO 20, och eftersom bägge noterar partikelförbindelserna som sublemman framstår de tydligare som egna lexikala enheter. I NKO noteras *följas åt* något avvikande; samma stil används som för sublemman, men förbindelsen noteras under *följas* och förbindelsen är inte utdragen till vänsterkanten som sublemman normalt i ordboken. Förbindelsen *skilja sig åt* noteras inte i en enda av ordböckerna.

Vad gäller NSO kan man notera att kolumnerna ser exakt likadana ut som för NEO utom att konstruktionsuppgiftskolumnen saknas.

För förbindelsen *säga åt* är Parolesiffran ungefärlig, eftersom jag bara räknat belägg med betydelsen 'uppmana, förmana'. Vid betydelsen 'informera, konstatera' är det osäkert om kombinationen ska räknas som partikelförbindelse, och dessutom är en del av beläggen ambigüosa i fråga om betydelse. (Se närmare Martola 2005.)

Den vanligaste förbindelsen i mitt material är *göra åt* och den är som ovan nämnts problematisk. Använder man enbart betoningskriteriet för att identifiera verbpartiklar klassar man kombinationen som partikelförbindelse, men synar man den syntaktiskt räcker betoningskriteriet inte till. I en partikelförbindelse *göra åt* måste objektet i svenskan stå efter *åt:et*: *göra åt ngt*. Det här kriteriet uppfylls vid betydelsen 'förbruka' (*familjen gjorde åt en limpa om dagen* – jfr illustration 4), men däremot inte av frasen *göra någonting åt något*, där ju objektet står före *åt*. SAG (3: 428) påpekar explicit att *åt:et* i förbindelsen *göra någonting åt något* optionellt kan vara betonat men att *åt*-frasen utgör ett bundet adverbial (*åt:et* är med andra ord inte ett partikeladverbial).

NKO tar upp hela frasen *göra någonting åt något* som sublemma med betoningsangivelser för *åt*, och kombinationen har förmodligen räknats som partikelförbindelse eftersom den alfabetiskt står på platsen för *göra åt*, efter *göra upp* (illustration 4).

Både NEO och SSB ansätter kombinationen *göra åt* som lemma, och i båda ordböckerna anges 'förbruka' som första betydelse (illustration 4). Den betydelsen har jag inga belägg för i mitt korpusmaterial. Som andra betydelse – i NEO i eget moment och i SSB efter asterisken för idiom – anges 'avhjälpa, åtgärda'. I NEO står dessutom den något anmärkningsvärda konstruktionsuppgiften *göra ngt åt ngt*. Lemmat har alltså delats upp och objektet lagts in mellan verb och

partikel, där det inte finns plats för objekt i en partikelförbindelse. Dessutom står det första *ngt* inte för ett mer eller mindre fritt utbytbart element, vilket är det normala i konstruktionsuppgifter, utan i huvudsak för just *något/någonting*. De två pronomenen står för hela 90 % av objekten i beläggen på uttrycket i Parolekorpusen. Därtill förekommer *inget/ingenting* och i någon mån kvantifierande uttryck som (*inte*) *mycket*, *inte ett dugg* o.likn.

Noteras kan slutligen att både NEO och SSB tar upp frasen *göra något åt något* också i artikeln för simplex verbet *göra*. Den förekommer som idiom i det första betydelsemomentet ('vara sysselsatt'). I NEO är exempelfrasen (*det är*) *inget att göra åt* och i SSB *vad kan/ska ngn göra åt det*. Det här är den syntaktiskt adekvata placeringen av uttrycket.

<p><b>NEO</b></p> <p><b>göra å t</b> verb <i>gjorde gjort</i>, pres. <i>gör</i>  ORDLED: <i>gjo-r-de</i>  SUBST.: <b>åtgörande</b> (till 2)  <b>1</b> förbruka: <i>familjen gjorde åt en hel limpa bröd om dagen</i>  KONSTR.: ~ <i>ngt</i>  HIST.: åtm. sedan 1950-talet  <b>2</b> (i vissa uttr.) (avhjälpa genom att) vidta lämpliga åtgärder med avs. på ngt trasigt el. felaktigt: <i>har du gjort någonting åt det trasiga trappsteget ännu?; det är ingenting att ~</i>  KONSTR.: <i>göra ngt åt ngt</i>  HIST.: sedan 1526</p>	<p><b>SSB</b></p> <p><b>göra åt:</b> <i>göra åt ngt</i> vard. se till att ~ åt (äta upp) påronen innan de ruttar * <i>det är inte mycket att göra åt</i> el. <i>det är ingenting att göra åt</i> inför ngt som inte kan åtgärdas, uppgivet hon kommer inte på festen, det är inte mycket att ~ åt</p>
	<p><b>NKO</b></p> <p><b>göra 'upp eld</b>  tända en eld</p> <p><b>göra någonting 'åt något</b>  ändra eller påverka något ♦ <i>Om du är missnöjd med kursen ska du försöka göra någonting 'åt det.</i>  ♦ <i>Det är ingenting att göra 'åt. (=Det är omöjligt att ändra eller påverka.)</i></p>

Illustration 4. Artikeln för *göra åt* i NEO, SSB och NKO

## 6 Uppsummering

I denna studie har jag undersökt hur konstruktionsuppgifter för verb och partikelförbindelser behandlas i svenska ordböcker med konstruktioner med *åt* som exempel. Undersökningen har visat att ordböckerna i förhållande till sin omfattning visserligen förhållandevis väl täcker de vanligaste konstruktionerna om man jämför med ett autentiskt språkmateriale, men att genomförandet överlag verkar lite inkonsekvent. De största bristerna uppvisar ordböckerna i fråga om behandlingen av partikelförbindelser. SSB och NKO ansätter partikelförbindelserna som sublemman, vilket är en bra lösning. De andra tre behandlar partiklar och valensprepositioner integrerat, vilket gör det svårt för användarna att urskilja vad som är partikelförbindelser med *åt* och vad som är verb med *åt*-frasbestämning.

Det är å ena sidan förståeligt att behandlingen av partikelförbindelser haltar,

eftersom det handlar om ett ytterst intrikat och mångfasetterat problem i svenskan. Å andra sidan är det just därför ett område som många användare kan tänkas behöva hjälp med, i all synnerhet sådana som inte har svenska som modersmål. Man kan visserligen försvara ordböckerna med att de, förutom NKO, inte är avsedda för icke-modersmålstalare. Det är dock en klen ursäkt, för ordböckerna ger en mängd andra uppgifter som svenska användare också kan antas känna till.

De svenska ordböckerna utgör utgångspunkt och stöd för redigeringen av bilingvala ordböcker med svenska som källspråk. Att redovisa svenska partikelförbindelser och deras ekvivalenter för talare av andra språk är en vansklig uppgift. Författarna till bilingvala ordböcker har dessutom mestadels målspråket som modersmål. Deras arbete skulle underlättas väsentligt om behandlingen av partikelförbindelser vore åskådlig och klar i källorna.

## Litteratur

### *Undersökta ordböcker*

BSO = Malmström, Sten & Györki, Iréne & Sjögren, Peter A.: *Bonniers svenska ordbok*. Stockholm: Albert Bonniers Förlag AB 2002.

NEO = *Nationalencyklopedins ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. Bd 1–3. Höganäs Bra böcker 1995–1996. – *Nationalencyklopedins ordbok*. <http://www.ne.se>, anv. maj 2007.

NKO = *Natur och kulturs svenska ordbok*. Författare: Per Olof Köhler, Ulla Messelius, Birgitta Hene. Huvudredaktör: Birgitta Hene. Stockholm: Natur och kultur 2001.

NSO = *Svensk ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. Stockholm: Norstedts Ordbok 1999.

SSB = *Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Utarbetad av Svenska språknämnden. Stockholm 2003.

### *Övriga ordböcker*

SAOB = *Svenska Akademiens ordbok*. [g3.spraakdata.gu.se/saob](http://g3.spraakdata.gu.se/saob), anv. 2004–2007.

SAOL12 = *Svenska akademiens ordlista över svenska språket*. 12. uppl. Stockholm 1998.

SAOL13 = *Svenska akademiens ordlista över svenska språket*. 13. uppl. Stockholm 2006.

SHO = *Svensk handordbok. Konstruktioner och fraseologi*. Ture Johannisson & K. G. Ljunggren (red.). Stockholm: Läromedelsförlagen 1970.

### *Annan litteratur*

Booij, Geert 2002: Separable complex verbs in Dutch: A case of periphrastic word formation. I: Nicole Dehe & al. (red.): *Verb particle explorations*. Berlin: Mouton de Gruyter, 21–41.

Martola, Nina 2005: Vad säger man åt folk? Kombinationen *säga åt* i finlandssvenskan och sverigesvenskan. I: Gunilla Byrman & al. (red.): *Svenskans beskrivning 27*. Växjö: Växjö University Press, 181–192.

Martola, Nina 2007: *Konstruktioner och valens. Verbfraser med åt i ett jämförande perspektiv*. Nordica Helsingiensia 10. Helsingfors: Institutionen för nordiska språk och nordisk litteratur.

Norén, Kerstin 2005: Lexikalisering av partikelverb. I: Gunilla Byrman & al. (red.): *Svenskans beskrivning 27*. Växjö: Växjö University Press, 246–256.

SAG = Hellberg, Staffan & Andersson, Erik & Teleman, Ulf 1999: *Svenska akademiens grammatik*. Stockholm: Svenska Akademien/ Norstedts ordbok.

*Svenskan i Finland. Syntaktiska drag i ett jämförande perspektiv*. Projektets slutrapport 2004–2006. <http://www.sls.fi/media/pdf/svenskanifinland.pdf>

*Korpusmaterial*

Parole-materialet. <http://spraakbanken.gu.se/parole>. Januari 2004.

Språkbanken i Finland. [www.csc.fi/kielipankki/index.phtml.sv](http://www.csc.fi/kielipankki/index.phtml.sv). Parolekorpusen. Anv. 2004–2007.

Nina Martola  
 forskare, f. 1955  
 Forskningscentralen för de inhemska språken  
 Sörnäs strandväg 25  
 FIN-00500 Helsingfors  
[nina.martola@focis.fi](mailto:nina.martola@focis.fi)





HENRIK NIKULA

## Valensinformation i ordböcker ur användarens synvinkel

Valency information is of central importance when describing verbs in dictionaries. An essential problem for valency theory is the distinction between complements and adjuncts. In this paper it is argued that ideas from construction grammar could be useful when trying to solve problems of this kind. Although construction grammar is a well-established linguistic theory, it has, as yet, not had a very strong influence on valency theory. This is rather remarkable, because the concept of valency itself plays an important role in most construction grammars. It is shown that construction grammar can be helpful not only in solving some central problems of valency theory but that it may also contribute to a better structure of lexical entries in dictionaries.

### Inledning

Av alla valensrelevanta ordklasser är verbet den viktigaste eftersom verbet genom sin valens präglar satsens grundstruktur. Därför kommer jag i detta sammanhang att begränsa mig till verbvalensen. På grund av verbvalensens inflytande på syntaxen har naturligtvis alla lexikografer med större eller mindre framgång och mer eller mindre explicit beskrivit verbets valens, helt oberoende av om de stött sig på någon valensteori i begreppets striktare bemärkelse eller inte. Den första "rena" valensordboken torde vara Gerhard Helbigs och Wolfgang Schenkels *Wörterbuch zur Valenz und Distribution deutscher Verben* från år 1969. Den andra upplagan av denna ordbok, Helbig/Schenkel (1973) (=HS) är en förbättrad version av den första upplagan och ändringar har inte gjorts i senare upplagor. Ett konkurrerande projekt till DDR-ordboken uppstod så småningom i väst vid Institut für deutsche Sprache i Mannheim, vilket resulterade i *Kleines Valenzlexikon deutscher Verben* år 1976 (=KVL), i *Verben in Feldern. Valenzwörterbuch zur Syntax und Semantik deutscher Verben* 1986 (=ViF) och i *VALBU – Valenzwörterbuch deutscher Verben* 2004 (=VALBU). Vid institutet initierades vidare ett antal kontrastiva projekt mellan tyskan och andra språk, dock ej nordiska, med tvåspråkiga valensordböcker som resultat. Valenslexikografin tycks således har varit, och också ännu vara, en framför allt tysk företeelse, även om valensen som teoretiskt begrepp fått spridning inom olika lingvistiska teoribildningar också på många andra språkområden. Det är sålunda inte någon överraskning att det första stora engelska valenslexikonet *A Valency Dictionary of English. A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives* (=VALENCY)

huvudsakligen kom till vid universitetet i Erlangen-Nürnberg, jfr VALENCY s. xxii, fotnot 3. Typiskt för de flesta valensordböcker är att de tyngs av en mer eller mindre tung lingvistisk beskrivningsapparat, allra mest kanske ViF, medan utgångspunkten mer eller mindre explicit är att ordböckerna ska vara till praktisk nytta för mera avancerade språkstudier, för lärare, läromedels- och ordboks-författare. Så sägs det t.ex. i inledningen till VALENCY s. vii:

*The Valency Dictionary of English* is intended to enable

- **linguists** to carry out research in the field of verb, adjective and noun complementation, in particular to extend the analyses presented in this book to investigate parallels between the syntactic and semantic properties of words, to consider the character of valency phenomena and the place they ought to occupy in more general and comprehensive models of language
- **applied linguists, grammarians and lexicographers** to develop new teaching materials
- **advanced foreign learners** to write grammatically correct and idiomatic English because it shows them, which constructions a word can be used with
- **non-native teachers of English** to mark students' papers because it tells them whether a construction a student has used is found frequently, occasionally or not at all in present-day English

VALENCY är säkerligen användbart för samtliga nämnda målgrupper, men jag är tämligen säker på att åtminstone de två sistnämnda föredrar ett vanligt lexikon. Av alla ovan nämnda valensordböcker är det nog bara VALBU som är tillräckligt användarvänligt för ickefackmän. Ett så specialiserat lexikon kan emellertid oberoende av användarvänlighet av rent praktiskt-ekonomiska skäl knappast få någon större spridning.

Nedan kommer jag först att kort diskutera några användarproblem beträffande valensbeskrivningar i "vanliga" ordböcker, varefter jag går in på frågan i vad mån och på vilket sätt valensordböcker och valensteorin överhuvudtaget kan bidra till att förbättra de lexikografiska beskrivningarna av verb.

## 2. Exempel på "otillräcklig" valensbeskrivning i ordböcker

Ett typiskt "valensproblem" som ofta uppstår då man som svenskspråkig skriver en text på tyska eller översätter en text till tyska är att man är osäker på vilken kasus och/eller preposition ett fyllnadsled ska ha. För att lösa problemet slår man kanske upp i en svensk-tysk ordbok, eller om man är mera avancerad, i ett enspråkigt tyskt lexikon. Lyckas inte det kan olika korpusar i Internet användas, eller rentav en vanlig sökmaskin som t.ex. Google. Eller så ett valenslexikon, om man har ett sådant till hands, vilket i fråga om de flesta användare är tämligen osannolikt. Jag kommer i det följande att granska beskrivningen av några verb,

först i några vanliga ordböcker och sedan jämföra med beskrivningen i valensordboken VALBU. Poängteras bör kanske att syftet inte är att kritisera enskilda ordböcker utan att utgående från några exempel belysa vissa svårigheter och problem. Noteras bör också att framställningen bygger på egna användarerfarenheter och att problemen sålunda i första hand inte är teoretiska utan praktiska.

Ett fiktivt exempel: Jag skriver ett brev till en tysk vän och funderar på om satsen jag just skrivit: *Er freute sich über das Gemälde in der Ausstellung* är korrekt. På något sätt stämmer inte *sich freuen über* här, även om satsen grammatiskt sett är korrekt. Jag slår upp närmaste svenska motsvarighet *glädja sig* i *Stora svensk-tyska ordboken* (=STORA) där jag bl. a. finner följande:

II rfl ~ *sig sich freuen* [*ät* (*över*) ngt *über* ack.; ngt *framtida auf* ack.]; ~ *sig ät god hälsa* (*gott rykte*) *sich guter Gesundheit* gen. (e-s *guten Rufes*) *erfreuen*

Jag är ändå inte riktigt nöjd och slår upp i DUDEN där jag naturligtvis finner liknande uppgifter, men också följande:

**freuen** [---] 1. <f. + sich> [---] *sich seines Lebens* f. (*sein Leben genießen*); *sich an* [den] *Blumen* f. (*seine Freude daran haben*)

De angivna betydelsenyanserna med genitiv eller med prepositionen *an* förefaller närma sig den sökta, men för säkerhets skull slår jag upp i VALBU, där beskrivningen av det reflexiva *sich freuen* omfattar drygt en sida med tre olika betydelsevarianter och därmed förknippade valensstrukturer:<sup>1</sup>

**SBP sich [A] freuen 1 auf** NomE PräpE

**BED** etwas mit Freude erwarten. (1) Ich freue mich schon auf meinen nächsten Urlaub.

**SBP sich [A] freuen 2** NomE (PräpE)

**BED** über etwas, das einem Zuteil wird, Freude empfinden. [---] Über eine großzügige Sachspende freut sich das Institut für optische Nachrichtentechnik und Mikrowellentechnik. (nach MM, 28.1.88, S. 19)<sup>2</sup>

**W.BED sich [A] freuen 1 <an>** NomE GenE/PräpE; PräpE: *an* + D; PräpSE mit obl. Korrelat *daran* jemand hat am Vorhandensein von jemandem/etwas [keine Restr.] seine Freude [---] (30) Das Publikum im nicht ausverkauften Saal freute sich an der Musik, auch an dem einen oder dem anderen Gag. (MM, 25.4.88, S. 28)

Resultatet av efterforskningarna blir att den ursprungliga "problemsatsen" visserligen är grammatiskt korrekt, men att den inte förmedlar den mera person-

1 Inte heller ordboken WDaF, se litteraturförteckningen, som är avsedd för tyska som främmande språk och därför ger tämligen utförliga syntaktiska beskrivningar, ger tydligare uppfattning om den betydelsenyans användningen av konstruktionen med *an* vid *sich freuen* innebär.

2 MM = dagstidningen *Mannheimer Morgen*.

liga känsla som avsetts, varför prepositionen *an* med dativ, *an dem Gemälde*, alternativt eventuellt genitiv, *des Gemäldes*, vore adekvatare. För det slutgiltiga valet är VALBU avgörande, dels på grund av den systematiska, välstrukturerade beskrivningen, dels – och inte minst – på grund av det stora antalet exempel-satser av vilka nästan alla är obearbetade belägg. Den i det här sammanhanget största skillnaden mellan VALBU och de båda andra ordböckerna är sättet att strukturera informationen. I STORA kan under *glädja sig* ingen tydlig struktur skönjas, vilket kanske är försvarbart p.g.a. lexikonartikelns ringa längd. Detsamma gäller DUDEN och många andra jämförbara ordböcker. I VALBU däremot, liksom i valensordböcker överhuvudtaget, är lexembeskrivningen tydligt strukturerad utifrån beskrivningen av den syntaktiska valensen vilket bidrar till att göra den omfattande beskrivningen överskådlig. Vad har då detta med frågan om verbets semantik att göra? Jo, redan den till synes obetydliga valensvariationen mellan *über* + akkusativ och genitiv alternativt *an* + dativ är förknippad med en liten men tydlig betydelskillnad som i vissa kontexter kan vara helt avgörande. Det är i sig en gammal insikt inom valensteorin att det alltid finns en mer eller mindre tydlig relation mellan valensen och betydelsen hos valensbäraren, vilket inte bara kommer till synes genom olika grad av ställighet, jfr t.ex. *äga* som är tvåställigt (någon äger någonting) och *skänka* som är treställigt (någon skänker någon någonting), utan också i fråga om mindre morfologiska variationer i den syntaktiska valensen.

### 3. Problem

Inom valensteorin är möjligheten att dra gränsen mellan valensbundna fyllnadsled och fria tilläggsled en central fråga som i många avseenden kan vara problematisk och dessutom leder till varierande beskrivningar av samma valensbärare i olika valensordböcker. Eftersom det finns ett uppenbart samband mellan betydelse och valens borde man ju i så fall kunna utgå från att detta innebär olika betydelsebeskrivningar av valensbärarna i fråga. Ett mycket omdiskuterat problemfall utgör beskrivningarna av verbet *kaufen* 'köpa' och andra med köp- och säljprocessen förknippade verb såsom *verkaufen*, *bezahlen*, *kosten* osv. I ViF och i VALBU beskrivs *kaufen* som fyrställigt, NomE AkkE (PräpE) (AdvE). I HS beskrivs däremot *kaufen* som obligatoriskt tvåställigt, NomE AkkE.<sup>3</sup> Gemensamt för beskrivningarna är att subjekt och objekt, dvs. KÖPARE och VARA är obligatoriska fyllnadsled. Det är tydligt att verbet *kaufen*, liksom svenskans *köpa* fokuserar, eller bättre, "perspektiverar" KÖPAREN och VARAN, medan t.ex. *verkaufen* och det svenska *sälja* perspektiverar SÄLJAREN och VARAN. Perspektiveringen kan

3 I VALENCY beskrivs den nära synonymen *buy* som maximalt treställigt. Eftersom beskrivningsapparaten starkt skiljer sig från den i VALBU och dessutom de teoretiska utgångspunkter till en del är olika, är en närmare jämförelse inom ramen för detta bidrag inte möjlig.

betraktas som en viktig del av verbens betydelse. Men också PRIS och SÄLJARE är onekligen viktiga element eller "semantiska roller" associerade med betydelsen hos *kaufen* 'köpa', ett faktum som på något sätt måste beaktas vid beskrivningen av verbets valens, t.ex. på det i VALBU tillämpade "maximalistiska" sättet. I allmänna ordböcker kan skillnaden mellan perspektiverade och övriga fyllnadsled åskådliggöras genom väl valda exempelsatser och torde inte bereda större problem.<sup>4</sup>

Rollen PRIS kan emellertid förekomma vid alla möjliga verb. Man kan t.ex. bygga ett hus, slå ihjäl någon, reparera en bil för en viss summa pengar, "Han reparerade bilen för 500 €". PRISSET i form av *förN* kan säkerligen inte i någon mening betraktas som något slag av fyllnadsled till *reparera*, även om det utgör en mycket "naturlig" del av den runt predikats verbet realiserade konstruktionen. Priset kan också förekomma vid "abstrakta" verb som i kontext kan ersätta t.ex. *köpa*, jfr följande satser:

- 1a) Han köpte lampan för 20 € hos Wiklunds.
- 1b) Han fick lampan för 20 € hos Wiklunds.

I kontext kan satserna förmedla ungefär samma innebörd vilket inte nödvändigtvis innebär att man för verbet *få* kan anta en betydelsevariant med betydelsen 'köpa'. Verbet *få* anger inte aktivitet hos subjektet vilket däremot *köpa* gör. Denna betydelseskilnad innebär att man inte utan vidare kan förutsätta samma valensegenskaper hos *få* som hos *köpa* trots att båda verben kan förekomma som predikat i identiska syntaktiska konstruktioner. Utgående från detta kan man försvara det faktum att verbet *få* varken i SOB eller i SSB förekommer i kontexter liknande 1b).

Den tyska översättningsekvivalenten till *få*, dvs. *bekommen*, får däremot i valensordboken VALBU bl.a. följande betydelsevariant, II 6: 'gegen eine Zahlung von irgendetwas oder kostenlos in den Besitz, in den Genuss von etwas kommen', dvs. 'mot betalning av någonting eller gratis komma i besittning, i åtnjutande av någonting'. Valensschemat är: NomE AkkE (AdvE), alltså KÖ-PARE, VARA och PRIS. Här saknas alltså SÄLJAREN, trots att ett exempel på en sådan konstruktion anförs med säljaren inom hakparentes: "(68) Gegen einen kleinen Unkostenbeitrag können die Kinder die Adventskerzen [von der

4 Intressant i detta sammanhang är följande konstaterande gällande finskan i Hakulinen *et al.* (2004: 439) „Lähimpänä määritteitä ovat sellaiset täydennykset, jotka pannaan ilmi vain, jos tarkoite halutaan erityisesti spesifoida. Esimerkiksi *ostaa-* ja *myydä-*verbien tarkoittamaan tapahtumaan kuuluvat myyjä, ostaja, vaihdettava tavara ja siitä maksettu rahasumma, mutta tavallisesti niistä kielennetään lauseessa vain osa [...]” [Mera lika tiläggsleden är sådana fyllnadsled som endast realiseras då man önskar specificera referenten närmare. T.ex. till det med verben *sälja* och *köpa* förknippade skeendet hör säljare, köpare, varan och den för varan betalade penningssumman, men vanligen verbaliseras endast en del av dessa i satsen. (övers. H. Nikula)]

Pfarrei] bekommen.” Detta är kanske inte så väsentligt. Det problematiska är att om alla för ett visst verb mer eller mindre typiska konstruktioner uppfattas som valenskonstruktioner så uppstår lätt en stark polysemi. För verbet *bekommen* II uppförs sålunda i VALBU 18 betydelsevarianter.<sup>5</sup> En sådan beskrivning är rent teoretiskt sett något “suspekt” och trots att den är välstrukturerad blir den i praktiken något oöverskådlig. Beskrivningen av *bekommen* II i VALBU upptar t.ex. två och en halv sidor i två spalter.

Det finns ytterligare många typer av konstruktioner av generell karaktär som regelbundet uppträder vid verb utan att ändå ingå som element i själva valensstrukturen och därmed utgöra element i verbbetydelsen. Sådana konstruktioner är de i konstruktionsgrammatiken ofta beskrivna “ditransitiva” konstruktionerna och liknande konstruktioner, jämför nedanstående exempel ur Goldberg (2006: 7):

- |   |                 |
|---|-----------------|
| 2) He sliced the bread.                   | (transitive)    |
| 3) Pat sliced the carrots into the salad. | (caused motion) |
| 4) Pat sliced Chris a piece of pie.       | (ditransitive)  |
| 5) Pat sliced the box open.               | (resultative)   |

*Slice* ‘skiva’ innebär ju (i sin primära betydelse) att någonting skärs i skivor med ett vasst verktyg, varför endast subjektet och det direkta objektet kan anses vara valensbundna medan de övriga satsdelarna i exemplen 3), 4) och 5) ingår i en större konstruktion, som visserligen kan motiveras genom verbet och dess betydelse, men som inte ingår i verbbetydelsen. För en ordboksbrukare kan emellertid sådan information vara viktig, eftersom den kan stöda tolkningen av verbbetydelsen genom att den ger information om i vilka vidare konstruktioner verbet kan ingå.

#### 4. Valens och konstruktion

Valensbeskrivningen utgör sålunda som ovan angivits en del av betydelsebeskrivningen av valensbärare, något som osökt leder tanken till den s.k. “konstruktionsgrammatiken”. Valensbeskrivningen beskriver ju ett slags syntaktiska konstruktioner. En central tanke inom konstruktionsgrammatiken är att konstruktionerna själva utgör ett slags tecken på samma sätt som orden är tecken, eller som Adele E. Goldberg uttrycker saken, de utgör “conventionalized pairings of form and function” (Goldberg 2006: 3). Vidare spelar begreppet valens en central roll i åtminstone de flesta varianter av konstruktionsgrammatiken.<sup>6</sup> De

5 *bekommen* I är homonym till *bekommen* II med starkt avvikande betydelse, vilket följande exempel under *bekommen* I tydligt visar, VALBU S. 227: “Das Essen ist mir [nicht] gut bekommen.” För *bekommen* I anges inga betydelsevarianter.

6 T.ex. James Fillmore, en av de ledande lingvisterna inom konstruktionsgrammatiken, har genom sina bidrag till kasusteorin haft ett mycket starkt inflytande på valensteorin. Det är därför tämligen märkligt att hans bidrag till konstruktionsgrammatiken hittills be-

“satsmodeller” man finner i valensordböcker och grammatikor kan helt uppenbart betraktas som konstruktioner i konstruktionsgrammatisk mening, jfr t.ex. satsmodellerna i VALBU s. 45 ff. Satsmodellerna bildas utgående från valensen hos klasser av verb med samma valens och sålunda med mer eller mindre gemensamma betydelse drag. Sambandet mellan satsmodell och betydelse blir naturligtvis ännu större om den semantiska valensen beaktas, t.ex. semantiska “kasus” eller “roller”, liksom även s.k. semantisk selektion, jfr t.ex. Helbig (1992: 143 ff.). Med hänvisning till detta förefaller det finnas anledning att undersöka om inte insikter hämtade från valensteori och konstruktionsgrammatik kunde utnyttjas som hjälp vid struktureringen av beskrivningen av valensbärare i vanliga ordböcker. I första hand är det enspråkiga ordböcker som av naturliga skäl kan komma i fråga. En explicit angivelse av “konstruktioner” kan man i själva verket finna i t.ex. engelskspråkiga ordböcker som COLLINS COBUILD i en särskild “kolumn”, jfr COLLINS COBUILD s. xi: “The extra column in this dictionary is used to give two different types of information: grammar notes and semantic relationships.” Det är närmast den grammatiska informationen som kan betraktas som angivelse av konstruktion; vid verb och andra valensbärare är det huvudsakligen fråga om valensinformation.

Konstruktionsgrammatiken kan få en viktig betydelse för valensteorin och därmed också för lexikologin och lexikografin genom att man kan skilja på element som entydigt kan klassificeras som valensbundna och därmed bör ingå i beskrivningen av valensbärare, och element, som visserligen ingår i konstruktioner som av olika anledningar visserligen är typiska för en viss valensbärare men som ändå inte ingår i valensbärarens betydelsestruktur och därmed inte kan betraktas som valensbundna.

## 5. Sammanfattning och diskussion

Som ordboksbrukare slår man upp ord i ordböcker för att få en uppfattning om deras betydelse, om vilka konstruktioner orden kan ingå i och om hur de används. Ordböckerna kan aldrig ge fullständig information, det ligger i sakens natur, eftersom språket till sin karaktär är “dynamiskt”, men de kan ge oss en anvisning i vilken riktning vi kan gå. Språkets dynamiska karaktär gör det svårt att strukturera ordboksartiklarna så att en användare snabbt och enkelt får den information hon eller han behöver för att få en tillräckligt god uppfattning om användningen av det ord hon söker. Tanken bakom föreliggande bidrag är att insikter inom valensteori och konstruktionsgrammatik kunde utgöra ett hjälpmedel för lexikografer att förbättra ordboksartiklarnas innehåll och struktur. Valensordböckernas historia visar att vägen från lingvistisk teori till en användbar ordbok är lång och svår dels beroende på problemens svårighetsgrad, dels på att lingvistisk och lexi-

---

aktats i så relativt ringa grad inom valensteorin, jfr t.ex. FISCHER/STEFANOWITSCH (2006), Östman/Fried (2006), liksom även Nikula (2007a o. b).

kografisk kunskap och erfarenhet inte alltid kombineras med varandra. Den första för den vanlige ordboksanvändaren för praktiskt bruk användbara valensordboken är så vitt jag kan se den tyska VALBU, medan den engelska VALENCY nog närmast är att beteckna som en ordbok för professionella användare.

Makrostrukturen i de enskilda lexikonartiklarna i VALBU är föredömligt överskådlig genom att den vid indelningen i betydelsegrupper utgår från valensen. För en vanlig ordbok, där också andra ord än valensbärande ord beskrivs, vore det sannolikt fördelaktigare att såsom i COLLINS COBUILD använda sig av en särskild kolumn för valensangivelserna. I VALBU utgår man från ett slags "maximalistisk" uppfattning om valensen, dvs. om något element eventuellt kunde vara valensbundet så beskrivs det som sådant. På detta sätt erhåller man naturligtvis en mycket ingående beskrivning av de olika verben, i många fall med ett stort antal betydelsevarianter. På rent teoretisk-lingvistiska grunder kan man säkerligen diskutera om en sådan beskrivning är optimal. Ett i det här sammanhanget större problem är hur sådana omfattande valensbeskrivningar ska kunna integreras i en vanlig ordbok där språkets samtliga lexem måste beskrivas.

En lösning på problemet med "verbtypiska" men icke valensbundna led kunde man kanske finna genom en integration av valensteorin med konstruktionsgrammatiken. Verbet med sina fyllnadsled utgör en typ av konstruktioner i konstruktionsgrammatisk mening, men dessa konstruktioner kan utvidgas genom konstruktioner som kan vara mer eller mindre typiska för verbet i fråga och därmed kan vara viktiga att beskriva i ordboksartikeln. En möjlighet att undvika allför stor polysemi i ordböcker utan att ge avkall på den strukturering som ett explicit beaktande av valensen ger möjlighet till kunde vara att beskriva valensen något mindre "minimalistiskt" än t.ex. i VALBU. En sådan valensbeskrivning med tydliga syntaktiska angivelser kunde utgöra basen för indelningen i betydelsevarianter, medan sådana verbtypiska konstruktioner av olika slag som inte innehåller tydligt valensbundna element kunde åskådliggöras genom exempel. Mer eller mindre tydliga tendenser i den riktningen finns i de flesta bättre ordböcker, men utan den systematik som ett beaktande av valensteori och konstruktionsgrammatik förhoppningsvis kunde bidra till.

## Litteratur

### *Ordböcker*

#### *Valensordböcker*

HS = Helbig, Gerhard/Schenkel, Wolfgang 1973: *Wörterbuch zur Valenz und Distribution deutscher Verben*. (2., überarb. u. durchges. Aufl.). Leipzig: VEB Bibliographisches Institut.



- KVL = Engel, Ulrich/Schumacher, Helmut 1978: *Kleines Valenzlexikon deutscher Verben*. (2., durchges. Aufl.) (= Forschungsberichte des Instituts für Deutsche Sprache 31). Tübingen: Verlag Gunter Narr.
- VALBU = Schumacher, Helmut/Kubczak, Jacqueline/ Schmidt, Renate/de Ruyter, Vera 2004: *VALBU – Valenzwörterbuch deutscher Verben*. (= Studien zur deutschen Sprache, Band 31) Tübingen: Verlag Gunter Narr.
- VALENCY = Herbst, Thomas/Heath, David/Roe, Ian F./Götz, Dieter 2004: *A Valency Dictionary of English. A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Berlin: Walter de Gruyter.
- ViF = Schumacher, Helmut (utg.) 1986: *Verben in Feldern. Valenzwörterbuch zur Syntax und Semantik deutscher Verben*. Berlin, New York: de Gruyter.

#### Övriga ordböcker

- COLLINS COBUILD = *Collins Cobuild. English Language Dictionary* 1993: Utg. John Sinclair *et al.* London: HarperCollins Publisher.
- DUDEN = *Duden. Deutsches Universalwörterbuch* 2001: Utg. Dudenredaktion. Mannheim etc.: Dudenverlag.
- STORA = *Stora svensk-tyska ordboken* 1989. Utg. Berglund, Britt-Marie *et al.* Stockholm: Esselte Studium AB.
- SOB = *Svensk ordbok* 1986: Utg. Sture Allén *et al.* Uppsala: Esselte.
- SSB = *Svenskt språkbruk. Ordbok över konstruktioner och fraser* 2003: Utg. Ulla Clausén *et al.* Stockholm: Norstedts.
- WDaF = Kempcke, Günther *et al.* 2000: *Wörterbuch Deutsch als Fremdsprache*. Berlin, New York: de Gruyter.

#### Övrig litteratur

- FISCHER, KERSTIN/STEFANOWITSCH, ANATOL 2006: *Konstruktionsgrammatik: Ein Überblick*. I: KERSTIN FISCHER/ANATOL STEFANOWITSCH (UTG.): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. Tübingen: Stauffenburg Verlag, 3–17.
- Goldberg, Adele E. 2006: *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Hakulinen, Auli/Vilkuna, Maria/Korhonen, Riitta/Koivisto, Vesa/Heinonen, Tarja Riitta/Alho, Irja 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Helbig, Gerhard 1992: *Probleme der Valenz- und Kasus-theorie*. Tübingen: Niemeyer.
- Nikula, Henrik 2007a: Aus der Perspektive der Valenz. I: Hartmut E. H. Lenk/ Maik Walter (utg.): *Wahlverwandtschaften. Valenzen – Verben – Varietäten. Festschrift für Klaus Welke zum 70. Geburtstag*. (= *Germanistische Linguistik* 188–189). Hildesheim, Zürich, New York: Georg Olms Verlag, 55–64.
- Nikula, Henrik 2007b: Valenz, Satzmodell und Konstruktion. I: *LSP and Translation. 27th VAKKI Symposium*. Vasa 9–10.2.2007. Vasa: Vasa universitet, 200–211.
- ÖSTMAN, JAN-OLA/FRIED, MIRJAM 2005: *The Cognitive Grounding of Construction Grammar*. I: JAN-OLA ÖSTMAN/MIRJAM FRIED (UTG.): *Construction Grammars. Cognitive Grounding and theoretical Extensions*. Amsterdam, Philadelphia: John Benjamins Publishing Company (= *Constructional Approaches to Language* 3), 1–13.

336 HENRIK NIKULA

Henrik Nikula  
prof. (emeritus), f. 1942  
Puolalaparken 1 b B 12  
FI-20100 ÅBO  
hennik@utu.fi

KRISTINA NIKULA

## *Pensionär – aldrig i livet*

### De äldre i ordböckerna

Today elderly people in many ways radically differ from those of former generations. One consequence of this is that terms like *pensioner* (Sw. *pensionär*) with their connotations are often not felt to be adequate any more. The same concerns the stereotype of old people as generally poor, ill and disabled. The lexicographic examples in the entries of Swedish monolingual dictionaries to an astonishingly great extent repeat this stereotype. In this paper the future of existing terms denoting old people and the introduction of new words are discussed. Further the function of lexicographic examples is discussed and the question is raised, whether the one-sided description of the elderly can be defended ethically.

En grundläggande, gemensam egenskap för oss människor är att alla varit unga och att alla blir äldre och många till och med gamla. De så kallade fyrtioåringarna börjar nu vara till åren komna och andelen pensionärer av befolkningen växer rekordartat. I Finland beräknas 26 % av befolkningen år 2030 vara över 65 år (Internetkälla 1) och utvecklingen är densamma i hela västvärlden. Denna tillväxt anses medföra problem av skilda slag, till och med språkliga. Speciellt problematiskt finner reklammakarna det vara hur man i reklamen verbalt och visuellt skall appellera och referera till äldre, men inte heller folk i allmänhet är tillfreds med existerande beteckningar för de äldre.

Målet med denna artikel är:

1. att redogöra för beteckningar för de äldre och beteckningarnas användbarhet,
2. att diskutera en förändrad betydelse av existerande beteckningar och lanseringen av nya,
3. att redogöra för vilken bild av de äldre som ges i språkproven i ordböckerna och
4. att studera språkproven ur funktionell och etisk synvinkel.

Materialet till undersökningen, vilket inte gör anspråk på att vara uttömmande, har främst excerperats i sådana enspråkigt svenska ordböcker (se litteraturförteckningen)<sup>1</sup> som de flesta ordboksanvändare kan antas ha tillgång till – av *Svensk ordbok* har för jämförelsens skull använts både den första (SOB 86)

1 Merparten av ordböckerna är såväl receptions- som produktionsordböcker.

och den tredje reviderade upplagan (SOB). Enstaka beteckningar har påträffats enbart i dags- (*oldies*) och veckotidningar (*mappies*). För att en beteckning skall tas med i undersökningen krävs att denna i sin grundbetydelse (gamling) eller utvidgade betydelse (*fossil*) är markerad för särdraget /gammal/ i det semantiska fält som dessa beteckningar bildar. I text är variationen bland substantiv som refererar till äldre stor, men eftersom tillfälliga beteckningar som *livsnjutare* och *offer* primärt inte är markerade för ålder tas de inte med i undersökningen.

Metoden är deskriptiv och kvalitativ. Betydelsebeskrivningen av fältet /äldre/ i bilagan bygger på komponentanalys. Ett plustecken anger att beteckningen i fråga är markerad för en viss egenskap, minustecken anger motsatt egenskap, t.ex. ett minus för *uppskattning* innebär *nedvärdering*. Varje ordboksartikel betraktas som en text. Språkproven har studerats mot bakgrund av den kritiska lingvistik (Fowler 1991) och den nära besläktade kritiska diskursanalysen (Fairclough 1992). Text utgör enligt Fairclough (1992: 71ff.) resultatet av en kommunikativ händelse i vilken ingår såväl diskursiv som social praktik. Vid analysen av textens egenskaper fokuseras på både produktion och konsumtion, dvs. på den diskursiva praktiken. Genom den diskursiva praktiken medieras kontakten mellan texten och den sociala praktiken, vilket innebär att språk och samhälle står i ett dialektiskt förhållande till varandra och att en förändring i det ena kan medföra en förändring i det andra. (Winter Jørgensen & Phillips 1999: 70ff.) – Inom den nordiska språkforskningen är de äldre fortfarande förvånansvärt osynliga. I Tyskland har man däremot allt sedan mitten av 1990-talet medvetet satsat på gerontologisk språkforskning efter brittisk och amerikansk förebild. En överblick över centrala intresseområden finns bland annat i Fiehler & Thimm (utg.) 1998: *Sprache und Kommunikation des Alters* och Filipp & Meyer 1999: *Bild des Alters*.

## 1. Den åldrande människan

Ålder är av två slag: numerisk och social. Med *numerisk* ålder avses hur gammal en person är räknat i år, månader och dagar. Hittills har som *yngre äldre* räknats personer i åldern 65–79 år och som *äldre äldre* personer i åldern 80 år och uppåt. Med *social* ålder avses hur gammal en person känner sig eller anses vara av andra. Ålder kan därför i likhet med till exempel kön ses som socialt konstruerad och som en egenskap som delvis är föreställd. Social ålder, som är oberoende av numerisk ålder, hänger inte samman med enbart samhälleliga föreställningar och normer utan bestäms också av situativa faktorer och kan därför uppfattas som olika med hänsyn till varje referentgrupp.

Vid sidan av indelningen i numerisk och social ålder förekommer också andra. Inom marknadsföringen börjar man inse att dagens äldre är köpstarka och att det därför vore skäl att satsa på dessa i reklamen. "Seniorgenerationen är aktiv, självsäker och levnadsglad. Den identifierar sig inte med den traditionella reklamklichén, typ blid mormor i gungstolen", konstaterar enligt Hufvudstadsbladet

(Hbl) vd för en stor tysk reklamagentur (Karlberg 2006). Denna insikt har lett till att man inom marknadsföringen börjat skapa beteckningar som bättre anses passa dagens äldre än de redan existerande. I en rapport som Sifo (Internetkälla 2) sammanställt som underlag för marknadsundersökningar betecknas kunder från 55 år och uppåt som *mogna människor*. De äldre är likväl inte en homogen grupp. På basis av olika karakteristika delas därför personer i åldern mellan 55 och 79 år in i fem olika grupper som getts olika beteckning<sup>2</sup>. I Tyskland har man på senare år börjat använda beteckningen *neue Alte*, en beteckning som bättre än *junge Alte* anses referera till dagens aktiva äldre<sup>3</sup> (Thimm 2000: 47).

### 1.1 Dagens äldre

Våra dagars äldre väcker uppmärksamhet kanske först och främst därför att de håller på att bli så många. I jämförelse med tidigare generationer är de tack vare god hälsovård dessutom friska och har därför krafter och lust att resa, placera och konsumera. Efter pensioneringen har de äldre dessutom tid och har i jämförelse med tidigare generationer också bättre råd att förverkliga sina drömmar. Den generation som nu håller på att bli äldre och pensioneras skiljer sig med andra ord väsentligt från tidigare pensionärsgenerationer.

Man blir i dag också med jämna mellanrum påmind om att nutida äldre eller de *nygamla*, som jag i fortsättningen valt att kalla dem, skiljer sig från de äldre i tidigare generationer. Den relativt nystartade damtidningen *M-magasin* (2006/1: 24f.) vars målgrupp är äldre beskriver denna på följande vis: "Vi vill ha allt nu! Vi är många, vi är krävande, vi har pengar och vi vill fortsätta bestämma" och i den finlandssvenska dagstidningen *Borgåbladet* (Andelin 2007) beskrivs "ålderdomskulturen" som en samhällsboom där "De deltidspensionerade 60-åringarnas rockband och disconätterna à la sjuttitalet är bara början för vad seniorgenerationen kommer att ställa till med".

### 1.2 Kritik av beteckningar för äldre

Existerande beteckningar för äldre har mot bakgrund av att en ny typ av äldre håller på att växa fram efter hand börjat upplevas som otidsenliga. Vissa beteckningar som *käring* 'gammal (och ful) kvinna' (SOB) och *gubbstrutt* '(skröplig) gubbe' (SOB 86) är till sin innebörd därtill klart nedvärderande och andra, som exempelvis *fossil* (SOB) 'helt otidsenlig person' och *kalkskuta* 'gammal kvinna', är rena invektiver och till följd av detta olämpliga. Till synes neutrala beteckningar för enskilda äldre och för äldre som kollektiv saknas inte men upplevs i många fall av olika orsaker också som olämpliga, och har på sistone kritiserats från flera håll. Också vissa nymodigheter som 50+ har mötts med en viss skepsis. I en enkätundersökning, *Ages and Attitudes*, som genomfördes inom EU redan 1993 sade

2 *De krismedvetna* (75–79 år), *Krigsbarnen* (70–74 år), *Den första tonårsgenerationen* (65–69 år), *Protestgenerationen* (60–64 år) och *Flower-Powergenerationen* (55–59 år).

3 I tyskan skiljer man mellan *junge Alte*, *Ältere* och *Alte* (Thimm 2000: 47).

sig de tillfrågade föredra beteckningar som *senior citizen* eller *older people* medan man ställde sig negativ till eufemismer som *golden age*, *oldies* och *50+* (Citerat enl. Kramer 1998: 273). I Hbl frågade sig 2004 en skribent (Jernvall-Ingman 2004) vad man skall kalla olika åldersgrupper och när en viss benämning är på sin plats. I Sverige publicerades året därpå en serie intervjuer<sup>4</sup> med personer som är födda på 1940-talet om deras syn på sitt liv och sin framtid. De intervjuade, som alla vid tidpunkten för intervjun var verksamma inom olika yrken, tog också ställning till vad man vill kallas som äldre. Få av de tillfrågade ville kallas *pensionär*, själva ordet gav upphov till kommentarer som: "Det låter jävligt tråkigt [...]", "Inget kul ord. Det får vi ändra på." och "Jag tror att det skulle vara bra om man hittade ett nytt ord" (Kristoffersson & Lagher 2005: 32, 84, 121). Bristen på adekvata beteckningar för dagens äldre är enligt skribenten i Hbl uppenbar, eftersom *pensionär* kan beteckna personer i mycket varierande ålder, *aldrig* inte upplevs som lämpligt och eufemismer som *senior* enligt vissa verkar onaturliga. Artikeln i Hbl ledde med tiden till en enkätundersökning bland finlandssvenskarna, vilken bland annat visar att dessa likväl inte har något emot att kallas *pensionär* som äldre (Internetkälla 3).

### 1.3 Åldersstereotyper

Att vara gammal är i dagens samhälle ingen merit utan det är ungdom och positiva konnotationer i anslutning till ungdom som utgör normen. Föraktet för ålder är emellertid inte kännetecknande enbart för vår tid utan är urgammalt samtidigt som aktning också visats för ålder och vishet. Man har i flera av varandra oberoende undersökningar kunnat konstatera att de äldre vanligen beskrivs som fattiga, ensamma, sjuka och deprimerade (se Thimm 2000: 61ff.). Både 1982 och drygt tjugo år senare (2005) ansåg 90% i en undersökning som genomfördes i Sverige att pensionärerna lider av ensamhet och 50% ansåg att de dessutom har tråkigt (Internetkälla 4). Om man förknippar en kategori människor med starkt förutsägbara attribut utvecklas småningom en stereotyp eller en extremt förenklad mental modell som inte ser människan utan de egenskaper man är van att tillskriva typen (Fowler 191: 92), det vill säga i och med att en person är äldre antas denne automatiskt vara ensam och sjuk. Denna stereotypa uppfattning vidmakthålls i hög grad av pressen som i sin bevakning av de äldres situation valt att i huvudsak fokusera på (bristfällig) vård (Nikula, K. 2006, 2007). Samma stereotypa uppfattning om de äldre gör det därtill viktigt att påpeka det oväntade, vilket man också finner goda exempel på i pressen där uttryck som "glada pensionärer", "pigg pensionär" eller "pigg och kry 76-åring" o.d. är ständigt återkommande (se Nikula, K. 2007)<sup>5</sup>. Eftersom etiketter kan fungera som utlösare av stereotypa

4 Intervjuerna finns samlade i Kristoffersson, B. & Lagher, H. (2005). *Pensionär – aldrig i livet! Elva 40-talister om sina liv och sin framtid*.

5 Kunde kanske också betraktas som en typ av "konversationella implikaturer", jfr Grice 1975.

uppfattningar om referenten, något som i sin tur kan påverka inställningen till de individer och den kategori som avses, är det sålunda inte enbart av språkligt intresse att granska beteckningarna för äldre utan också relevant ur samhällelig synvinkel (Öberg 1997: 23).

Också ordböckerna spelar roll i detta sammanhang. Stereotypier utgör inte en del av det språkiga tecknets betydelse utan är förgivettaganden (inte att förväxla med t.ex. lexikala presuppositioner, Svensén 2004: 298) eller gängse uppfattningar om en företeelse. Stereotypierna återspeglas däremot i språket. Valet av språkprov saknar därför inte betydelse då det gäller att skapa en bild av referenten – det är ju därför de är till. Utifrån den diskursanalytiska tanken att språk och samhälle hänger ihop på ett dialektiskt sätt kan språkprov som till sitt innehåll förmedlar en stereotyp uppfattning om de äldre cementera och föra dessa föreställningar vidare (Milles 2007: 173). Ordböckerna borde, utan att för den skull pruta på sanningen, motverka en sådan utveckling och i stället vara lyhörda för de förändringar i samhället som återspeglas i språket.

## 2. Beteckningar för äldre

I bilaga 1 har samlats ett antal beteckningar för äldre. Av de beteckningar som inte är markerade för kön denoterar i åldersordning *femtiopluss(are)*<sup>6</sup>, *senior*, *äldre*, *åldring* och *gamling* främst ålder och kunde därför teoretiskt avlösa varandra som beteckningar för den åldrande människan. Övriga beteckningar är förutom för kön ytterligare markerade för ett eller flera särdrag (se bilagan). Markeringarna bygger på betydelseuppgifterna i SOB, saknas lemmat har andra ordböcker tagits till hjälp. Strömarkeringar har sammanförts i särdraget Övrigt. För *gamling* anges i SOB könsmarkeringen som fakultativ, vilket gör att beteckningen också kunde fungera som benämning på äldre i allmänhet men uppges i den nyare *Norstedts svenska ordbok* (NSOB) ha enbart maskulin referens. Beteckningen *farbror* används utom som släktskaps- och tilltalsord enligt SOB också i betydelsen 'äldre manlig person' och *gumma* uppges betyda 'gammal kvinna'. Både *tant* och *gubbe* innehåller betydelsekomponenter som gör dem olämpliga som beteckningar för äldre i allmänhet eftersom med *tant* framhävs 'något förlegat' och med *gubbe* antyds 'svikande hälsa'. Övriga beteckningar är till sin grundbetydelse nedvärderande (*gubbjävel*, *käringskrälle*) eller för specialiserade till sin betydelse för att kunna användas om äldre individuellt eller om någotdera könet generellt.

Beteckningen *pensionär*, som enligt SOB innebär 'person som pensionerats av åldersskäl' och som i Svenska Akademiens ordlista (SAOL) ges betydelsen 'pensionstagare', fokuserar på att referenten lyfter pension och inte (längre) är förvärvsarbete. *Pensionär* är sålunda inte markerat för ålder även om de flesta pensionärer i realiteten är äldre. Beteckningen har uppförts i bilagan och tagits med i undersökningen därför att den med tiden blir användbar för alla som upp-

6 Skrivs i SAOL *femtioplussare* i andra sammanhang vanligen *femtioplussare*.

nått en viss ålder. *Pensionär* medtas också därför att den negativa inställningen till beteckningen (jfr rubriken) riktat uppmärksamheten på de problem som uppstått när det gäller att benämna och beskriva äldre människor och att hitta en beteckning som accepteras av dessa.

I SAOL redovisas sjutton sammansättningar med *pensionär(s)*- som förled bland vilka största delen direkt eller indirekt antyder verksamhet av något slag, till exempel *-biljett*, *-gymnastik*, *-kurs*, *-träff* och *-verksamhet*. Med språkproven "en glad pensionär; gammaldans för pensionärer" belyses beteckningens betydelse ytterligare. I sammansättningarna med *pensionär(s)*- ingår inte några negativa betydelsekomponenter som kunde förklara varför beteckningen blivit impopulär. Det faktum att de flesta pensionärer är ålderspensionärer och den negativa synen på ålder i allmänhet, men kanske framför allt de nygamlas bättre utgångsläge i jämförelse med tidigare pensionärsgenerationers, har sannolikt gjort att beteckningen uppfattas som otidsenlig.

Beteckningen *50+/femtiopluss(are)* är en eufemism som nått rätt stor spridning. *Femtiopluss(are)*, som för första gången dök upp i svensk press 1995 (Nyordsboken), finns i varken SOB eller *Bonniers svenska ordbok* (BSO), men i SAOL med tillägget "vardagligt". För att beteckningen *femtiopluss(are)* skall vara gångbar krävs att referenten åtminstone har fyllt femtio, men hur länge beteckningen är användbar och om *50+* kommer att följas av *60+* är ovisst. Redan nu upptar BSO beteckningen *55+*. Även om beteckningen *50+* har fått stor spridning i olika språk har den i enkätundersökningar inte visat sig vara särskilt populär (Kramer 1998: 237). Att det är oklart hur gammal man kan bli och fortfarande kallas *femtiopluss(are)* och att beteckningen utgör en eufemism som dessutom så tydligt associerar till reklam gör antagligen att den inte uppfattas som seriös och därmed inte vinner gehör bland alla språkbrukare.

Beteckningen *senior* 'som är äldre och har erfarenhet' (Nyordsboken) har förekommit i svenskan allt sedan 1993 och den ersätter numera ofta *pensionär* i betydelsen 'ålderspensionär'.<sup>7</sup> *Senior* förekommer inte i denna betydelse i SOB och BSO, däremot både i de nyare *Nationalencyklopedins ordbok* (NEO) och NSOB med betydelsen 'gammal person, vanligen pensionär' och i SAOL 'även om äldre person'. Hur långt upp i åren *senior* är gångbart är ovisst till följd av att ordböckerna ger lite olika betydelse. Beteckningen får en snävare betydelse i SAOL än i NEO som anger betydelsen 'gammal' eftersom en *äldre* person i själva verket är yngre än en *gammal* person. NEO jämför *senior* med beteckningen *pensionär* som är ju användbar alltifrån pensioneringen och livet ut. I tyskan används *Senior* främst under den tioårsperiod som följer efter pensioneringen (Thimm 2000: 58), som vanligen sker i 65-årsåldern. Beteckningen *senior* ingår numera i namnet på flera pensionärssammanslutningar i Finland och beteckningen verkar inte vidkännas några tydliga begränsningar i fråga om ålder. *Senior* är också frekvent i

7 Flera pensionärssammanslutningar går numera under namnet *Senior-*, t.ex. *Seniorförbundet*, *Svenska seniorer i Finland* och *Hufvudstadsbladets seniorer*.



nya sammansättningar, t.ex. *seniorboende*, *seniorbostad* (SAOL), *seniorhus*, *senior-dans* o.a. Med sin starka förankring i idrottspråket upplevs *senior* kanske bättre än *pensionär* motsvara den bild av aktiva äldre som är på väg att uppstå även om beteckningen inte tilltalar alla. Stort uppslagna rubriker som “Seniorerna konsumerar och njuter” (Hbl 9.4.2006) och “Allt fler satsar på de köpstarka seniorerna” (Hbl 10.3.2005) bidrar säkerligen också till att förmedla bilden av aktiva, levnads-glada äldre.

*Äldre* ges i SOB betydelsen ‘förhållandevis gammal, <ibl. ngt förskönande>’ och samma betydelse ges också i NEO med tillägget “ä. i absolut användning om person, utan ngn jämförelse, förhållandevis gammal”. Till följd av att alla kan vara *äldre* i relation till någon annan och *äldre* därmed inte kan knytas till en viss högre, kronologisk ålder besväras beteckningen inte av negativa konnotationer. Beteckningen är därför användbar i de flesta sammanhang och kan sannolikt accepteras av de flesta.

Som sammansättningsled ingår *äldre-* i ord som *äldreboende*, *äldreomsorg*, *äldresjukvård* och *äldrevård*. (SAOL). Detta innebär att de sammansättningar med *äldre-* som anförs huvudsakligen fokuserar på vård. Följande språkprov visar bland annat att *äldre* är gångbart om båda könen: “en distingerad [stilfullt förnäm] äldre herre; han gav sin sittplats åt en äldre dam i persianpäls” (SOB, NEO). Av språkproven framgår däremot inte att *äldre* kan användas substantiviskt för att beteckna äldre personer kollektivt, *de äldre*. Lemmat *äldre* ger säkerligen upphov till olika associationer, men särskilt välvalt är inte det senare språkprovet som verbalt återger en niddbild av (förmögna?), äldre damer. Trots att exemplet från början är olyckligt valt kvarstår detta i NEO. I NSOB ges enbart exemplet “en äldre herre”. Exempel på ett ur könssynvinkel positivt försök till nytänkande finns i *Svenskt språkbruk* (SSB) där *äldre* följs av språkprovet “*hon hjälpte en äldre dam över gatan*” (min kursivering).

*Åldring* är enligt SOB det samma som ‘gammal person’, enligt NSOB ‘en mycket gammal person’. Att en beteckning så tydligt implicerar hög ålder gör säkerligen att få för egen del gärna vill vidkännas denna. De negativa konnotationerna i anslutning till *gammal* låter sig knappast förenas med de äldres självbild; det är inte vi utan i regel ”de andra” som är gamla. Undersökningar har visat att självbilden och de andras bild av de äldre är tämligen olika men inte oberoende av varandra, eftersom de äldres självbild kan påverkas av omvärldens syn på dem. I stället för att uppfatta ålderdomen som en tid av avveckling ser de äldre själva ålderdomen ofta som en tid av lugn och ro med gott om tid och möjlighet att upprätthålla kontakt med andra. (Thimm 2000: 42)

*Åldring* ingår som förled i t.ex. *åldringbrott*, *åldringssjukvård*, *åldringsvård* (SAOL), *åldringbostad*, *åldringsfrågor* (BSO) och i följande längre språkprov: i SOB som första exempel “åldringar på långvården”, följt av “en vitskäggig, vördnadsbjudande åldring” och “pigga åldringar”. Två av de tre sammansättningarna fokuserar på de äldres behov av vård och den stereotypa uppfattningen att äldre är

sjuka återspeglas också i språkprovet "åldringar på långvården". "Pigga åldringar", som vid det här laget nära på fått karaktären av en verbal stereotyp eller korpusorienterad kollokation (se Svensén 2004: 208 ff.), har som förtäckt utgångspunkt samma uppfattning om de äldre, om inte vore det ju knappast motiverat med ett ständigt påpekande att en åldring är pigg och uttrycket hade knappast fått hög frekvens. I nyare ordböcker som NEO har exemplet med långvården utgått och i NSOB ingår enbart exemplet "en pigg åldring".

*Gamling* 'gammal (mans)person' (SOB) betecknas av SAOL som vardagligt. Avledningar på *-ing* som i *gamling* betecknar en person som kännetecknas av den egenskap som adjektivbasen anger. Som bas tjänar ofta adjektiv med pejorativ betydelse (Thorell 1981: 118). *Gammal* används nu för tiden i positiv bemärkelse främst om artefakter som *gamla* dvs. antika *möbler* och *gammal* dvs. lagrad *ost/konjak*, mera sällan om människor.

I SOB belyses användningen av *gamling* med följande exempel: "en orkeslös *gamling*; *gamlingar* på vårdhem; några *gamlingar* satt i parken och matade duvor"! Den stereotypa uppfattning om äldre som avspeglas i dessa språkprov är kränkande eftersom den ensidigt fokuserar på åldrande som avveckling och därtill framställer de äldre i en löjlig dager. Som konstaterats tidigare sammanfaller del äldres självbild inte alltid med de andras bild av de äldre, vilket gör att beteckningen *gamling* också används av de gamla själva i uttryck som *vi gamlingar* ... ofta väl skämtsamt och utan negativa konnotationer. Beroende av kontext kan i princip vilken beteckning som helst få till och med positiv innebörd.

### 2.1 Nya beteckningar för äldre?

Det är ingen nyhet att ord i ett språk efter hand börjar anses olämpliga antingen för att de upplevs som socialt stigmatiserade eller framhäver en inte önskvärd sida hos det betecknade. Kända exempel är bland annat *dårhus*, *städerska* och *kalhygge* som efter hand har ersatts med andra ord, *dårhus* till och med flera gånger. För att möta en sådan utveckling har med tiden uppstått olika strategier varav användningen av så kallade övertalningsdefinitioner är en. En övertalningsdefinition innebär att det betecknade ses i ett nytt perspektiv. I stället för att till exempel använda beteckningen *vårdmottagare* ses referenten som *kund* eller *klient*, eftersom denne ju betalar för tjänsten. En annan strategi är att ge ny form åt en verklighet som är behäftad med negativa associationer och att ge den en ny beteckning som inte säger någonting om verkligheten bakom, t.ex. *servicehus med helinackordering* 'boendeform för äldre'. En term som blivit belastad kan också avskaffas utan att en ersättning kommer i dess ställe. Slutligen kan ordet också behållas medan verkligheten förändras eller rehabiliteras, t.ex. *penna* (Allén & Gellerstam & Malmgren 1989: 87ff.).

Beteckningarna för äldre kan naturligtvis inte avskaffas med mindre än att nya ord kommer till men vilken strategi som kommer att följas är ovisst. Sifos användning av *mogna människor* (se ovan s. 337) kan betraktas som ett försök till en

övertalningsdefinition. Att utrota negativa stereotyper och samtidigt introducera en ny beteckning ställer sig svårt i synnerhet på kort sikt. Detta förutsätter att den bild av de så kallade nygamla som nu håller på att marknadsföras blir dominerande och att någon av de nya beteckningarna samtidigt blir allmänt accepterad. Eftersom det språkliga tecknet i princip är arbiträrt vore en ombenämning i teorin förhållandevis enkel medan det däremot visat sig vara svårt att få en sådan accepterad i praktiken (Allén & Gellerstam & Malmgren 1989: 97 ff.). Eftersom de nygamla skiljer sig från de äldre i tidigare generationer är det också tänkbart att åtminstone vissa av beteckningarna ovan bibehålls men med andra konnotationer och befriade från de stereotypa uppfattningar som nu åtföljer dessa. Om de nygamla aktivt gick inför att kalla sig till exempel *gamlingar* skulle detta innebära en rehabilitering av den gamla beteckningen (jfr *bög*) något som väl knappast är att räkna med. Risken är under alla omständigheter stor att man i strävan att avskaffa negativa stereotyper samtidigt skapar nya och att det på sikt blir frågan om ordmagi (jfr Grønvik 2002: 161).

Inom reklamen, men också i andra sammanhang, har man inte varit sen att skapa nya beteckningar för de nygamla. Förutom med den tidigare nämnda *oldies*, refereras till denna ålderkategori med beteckningar som *mappies* (M-magasin 2006/1: 25), *grumpies*, *woopies*, *harvest agers* och *best agers* (Hbl 9.4.2006). Förutom *oldies* och *harvest agers* är dessa engelska skapelser förkortningar av längre uttryck: *grumpies* uppges komma av "grown off older people", det vill säga 'fullvuxna, mogna människor' och *woopies* av "well off older people" eller 'välbärgade, äldre människor'. Den tidigare nämnda damtidningen *M-magasin* har alltifrån sitt första nummer hösten 2006 gått in för att kalla sin målgrupp *mappies*. Till grund för denna beteckning uppges ligga "mature affluent pioneering people" eller 'mogna attraktiva pionjärer'. (M-magasin 2006/1: 25) Också *golden age* och *generation silver* (Hbl 9.4.2006) har nämnts i sammanhanget. Förslag till svenska beteckningar är bland annat *rekordgenerationen* och de *mest erfarna* (Kristoffersson & Lagher 2005: 63, 68).

Uppgiften att hitta lämpliga beteckningar för de äldre påminner i stort om försöken att hitta ett neutralt ord för flickornas könsorgan; en lexikal lucka på systemnivå är det inte frågan om för beteckningar för äldre finns, utan problemet rör ordens lämplighet (jfr Milles 2006: 149). Eftersom det finns ord som står i lexikon och sådana som tillhör talspråket och eftersom man nu är i full färd med att skapa nya ord kan det småningom till och med bli frågan om överlexikalisering. Enligt Fowler (1991: 85) är det påfallande att förhållanden och personer som i en kultur i något avseende tilldrar sig intresset eller upplevs som ett problem överlexikaliseras eller omtalas med kvasisynonyma beteckningar.

### 2.3 Språkprovens funktion

Det finstilla i ordboksartiklarna eller definitionstillägget bidrar utan att direkt höra till definitionen till den semantiska beskrivningen. (SOB, XVIII ff.) Det

samma gäller också språkproven. På språkproven som visar hur uppslagsorden konstrueras i satser eller delar därav kan ställas många och till och med motstridiga krav (Nikula, H. 1995: 312ff.). Språkproven kan enligt Malmgren (1994: 109f.) bidra med semantisk, hit räknas också morfologiska exempel, syntaktisk, konnotativ och encyklopedisk information. De semantiskt motiverade språkproven utgör ett omedelbart förtydligande av en given betydelseuppgift medan till de syntaktiska språkproven räknas uppgifter om bland annat valens. Vid sidan av det språkliga tecknets denotation utgör konnotationerna ett viktigt betydelseelement. Med konnotationer avses de associationer och stämningar som lemmat konventionellt förknippas med. Encyklopediska språkprov ger tilläggsinformation som i bästa fall kan bidra till förståelsen av ett ord. Till dessa funktioner fogar Svensén (2004: 350) pragmatisk och filologisk funktion. Den pragmatiska funktionen ger information om användningen inom fackområden och stilnivåer. Den filologiska funktionen fullgörs av belägg.

Ur lexikalisk synvinkel är språkprov som "glad pensionär", "pigg pensionär" och "pigg åldring" tämligen triviala och säger ytterst lite om det lemma vars betydelse de valts att illustrera, eftersom i princip vem som helst kan vara både pigg och glad. Ett betydligt mer informativt bidrag till den semantiska beskrivningen hade varit *gammal pensionär* av vilket hade framgått att en pensionär inte automatiskt behöver vara gammal. Ses de tre språkproven i relation till den stereotypa uppfattningen om äldre, är användningen som exempel förklarlig. För "pigg åldring" som språkprov har antagligen därtill talat såväl frekvens som likheten med en kollokation.

De sociologiskt administrativa termerna *yngre äldre* och *äldre äldre* gör enligt SSB skillnad mellan gruppen yngre, pigga pensionärer och gruppen äldre, svagare och sjuka. Förklaringen stämmer illa eftersom inte i första hand hälsotillståndet utan ålder legat till grund för indelningen. Termerna åtföljs av språkprovet "modellen som nu växer fram för moderna pensionärer är de pigga yngre äldre". I samma ordbok beskrivs *grå panter* på följande vis: "om pensionär, vard. skämts. jag känner mig som om jag vore sjutti, färdig för servicehuset, kondis som en trött grå panter". Oberoende av vilken funktion språkprovet är avsett att ha är detta mera förvirrande än upplysande i synnerhet om man utgår ifrån att *grå panter* har betydelsen 'person som är vital och aktiv' (Nyordsboken) eller 'vital pensionär' (NEO). I anslutning till *gubbe* ges i SSB språkprovet "på bänken satt en gammal gubbe och matade duvorna", men i vilken funktion? och i NSOB "några gubbar med käpp vilade sig på en parkbänk", i SOB 86 ytterligare "i en koja i skogen bodde en klok gubbe". Om det första exemplet är avsett att ge en bild av verkligheten är det förfelat för äldre damer och små barn matar sannolikt duvorna i ännu större utsträckning än äldre män. Språkprovet finns med något annan ordalydelse också i SOB men här under uppslagsordet *gamling*: "några gamlingar satt i parken och matade duvor". Sagans "kloka gubbar", som kanske bodde granne med Rödluvans farmor, tillhör ett utdöende släkte. Det samma gäller för

*gummor* sådana de beskrivs i språkprovet i SOB 86: “några tandlösa gummor i svarta sjalar”. Beskrivningen leder snarare leder tankarna till sagans häxa än till “en vithårig liten gumma” (SOB, NSOB) som närmast konnoterar vänlighet och rundhet, jfr *lilla gumman* ‘min lilla vän’ och *gumman* ‘min hustru’. Förmodligen har språkproven utgått ur SOB, NEO och NSOB just därför att de upplevts som otidsenliga.

Det är många gånger svårt att avgöra vilken eller vilka funktion(er) ett språkprov är avsett att ha. Språkprovet “de vithåriga patriarkerna på ålderdomshemmet” i anslutning till *patriark* ‘ärevördig åldring’, har i likhet med *åldring* ‘gammal person’ och “åldringar på långvården” (SOB) sannolikt främst konnotativ funktion. Äldre människor placeras i språkproven påfallande ofta på ålderdomshem, långvård, vårdhem eller i servicehus. Verkligheten må vara sådan, men ett stereotypiskt tänkande bidrar säkert också till att t.ex. “en vithårig patriark” inte placerats på en mera upphöjd plats, till exempel i högsätet eller bordsändan.

#### 2.4 Ensidiga språkprov?

Med en nött metafor sägs språket spegla samhället. De språkprov som diskuteras ovan kan eventuellt försvaras med att exemplen antingen är autentiska eller semiautentiska<sup>8</sup> och hämtade ur korpusar. Oberoende av om ett språkprov är autentiskt eller inte och om det förmedlar gängse åsikter eller inte, bör språkprovet bedömas inte enbart utifrån hur detta fyller sin funktion i ordboken utan också med hänsyn till om den information exemplet förmedlar öppet eller förtäckt är etiskt försvarbar. Det finns enligt Malmgren ypperliga, autentiska exempel som av etiska skäl inte kan tas med i en ordbok, som exempel nämns språkprov med rasistiskt innehåll (Malmgren 1994: 111f.). Flera av språkproven i anslutning till beteckningarna för äldre ligger sedda ur denna synvinkel i gråzonen för vad som är att rekommendera eftersom de ger uttryck för en stereotyp negativ uppfattning om referenten.

När kritik riktas mot exemplen i en ordbok är det enligt Grönros inte alltid lätt att avgöra om kritiken riktar sig mot exemplet eller mot den verklighet och det språkbruk som exemplet beskriver. Som exempel tar Grönros det finska ordet för *styvmor* (*äitipuoli*) och språkprovet “den elaka styvmodern i sagorna” (Grönros 2003: 103) som kritiserats för att ge en alltför negativ beskrivning av styvmödrarna. Sociala stereotyper kan rikta sig mot i stort sett vilka kategorier som helst men vanligen är det minoriteter av olika slag (t.ex. etniska och sexuella) och svagare parter i samhället, dit de äldre åtminstone hittills har kunnat räknas, som drabbas. Språkproven såväl som sammansättningarna i anslutning till beteckningar för äldre fokuserar rätt ensidigt på åldrandet som avveckling och ger därmed en onödigt negativ bild av referenten.

8 I ordböckerna framgår inte vilka exempel som är autentiska. Malmgren bedömer att en klar minoritet av exemplen i SOB är autentiska. BSO bygger i sin tur inte på en datorbaserad korpus. (Malmgren 1994: 111f.)

En kritisk syn på språkproven innebär inte att en ordbok skall vara "politiskt korrekt", något som väl de flesta seriösa lexikografer med rätta skulle motsätta sig. Men eftersom man normalt utgår ifrån att en ordbok förutom ett ords betydelse också visar lemmat i kontext, kan man inte bortse ifrån att språkproven i ordboksartiklarna bidrar till att hos ordboksanvändarna skapa eller vidmakthålla vissa attityder (Kram 1998: 168f.), i detta fall "ålderistiska" innebörder. När SOB för första gången utkom på 1980-talet var det relativt långt till dess att samhället "skulle drabbas av en åldringsboom" och det skulle dröja ända till 1997 innan ordet *ålderism*<sup>9</sup> dök upp i svenskan. Det är därför förstäeligt att språkprov som vidarebefordrar och stärker en negativ uppfattning om de äldre ingår i den första upplagan, eftersom stereotyper är svåra att upptäcka då de ju är uttryck för gemensamma "kunskaper". Pensionärsorganisationerna hade visserligen under 70- och 80-talen lagt ner mycket arbete på att motverka fördomsfulla och stereotypa bilder av äldre människor (Internetkälla 6), men det är väl egentligen först i våra dagar som åldersfascism och diskriminering av äldre, det vill säga ålderism, blivit ett tema också utanför dessa kretsar.

### Sammanfattning och avslutning

På sistone har uppstått en efterfrågan på nya beteckningar för äldre eftersom existerande beteckningar upplevs vara otidsenliga och illa motsvara den bild av de äldre som nu är på väg att växa fram. Vissa nya beteckningar har också sett dagens ljus (*femtipluss*) och andra har fått en utvidgad betydelse (*senior*). Om den stereotypa uppfattningen om äldre inte förändras kommer dessa beteckningar sannolikt att möta samma öde som många andra ombenämningar, det vill säga *pensionär* blir *senior* som i sin tur blir något annat. Att beteckningar som *åldring* eller *gamling* skulle bli allmänna förefaller föga troligt. De anglicismer som dykt upp inom marknadsföringen har åtminstone inte heller hittills fått någon större spridning. *Äldre* (*äldre x.*, *de äldre*) är såväl komparativt som absolut sannolikt den enda beteckning som de flesta kan omfatta.

Samhällets negativt stereotypa syn på åldrandet reflekteras i stor utsträckning i ordböckernas språkprov och i valet av sammansättningar där beteckningar för äldre ingår, vilket framgått vid genomgången av ett antal lättillgängliga ordböcker. En deskriptiv ordbok skall självfallet spegla språkbruket och lexikograferna kan numera hänvisa till korpusar om innehållet i en ordbok ifrågasätts. Även om det är svårt, är det likväl viktigt att skilja mellan å ena sidan ordbetydelser och språkliga konstruktioner och å andra sidan sociala stereotyper och encyklopedisk kunskap. Icke desto mindre vore en mera balanserad syn på åldrandet att önska även om de icke-stereotypa språkproven i en korpus inte uppnår samma

9 *Ålderism* (eng. Ageism) användes för första gången 1997 av Lars Andersson, professor i *Tema äldre och äldrane* vid Linköpings universitet. Den engelska termen skapades av Robert Butler redan år 1969 (Internetkälla 5).

frekvenser som de stereotypa. Ordspråket säger att hälsan tiger still, något som säkert också ger utslag i en textkorpus.

Språkproven är i stort sett de samma i de undersökta ordböckerna. I de fall ett negativt, stereotypt språkprov utgått ur en nyare upplaga, förblir det oklart på vilka grunder detta skett. Förorden är synnerligen knappa på information och talar vanligen i allmänna ordalag om rättelser och tillägg. Det är också möjligt att exemplen strukits till följd av att de inte fyllt någon egentlig funktion, kanske upplevts som otidsenliga eller helt enkelt till följd av brist på utrymme då nya uppslagsord kommit till.

## Litteratur

### Ordböcker

- BSO 2006 = *Bonniers svenska ordbok*. 9 uppl. Stockholm: Bonniers.
- NEO 1995–1996 = Nationalencyklopedins ordbok. Utarbetad vid Språkdata vid Göteborgs universitet. Höganäs: Bra Böcker.
- NSOB 2004 = *Norstedts svenska ordbok*. Stockholm: Språkdata och Norstedts Förlag. Nyordsboken. Med 2000 nya ord in i 2000-talet. 2000. Stockholm: Svenska språknämnden och Norstedts Ordbok.
- SAOL 2006 = *Svenska Akademiens ordlista över svenska språket*. Stockholm: Svenska Akademien o. Norstedts Akademiska Förlag.
- SOB 86 = *Svensk ordbok 1986*. Stockholm: Språkdata & EsselteStudium.
- SOB = *Svensk ordbok 1999*. 3 reviderade uppl. Stockholm: Norstedts Ordbok
- SSB 2004 = *Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Stockholm: Svenska språknämnden.

### Övrig litteratur

- Allén, Sture & Gellerstam, Martin & Malmgren, Sven-Göran 1989: *Orden speglar samhället*. Stockholm: Allmänna Förlaget.
- Andelin, Jan-Erik 2007: *Mommo har fallit – var är geriatrikern?* (Borgåbladet 5.5.2007)
- Fairclough, Norman 1992: *Discourse and Social Change*. Cambridge & Oxford: Polity Press.
- Fiehler, R. & Thimm, C. (Hrsg.) 1998: *Sprache und Kommunikation im Alter*. Wiesbaden: Westdeutscher Verlag, 257–278.
- Fowler, Roger 1991: *Language in the News. Discourse and Ideology in the Press*. London & New York: Routledge.
- Grice, H[erbert] P[aul] 1975: *Logic and Conversation*. I: Cole, P & Morgan, J.L. (Ed.) *Syntax and Semantics 3. Speech Acts*. New York: Academic Press.
- Grönros, Eija-Riitta 2003: *Ordboken och värderingarna*. I: *Nordiske studier i leksikografi 6*. Tórshavn: Nordisk forening for leksikografi, 101–107.
- Grønvik, Oddrun 2002: *Er de stygt å seie neger?* Om konnotasjonar, bruksmarkering og nemne for folkegrupper. I: *LexicoNordica 9*, 149–163.

- Jernvall-Ingman, Birgitta 2004: *Medelålders, äldre – men inte åldring*. (Hbl. 2.12.2004)
- Karlberg, Peter 2006: *Seniorerna konsumerar och njuter*. (Hbl 9.4.2006)
- Kram, Kristina 1998: "... blott rent undantagsvis om kvinnor. Om representationen av kvinnan i moderna ordböcker. I: *Språk & Stil* 8, 159–171.
- Kramer, Undine 1998: AGEISMUS – Zur sprachlichen Diskriminierung des Alters. I: Fiehler, R. & Thimm, C. (Hrsg.): *Sprache und Kommunikation im Alter*. Wiesbaden: Westdeutscher Verlag, 257–278.
- Kristoffersson, Birgitta & Lagher, Håkan (red.) 2005: *Pensionär – aldrig i livet! Elva 40-talister om sina liv och sin framtid*. Lidingö: Bokförlaget Langenskiöld.
- M-magasin* 1/2006. Stockholm: amelia Publishing Group.
- Malmgren, Sven-Göran 1994: Språkprovets form och funktion i svenska betydelsordböcker från Östergrens Nusvensk ordbok till Svensk ordbok. I: *LexicoNordica* 1, 107–117.
- Milles, Karin 2006: En jämställd kroppsdel? Om lanseringen av ett neutralt ord för flickors könsorgan. I: *Språk och stil* 16, 149–177.
- Nikula, Henrik 1995: Exemplens funktion i ordböcker. I: *Nordiske studier i leksikografi* 3. Reykjavik: Nordisk forening for leksikografi, 311–320.
- Nikula, Kristina 2006: JAG EN GERONT? Åldersdiskursen i finlandssvensk press. I: *Svenskan i Finland* 9. Jyväskylä: Jyväskylä universitet, 161–171.
- Nikula, Kristina 2007: Den åldrande kvinnan och pressen. Under utgivning i rapport från konferensen *Språk och kön* 6. Uppsala: Uppsala universitet.
- Svensén, Bo 2004: *Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Norstedts Akademiska Förlag.
- Thimm, Caja 2000: *Alter – Sprache – Geschlecht. Sprach- und kommunikationswissenschaftliche Perspektiven auf das höhere Lebensalter*. Frankfurt & New York: Campus Verlag.
- Thorell, Olof 1981: *Svensk ordbildningslära*. Lund: Esselte Studium.
- Winther Jørgensen, Marianne & Phillips, Louise 1999: *Diskursanalys som teori och praktik*. Lund: Studentlitteratur.
- Öberg, Peter 1997: *Livet som berättelse. Om biografi och åldrande*. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 62. Uppsala: Uppsala universitet.

#### *Elektroniska källor*

**Internetkälla 1** = Statistikcentralen

<[http://www.stat.fi.tup/suoluk/suoluk\\_vaesto\\_sv.html](http://www.stat.fi.tup/suoluk/suoluk_vaesto_sv.html)> Hämtat 24.7.2005

**Internetkälla 2** = *Nyhetsarkiv. Mogna människor*. Research International, Sifo. <<http://www.sifo.se/Public/News/NewsArchive.aspx?ArticleId=439062fb-0963-49d1-8d>

Hämtat 17.4.2007

**Internetkälla 3** = Livsfasernas förändring

<[http://64.233.183.104/search?q=cache:5TwjmpW1VVcJ:www.fskompetenscentret.fi/pdf/150306HERBERTS.pdf+Herberts+livsfasernas+f%C3%B6r%C3%A4ndring&hl=sv&ct=clnk&cd=1&gl=fi&lr=lang\\_sv](http://64.233.183.104/search?q=cache:5TwjmpW1VVcJ:www.fskompetenscentret.fi/pdf/150306HERBERTS.pdf+Herberts+livsfasernas+f%C3%B6r%C3%A4ndring&hl=sv&ct=clnk&cd=1&gl=fi&lr=lang_sv)> Hämtat 8.8.2007



- Internetkälla 4** = Tornstam, Lars (2007). Stereotypes of Old People Persist: A Swedish "Facts on Ageing Quiz" in a 23-year Comparative Perspective. I: *International Journal of Ageing and Later Life*. Linköping University Electronic Press 2007 2:1. 33-59.  
<<http://www.ep.liu.se/ej/ijal/>> Hämtat 15.8.2007.
- Internetkälla 5** = Agebjörn, Anika: *Åldrandets sociala dimensioner* <<http://www.liu.se/forskning/portratt/Lars+Andersson>>. Hämtat 29.3.2007
- Internetkälla 6** = Karlsson, Lena: *Offrer eller livsnjutare?* <[http://www3.lu.se/info/lum/LUM\\_05\\_01/07\\_offer.html](http://www3.lu.se/info/lum/LUM_05_01/07_offer.html)>. Hämtat 29.3.2007

Kristina Nikula  
professor (emeritus), f. 1943  
Puolalaparken 1 b B 12  
FI-20100 ÅBO  
kristina.nikula@uta.fi

## Fältet /äldre/

## BILAGA

	1	2	3	4	5	6	7	8	9
emeritus	+	+							
emerita	+		+						
farbror	+	+							
femtiopluss	+								
fossil	+						+		
gamling	+(+)								
gamyl*	+								
geront	+	+							
grande dame	+		+		+	+			
grand old man	+			+	+				
grågosse	+	+			+				
gråhårsman	+	+							
grå panter	+	+			+				
gubbe	+	+						-	
gubbdjäväl	+	+					-		
gubbskrälle	+	+						-	
gubbstackare	+	+					-		
gubbstrutt	+	+					-		
gubbtjuv/tjyv	+	+					+		
gumma	+		+						
kalkskuta	+		+				-		
krutgubbe	+	+						+	
käring	+		+			-		-	
käringskrälle	+		+					-	-
nestor	+	+							
pantertant	+		+		+				
patriark	+	+					+		
pensionär									
senior	+								
stofil	+	+		+					
tant	+		+	(+)					
veteran	+	+					+		
åldring	+								
äldre	+								

\*finlandssvenska 'äldre medlem på studentnation', äv. 'gammal person'

1 = ålder

2 = mask.

3 = fem.

4 = ålderdomlighet

5 = aktivitet

6 = uppskattning

7 = hälsa

8 = utseende

9 = övrigt

## DanNet: udvikling og anvendelse af det danske wordnet

A wordnet for Danish is under compilation as a joint project between the Centre for Language Technology at the University of Copenhagen and the Society for Danish Language and Literature. Both partners have recently been involved in the development of relevant lexical resources which are utilized as an important part of the current project, most importantly The Danish Dictionary, a corpus-based dictionary of modern Danish, and the international SIMPLE project. This article describes how the existing data are reused to create a much sought-after resource within Danish language technology. Typical examples and problems faced during the editing process are presented, with focus on polysemy, synonymy and semantic classification. Finally, we outline some perspectives for lexicographic products aimed at human users. Specifically, the development potential for the productive function of dictionaries is discussed, as are ways of improving the pedagogical function of learners' dictionaries.

### 1. Indledning

Det første wordnet, Princeton WordNet (Fellbaum 1998), blev udviklet i et psykolingvistisk forskningsmiljø i et forsøg på at afbilde det mentale leksikon, altså den orden som vi antager at ordene indtager i vores bevidsthed når vi forstår og producerer sprog. Men ud over at være en god platform for psykolingvistiske eksperimenter udgør wordnets også en interessant resurse for teknologisk forskning og udvikling inden for sprogteknologi og kunstig intelligens. I disse fagmiljøer er manglen på ordsemantik i stor målestok nemlig ofte den barriere der vanskeliggør opskalering af en prototype til et stort velkørende system. Også for leksikografien er wordnets interessante; de er i grunden blot ordbøger udformet ud fra en onomasiologisk tankegang med orddefinitionen som det væsentlige omdrejningspunkt, og derfor kan de give ny inspiration til hvordan især *digitale* ordbøger kan opbygges og anvendes i fremtiden.

I denne artikel giver vi først en introduktion til DanNet-projektet og beskriver i hvilken ramme det realiseres, nemlig som et samarbejdsprojekt mellem et sprogteknologisk miljø og et leksikografisk miljø (afsnit 2). Der er i høj grad tale om genbrug af eksisterende resurser, først og fremmest i form af genbrug af *Den Danske Ordbog*. Dernæst giver vi i afsnit 3 en række eksempler fra selve udviklingsarbejdet hvor polysemi, synonymi og semantisk klassifikation udgør nogle af de typiske problemstillinger der er i fokus når en traditionel ordbog skal genanvendes i form af et wordnet. Endelig kommer vi i afsnit 4 ind på an-

vendelser af wordnettet i ordbogsammenhæng. Vi ser blandt andet på hvordan et wordnet kan anvendes til at udvikle ordbøgers *sprogproduktive* funktion, samt på hvordan de kan bruges til at forbedre learnerordbøgers *pædagogiske* funktion.

## 2. DanNet – et leksikalsk-semantic wordnet for dansk

### 2.1. Et samarbejdsprojekt

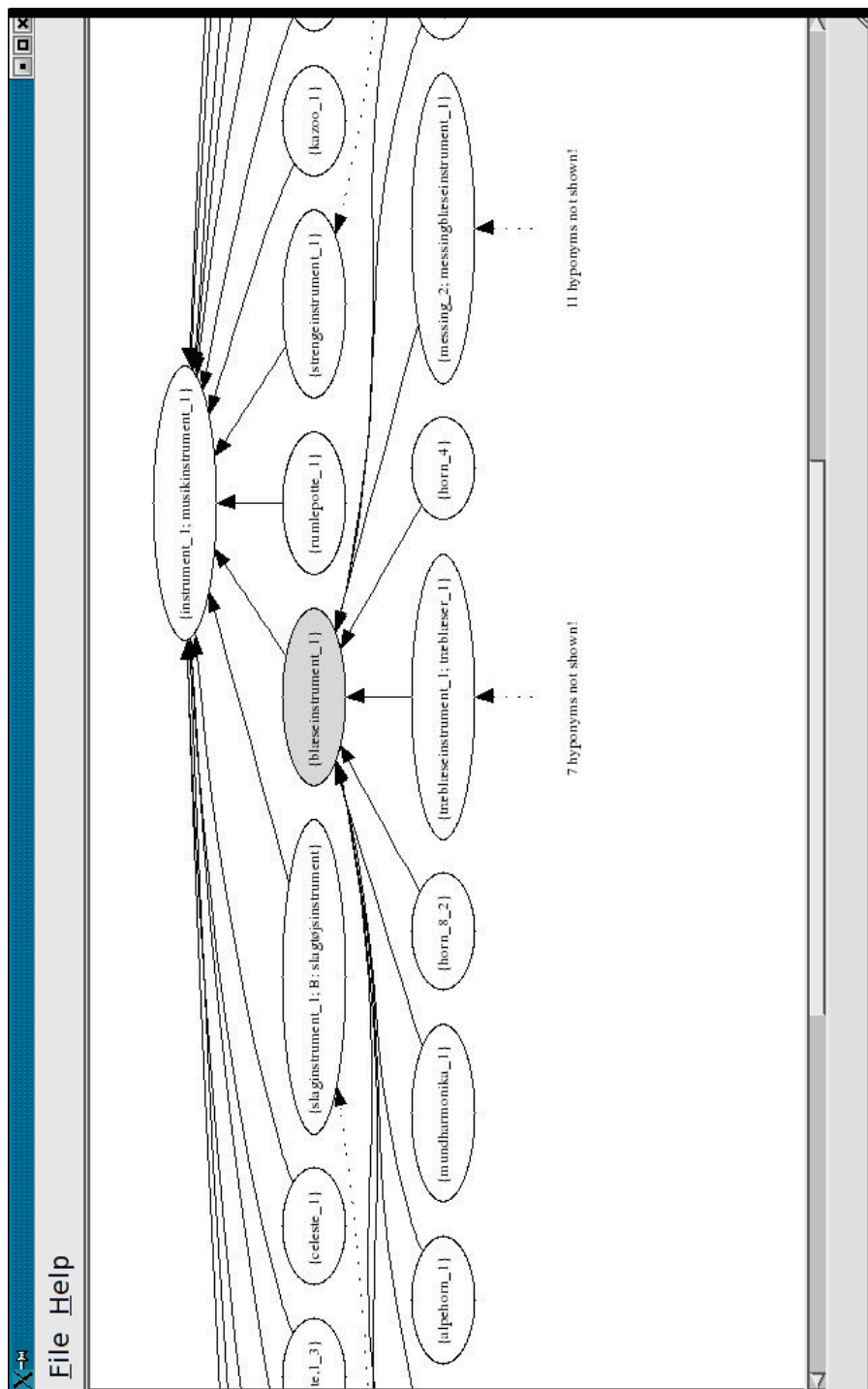
DanNet-projektet er et samarbejdsprojekt mellem Center for Sprogteknologi, Københavns Universitet (CST) og Det Danske Sprog- og Litteraturselskab (DSL). Hvert af disse to miljøer har inden for de senere år afsluttet et omfattende leksikalsk projekt hvis resultater nu tilsammen udgør et væsentligt udgangspunkt for udviklingen af det danske wordnet: *Den Danske Ordbog* (DDO) er en stor ordbog der på baggrund af korpusundersøgelser beskriver ords betydninger, primært ved hjælp af definitioner og brugseksempler (Lorentzen 2004). CST har deltaget i det internationale projekt SIMPLE (Semantic Information for Multifunctional, Plurilingual Lexica, jf. Lenci m.fl. 2000), hvor ontologiske modeller for formel betydningsbeskrivelse er blevet afprøvet og konsolideret på et bredt ordmateriale; en afprøvning på 10.000 begreber i dansk er udført i den danske del af projektet, SIMPLE-DK (Pedersen og Paggio 2004). Erfaringerne fra disse to projekter i henholdsvis det leksikografiske og det sprogteknologiske miljø har givet gode forudsætninger for at udarbejde et wordnet for dansk.

DanNet-projektets første fase er finansieret af Forskningsrådet for Kultur og Kommunikation og løber i perioden 2005–2008. Målet er i denne fase at nå op på en beskrivelse af 40.000 betydninger i dansk samt at få en delmængde af disse forbundet med det engelske Princeton WordNet via tværproglige links. På længere sigt er det målet at nå op på en dækningsgrad der svarer til hele *Den Danske Ordbogs* ca. 100.000 betydninger. En række publikationer beskriver de forskellige stadier af projektet; jf. Pedersen & Asmussen (2006), Pedersen m.fl. (2006), Pedersen & Sørensen (2006) & Asmussen m.fl. (2007); se i øvrigt projektets hjemmeside [www.wordnet.dk](http://www.wordnet.dk).

Et af målene med wordnettet er at det skal kunne anvendes i forskellige sprogteknologiske værktøjer; her tænkes især på begrebsbaseret søgning, men også på værktøjer der kan resumere og indeksere tekst på en mere intelligent måde end de gængse, mønsterbaserede værktøjer er i stand til.

### 2.2. Basisbegreber der anvendes i DanNet

Som det er tilfældet i både Princeton WordNet og i EuroWordNet (Vossen 1999), anvender vi i DanNet såkaldte *synsets* som grundlæggende enheder i netværket. Et synset – eller synonymsæt – er det sæt af synonymmer der tilsammen udgør et begreb. Fx kan man i figur 1 se at *slaginstrument* og *slagtøjsinstrument* optræder i samme synset og dermed opfattes som refererende til samme begreb.



Figur 1. Et wordnet er et netværk af relationer mellem begreber, eller såkaldte 'synsets'

Som udgangspunkt for den praktiske udvikling af wordnettet har vi udtrukket grupper af ord med samme overbegreb fra Den Danske Ordbog, som det fx ses i figur 2 for overbegrebet *redskab*. Herudfra bygges de over- og underbegrebs-hierarkier som danner wordnettets grundstruktur.

Focus on table:				Temporary selection of senses (not locked to user)
Expanded item	Pos	GenProx	Definition	
1 afrefter_1	sb	redskab	redskab til at gøre en flade el. linje lige med	
2 applikator_1	sb	redskab	lille pensel- el. spatellignende redskab der fx bruges når man skal læ	
3 bestik_1_1	sb	redskab	redskaber til at spise med el. til at tage mad fra fæde, gryder m.m.	
4 blyantspidser_	sb	redskab	redskab til at spidse blyanter med	
5 bor_1_1	sb	redskab	redskab til at bore huller med, ofte udformet som en rund metalstang	
6 bræt_2_3	sb	redskab	sports- el. legeredskab i form af en plade som man står på mens det	
7 bue_1_4	sb	redskab	redskab hvormed man stryger hen over strengene på et strygeinstru	
8 børste_2_1	sb	redskab	redskab der består af en plade af træ, plast e.l. hvorpå der er fastgjor	
9 bådshage_1	sb	redskab	redskab der består af et langt skaft med en krog og en spids af metal	
10 bære_1	sb	redskab	redskab til at bære el. køre en syg, tilskadekommen el. død person i	
11 cigarklipper_1	sb	redskab	redskab til at klippe spidsen af en cigar	
12 donkraft_1	sb	redskab	redskab el. anordning der mekanisk el. vha. hydraulik er i stand til at	
13 dorn_1	sb	redskab	redskab, maskindel e.l. i form af en (lille) rund metalstang	
14 dorn_1_1	sb	redskab	redskab i form af en lille metalstang som i den ene ende er tilspidse	
15 drejestål_1	sb	redskab	skærende værktøj på en drejebænk, i form af en metalstang med et	
16 dåseåbner_1	sb	redskab	redskab til åbning af konservesdåser	
17 ekspander_1	sb	redskab	redskab som består af en stram fjeder med to håndtag for enderne,	
18 flaskerenser_1	sb	redskab	redskab til at rense flasker med i form af et stykke kraftig metaltråd n	
19 flaskeåbner_1	sb	redskab	redskab hvormed man kan åbne en flaske som er lukket med kapse	
20 fluesmækker_1	sb	redskab	redskab der består af et langt, tyndt, fjedrende skaft af plastic el. stål	
21 fork_1	sb	redskab	gaffelformet redskab der har to (el. tre) spidse tænder og et langt ska	
22 fugeske_1	sb	redskab	meget smal murske hvormed fugerne i et murværk kan udfyldes me	
23 gastænder_1	sb	redskab	redskab der vha. en gnist el. en glødetråd kan antænde gassen i et c	
24 grejer_1	sb. p	redskab	sammenhørende redskaber, hjælpemidler el. ejendele	
25 gymnastikredsk	sb	redskab	redskab der bruges til gymnastik	
26 hakke_1_1	sb	redskab	redskab som består af et langt, buet metalhoved, ofte spidst i den er	
27 hammer_1	sb	redskab	redskab som typisk består af et træskaft med en tværstillet, kort, tyk	
28 haspe_1_2	sb	redskab	redskab til at vinde gam el. tråd op på, bestående af en aksel med u	
29 hegle_1_1	sb	redskab	redskab som består af et bræt med rækker af spidse stål- el. jerntær	

Figur 2. Genus proximum-oplysninger udtrukket fra Den Danske Ordbog

Ud over relationer mellem over- og underbegreber etableres der også en lang række andre relationer mellem de enkelte synsets. I figur 3 ses en oversigt over de relationer der tages i anvendelse. Vi har som udgangspunkt valgt at fokusere på de relationer der angives i den almindelige ordbogsdefinition; hvis det her er blevet skønnet relevant at angive hvilket materiale noget er lavet af, eller hvad noget bruges til, etableres disse relationer også i DanNet (via relationerne *has\_holo\_madeof* og *for\_purpose\_of*). Mange af relationerne nedarves desuden fra overbegreberne; *for\_purpose\_of* *spise* nedarves fx til alt spiseligt i hierarkiet.

Relationer i DanNet	Eksempel
<i>concerns</i>	<i>fodboldmål concerns sport</i>
<i>for_purpose_of</i>	<i>hammer for_purpose_of hamre</i>
<i>fpo_object</i>	<i>klipse fpo_object klips</i>
<i>has_holo_madeof</i>	<i>mel has_holo_madeof brød</i>
<i>has_holo_member</i>	<i>partimedlem has_holo_member parti</i>
<i>has_holo_location</i>	<i>oase has_holo_location ørken</i>
<i>has_holo_part</i>	<i>øje has_holo_part ansigt</i>
<i>has_hyperonym</i>	<i>birketræ has_hyperonym træ</i>
<i>has_orthohyperonym</i>	<i>vejtræ has_orthohyperonym træ</i>

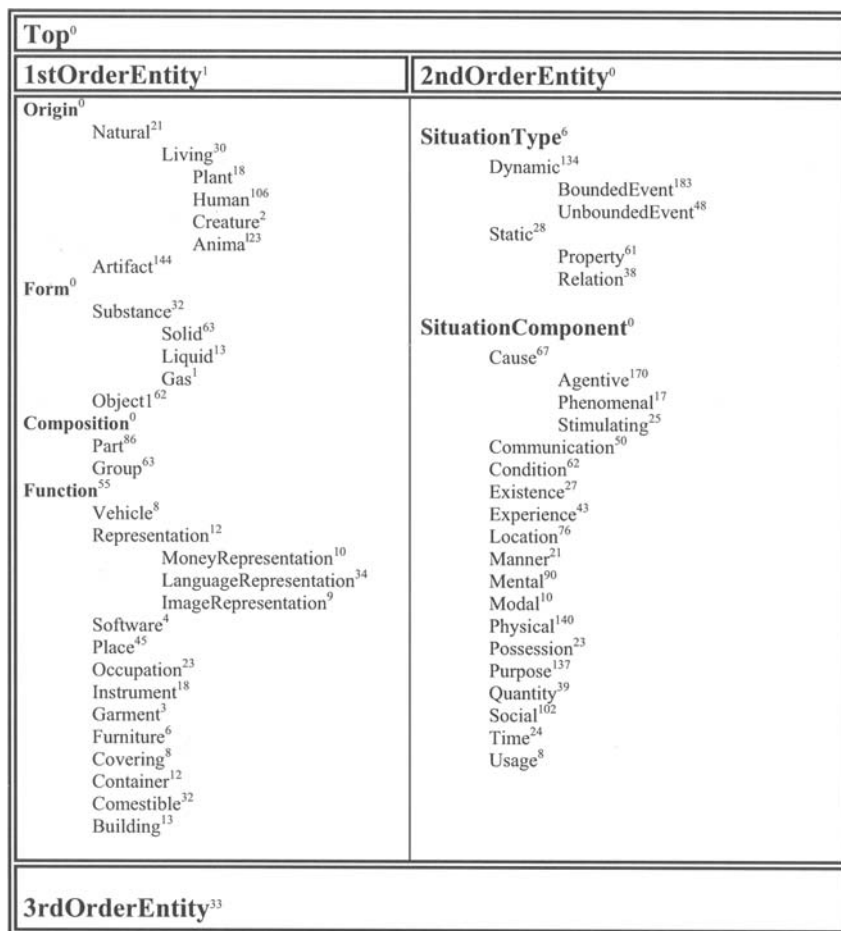
has_hyponym	<i>træ</i> has hyponym <i>birke</i> <i>træ</i>
has_mero_madeof	<i>brød</i> has_mero_madeof <i>mel</i>
has_mero_member	<i>parti</i> has_mero_member <i>partimedlem</i>
has_mero_part	<i>hånd</i> has_mero_part <i>finger</i>
has_mero_location	<i>ørken</i> has_mero_location <i>oase</i>
role_agent	<i>passager</i> role_agent <i>rejse</i>
role_patient	<i>modtager</i> role_patient <i>modtage</i>
made_by	<i>bagværk</i> made_by <i>bage</i>
near_synonym	<i>køkkentøj</i> near_synonym <i>kogegrej</i>
xpos_near_synonym	<i>behandle</i> xpos_near_synonym <i>behandling</i>

Figur 3. Relationer der anvendes i DanNet

Endelig forsynes alle begreber i DanNet med en top-ontologisk type. Her anvender vi EuroWordNets flerdimensionelle top-ontologi byggende på Lyons entiteter af 1., 2. og 3. grad, jf. figur 4. Fordelen ved at angive top-ontologiske typer er blandt andet at man kan identificere mængder af begreber der deler visse fælles egenskaber, men som ikke nødvendigvis er placeret under det samme danske overbegreb. Planter og dyr som mennesker spiser, er fx ikke placeret under det danske begreb *fødevarer*, men derimod under henholdsvis *planter* og *dyr*. Via den ontologiske type Comestible kan man imidlertid identificere dels tilberedte madvarer, dels de dyr og planter som mennesker typisk spiser. Man kan også have brug for at fremfinde en delmængde af de begreber der hører under et dansk overbegreb. Fx kan man udtrække alle jobfunktioner (som fx *maler* og *skolelærer*) via den ontologiske type Human+Object+Occupation. Disse er alle listet under det danske overbegreb *person*, men her finder man også *spanier*, *demokrat* og *fodgænger*.

### 3. Opbygningen af det danske wordnet med udgangspunkt i DDO

Selve opbygningen af det danske wordnet foregår som nævnt med udgangspunkt i DDO's betydningsinddeling og definitioner. I arbejdet udnytter vi dels at hver enkelt betydning i ordbogen kan udtrækkes som en entydigt nummereret enhed, dels at genus proximum for hver af disse betydninger kan udtrækkes direkte. Overbegrebet, der samtidig er et ord fra selve definitionen, er nemlig markeret i et selvstændigt felt i ordbogsstrukturen. I figur 5 ses et udsnit af ordbogsstrukturen for substantivet *flaskeåbner* med dets entydige betydningsnummer samt ordet *redskab* som genus proximum, markeret i feltet Genprox.



Figur 4: Top-ontologien der anvendes i DanNet (Vossen 2005)

```

flaskeåbner, sb.
<-Semdel>
  <-Semem>
    <-Denbet DanNetSemID="21020396">redskab hvormed man kan åbne en flaske som er
    lukket med kapsel el. prop
    <-Genprox>redskab

```

Figur 5. Udsnit af DDO's ordbogsstruktur med betydnings-id, definition og genus proximum

Wordnettet opbygges ved automatisk at udtrække alle de ord fra DDO der deler genus proximum, og derefter vurdere og evt. tilrette de data der fremfindes. I figur 2 ses fx et udsnit af udtrækket på overbegrebet *redskab*.



### 3.1. Efterredigering af DDO's genus proximum-oplysninger

Genus proximum-oplysningen fra DDO er dog langt fra altid entydig idet der i notationen hverken er skelnet mellem homografi eller et overbegrebs eventuelt flere betydninger. Derfor er der ofte tale om et større udredningsarbejde når man udtrækker et antal ord med samme overbegreb. I figur 6 ses et eksempel på hvordan samme overbegreb, i dette tilfælde *blad*, er brugt for ord af helt forskellig semantisk type, og hvor vi derfor i hvert enkelt tilfælde skal beslutte hvilken af *blads* mange betydninger det drejer sig om.

	Expanded ler	Pos	Genf	Definition
4	basilikum_1_1	sb.	blad	friske el. tørrede blade fra denne p
5	højblad_1	sb.	blad	blad som sidder højt oppe på stilk
6	billedblad_1	sb.	blad	blad med mange billeder og unde
7	koriander_1_2	sb.	blad	blade fra denne plante, brugt som
8	kirkeblad_1	sb.	blad	blad som regelmæssigt udgives til
9	klinge_1_1	sb.	blad	skarpt el. spidst blad på et (større)
10	koka_1_1	sb.	blad	blade fra denne busk der, ved tyg
11	pomoblاد_1	sb.	blad	blad der især indeholder pomogra
12	bog_1_1	sb.	blad	trykte el. beskrevne blade af papi
13	timian_1_1	sb.	blad	friske el. tørrede blade fra denne p
14	småblad_1	sb.	blad	lille, fint blad på en plante
15	bog_1_3	sb.	blad	indbundne el. sammenhæftede bl
16	småblad_2	sb.	blad	blad el. avis af et lille format
17	gratisblad_1	sb.	blad	blad der finansieres gennem anno

Figur 6. Ordene i venstre kolonne har alle samme genus proximum i DDO, nemlig det polyseme substantiv *blad*, og kræver derfor efterredigering

Ud over problemet med at udrede homografer og polyseme udtryk er det genus proximum der er angivet i DDO, ikke altid det systematisk set nærmeste overbegreb, idet DDO's primære formål er at formidle på den i hvert enkelt tilfælde mest hensigtsmæssige måde til ordbogsbrugeren snarere end at angive en stringent systematik inden for et semantisk felt. Fx har substantiverne *kage-spade*, *ostehøvl*, *osteskærer*, *persillehakker*, *si* og *ske* det meget generelle overbegreb *redskab* i DDO, hvorimod substantiverne *dørslag*, *hjulpisker*, *hulske*, *hvidløgspresser*, *potageske*, *rivejern* og *æggedeler* alle har det systematisk set mere korrekte overbegreb *køkkenredskab*. I sådanne tilfælde vælger vi i DanNet at se bort fra DDO's overbegreb og markerer i stedet alle ordene som underbegreber til *køkkenredskab*, jf. figur 7.

Der er dog masser af tilfælde hvor det i høj grad kan diskuteres hvilket overbegreb der er mest passende. Fx har substantiverne *isspand*, *kedel*, *litermål*, *osteklokke*, *randform* og *saltkar* alle overbegrebet *beholder* i DDO, og substantiverne *foodprocessor*, *kogeapparat*, *sandwichriste* og *sodavandsmaskine* har overbegrebet

*apparat*. Men begge disse grupper af ord kunne lige så vel grupperes som køkkenredskaber på linje med substantiverne nævnt ovenfor. Det er altså meget ofte et spørgsmål om hvordan man vælger at opbygge wordnettet, og der er i rigtig mange tilfælde mere end én måde at inddеле kategorierne af betydninger på. I de ovennævnte tilfælde har vi valgt at samle mange af betydningerne under det fælles overbegreb *køkkenredskab* og har derfor valgt at se bort fra DDO's genus proximum'er. I figur 7 ses nogle eksempler.



Figur 7. Udsnit af DanNet. Overbegrebet *køgegrej/køkkengrej/køkkenredskab/køkkentøj* med nogle af dets underbegreber efter færdigredigering og ændring af forskellige genus proximum'er fra DDO. Indramning viser hvor der under redigeringen af DanNet er ændret i genus proximum.

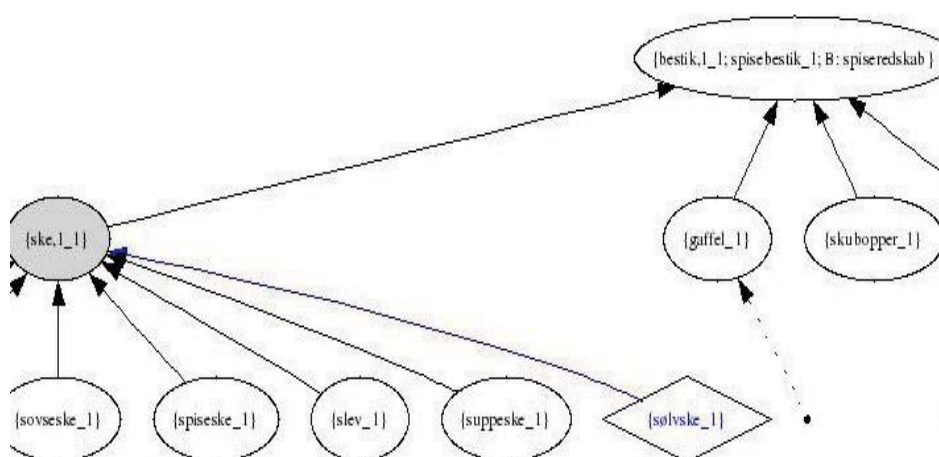
### 3.2. To typer af underbegreber

Et interessant problem som vi er stødt på i vores arbejde med at gruppere ordene i overbegreber og underbegreber, er de mange tilfælde hvor to ord ganske vist har samme overbegreb, men hvor de grupperer sig i forskellige slags underbegreber der er kompatible på tværs af grupperingerne, men udelukker hinanden inden for den enkelte gruppering (se også Pedersen & Sørensen 2006). Fx kan en stor gruppe af underbegreber have det tilfælles at de alle er et bestemt eksempel på overbegrebet med særlige fysiske kendetegn. Overbegrebet *træ* har således en masse træsorter som systematiske (taksonomiske) underbegreber, fx *bøgetræ*, *el* og *gran*. Et eksempel på et underbegreb til *træ* der går på tværs af denne store gruppe af taksonomiske underbegreber, er *vejtræ*. Et vejtræ er nemlig ikke en bestemt slags træ, men snarere et hvilket som helst træ der står langs en vej. Begrebet betegner i dette tilfælde funktionen af træet og ikke en specifik type træ

med særlige karakteristika, og et vejtræ kan sagtens også være et bøgetræ, hvori-  
 mod et grantræ ikke samtidig kan være et bøgetræ. Hvis vi betragter figur 8, ser  
 vi et andet eksempel på det samme fænomen, her ved en række underbegreber  
 til *ske*. De fleste underbegreber angiver i dette tilfælde i stedet hvilken **funktion**  
 den pågældende ske har, men en af sketyperne skiller sig ud, nemlig *sølvske*, der i  
 stedet betegner hvilket **materiale** skeen er lavet af. En sølvske kan godt også være  
 en suppeske, hvorimod fx en dessertske ikke samtidig kan være en suppeske.  
 I arbejdet med opbygningen af DanNet har vi valgt at skelne mellem den tak-  
 sonomiske eller systematiske gruppe af underbegreber til et bestemt overbegreb  
 og de underbegreber der går på tværs af den systematiske gruppe. De første er  
 umarkerede; de sidste markerer vi med et særligt træk 'orthogonal' i databasen,  
 visualiseret med en rombe i figur 9.

- {ske\_1\_1} (Instrument+Artifact+Object): redskab af metal, træ, plastic e.l. bestående af et sk
- {barneske\_1} (Instrument+Artifact+Object): lille ske til små børn
- {dessertske\_1} (Instrument+Artifact+Object): ske der er større end en teske og mindre e
- {grydeske\_1} (Instrument+Artifact+Object): stor ske af træ el. plastic der er beregnet til o
- {måleske\_1} (Instrument+Artifact+Object): ske som har et bestemt rumfang, og som brug
- {slev\_1} (Instrument+Artifact+Object): stor (træ)ske til køkkenbrug, ofte en grydeske
- {sølvske\_1} (Instrument+Artifact+Object): ske af sølv
- {sovseske\_1} (Instrument+Artifact+Object): dyb ske der bruges til at øse sovs op med
- {spiseske\_1} (Instrument+Artifact+Object): ske som man bruger til at spise fx grød el. sup
- {suppeske\_1} (Instrument+Artifact+Object): ske med et større hoved (laf) end en almind
- {teske\_1} (Instrument+Artifact+Object): lille ske som man fx bruger til at røre sukker rundt

Figur 8. Forskellige typer af underbegreber til ske. Der skelnes mellem to typer  
 overbegrebsrelationer, og *sølvske* markeres derfor efterfølgende med et særligt træk, jf.  
 figur 9.



Figur 9. Sølvske markeret som en særlig type underbegreb til *ske*

For de konkrete substantivers vedkommende er denne skelnen mellem en systematisk gruppe af undergreber og en gruppe der adskiller sig fra denne, forholdsvis uproblematisk. Det er straks mere kompliceret for verbernes og verbalsubstantivernes vedkommende, og et af de fremtidige forskningsområder i DanNet er da også at undersøge om man kan opstille lignende skel mellem disse ordgrupper typer af forskellige underbegreber.

### 3.3. Synonymi i DanNet

En sidste meget vigtig ting der skal tages stilling til i den konkrete opbygning af wordnettet ud fra DDO's betydninger, er synonymi: Hvornår skal betydningerne fra DDO slås sammen i et enkelt synset (synonym-sæt)? Vi arbejder i wordnet med et mere udvidet synonymi-begreb end man gør i DDO, og der er derfor en lang række tilfælde hvor begreber er synonyme i DanNet, men ikke i DDO. Fx ser man i DanNet helt bort fra ordenes valør og opfatter ord som *bil* og *spand* som synonyme. Derudover har vi følgende retningslinjer: Hvis ordene er synonyme i DDO (markeret med SYN i den trykte ordbog), tilhører de samme synset i DanNet. Derudover kan de, hvis de har næsten enslydende definitioner i DDO og fx er angivet som nærsynonymer – i den trykte ordbog angivet vha. JF ('jævnfør') – også komme i samme synset i DanNet. Et eksempel på dette er substantiverne *køkkengrej*, *kogegrej*, *køkkenredskab* og *køkkentøj*, se figur 7 øverst. De har nemlig i DDO næsten ens betydningsbeskrivelser og er for nogles vedkommende angivet som nærsynonymer til hinanden:

- køkkengrej:** redskaber der bruges ved madlavning JF husgeråd, køkkentøj
- kogegrej:** redskaber til at lave mad med i et køkken
- køkkenredskab:** redskab der bruges til madlavning JF køkkengrej
- køkkentøj:** redskaber der bruges til madlavning og ved spisning SYN køkkenting, JF køkkengrej.

## 4. Anvendelser i ordbogssammenhæng

Wordnets bruges primært i sprogteknologisk sammenhæng og har bl.a. fundet anvendelse i forbindelse med forbedring af emnesøgning i søgemaskiner og til udvikling af internettets næste fase, det såkaldte Semantic Web. Men også inden for ordbøger til menneskelige brugere rummer det visse interessante muligheder. Vi vil her pege på tre områder hvor vi kan se et leksikografisk perspektiv:

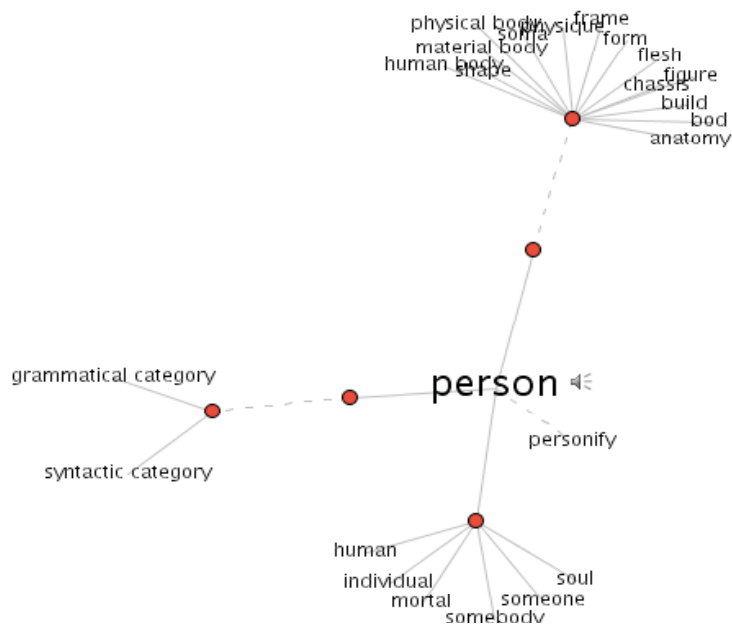
1. udvikling af ordbøgers sprogproduktive funktion
2. forbedring af learnerordbøgers pædagogiske funktioner
3. underholdningsfunktion, især krydsordshjælp

Man kunne hertil føje et fjerde område, viden om sproget, men fordi det er så generelt et formål at det næsten altid kan bringes i anvendelse, vil det ikke få selvstændig behandling.

#### 4.1. Sprogproduktion: nuancer og variation

En leksikografisk resurse hvor ordforrådet ordnes efter dets begrebsindhold (onomasiologisk) frem for efter dets benævnelser (semasiologisk), er ikke nogen ny idé. Eksemplet par excellence er P.M. Rogets *Thesaurus of English Words and Phrases* fra 1852, og siden er der udsendt mange lignende begrebsordbøger, ikke mindst inden for fagsproglige områder. DanNet har meget tilfælles med begrebsordbøgerne: den systematiske ordning af ordforrådet, den onomasiologiske tilgang og vægtningen af relationer mellem begreberne i synonyme, antonymer, meronymer, hyperonymer og hyponymer. Ganske vist er kodningen i et wordnet foretaget med henblik på sprogteknologisk brug, dvs. med computere som "målgruppe", men med en vis redaktionel bearbejdning vil DanNet også kunne finde anvendelse som en selvstændig ordbog for menneskelige brugere. På internettet kan man også se eksempler på hvordan man har tilgængeliggjort wordnets for menneskelige brugere ved at knytte en særlig brugergrænseflade til ressourcen. Der er selvfølgelig en sammenhæng mellem graden af bearbejdning til menneskelige brugere, grænsefladen og kvaliteten af det leksikografiske produkt; eksempler kan ses på <http://wordnet.princeton.edu/perl/webwn> og [www.online-thesaurus.net](http://www.online-thesaurus.net), mens en grænseflade med dansk hjælpetekst findes på [www.onlineordbog.dk/wordnet/da](http://www.onlineordbog.dk/wordnet/da). Sidstnævnte tilbyder også norsk og engelsk hjælpetekst.

En anden mulighed er at udnytte wordnettet ved at knytte det sammen med en anden ordbog. I Det Danske Sprog- og Litteraturselskabs projekt *ordnet.dk* vil DanNets oplysninger med tiden blive integreret for at kunne forbedre visse funktioner (jf. Lorentzen & Trap-Jensen 2006). Det er ikke mindst nærliggende fordi DanNet i forvejen henter sine oplysninger fra *Den Danske Ordbog*, som indgår i *ordnet.dk*. I *Den Danske Ordbog* bruges som nævnt – og bevidst – et snævert synonymibegreb, men et bredere udvalg af muligheder kan være en fordel for den der har brug for overblik over sprogets nuancer og variationsmuligheder i forbindelse med tekstproduktion (jf. Bergenholtz & Vrang 2005, Trap-Jensen 2005b). DanNets kodning af synonyme og nærsynonymer udgør netop et sådant bredere udvalg, og det kan man udnytte ved at gøre disse oplysninger tilgængelige i ordbogen, helst som en yderligere, klikbar mulighed ("se et større udvalg"), hvorved man kan bevare fordelene af både det snævre sæt synonyme (til receptions- og vidensformål, også for ikke-modersmålsbrugere) og det bredere sæt (til tekstproduktion, for brugere med modersmåls-lignende kompetence). På tilsvarende vis kan der ved en given betydning oplyses om alle øvrige relationer til betydninger der er registreret i DanNet.



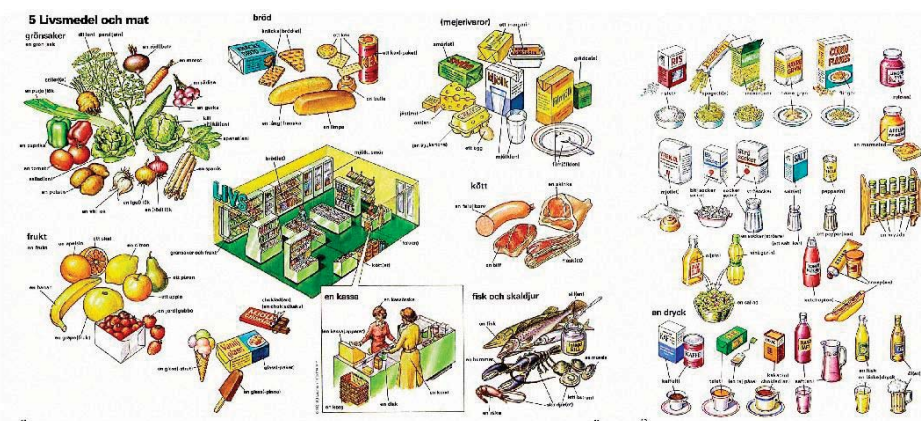
Figur 10. Visuel fremstilling af søgeordet *person* fra [www.visualthesaurus.com](http://www.visualthesaurus.com)

Brugen af DanNet som begrebsordbog kan også tilbydes som en selvstændig søgefacilitet inden for *ordnet.dk*. Muligheden for at foretage begrebsorienterede søgninger i DDO på nettet har lige fra begyndelsen været en vigtig anledning for DSL til at involvere sig i projektet (jf. Trap-Jensen 2005a). Det der kræves for at gøre det muligt, er dels at der oprettes et særligt felt til søgning på begreber, dels at der udvikles en grænseflade der kan vise resultatet på en hensigtsmæssig måde. En udfordring her er især hvordan man præsenterer meget store søgeresultater. En søgning som “vis alle ord der betegner en *person*” giver fx flere tusind resultater som det kan være svært at få overblik over. Det problem kan man forestille sig løst dels ved at folde hyponymgrupper sammen så de kun vises i deres helhed når man klikker på dem (fx ‘mand’, ‘kvinde’, ‘idrætsudøver’, ‘kunstner’), dels ved at ordne de resterende forekomster i meningsfulde grupper efter karakteristiske fællestræk. En metode til automatisk gruppering efter et sådant princip (kaldet *feature detection method*) er beskrevet i Asmussen (2004).

Endelig kan man hjælpe brugeren til at få overblik over et betydningsområde ved at give en visuel fremstilling af det. Også det kan man finde eksempler på i eksisterende netpublikationer. I figur 10 ses et eksempel fra [www.visualthesaurus.com](http://www.visualthesaurus.com).

#### 4.2. Pædagogiske anvendelser

Mange elektroniske learnerordbøger har efterhånden ud over opslagsdelen også en afdeling med forskellige ordøvelser, grammatik og systematisk glosetræning. Hvis ordbogen publiceres som cd-rom, er der desuden mulighed for at lagre oplysninger lokalt på brugerens harddisk, og man kan derfor operere med en personlig tilpasning af ordbogen så brugeren kan afstemme den efter sit eget behov og niveau, markere bestemte ord som særlig vigtige, tilføje nye artikler osv. Fordi et wordnet til en vis grad også opdeler det sproglige felt i systematiske betydningsområder, kan det være interessant som leverandør af input til den slags betydningsoversigter. Det kræver naturligvis en videre forarbejdning så stoffet svarer til målgruppen. Man kan også peppe præsentationen op med billeder, lyd eller animationer. Et eksempel med billeder er vist i figur 11.



Figur 11. Oversigt over levnedsmidler fra svensk Lexin

Tilsvarende oversigter, men i mindre format, kan man give under de enkelte betydninger. I læringsmæssig sammenhæng er det vigtigt at man ikke nøjes med at tilegne sig det enkelte ord og dets betydning i isolation, men også er opmærksom på hvordan det adskiller sig fra andre betegnelser inden for samme betydningsområde. Det kan man gøre ved at tilbyde en mulighed for at se en samlet oversigt over ko-hyponymer, altså ord eller betydninger som har samme overbegreb, fx strengeinstrumenter, køkkenredskaber, ordklasser eller sygdomme.

#### 4.3. Underholdning – krydsordshjælp

En hjemmeside der tilbyder en netordbog, er i konkurrence med alle andre hjemmesider på internettet og er derfor underlagt de samme betingelser som gælder

for nettet generelt. Hvis man vil fange og fastholde opmærksomheden hos netbrugere der blot kommer forbi siden mere eller mindre tilfældigt, må man benytte sig af de midler der egner sig til formålet. Dertil hører forskellige former for underholdning og spil. En oplagt anvendelsesmulighed er at bruge DanNet i kombination med *Den Danske Ordbog* som hjælp til krydsordsløsere. Man kan allerede finde forskellige former for krydsordshjælp på nettet, men ofte er de ikke særlig raffinerede. Som regel fungerer de ved at brugeren kan udfylde de dele af ordet han/hun kender i forvejen, og hjælpefaciliteten undersøger derefter hvilke ord i basen der matcher den indtastede streng. Søgfeltet kan være opdelt i felter som i krydsordsforlægget eller være et almindeligt søgefelt hvor forskellige jokertegn erstatter ét eller flere bogstaver.

Når denne hjælp ikke forekommer særlig avanceret, skyldes det dels at man ikke har mulighed for at angive om ordet optræder i bøjet form, dels at man ikke kan begrænse antallet af søgeresultater ved at oplyse hvilken type ledeord der er tale om. Enhver krydsordsløser ved at begge dele er helt afgørende når man løser krydsord. Det vil derfor være en stor forbedring hvis man kan supplere krydsordshjælpen med ordbogens oplysninger om ordenes bøjningsformer og wordnettets opdeling af ordforrådet i begrebsområder. Et enkelt eksempel kan illustrere hvordan.

Lad os antage at man er gået i stå i sin krydsord ved et ord på seks bogstaver. De eneste oplysninger man har, er at det slutter på *s*, og at ledeordet er "sport". Taster man de oplysninger ind i den krydsordshjælp der findes hos forskellige eksisterende udbydere, giver det fx hos Caplex 2244 resultater (se figur 12) og hos Nationalencyklopedins internettjenst 1600. Den Danske Online Ordbog oplyser ikke antallet af resultater, men viser den første side med 50 forekomster. Ved at bladre frem kan man finde ud af at der er i alt 14 sider resultater, dvs. ca. 700 forekomster. Det siger sig selv at det er alt for mange og alt for uoverskueligt et antal resultater. En søgning i DanNet viser derimod hurtigt at der kun er to sportsgrene der opfylder kriteriet: *isdans* og *tennis* (evt. *diskos* opfattet som en disciplin).

DanNet kan således bidrage væsentligt til forbedring af krydsordshjælp – hvad enten det er en integreret hjælp som i eksemplerne her eller ved udarbejdelse af et selvstændigt krydsordsleksikon, som jo traditionelt også indeholder systematisk ordnede oversigter med ordene fordelt efter antallet af bogstaver. Selvom der kan opnås en betydelig teknisk forbedring af krydsordshjælpen med resurser som DanNet og DDO, bør det dog retfærdigvis understreges at den underliggende base i andre henseender er mangelfuld. Det gælder ikke mindst med hensyn til navnestof og andet encyklopædisk materiale, som ofte udgør en stor del af krydsordenes ordforråd. Det er også grunden til at man især finder funktionen hos forlag der råder over både encyklopædi og ordbøger.





egenskaber samt korpuseksempler knyttes til de enkelte synsets.

## Litteratur

### *Ordbøger m.m.*

Caplex: [www.caplex.no](http://www.caplex.no). Oslo.

DanNet: [www.wordnet.dk](http://www.wordnet.dk). Center for Sprogteknologi og Det Danske Sprog- og Litteraturselskab 2005–2008. København.

DDO = *Den Danske Ordbog. Bind 1-6*. Udgivet af Det Danske Sprog- og Litteraturselskab. Hovedredaktører: Ebba Hjorth og Kjeld Kristensen. København: Gyldendal 2003-2005.

Den Danske Online Ordbog: [www.ddoo.dk](http://www.ddoo.dk)

Lexin: <http://lexin.nada.kth.se/lexin.html>. Stockholm.

Nationalencyklopedins internettjänst: [www.ne.se](http://www.ne.se). Göteborg.

SIMPLE: <http://cst.dk/simple>. København.

### *Anden litteratur:*

Asmussen, Jørg 2004: Feature Detection – A Tool for Unifying Dictionary Definitions. I: Williams, Geoffrey and Vessier, Sandra: *Proceedings of the 11th EURALEX International Congress*. Lorient, 63–69.

Asmussen, Jørg, Bolette S. Pedersen & Lars Trap-Jensen 2007: DanNet: From Dictionary to WordNet. I: Kunze, Claudia, Lemnitzer, Lothar & Osswald, Rainer (eds.): *GLDV-2007 Workshop on Lexical-Semantic and Ontological Resources*. Universität Tübingen, 1–11.

Bergenholtz, Henning, & Vibeke Vrang 2005: Den Danske Ordbog bind 2 (E–H) og 3 (I–L) – en ordbog for folket eller for akademikere? I: *LexicoNordica 12*, 169–187.

Fellbaum, Christiane (red.) 1998: *WordNet: An Electronic Lexical Database*. Cambridge, MA: The MIT Press.

Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas & Antonio Zampolli 2000: SIMPLE – A General Framework for the Development of Multilingual Lexicons. I: *International Journal of Lexicography 13*, 249–263.

Lorentzen, Henrik 2004: The Danish Dictionary at large: presentation, problems and perspectives. I: *Proceedings of the 11th EURALEX International Congress*. Lorient, 285–294.

Lorentzen, Henrik & Lars Trap-Jensen 2006: *ordnet.dk* – et nyt sprogligt opslagsværk på internettet. I: *Nordiske Studier i Leksikografi, NFL-skrift nr. 9*. København, 253–264.

Pedersen, Bolette S. & Jørg Asmussen 2006: DanNet – Fra ordbog til et leksikalsk-semantic WordNet for dansk. I: *LEDA-Nyt 42*, 3–12.

Pedersen, Bolette S., Sanni Nimb, Jørg Asmussen, Nicolai H. Sørensen, Lars Trap-Jensen & Henrik Lorentzen 2006: DanNet – a WordNet for Danish. I: *Proceedings of the Third International WordNet Conference*. Jeju, 329–331.

- Pedersen, Bolette S. & Patrizia Paggio 2004: The Danish SIMPLE Lexicon and its Application in Content-based Querying. I: *Nordic Journal of Linguistics* 27(1), 97–127.
- Pedersen, Bolette S. & Nicolai H. Sørensen 2006: Towards Sounder Taxonomies in Wordnets. I: Oltramari, Alessandro, Chu-Ren Huang, Alessandro Lenci, Paul Buuitleaar og Christiane Fellbaum (eds.): *Ontolex 2006* at 5th International Conference on Language Resources and Evaluation. Genova, 9–16
- Trap-Jensen, Lars 2005a: Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: *LexicoNordica* 12, 109–122.
- Trap-Jensen, Lars 2005b: Kommentar til Henning Bergenholtz og Vibeke Vrang: Den Danske Ordbog bind 2 (E-H) og 3 (I-L) – en ordbog for folket eller for akademikere? I: *LexicoNordica* 12, 189–198.
- Vossen, Piek (red.) 1999: *EuroWordNet. A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.
- Vossen, Piek (red.) 2005: *EuroWordNet General Document*. University of Amsterdam.

Sanni Nimb  
seniorredaktør, f. 1961  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
sn@dsl.dk

Bolette Sandford Pedersen  
seniorforsker, f. 1962  
Center for Sprogteknologi, Københavns Universitet  
Njalsgade 80  
DK-2300 København S  
bspedersen@hum.ku.dk

Lars Trap-Jensen  
ledende redaktør, f. 1960  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
ltj@dsl.dk



## Leksikografi i pedagogisk perspektiv

### Ein presentasjon av Norsk Ordbok for skoleelevar og studentar

Although publishing houses ought to be an important participant in the marketing of dictionaries, necessary qualifications are to be found in the editorial dictionary staff. In particular this is the case when it comes to large multi-volume national dictionaries. In spite of this, it is hard to find descriptions of dictionary presentations intended for pupils and students. An investigation of curricula in compulsory school and upper-secondary school in the Scandinavian countries reveals big differences between the countries. The Danish compulsory school curriculum shows thorough use of dictionaries, whereas the subject is not mentioned at all in the Swedish curriculum. Taking a presentation of *Norsk Ordbok* for pupils and students as a starting point, the authors show how the use of *Norsk Ordbok* can be integrated in mother tongue teaching to expand not only word knowledge, but also knowledge of language history and dialectology. By use of auditive and visual effects pupils and students can be motivated to use a complex dictionary. An evaluation of the presentation shows that pupils and students find it useful and instructive.

#### 1. Innleiing

Elevar i grunnskolen og i den vidaregåande skolen bør vere ei viktig brukargruppe for mange typar ordbøker. Men i kva grad forsøker ordboksredaksjonar eller -prosjekt å marknadsføre seg for denne brukargruppa? Prøver dei i det heile å nå aktuelle brukargrupper med produktet sitt? Kanskje stiller vi feil spørsmål. Kanskje skulle ein heller spørje om dette er ei oppgåve for dei som utviklar og skriv ordboka. Er ikkje dette ei oppgåve som utgjevaren, forlaget, skulle ha eineansvaret for?

Svaret på dette spørsmålet er ikkje avgjerande for denne framstillinga, så vi treng for så vidt ikkje ta stilling til problemet no. Det som er sikkert, er at *kompetansen* for å kunne gjere eit toverretta arbeid mot aktuelle brukargrupper, ligg i ordboksredaksjonen. Iallfall må ein slå fast at det er tilfellet for store nasjonale ordboksverk. Derfor bør ordboksredaksjonar i alle høve vere budd på å presentere ordboka eller ordboksverket for ålmenta. I tillegg er det òg eit moment at ein gjennom ein slik prosess, gjer redaksjonen meir bevisst på kven som skal lese det ein skriv. Dessutan må det vere motiverande for redaktørane at det òg blir arbeidd med å gjere ordboka kjend for omverda. Sist, men ikkje minst, er det sjølv sagt også i forlaget si interesse at ordboksproduktet blir gjort kjent for potensielle kjøparar.

## 2. Litteratur om emnet

Vi har prøvd å finne ut kva som har vore gjort på dette feltet i Norden, men har funne svært lite relevant. Det meste som har vore skrive, dreier seg om leksikografi som akademisk emne ved universitet og høgskole, og det er ikkje særleg relevant i vår samanheng. Vi har likevel merka oss to mindre arbeid av Ruth V. Fjeld (Fjeld 1997 og 1999). Perspektivet er bruken av ordboka i ein fagleg samanheng, og i tillegg er det ein del om ordbokstypologi. Utifrå det tek ho opp kva slags ordbok som høver best til det einskilde konkrete føremålet. Ho kjem i liten grad inn på leksikografi som emne, og seier ingenting om dei ressursane som ordbøkene byggjer på. I begge tilfella er det lærarane som er den direkte målgruppa, men indirekte er det elevane sin bruk av ordboka det dreier seg om.

Lærarane er vel òg den viktigaste målgruppa for Oddrun Grønvik (Grønvik 1992) i ein artikkel der ho drøftar kva skoleelevar må kunne for å ha nytte av ei definisjonsordbok. I artikkelen kjem ho inn på korleis definisjonane bør vere utforma for at det skal vere enkelt for elevane å forstå dei. Ho tilrår likevel ikkje noka bestemt ordbok til bruk i skolen, men etterlyser heller ei god barneordbok på norsk.

Eit siste døme er Karsten Schou (Schou 1999), som skriv om tospråklege ordbøker, og målgruppa er engelskstudentar ved universitetet som treng rettleiing i eigen ordboksbruk. Slike meir eller mindre private rettleiingar frå lærarar til studentar vil vi tru ikkje er heilt uvanlege, sjølv om vi ikkje har greidd å spore opp fleire.

## 3. Læreplanar

Ei undersøking av læreplanane i grunnskolen og i den vidaregåande opplæringa i dei skandinaviske landa (Danmark, Noreg og Sverige) viser store forskjellar landa imellom når det gjeld i kva grad ein legg vekt på bruk av ordbøker i undervisninga.

### 3.1 Danmark

Den gjeldande læreplanen i Danmark, Fælles Mål (Det danske Undervisningsministeriet), frå 2003, dekkjer danskfaget frå første til niande årstrinn. Dokumentet er delt i ein læreplan (som mellom anna inneheld beskriving av faget, formål, slutt mål, trinnmål, progresjon og rettleiande leseplan) og ei undervisningsrettleiing (med råd og vink til undervisninga). Bruk av ordbok er nemnd mange gonger i den danske læreplanen. I fleire delar av planen og så å seie på alle årstrinn inngår bruk av ordbok som ein del av undervisningsopplegget. Allereie i beskrivinga av undervisninga for første og andre årstrinn er det lagt vekt på at elevane skal bruke ordbøker og leksikon som er lagde til rette for alderstrinnet. I trinnmål etter fjerde årstrinn er det sagt at elevane skal kunne bruke ordbok som opplyser om bøyingsformer. Deretter er det lagt vekt på bruk av ordbok på alle

årstrinn. Det er òg lagt inn ein viss progresjon ved at ein først berre krev bruk av oppslagsdelen av ordboka (6. årstrinn), mens ein seinare ventar at elevane skal vere i stand til å bruke både oppslags- og innhaldsdelen av ordboka (9. årstrinn). Meldinga til lærarane (og dei føresette) er at det skal trenast på å bruke ordbok jamt, både i lese- og skriveprosessen. I undervisningsrettleiinga kjem vektlegg- inga av ordbøker enda klarare fram. I avsnittet om staving heiter det til dømes at

Brugen af ordbog bør have en høj placering i undervisningen i retstavning. Ordbogen må være tilpas let at finde rundt i for hver enkelt elev, sådan at de alle betragter den som et hjælpemiddel, det kan betale sig at bruge. Derfor må klassen have forskellige ordbøger til rådighed, fra den enkleste til den mest udbyggede.

Og vidare i avsnittet om språk generelt:

At der er en ordbog med ordforklaringer, bør være en selvfølge, men også ordenes udtale, bøjning og etymologi, synonymer, almindelige forkortelser, stave- og tegnsætningsregler og forklaringer på særlige ord som fremmedord eller sted- og personnavne kan der være brug for at slå op. Enkelte ordbøger samler alle disse oplysninger og kan bruges i klassens håndbibliotek, og spesielle ordbøger, opslagsværker og tidsskrifter om sprog og antologier med fx ordsprog, citater og bevingede ord bør kunne findes på skolebiblioteket. For at vænne eleverne til at bruge håndbøgerne må læreren selv hyppigt gøre brug af dem, både ved selv at slå op og især ved at opfordre eleverne til at slå op.

Når det gjeld den vidaregåande opplæringa i Danmark, er læreplandokument- makarane mykje mindre ambisiøse med omsyn til bruk av ordbok. I 2003 kom det ei reform av dei vidaregåande utdanningane ('de gymnasiale uddannelser') i Danmark. Reforma omfatta både studenteksamen, høgare handelseksamen, høgare teknisk eksamen og høgare førebuingseksamen. Det mest konkrete dømet på at bruk av ordbok er nemnd i læreplanane for gymnaset, finn ein i undervisningsrettleiinga til dansk A for elevar som skal ta studenteksamen (ny ordning). Her rår ein læraren til å gjere elevane fortrulege med bruk av hjelpemiddel for å skrive korrekt, til dømes ulike ordbøker, oppslagsverk og nettsida til Dansk Sprognævn.

### 3.2 Noreg

Den gjeldande læreplanen i Noreg, Kunnskapsløftet, avløyste Læreplanverket for den tiårige grunnskolen frå 1997 (L97) i 2006 (Det norske utdanningsdirektoratet). I morsmålsfaget i grunnskolen er det dessverre ei negativ utvikling når det er tale om ordboksbruk frå det gamle læreplandokumentet til det som gjeld i dag. Etter L97 skulle elevane arbeide med rettskriving i egne tekstar, mellom anna ved hjelp av ordbok, og elles gjere seg kjende med både ein- og tospråklege

ordbøker (femte klasse). Også i måla for tiande klasse var bruk av ordbok nemnd. I Kunnskapsløftet er dette redusert til at eit mål for opplæringa etter sjuande årstrinn er at elevane skal bruke oppslagsverk og ordbøker. I engelskfaget er det eit mål at elevane skal bruke ordbøker og andre hjelpemiddel i eiga språklæring etter fjerde årstrinn.

Også i den vidaregåande opplæringa i Noreg blei læreplanane reviderte i 2006. Kunnskapsløftet overtok då for Reform 94 (R94). Vektlegginga av ordbøker må seiast å vere nokolunde den same i desse dokumenta. I R94 heitte det i felles mål for modul 1 og 2 at elevane skulle kjenne til og kunne bruke hjelpemiddel som ordbøker, leksika og andre oppslagsverk, mens det i Kunnskapsløftet heiter at elevane etter Vg1 skal kunne bruke "bibliotekets sentrale databaser og andre faglige kilder, både tradisjonelle og elektroniske, i egne arbeider". Det skulle vel òg inkludere ordbøker.

### 3.3 Sverige

Den siste læreplanen for den niårige obligatoriske skolen i Sverige, Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet (Lpo94) (Det svenska Skolverket), tok til å gjelde i 1994. Han er felles for grunnskolen, sameskolen, den obligatoriske særskolen og spesialskolen, og blei i 1998 tilpassa til òg å gjelde førskolen og skolefritidsordninga (fritidshemmet). I tillegg til læreplanen finst det i kvart fag ein nasjonalt fastsett kursplan. Bruk av ordbok er ikkje nemnd i korkje læreplan eller kursplan for noko årstrinn. Dette gjeld både for kursa svensk, svensk som morsmål og morsmål (for elevar med andre morsmål enn svensk). Leitar ein i kursplanen for framandspråk, finn ein denne passusen under mål for engelsk etter niande skoleår: "kunna välja och använda hjälpmedel vid textläsning, skrivning och andra språkliga aktiviteter". Sjølv om det ikkje er nemnt eksplisitt, må vi tru at ordbøker kan vere eit slikt hjelpemiddel.

Den same tendensen finn vi òg i den vidaregåande opplæringa i Sverige (gymnasial utdanning). Det finst ingenting om ordboksbruk i emna svensk eller svensk som andrespråk. Ikkje eingong i kursplanen for engelsk har svenskane nedfelt bruk av ordbok som eit mål i undervisninga.

### 3.4 Læreplanar – konklusjon

Kva som er nedfelt i læreplanane, er sjølvstøtt viktig for kva som er eller kan bli praksis i skolen. Vi har ikkje funne nokon undersøkingar om kva som er praksis når det gjeld bruk av ordbøker i skolen i dag, men vårt inntrykk er at det er eit undervurdert og uutnytta hjelpemiddel i skolekvardagen, iallfall i den norske skolen. Danskane er definitivt best i klassen. Gjennom heile læreplandokumentet for morsmålsfaget i grunnopplæringa er det lagt vekt på at elevane skal bruke ordbøker aktivt. I undervisningsrettlegginga for læraren er det eksplisitt streka under at elevane skal ha tilgang til ulike typar ordbøker, og at læraren sjølv skal vere eit føredøme for elevane ved sjølv å bruke ordbok aktivt. Ordboksbruk er tona



kraftig ned i læreplanen for gymnaset, så ein får håpe at elevane tek med seg gode ordboksvanar over i den vidaregåande skolen. At bruk av ordbok i morsmålsfaget ikkje er nemnd i dei svenske læreplanane i det heile, er overraskande. Det kan sjølvstund ha noko å gjere med kor langt ned på detaljplanet læreplanen går, men det er verdt å merke seg at ordboksbruk heller ikkje er nemnd i kursplanane, som vel er meint å vere ei utdjuping av læreplanen. Den norske læreplanen for grunnskolen plasserer seg mellom den danske og den svenske når det gjeld vekt på ordboksbruk. Det oppsiktsvekkjande er at bruk av ordbok er tona ned i den nye læreplanen frå 2006.

#### 4. Korleis kan ordbøker – og ordbokskjelder – brukast i undervisninga?

##### 4.1 Ordboka – ikkje berre hjelp til å stave riktig

Det å aktivt bruke ordbok i alt tekstarbeid, også når det gjeld morsmålet, bør vere eit sjølvstund mål i undervisninga allereie tidleg i grunnskolen. Men i tillegg til å vere eit viktig hjelpemiddel i sjølv skriveprosessen, kan nasjonale ordboksverk som *Norsk Ordbok* vere ein interessant inngangsport til ulike emne som læreplanane i morsmålet seier noko om. Vi skal mot slutten kome tilbake til ein diskusjon om kva for aldersgrupper og klassetrinn dette kan vere mest aktuelt for.

Store fleirbandsverk som *Norsk Ordbok* eignar seg ikkje først og fremst i det daglege arbeidet med stavekontroll og rettskriving i skolen eller heime. Dette er heilt sjølvstund når det gjeld papirversjonen, men ei elektronisk utgåve vil vere betre eigna. Papirversjonen er stor og tung, og å slå opp ei side der vil truleg mest verke forvirrande på grunn av den fortetta mengda av forkorta informasjon. Like fullt vil dei ulike delane av artikkelen innehalde verdfull informasjon av ulikt slag, og det er derfor nyttig og viktig at elevane blir kjende med kva for informasjon som finst, og kor i artikkelen dei kan finne han. Når det gjeld den elektroniske utgåva, blir det også viktig med kunnskap om korleis ein kan gjere søk i dei einskilde delane av artikkelen. *Norsk Ordbok* er til dags dato enno ikkje gjort tilgjengeleg elektronisk i den forstand at det er mogleg å slå opp eller gjere søk i dei ferdige artiklane, men grunnlagsmaterialet for ordboka er fulldigitalisert og søkbart i alle dei ulike delane. Med rettleiing her vil det vere mogleg å kunne finne fram til om lag den same informasjonen som er å finne i den ferdige artikkelen.

I kva norskfaglege emne kan så norske skoleelevar ha nytte av å gjere søk i *Norsk Ordbok*? Kunnskap om ulike ord, tydingar og bruksområde seier seg sjølv, spesielt knytt til den eine av dei offisielle målformene, nynorsk. I tillegg meiner vi at det kan vere mykje å hente først og fremst på områda språkhistorie og dialektlære.

#### 4.1.1 Språkhistorie

*Norsk Ordbok* er ikkje noka reindyrka historisk ordbok (jf. Vikør 2006), men der vi har informasjon om det, tek vi med opphavsspråk, eventuelt også kva for språk det har kome frå til norsk. Søk i denne delen av artiklane vil derfor kunne eksemplifisere kva for språk som særleg har vore långjevarar for norsk, til dømes i form av tal som kan samanliknast. I tillegg kjem kunnskap om opphavet til einskilde lånord.

Ein svært sentral del av nyare norsk språkhistorie slik ho blir framstilt i skolen, er framveksten og utviklinga av nynorsk. Her spelar Ivar Aasen ei stor rolle, og i *Norsk Ordbok* er vi nøye med å dokumentere alt Aasen hadde med i dei to utgåvene av ordboka si. Vi tek også med informasjon om stavemåten og det grammatiske systemet han gjorde framlegg om, i den grad det skil seg frå dagens nynorsknorm. Av dette kan elevane mellom anna sjå sjølve korleis nynorsk har endra seg frå Aasens tid, og relatere det til den andre offisielle målforma, bokmål.

Det talespråklege grunnlaget nynorsk vart tufta på, er ein annan viktig del av historia til språket. Å kjenne til dette er også rekna som viktig for å oppnå ei større forståing for at vi må lære to skriftspråk i Noreg (og det er dessverre nynorsk som oftast blir skyteskiva blant norske skoleelevar i den samanhengen). Her er det mykje å lære frå *Norsk Ordbok*, ettersom ho òg er ei dialektordbok. Artiklane er bygde opp slik at alle dokumenterte dialektformer av eit ord kjem lista opp – med heimfestingar – ganske rett etter det normerte oppslagsordet. Dermed kan elevane med eigne auge sjå kva for talespråkleg breidd som ligg under den normerte skriftforma, eventuelt også kva delar av landet som har talemål som ligg tett opp til nynorsk. Her kan det kome overraskande resultat for mange, og dei vil neppe vere til ulempe for nynorskens popularitet.

Elles kan ord med fleire tydingar i mange tilfelle brukast som døme på korleis ord utviklar seg semantisk over tid, og kanskje også på korleis det har seg at same ordet kan ha ulike tydingar på ulike stader i landet. Å finne ut noko om dette kan nok krevje meir styring frå læraren si side, men kan i neste omgang kome elevane til gode i form av dømestoff på språkutvikling.

#### 4.1.2 Dialektlære

Vi har allereie nemnt at *Norsk Ordbok* i tillegg til å vere ei litterær ordbok som dekkjer det nynorske skriftspråket, også er ei dialektordbok. Vi viser den dialektale variasjonsbreidda som finst av ord både i uttale, formverk, tyding og bruk, i den grad vi kan dokumentere dette i kjeldene våre. Her er det mykje å hente for skoleelevar i eit land som Noreg, der interessa for talespråk og dialektar er stor, og der dialektbruk framleis er relativt vanleg og akseptert. Informasjon frå målføredelen av artiklane kan i tillegg til kunnskapsverdien også fungere som motivasjonsfaktor i dialektlære, til dømes i form av leik- og konkurranseprega oppgåver om 'kva, korleis og kvar' når det gjeld einskildord.

Berre på eitt felt vil *Norsk Ordbok* i denne samanhengen kunne kome til kort. Ungdom er ofte svært interessert i nyutvikling innafor slangprega ungdomsspråk i norsk. Her har *Norsk Ordbok* den lina at vi ikkje tek med nye ord som ikkje er godt dokumenterte over ei viss tid. Vi vil helst ikkje ha med ord som ikkje lenger er i bruk eller er ukjende om nokre år. I presentasjonen vår, som vi snart skal seie meir om, prøver vi å grunngje kvifor det er slik, men vi tek òg med døme på at vi faktisk har fanga opp nyutviklingar som ungdom er opptekne av, og aktivt bruker.

#### 4.2 Døme frå presentasjonen av *Norsk Ordbok*

Mange vil hevde at det å skulle presentere eit leksikografisk emne for elevar eller studentar er ei pedagogisk utfordring i seg sjølv. Likevel meiner vi at vi i eit ordboksverk som *Norsk Ordbok* arbeider med emne og rår over eit materiale som absolutt kan eigne seg for elevar i vidaregåande opplæring og for studentar. Vi vedgår at det er ei pedagogisk utfordring, men ved å lage ei innpakning som appellerer til dei unge, kan ein likevel nå fram med budskapet.

For vår del vil det seie at vi har laga ein PowerPoint-presentasjon som i utgangspunktet skal vare 45 minutt, altså ein vanleg skoletime. Presentasjonen inneheld rundt 25 lysark og er delt i tre hovuddelar: ein historisk del, ein del der vi gjev døme på innhaldet i ordboka, og ein del der vi gjer greie for dei elektroniske ressursane vi rår over med døme på korleis dei kan nyttast i ein skolesituasjon. Ein profesjonell teiknar har illustrert nokre av artikkeldøma, og vi har aktivt brukt ein del audiovisuelle innslag for å bryte opp framstillinga, men samtidig slik at innslaga er knytte til det faglege innhaldet. Til dømes opnar presentasjonen med eit videoklipp frå den britiske serien *Black Adder* (med Rowan Atkinson i rolla som *Black Adder*) der prins George og *Black Adder* møter Dr. Samuel Johnson, redaktøren av den første store engelske ordboka, *A Dictionary of the English Language* (1755). (Dette møtet endar med at alle trur det einaste eksemplaret av ordboka, som det har teke Johnson 18 timar kvar dag i 10 år å få ferdig, brenn opp i peisen.)

Vidare er *Norsk Ordbok* sett inn i ein historisk samanheng og samanlikna med andre nasjonale europeiske ordboksverk. Så blir perspektivet snevra inn til den norske språksituasjonen midt på 1800-talet og forløparane til *Norsk Ordbok*, Ivar Aasen og Hans Ross. Det særleine med den norske språksituasjonen og statusen til talemåla står sentralt, og ikkje minst det faktum at *Norsk Ordbok* ved sida av det nynorske skriftspråket òg beskriv talespråket, dialektane.

I andre delen viser vi først med nokre døme korleis artiklar i *Norsk Ordbok* kan sjå ut. Det siste dømet, den lange artikkelen om verbet *gå*, er illustrert med eit videoklipp frå Monty Python, *The Ministry of Silly Walks*, med John Cleese i hovudrolla. Ei anna tilnærming til ordartiklar, med større vekt på deltaking frå salen, har vi i form av ein gjetteleik der vi berre gjev opp definisjonane og bruksdøma til eit ord, slik at elevane/studentane skal finne oppslagsordet. Eit auditivt

verkemiddel vi har brukt, er ein låt av ein norsk rappar (Ravi: *Dødsstøt*) som syng på dialekt, og der tittelen på låta illustrerer ei nyutvikling i språket med omsyn til samansetjingsstrategi.

I den tredje og siste delen viser vi korleis dei elektroniske søkjeværktøya til *Norsk Ordbok* også kan brukast av elevar, studentar og lærarar i undervisninga. Sjølv om det ikkje er mogleg å søkje i sjølve ordboka i dag, er det fullt mogleg å søkje i dei fleste kjeldene ordartiklane byggjer på. På nettstaden til Norsk Ordbok 2014 finn ein dei søkjeværktøya prosjektet har utvikla i samarbeid med Eining for digital dokumentasjon (EDD). Her skil vi mellom søk i *Metaordboka*, som er ei komplett liste over dei normerte oppslagsorda som finst i databasen (om lag 600 000 oppslagsord), og søk i setelarkivet, som er ei digitalisert samling av dei rundt 3,2 millionar (unormerte) papirsetlane (og ein del elektroniske setlar) med ordininformasjon, døme og uttaleopplysningar. I tillegg demonstrerer vi søk i tekstkorpuset til Norsk Ordbok 2014, som no er på meir enn 30 millionar ord. Gjennom følgjande illustrerande døme prøver vi å få elevane/studentane til å tenkje over kva skriftspråksnormering er, og forholdet mellom skrift- og talespråk:

I normeringa av nynorsk har ein tradisjonelt vore skeptisk til (somme) lågtyske prefiks og avleiingar (*an-*, *be-*, *-heit*, *-else*). Desse er likevel utbreidde i dialektane, og dukkar såleis ikkje sjeldan opp i elev- og studenttekstar. Mange lærarar rettar difor ofte ei form som *friheit* til *fridom* i tråd med gjeldande rettskriving. I eit deskriptivt ordboksverk som *Norsk Ordbok* syner realitetane med omsyn til bruk seg. Søk i setelmaterialet vårt viser at *friheit* er godt belagt, og at det særleg er utbreidd i talespråket. Men også søk i tekstkorpuset gjev mange tilslag (25). Kjende nynorskforfattarar som Arne Garborg og Aasmund Olavsson Vinje har brukt ordet i skrift. Vi nemner òg at Ivar Aasen sjølv hadde *friheit* som oppslagsord i 1850-utgåva av ordboka si. I 1873-utgåva var *friheit* ute som oppslagsord. Då hadde han teke inn *fridom*, men berre som tilvising til oppslaget *friskap* der han realdefinerer ordet. Der skriv han at forma *fridom* er sjeldan og at den vanlegaste forma er *friheit*.

Gjennom dette dømet får elevane demonstrert kva som kan syne seg som språklege realitetar bak ei rettskrivingsnorm. Med sakleg rettleiing frå læraren kan dømet vere grunnlag for ein diskusjon om kva som skal vere lov eller ikkje lov å skrive. Læraren kan på si side kome til å reflektere litt over sin eigen rette- og rettleiingspraksis i skriftleg norsk. Eit ordboksmateriale kan altså vere utgangspunkt for å handsame norskfagleg relevante tema, og det er ikkje vanskeleg å finne andre døme som på same måten kan kaste lys over sentrale emne i norskfaget og invitere til refleksjon og diskusjon.

## 5. Erfaringar og konklusjon

Utgangspunktet for at vi laga presentasjonen, var at vi skulle ha noko å by på dersom elevgrupper eller skoleklassar meir eller mindre tilfeldig skulle dukke opp

i redaksjonen i følge med ein spesielt interessert og/eller (over)ambisiøs lærar. Etter kvart har vi sett betre kva for moglegheiter som ligg i ei meir planmessig satsing mot denne målgruppa. Ikkje berre kan vi by på ressursar som elevar og lærarar kan nytte seg av i arbeidet med morsmålet, men vi kan også få den effekten at *Norsk Ordbok* blir gjort meir kjent for folk, og at nynorsk dermed blir meir synleg. I tillegg får vi spreidd kunnskap om kva leksikografi er, og kva som ligg bak "fasaden" til ei ordbok.

### 5.1 Erfaringar

Fram til no har vi testa presentasjonen tre gonger. I tillegg til ei jomfrufamsyning for største delen av NO-redaksjonen, har vi lagt han fram for ei forsamling på 150–200 lærarstudentar ved Høgskolen i Oslo. Vi har òg fått høve til å prøve oss for ein grunnkursklasse ved den vidaregåande skolen i Oslo.

Ved dei siste to tilfella delte vi ut eit evalueringsskjema og bad om respons frå tilhøyrarane. Svara på to av spørsmåla er særleg interessante i denne samanhengen. 70 av lærarstudentane fylte ut skjemaet. På spørsmål om kva for delar av norskfaget dei meinte presentasjonen var mest relevant for, merkte 48 av for nynorsk, 49 for dialektar og 39 for språkhistorie. 10 av dei kryssa av for "anna", og av desse presiserte 4 at det var mest relevant å lære å slå opp i ei ordbok. Av 27 elevsvar på tilsvarende spørsmål hadde 21 merkt av for nynorsk, 12 for dialektar og 13 for språkhistorie. Dette stadfester langt på veg den oppfatninga vi på førehand hadde om relevansen for språkhistorie og dialektlære, medan vi vart positivt overraska over at så mange understreka kor relevant dette er for nynorskdelen av norskfaget.

Lærarstudentane vart også bedne om å gje uttrykk for kva målgruppe dei meinte ville ha nytte av presentasjonen. Her kryssa 32 av for ungdomsskoleelevar, 33 for elevar i vidaregåande skole og 42 for lærarar. Berre tre meinte at det kunne vere nyttig for mellomtrinnet i grunnskolen, medan éin meinte det kunne passe for barnetrinnet. Av dei mange som hadde kryssa både for ungdomsskole- og vidaregåande elevar, var det ganske mange som kommenterte at det nok var best for dei i den vidaregåande skolen. Opphavleg var det også den gruppa vi hadde i tankane. At såpass mange meinte dette var nyttig for lærarane, var vi kanskje ikkje så budde på.

Utifrå desse evalueringresultata, og dei inntrykka vi sjølve sit att med etterframsyningane, trur vi helst at dette stoffet er best eigna for elevar i den vidaregåande skolen, og dessutan for lærarar og lærarstudentar. Av dei vidaregåande elevane er det truleg mest å hente for det øvste klassetrinnet, altså dei eldste elevane. Det trengst ein god del bakgrunnskunnskap for å forstå grunnleggjande ting som det elles vil ta for lang tid å forklare, og det er det nok berre dei eldste elevane som har. Om den faglege forenklinga og dei "gøyale" pedagogiske innslaga blir for omfattande, kan mykje sentralt og viktig stoff bli borte på vegen. Slik sett verkar det som vi råka godt då vi hadde elevar i vidaregåande skole i tankane ved den første utforminga av presentasjonen.

Den store interessa og gode responsen vi fekk frå lærarstudentane på Høgskolen i Oslo, gjer det også interessant å i større grad tenkje på komande lærarar som ei målgruppe. Dei bør ha dei beste faglege føresetnadene for å sjå kva moglegheiter som ligg i å nytte ressursane til *Norsk Ordbok* i norskundervisninga, og dei er, ved å vere pedagogar, godt skikka til å formidle denne innsikta vidare til elevane sine, gjerne i planmessig utforma pedagogiske opplegg. Det kan derfor godt hende at ei satsing på to frontar kan gje dei beste resultatata med tanke på kva vi ønskjer å oppnå med presentasjonen av *Norsk Ordbok*.

### 5.2 Ressursar

Det høyrer normalt ikkje med til kjerneoppgåvene for ein ordboksredaksjon å drive med formidlingsverksemd til ålmenta. Framdriftskrav og snaue tidsfristar gjev lite rom for slikt. Dette er også situasjonen i *Norsk Ordbok*, men vi trur det er fornuftig bruk av ressursar å utarbeide eit opplegg som vi meiner kan ha eit stort potensial i seg på område som vi no har presentert. Og kanskje er det ikkje så dumt å tenkje nytt på dette feltet. Store nasjonale ordboksverk av typen *Norsk Ordbok* vil aldri bli kioskveltarar, men kan vi gjere meir for å utvide marknaden, utan at det skjer på kostnad av vitskapleg standard og kvalitet, bør vi kunne vurdere det. For *Norsk Ordbok* sin del kan resultatet også bli større aksept både for sjølv prosjektet og for nynorsken. Det meiner vi må vere ein stor bonus. Og som vi sa i innleiinga: Den naudsynte kompetansen for å drive utoverretta verksemd ligg i redaksjonane, og denne kompetansen bør utnyttast.

### Kjeldeliste

- Fjeld, Ruth V. 1997: Bruk av tospråklige ordbøker. I: *Lingua. Tysk. Lærereens hefte*. Oslo: Universitetsforlaget.
- Fjeld, Ruth V. 1999: Hvilken ordbok bør elevene velge? Upublisert førelesing ved Fagleg-pedagogisk dag, Universitetet i Oslo, 5.1.1999.
- Grønvik, Oddrun 1992: Kva må ein kunna for å ha glede av ei ordbok? I: *Norsklæreren* nr. 4, 5–10.
- Norsk Ordbok 2014: <http://no2014.uio.no/>
- Schou, Karsten 1999: Ordbogsintroduktion for engelskstuderende. I: *Sprogforum* nr. 13, 53–61.
- Vikør, Lars S. 2006: Dei diakrone aspekta ved Norsk Riksmålsordbok og Norsk Ordbok. I: *LexicoNordica* 13, 7–23.

### Undervisningsplanar

Det danske Undervisningsministeriet (henta mai 2007)  
<http://www.faellesmaal.uvm.dk> (Fælles mål)  
<http://us.uvm.dk/gymnasie/vejl/index.htm?menuid=1540> (Vejledninger og læreplaner til fagene i de gymnasiale uddannelser)

Det norske utdanningsdirektoratet (henta mai 2007)

<http://www.utdanningsdirektoratet.no> (Kunnskapsløftet 2006, Læreplanverket for den tiårige grunnskolen, Reform 94)

Det svenska Skolverket (henta mai 2007)

<http://www.skolverket.se/sb/d/468> (Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet)

<http://www3.skolverket.se/ki03/front.aspx?sprak=SV&ar=0607&infotyp=15&skolform=11&id=2087&extraId=> (Grundskola, kursplan)

<http://www3.skolverket.se/ki03/front.aspx?sprak=SV&ar=0607&infotyp=7&skolform=21&id=S&extraId=> (Gymnasial utbildning)

Dagfinn Rødningen

forskar, redaktør, f. 1959

Norsk Ordbok 2014, Universitetet i Oslo

Postboks 1021 Blindern

NO-0315 Oslo

[dagfinn.rodningen@iln.uio.no](mailto:dagfinn.rodningen@iln.uio.no)

Knut E. Karlsen

forskar, redaktør, f. 1969

Norsk Ordbok 2014, Universitetet i Oslo

Postboks 1021 Blindern

NO-0315 Oslo

[k.e.karlsen@iln.uio.no](mailto:k.e.karlsen@iln.uio.no)





## Virksomhedsleksikografien viser tænder

### Leksikografiske løsninger i København Zoo og Fagerberg A/S

This article discusses the challenge of satisfying the lexicographic and text genre-related demands of employees in two Danish companies. These employees not only have cognitive and communicative lexicographic demands but also a demand for text genre-related guidelines to be able to communicate effectively. This article addresses these challenges and proposes two theoretical models developed during two R&D projects in Copenhagen Zoo and Fagerberg A/S. The first model is the lexicographic convergence model, which facilitates convergence of the lexicographic dictionary article and integrated searches in eight external function-supporting sources. The second model is the lexicographic document template model, which facilitates the convergence of the lexicographic article and text genre data. The article finally demonstrates how the models were implemented in practice in the corporate dictionaries ZooLeks and FagerbergLex.

#### 1. Indledning og problem

København Zoo er en selvejende kulturinstitution og er Danmarks tredjestørste turistattraktion. København Zoo er involveret i flere større internationale avls- og naturbevarelsesprojekter, og i 2003 fik København Zoo blandt andet til opgave at oprette en europæisk afdeling af CBSG, som er en verdensomspændende naturbevarelsesorganisation. Se også [www.zoo.dk](http://www.zoo.dk) for yderligere information.

Fagerberg A/S er et handels- og ingeniørfirma, der blev grundlagt i Sverige i 1897. Fagerberg startede i Danmark i 1970, tilbyder i dag kunderne et bredt program af ventiler, filtre, sikkerhedsudstyr, instrumenter og analyseudstyr samt service fra deres serviceafdeling. Fagerberg indgår i Indutrade AB, der er en svensk koncern med ca. 60 handelsselskaber inden for tekniske komponenter og serviceydelser. Se også [www.fagerberg.dk](http://www.fagerberg.dk) for yderligere information.

En analyse af de to virksomheder afslørede, at de potentielle brugergrupper var særdeles heterogene, og at de deraf følgende behov var lige så heterogene. Dette stillede krav om nytænkning, og følgende leksikografiske overvejelser blev udviklet.

Den første model, som er en videreudvikling af overvejelser fra et tidligere forsknings- og udviklingsprojekt i TDC A/S, jf. (Simonsen 2002), er den leksikografiske konvergensmodel. Denne model og de bagvedliggende overvejelser sikrer den potentielle bruger adgang til en lang række funktionsunderstøttende omtekster og konkordanslinjer inde fra den leksikografiske artikel.

Den anden model, som er baseret på overvejelser om genanalyse kombineret med virksomhedsleksikografi, er den leksikografiske dokumentskabelonmodel. Denne model var ligeledes motiveret af de heterogene brugergrupper samt disses meget forskellige behov, ligesom overvejelserne om de leksikografiske dokumentskabeloner var motiveret af ønsket om at kunne hjælpe brugerne til at kunne producere tekster, der i højere grad overholder de mest basale tekstkonventioner. Den leksikografiske dokumentskabelonmodel giver således den potentielle bruger adgang til at søge på en kombination af lemma og tekstgenre og således inde fra den leksikografiske artikel få adgang til en oversigt over den pågældende tekstgenres kommunikative formål, tekstkonventioner samt typisk anvendte retoriske strategier.

Med udgangspunkt i disse indledende overvejelser er artiklens forskningsspørgsmål:

1. Hvordan understøttes en sprogmedarbejders leksikografiske behov i forbindelse med f.eks. oversættelse af en obduktionsrapport til engelsk i København Zoo?
2. Hvordan understøttes en ingeniørs leksikografiske behov i forbindelse med f.eks. produktion af en manual på engelsk i Fagerberg A/S?
3. Hvordan understøttes begge type medarbejders tekstgenreorienterede behov i forbindelse med både oversættelse og tekstproduktion?

## 2. Empiri

Med udgangspunkt i forskningsspørgsmålene, er det nu tid til kort at diskutere det empiriske fundament for denne artikel. Det empiriske basis for de to forsknings- og udviklingsprojekter omfatter data indhentet fra hhv. virksomhedsanalyser, kravsaftækningsmøder, brugerprofiler, forskningsinterviews, korpusanalyser og omfattende dokumentanalyser. De empiriske data er primært kvalitative, og der er anvendt en hermeneutisk tilgang, hvor data fra de to virksomheder er blevet fortolket og forstået.

Virksomhedsleksikografi er bl.a. kendetegnet af at foregå inden for rammerne af en virksomhed, og det var nødvendigt at gennemføre en virksomhedsanalyse, hvilket i øvrigt også anbefales af (Kjøller 1997:101) som argumenterer “når du indkaldes som rådgiver i kommunikation og sprog, så kommer du absolut ingen steder uden et godt kendskab til kulturen. Kommunikation og sprogbrugen i virksomheden kan slet ikke forstås uden at man samtidig forstår kulturen omkring det sagte og skrevne...” (min understregning).

Derudover gennemførte jeg en række kravsaftækningsmøder, hvor krav, forventninger og ønsker blev diskuteret med en projektgruppe, og jeg fik dermed et godt indblik i netop den organisationskultur, som de to virksomhedsordbøger ZooLeks og FagerbergLex skulle være en del af. Endelig har jeg for København Zoos vedkommende gennemført en række interviews samt haft adgang til knap

12000 dokumenter, og jeg har på basis af disse kunnet foretage omfattende korpusanalyser. Disse dokumenter er endelig blevet analyseret med henblik på at skabe et overblik over de forskellige tekstgenrer og ikke mindst disse tekstgenrers karakteristika. Et samlet overblik over de to virksomheders tekstgenrer er forsøgt givet i den tekstgenretaksonomi, som er implementeret i de to søgeskærmbilleder.

### 3. Teori

Det teoretiske basis er dikteret af og tilpasset til den virkelighed, hvori virksomhedsordbøger indgår, og jeg har inddraget de teoretiske overvejelser, der har været nødvendige for at udvikle de to modeller. Virksomhedsordbøger er baseret på en teoretisk tilgang, der er så tilpas eklektisk, at løsningerne er værdiskabende for virksomheden som en helhed og ikke kun til gavn for en lille håndfuld sprogeksperter. Endelig tilbyder virksomhedsordbøger den potentielle bruger omfattende hjælp på såvel det leksikalske, syntaktiske og ikke mindst tekstlige plan – noget som konventionelle ordbøger ikke gør i tilstrækkelig grad.

Virksomhedsleksikografi defineres her som “den type funktionsbaserede leksikografiske overvejelser og løsninger som anvendes i en specifik virksomhed eller organisation i overensstemmelse med dennes virksomhedsorganisation og -kultur”, hvilket er en videreudvikling i forhold til tidligere definitioner i f.eks. (Simonsen 2002).

Virksomhedsleksikografi inddrager og triangulerer som udgangspunkt en række teoretiske overvejelser, herunder naturligvis primært teoretiske overvejelser om leksikografi og fagleksikografi, jf. f.eks. (Tarp 2006), andre typer af teoretiske overvejelser om unificerende teksteksempler jf. (Leroyer 2002), overvejelser om genreanalyse jf. f.eks. (Swales 1990) samt sekundært teoretiske overvejelser om f.eks. virksomhedens virksomhedskultur, jf. f.eks. (Schein 1994).

Tarp (2006: 100) diskuterer især ordbøgers funktioner og definerer en leksikografisk funktion således: “En leksikografisk funktion er tilfredsstillelse af de specifikke typer leksikografisk relevante behov, som kan opstå hos en specifik type potentielle brugere i en specifik type ekstra-leksikografisk situation” (min understregning).

Tarp diskuterer desuden “kommunikative leksikografiske funktioner” og “kognitive leksikografiske funktioner”, dvs. en potentiel bruger kan ifølge Tarp (2006: 104) i princippet kun enten have kommunikationsorienterede behov for data, der kan hjælpe ham med at kommunikere effektivt eller have kognitionsorienterede behov for at vide mere/lære mere om et specifikt emne. Som det vil fremgå af den efterfølgende diskussion, er denne stringente opdeling mellem kognitive og kommunikative funktioner ikke let at adskille i praksis, idet de to funktioner i mange tilfælde overlapper hinanden (man tilegner sig ny viden OG får kommunikative data til at kunne kommunikere bedre og omvendt).

Også overvejelser om “unificerende teksteksempler”, jf. (Leroyer 2002) har været relevante for udviklingen af de to modeller. Unificerende teksteksempler defineres af (Leroyer 2002) som “Cette démarche est essentiellement caractérisée par la sélection et par la structuration lexicographique. Des énoncés des textes techniques sous la forme d'exemples textuels comme catégorie sui generis, et comme support cohérent de données lexicographiques privilégiées”, hvilket peger i retning af mere hjælp på teksthiveau til de potentielle brugere.

Endelig var hele diskussionen om genre og genreanalyse særdeles relevant for udviklingen af de to modeller. De gamle græske filosoffer og talere var for tusinder af år siden klar over, at både indhold, formål, struktur og form var afgørende for, hvordan et andet menneske forstår og fortolker et budskab, og i den senere tid har især Swales (1990) sat dagsordenen i forbindelse med diskussion af genreanalyse. Især var genremodellen, som foreslået af Swales relevant for udviklingen af modellen om de leksikografiske dokumentkabeloner, og Swales diskuterer især tre vigtige elementer i forbindelse med analyse og produktion af genreadækvate tekster, nemlig kommunikativt formål, trækstruktur og retoriske strategier.

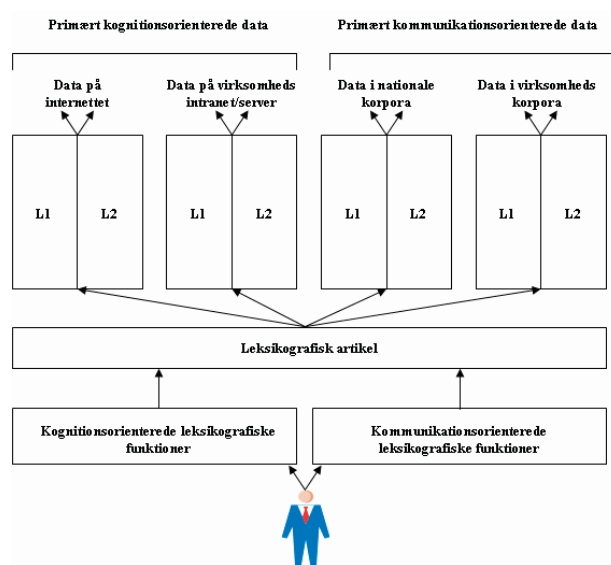
Endelig har teoretiske overvejelser om for det første virksomheders kultur og organisation spillet en rolle for udviklingen af de to modeller. En virksomheds kultur samt organisering har naturligvis stor indvirkning på en virksomhedsordbogs konception, indhold og brug, idet f.eks. en åben kultur kombineret med en flad struktur peger i en retning, hvorimod f.eks. en lukket kultur og meget hierarkisk struktur klart peger i en anden retning.

Med udgangspunkt i ovenstående teoretiske overvejelser blev først den leksikografiske konvergensmodel udviklet og derefter den leksikografiske dokumentkabelonmodel. Den leksikografiske konvergensmodel blev først omtalt hos Simonsen (2002), men er siden blevet væsentligt forbedret og videreudviklet.

Den leksikografiske konvergensmodel viser, hvorledes den potentielle bruger i en given brugersituation (hvor behovet er enten kognitionsorienterede data eller kommunikationsorienterede data) efter indtastning af et lemma i søgefeltet (eventuelt kombineret med en given dyreklasse (i ZooLeks) og et forretningsområde (i FagerbergLex) får adgang til en leksikografisk artikel. Inde fra denne artikel har brugeren direkte adgang til funktionsunderstøttende data i form af data på både  $L_1$  og  $L_2$  fra internettet, virksomhedens egne server/intranet, data i nationale korpora samt data fra virksomhedens egne korpora, se også (Simonsen 2007).

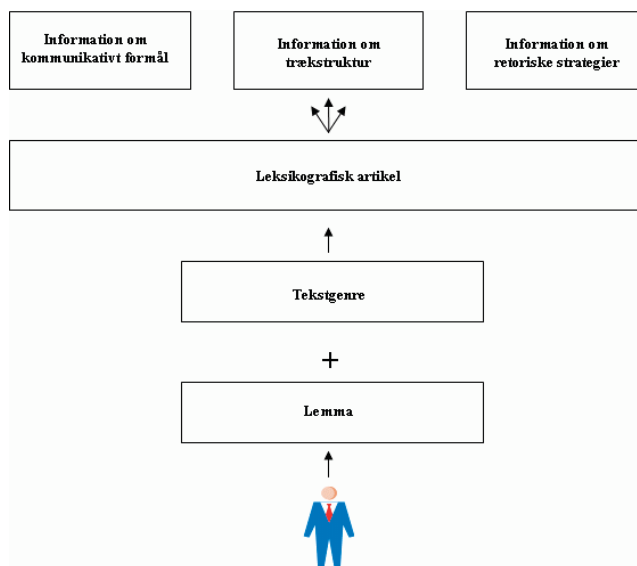
Hermed sikres optimal adgang til funktionsunderstøttende data om både lemma og ækvivalent, og både de kognitionsorienterede og kommunikationsorienterede behov tilfredsstilles. Modellen kan med fordel anvendes i monolingvale, bilingvale og multilingvale ordbogsprojekter, kan bruges på alle typer sprog og er så tilpas generaliserbar, at den sagtens kan anvendes i konventionelle leksikografiske projekter. Endvidere kan modellen videreudvikles, således

andre relevante funktionsunderstøttende omtækster kan konvergeres med den leksikografiske artikel.



Figur 1. Leksikografisk konvergensmodel

Den leksikografiske dokumentskabelonmodel blev udviklet med henblik på at understøtte den potentielle brugers behov for viden og data om den pågældende tekstgenre, noget som andre ordbøger ikke i tilstrækkeligt omfang tager højde for. Modellen om de leksikografiske dokumentskabeloner ser således ud:



Figur 2. Den leksikografiske dokumentskabelonmodel

Den leksikografiske tekstskabelonmodel viser, hvorledes den potentielle bruger i en given brugersituation efter indtastning af et lemma i søgefeltet kombineret med angivelse af den pågældende tekstgenre får adgang til en leksikografisk artikel. Inde fra denne artikel har brugeren direkte adgang til værdifuld genererelevant information om den pågældende tekstgenre på begge sprog. Brugeren får adgang til et prototypisk eksempel på den pågældende tekstgenre og nederst i dette eksempel vises information om det kommunikative formål med den pågældende tekstgenre, information om tekstgenreens trækstruktur og tekstkonventioner og endelig information om de retoriske træk, som tekstgenren typisk gør brug af. Endelig er der direkte forbindelse mellem det prototypiske tekstgenreeksempel og så selve virksomhedsordbogen, således brugeren ved at klikke på et lemma i tekstgenreeksemplet kan slå det op i ordbogsdelen og omvendt. Dermed er de leksikografiske tekstskabeloner integreret med ordbogsdelen og udgør et særdeles funktionsunderstøttende aktiv.

Som det fremgår, foreslås genremodellen operationaliseret i en leksikografisk sammenhæng vha. modellen om leksikografiske tekstskabeloner, og det argumenteres, at netop den lette adgang til tekstgenreorienterede data tilfredsstillende et leksikografisk relevant behov og dermed afsløres konturerne af en ny leksikografisk funktion. Og netop behovet for at vide mere om en given tekstgenre opstår i forbindelse med enten  $L_1$ -produktion eller  $L_2$ -produktion, hvorfor det er naturligt at tilbyde data om den pågældende tekstgenre i virksomhedsordbogen.

#### 4. Diskussion af de to modeller

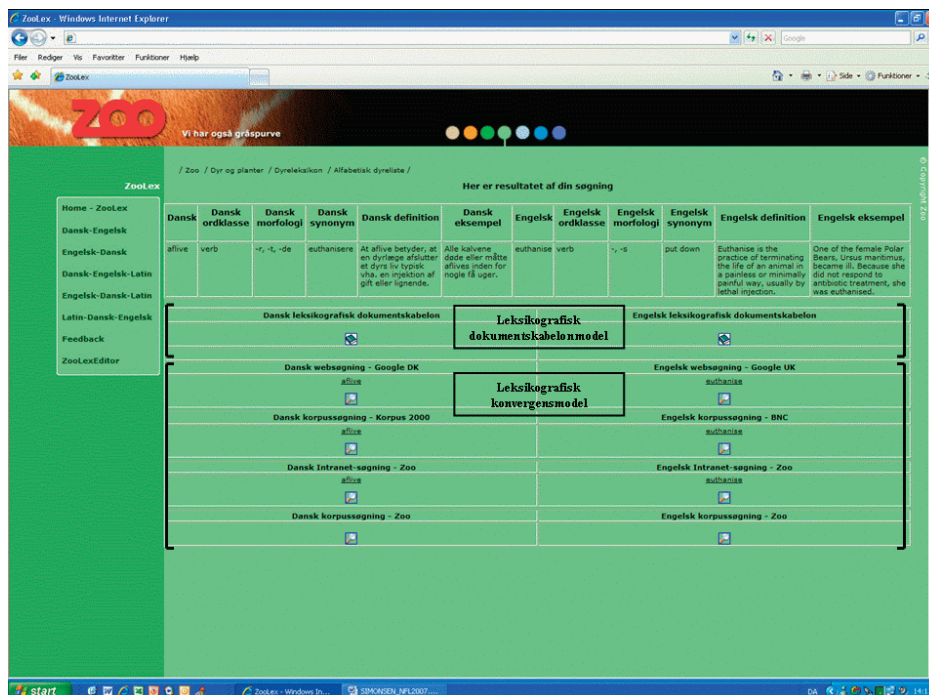
Lad os nu se på, hvorledes de to modeller er blevet implementeret i praksis og på den hjælp, som to typer potentielle brugere får i to brugssituationer. Den første brugssituation i København Zoo opsummeres i figur 3.

Tekstgenre	Obduktionsrapport
Opgave	Oversætte højt specialiseret fagsproglig tekst fra dansk til engelsk
Brugerkategori	Sprogmedarbejder
Funktioner	<p>Kognitiv leksikografisk funktion (primær) <i>Har et stort behov for at tilegne sig faglig viden for at kunne oversætte</i></p> <p>Kommunikativ leksikografisk funktion (sekundær) <i>Har et lille behov for at få kommunikative data for at kunne oversætte</i></p> <p>Tekstgenreorienterede funktion (primær) <i>Har et stort behov for at få tekstgenredata og -viden for at kunne oversætte</i></p>

Figur 3. Sprogmedarbejderens situation og behov i forbindelse med L1-L2 oversættelse

Sprogmedarbejderen i København Zoo konsulterer ZooLeks med henblik på at slå lemmaet *aflive* op, idet det synes at være et centralt ord i teksten. Han indtaster

lemmaet ”*aflive*” i et søgefelt, angiver tekstgenren ”obduktionsrapport” i en drop down-boks og får nu vist skærbilledet i figur 4.

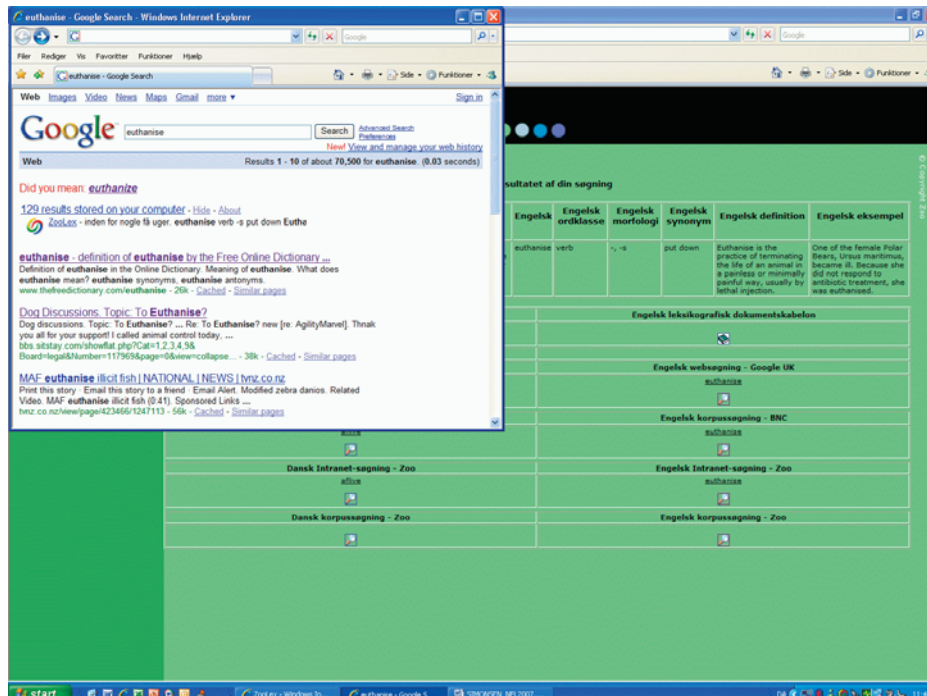


Figur 4. Søgeresultatskærbillede i ZooLeks

Øverst i figur 4 vises den konventionelle leksikografiske artikel i horisontal visning. Brugeren kan i øvrigt dobbeltklikke på et hvilket som helst ord overalt i ZooLeks og foretage en søgning efter den pågældende søgestreng på Google. I figur 4 vises ligeledes med klammer, hvor de to modeller har virkning i ZooLeks.

Brugerens kognitive leksikografiske behov dækkes for det første ved at læse de danske og engelske definitioner og eksempler i den leksikografiske artikel. Af den danske definition fremgår, at ”en dyrlæge afslutter et dyrs liv typisk vha. en injektion af gift eller lignende”.

Men brugeren ønsker yderligere kognitive leksikografiske data om lemmaet og klikker derfor på forstørrelsesglasikonet under ”Engelsk websøgning – Google UK” og foretager en søgning efter funktionsunderstøttende omtekster på internettet. Han får nu vist følgende skærbillede, som indeholder en række værdifulde tekster om lemmaet *aflive*, hvorved hans kognitive leksikografiske behov dækkes til fulde, idet han i det andet dokument i listen kan lære mere om *aflive*.

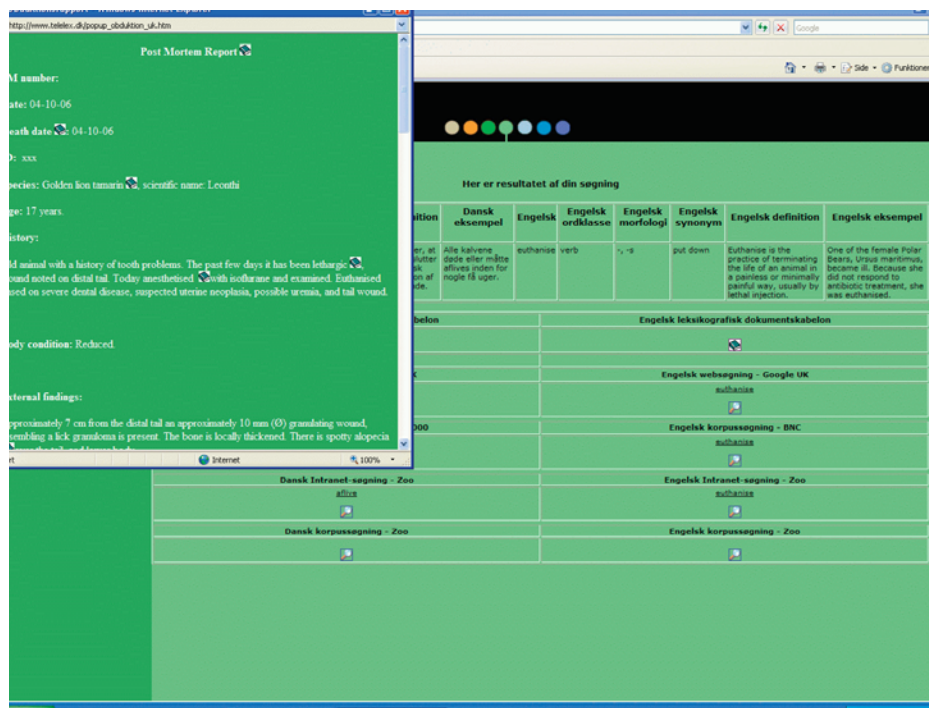


Figur 5. Leksikografisk konvergensmodel i ZooLeks

Brugerens behov for kommunikative leksikografiske data er mindre, idet han er sprogmedarbejder, men han har umiddelbart adgang til kommunikative leksikografiske data i felterne ordklasse og morfologi og han kan endelig via den leksikografiske konvergensmodel få yderligere kommunikative leksikografiske data ved at klikke på forstørrelsesglasikonet under overskrifterne “Engelsk korpusøgning – BNC” samt “Engelsk korpusøgning – Zoo”, hvis visning af pladsmæssige årsager dog udelades her.

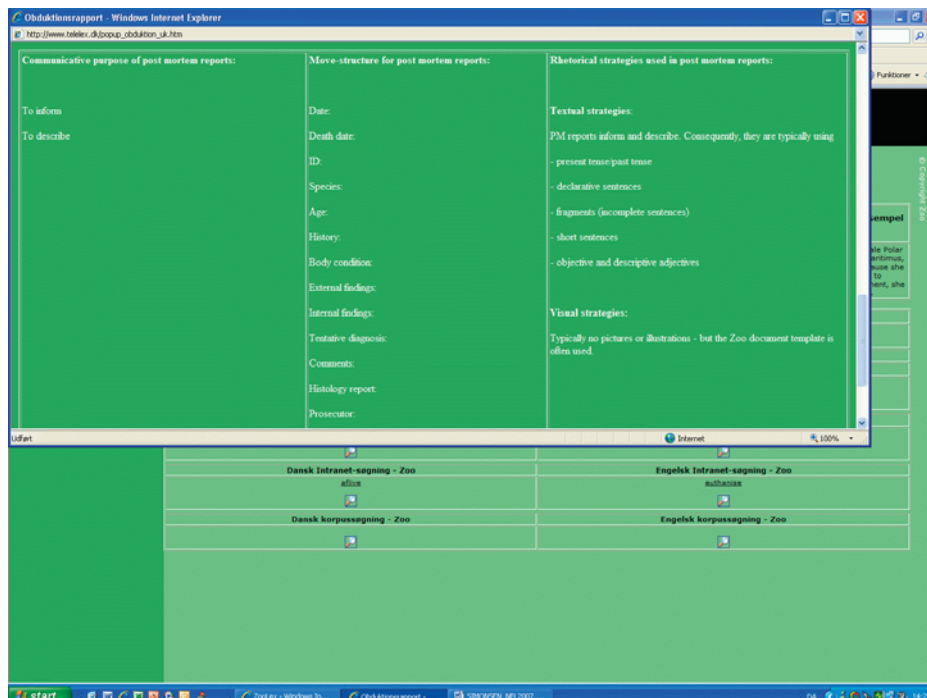
Brugerens behov for tekstgenredata er stort og ved at klikke på de to bogikoner får han adgang til værdifulde tekstgenredata. Brugeren er som sagt ved at oversætte fra dansk til engelsk, så derfor klikker han på bogikonet under overskriften “Engelsk leksikografisk dokumentkabelon“, se figur 6. I det skærbillede, der nu kommer op, vises en prototypisk obduktionsrapport og hvis brugeren scroller ned, vises resten af den leksikografiske tekstkabelon.





Figur 6. Leksikografisk dokumentkabelonmodel i ZooLeks – øverste del

Tekstgenrens kommunikative formål, trækstruktur og typisk anvendte retoriske strategier er opstillet i en tabel nederst i figur 7 således:



Figur 7. Leksikografisk dokumentskabelonmodel i ZooLeks – nederste del

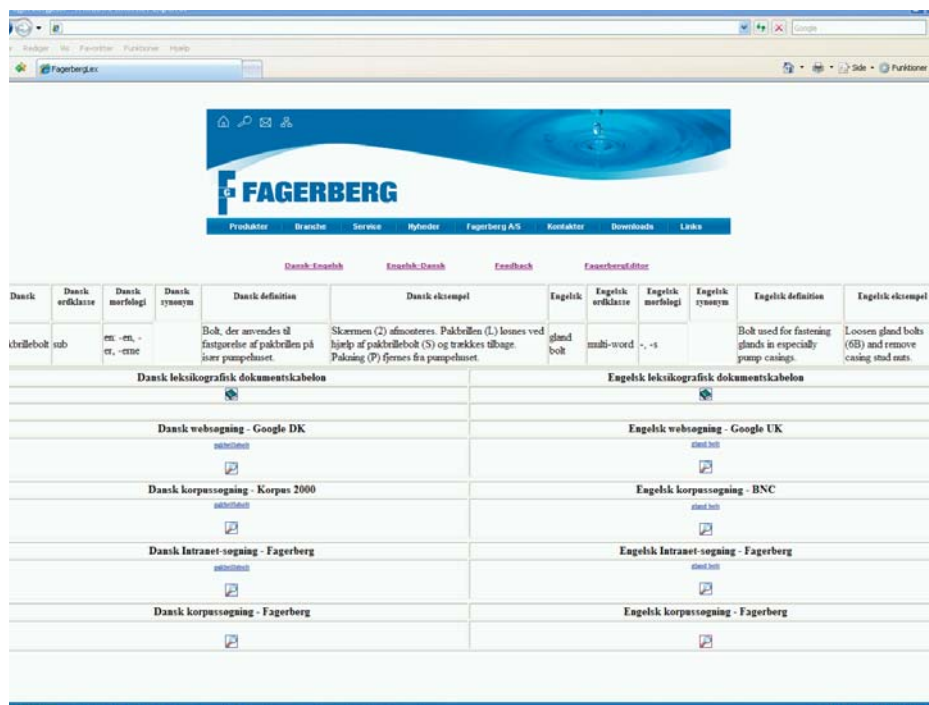
Brugeren er nu i stand til at producere en oversættelse, som er tilfredsstillende på såvel det leksikalske, syntaktiske og tekstlige plan.

Den anden brugssituation i Fagerberg A/S opsummeres i figur 8.

Tekstgenre	Manual
Opgave	Skrive højt specialiseret fagsproglig tekst på engelsk
Brugerkategori	Ingeniør
Funktioner	<p>Kognitiv leksikografisk funktion (sekundær) <i>Har et lille behov for at tilegne sig faglig viden for at kunne skrive</i></p> <p>Kommunikativ leksikografisk funktion (primær) <i>Har et stort behov for at få kommunikative data for at kunne skrive</i></p> <p>Tekstgenreorienterede funktion (primær) <i>Har et stort behov for at få tekstgenredata og -viden for at kunne skrive</i></p>

Figur 8. Fagekspertens situation og behov i forbindelse med L2-tekstproduktion

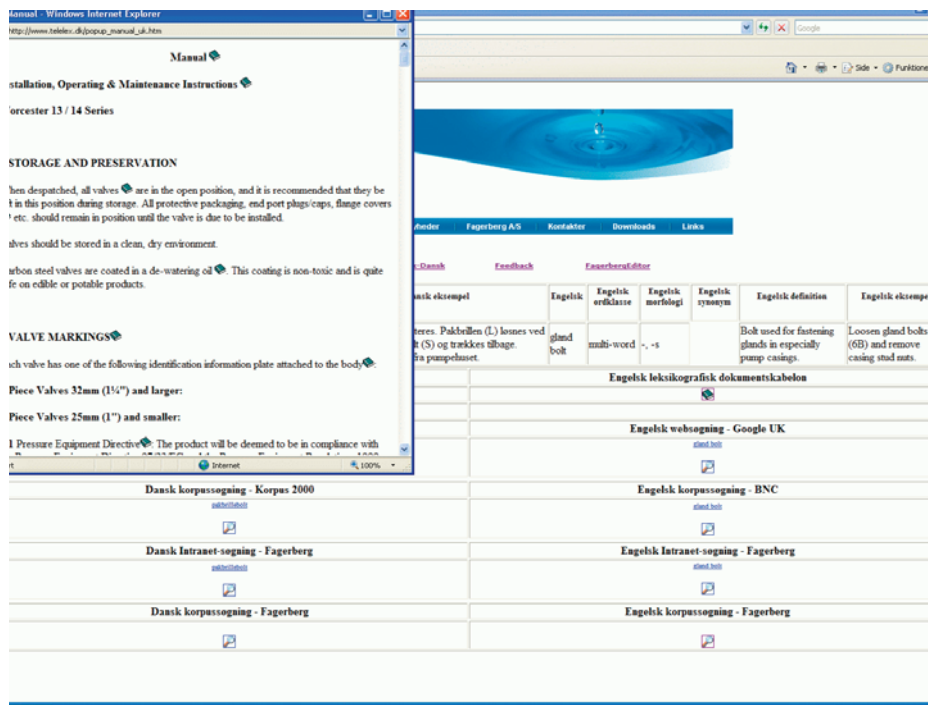
Fageksperten konsulterer nu FagerbergLex med henblik på at slå lemmaet “pakbrillebolt” op. Efter at have indtastet lemmaet i søgefeltet, angiver han tekstgenren i en drop down-boks, og han vælger “manual” i en liste med tekstgenrer. Efter at have klikket på knappen Søg vises søgeresultatbilledet, se figur 9.



Figur 9. Søgeresultatskærmbillede i FagerbergLex

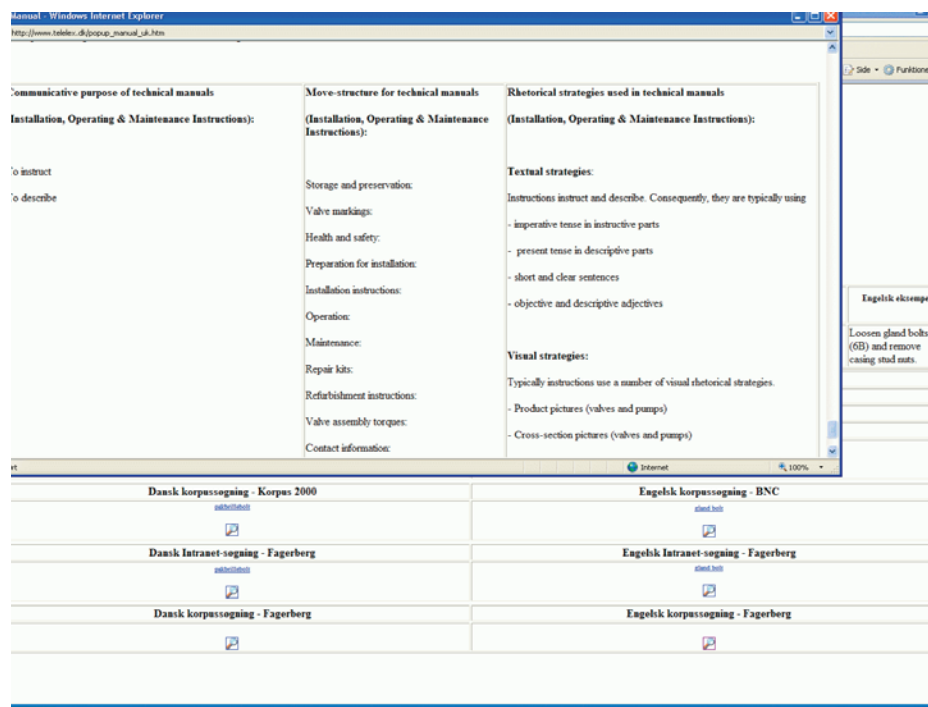
Brugerens kognitive behov er ikke ret stort, idet han er fagekspert, men han kan blive mindet om en pakbrillebolts funktion i de danske og engelske definitioner og eksempler i den leksikografiske artikel, ligesom han kan konsultere de relevante omtækster vha. konvergensprincippet og foretage søgninger efter mere viden om en pakbrillebolt. Brugerens behov for kommunikative leksikografiske data er derimod stort, idet vedkommende ikke er sprogligt uddannet, men han har heldigvis adgang til en masse kommunikative data i felterne ordklasse og morfologi og kan endelig via den leksikografiske konvergensmodel få kommunikative data ved at klikke på forstørrelsesglasikonet under overskrifterne “Engelsk korpusøgning – BNC” samt “Engelsk korpusøgning – Fagerberg”, hvis visning af pladsmæssige årsager udelades her.

Endelig er brugerens behov for tekstgenredata stort og ved at klikke på de to bogikoner får han adgang til tekstgenredata. Brugerens er ved at producere en manual på engelsk, så derfor klikker han på bogikonet under overskriften “Engelsk leksikografisk dokumentskabelon” og får vist et prototypisk eksempel på en teknisk manual, se også figur 10.



Figur 10. Leksikografisk dokumentkabelonmodel i FagerbergLex – øverste del

Tekstgenrens kommunikative formål, trækstruktur og typisk anvendte retoriske strategier fremgår af en tabel nederst i skærbilledet, se også figur 11.



Figur 11. Leksikografisk dokumentskabelonmodel i FagerbergLex – nederste del

Fageksperten er nu i stand til at skrive om pakbrillebolten, og han har nu viden om, hvad der konstituerer tekstgenren manualer.

De to modeller, som diskuteres i dette bidrag er relevante for leksikografisk grundforskning og er eksempler på specifikke teorier, der fokuserer på delområder inden for virksomhedsleksikografi. De to modeller kan, som det fremgår af denne diskussion, umiddelbart anvendes i praksis, og de tilfredsstiller helt sikkert “et leksikografisk relevant behov”, jf. Tarp (2006:100), som diskuterer, hvad der konstituerer en leksikografisk funktion. Dermed afsløres konturerne af en ny type leksikografisk funktion, der dog ikke helt passer ind i det stramme teoretiske paradigme, som opstilles af Tarp (2006: 105–106). Men som Tarp (2006: 105) netop selv angiver, så kan brugere sagtens få f.eks. kommunikative leksikografiske data også selvom de aktiverer den pågældende model i ZooLeks eller FagerbergLex for primært at tilegne sig mere viden, dvs. kognitive leksikografiske data, se f.eks. figur 10 og 11. Med andre ord er den sort-hvide opdeling ikke hensigtsmæssig, idet en potentiel bruger givetvis får både kognitive og kommunikative leksikografiske data samtidig, specielt i forbindelse med virksomhedsleksikografi. Den grundlæggende opdeling ml. kommunikative og kognitive leksikografiske funktioner er stadig helt rigtig og fuldstændig uundværlig. Men som det fremgår af følgende eksempler genereret fra ZooLeks vha. den leksikografiske konvergensmodel så er “sidegevinsten”, som Tarp (2006: 106) kalder

det, ganske betragtelig og bør ikke negligeres hverken teoretisk eller praktisk. Dette paradoks indikeres i figur 1 ved overskrifterne “Primært kommunikation-orienterede data” og “Primært kognitionsorienterede data”.

Lad os se på et eksempel, hvor en sprogmedarbejder i København Zoo slår lemmaet *bæverrotte* op i ZooLeks. Brugeren ønsker at vide mere om bæverrotter, f.eks. information om deres udseende og deres habitat. Et uddrag af søgeskærm-billedet fra ZooLeks ses i figur 12.

Her er resultatet af din søgning													
Dansk	Dansk ordklasse	Dansk morfologi	Dansk synonym	Dansk definition	Dansk eksempel	Engelsk	Engelsk ordklasse	Engelsk morfologi	Engelsk synonym	Engelsk definition	Engelsk eksempel	Klasse	Orden
bæverrotte	Sub	en: -en, -er, -eme	nutria, sumpbæver	Bæverrottens øjne er små og brune, ørene små og afrundede og på bagbenene har den svømmehud mellem tæerne.	Bæverrotten stammer fra Sydamerika, hvorfra den blev indført til Europa til pelsdyravl, nu lever den vidt: i både Europa, Sovjetunionen, Japan og Ostafrika.	coypu	Noun	-, -s		The coypu is a large semi-aquatic rodent which is native to South America.	The coypu and the brown rat are rather similar in appearance, hardly surprising as they are both rodents. But the blunt square shape of the coypu's muzzle and it's webbed feet clearly distinguish it, not to mention it's size.	Pattedyr	Gnavere

Figur 12. Lemmaet *bæverrotte*

Som det fremgår af den leksikografiske artikel i figur 12 får ordbogsbrugeren i den leksikografiske artikel både kommunikative og kognitive data, dog fordelt på forskellige felter. Sammenblandingen og de omtalte “sidegevinster” opstår, når brugeren aktiverer den leksikografiske konvergensmodel under ”Engelsk web-søgning – Google UK”, se også figur 5. Nu får brugeren vist en liste med funktionsunderstøttende dokumenter, og et uddrag af et dokument vises i figur 13.

<p>The Nutria (also called the coypu) is a large, semi-aquatic South American rodent that has webbed hind feet. It has been introduced to the USA, Asia, and Europe, as a result of fur farms (where the nutria is farmed for its luxurious coat). It lives in swamps, in marshes, and near lakes, streams, and rivers, digging burrows near water in rainforests, temperate deciduous forests, and grasslands. The nutria is nocturnal (most active at night); it rests in its burrow during the day.</p>
--

Figur 13. Kognitive og kommunikative data i funktionsunderstøttende omtækt

Uddrag fra <http://www.enchantedlearning.com/subjects/mammals/rodent/Nutria.shtml>

Uddraget viser tydeligt, at selvom brugeren først og fremmest var ude efter kognitive leksikografiske data, så får han adgang til mange relevante data (markeret med gråt i figur 13).

## 5. Konklusion

Denne diskussion har vist, hvorledes en sprogmedarbejders leksikografiske og tekstgenre-relaterede behov tilfredsstilles i forbindelse med oversættelse af en dansk obduktionsrapport til engelsk i København Zoo, samt hvorledes en fageksperts leksikografiske og tekstgenre-relaterede behov tilfredsstilles i forbindelse med produktion af en teknisk manual på engelsk i Fagerberg A/S. Den leksikografiske konvergensmodel og den leksikografiske dokumentskabelonmodel hjælper brugeren, og det argumenteres, at en udbredt brug af funktionsunderstøttende omtækster, som i den leksikografiske konvergensmodel og en integreret tilgang til tekstgenredata som i den leksikografiske dokumentskabelonmodel, optimerer brugerens vilkår og sætter ham bedre i stand til at producere tekster, som ikke kun indeholder de rigtige termer, men som også overholder de vigtigste tekstkonventioner.

## 6. Litteratur

- Kjøller, Klaus 1997: *Image – effektive råd til virksomheden om sprog og kommunikation*. København: Akademisk Forlag.
- Leroyer, Patrick 2002: *Les exemples textuels unificateurs – Fondements et applications pour le dictionnaire de traduction technique*. Ph.d.-afhandling. Århus: Handelshøjskolen i Århus.
- Schein, E. H. 1994: *Organisationskultur og ledelse*. København: Valmuen.
- Simonsen, Henrik Køhler 2002: *TeleLex - Theoretical Considerations on Corporate LSP Intranet Lexicography: Design and Development of TeleLex - an Intranet-based Lexicographic Knowledge and Communications Management System*. . Ph.d.-afhandling. Århus : Handelshøjskolen i Århus.
- Simonsen, Henrik Køhler 2007: ZooLex: Det er for vildt! I: *LEDA-Nyt* nr. 43 – marts 2007, 4–12.
- Swales, J. M. (1990). *Genre Analysis. English in Academic Settings*. , Cambridge: Cambridge University Press.
- Tarp, Sven 2006: *Leksikografien i grænselandet mellem viden og ikke-viden: Generel leksikografisk teori med særlig henblik på lørnerleksikografi*. Århus: Handelshøjskolen i Århus.

Henrik Køhler Simonsen  
Associate Professor, Industrial PhD, f. 1966  
Handelshøjskolen  
Dalgas Have 15  
DK-2000 Frederiksberg  
Department of International Language Studies and Computational Linguistics  
hks.isv@cbs.dk





MARTIN SKJEKKELAND

## Noreg og Nederland – eit tett samband i to hundre år

Det norske språket har tatt imot svært mange ord og ordlagingselement frå nedertysk. Låna kom inn som eit resultat av den kontakten som var mellom dei tyske hansakjøpmennene og nordmenn i byar sør og vest i Noreg. Midt på 1500-talet tek hollendarane, frisarane, engelskmennene og skottane over som handelspartnarar på sørvestkysten av Noreg. Også i denne perioden, fram til slutten av 1700-talet, kom det inn mange ord og uttrykk i det norske språket, og då særleg frå hollandsk. I artikkelen ser eg nærmare på den kulturelle og språklege påverknaden frå hollandsk på talemålet i Agder-fylka og i Noreg generelt.

### 1 Innleiing

Me kjenner alle det sterke handelsambandet som eksisterte mellom dei nordtyske hansabyane og byar i Skandinavia frå midt på 1300-talet og framover. Norsk (og skandinavisk) tok etter kvart opp mykje språkstoff frå nedertysk. Det er særleg i *ordforrådet* me ser den nedertyske påverknaden på dei skandinaviske språka. Store mengder nedertyske lånord finst både i skriftspråket og i dialektane. Eg nemner ord som *arbeid, å bli, bukse, bukt, fin, flink, frukt, frykt, gå, mektig, nytte, håpa, klokke, koka, lukt, måte, plaga, prata, snakka, skrika, skriva* osv. I tillegg til ord av dette slaget tok me mot mange ordlagingselement (affiks) frå nedertysk. Døme på ord med prefiks er: *angrep, anlegg, bedageleg, bedrag, bedømme, befala, erstatning, erfaring* osv., og med suffiks: *fruktbar, lasbar; betingelse, besvarelse, elendighet, sikkerhet* osv.

Ikkje noko anna språk har påverka norsk språk så mykje som nedertysk og hollandsk. Gustav Indrebø rekna i 1932 med at 30 % av orda våre kom frå nedertysk og hollandsk (Indrebø 1932:32, etter Sandøy 2000:28). Ord frå latin og gresk utgjer nok ei større mengd av ordmassen i norsk. Men dei latinske og greske orda er mindre brukte i daglegtalen. Dei nedertyske orda er i høgare grad kvardagsord i språket. Det kan elles ofte vera vanskeleg å avgjera om eit ord kjem frå nedertysk eller frå hollandsk. Grunnen er at desse språka var så like. Nederlendarane kalla sjølve sitt språk på denne tida for *nederduits* (Torp 1998:11). Torp har elles understreka det nære sambandet som finst mellom nederlandsk og nedertysk:

Grensa mellom tysk og nederlandsk er på tilsvarende måte et produkt av politiske og historiske tilfeldigheter, og ikke av virkelige språkgrenser i det folkelige talemålet. Reint språkleg er det f.eks. mindre skilnad mellom

dialekten i Nederland og Nord-Tyskland enn mellom dialekten nord og sør i Tyskland (Torp 1998:11).

### 1.1 Er den nedertyske påverknaden like stor i alle delar av Noreg?

Det er nok visse skilnader på dialektane i opptaket av ord frå nedertysk. I eit arbeid frå 1999 har eg granska bruken av nedertyske ord i Kvinesdal og jamført dette med tilhøva i innlandsbygda Bø i Telemark. Granskinga viste at fleire av dei nedertyske lånorda (“anbehetelse-orda”) synest å utgjera ein større og viktigare del av det allmenne ordtilfanget i Kvinesdal enn i Midtre Telemark (jf. Skjekkeland 1999: 76–80).

Dette resultatet må seiast å vera i samsvar med det mønsteret me er kjende med frå den språkhistoriske litteraturen, jf. t.d. Indrebø (1951:173): “Mest innverknad fekk tysk på tala vest og nord i landet.” Vestlege delar av Ytre Agder høyrer klart med til det området Indrebø omtalar. Det er også naturleg at kystbygdene tok imot mest språkgods frå utlandet. Det var folk som budde ved kysten som hadde mest samband med handelsmennene sør- og vestfrå.

## 2 Handelssamband mellom Ytre Agder og utlandet – “Hollendartida”

Midt på 1500-talet tek hollendarane, frisarane, engelskmennene og skottane over som handelspartnarar på Agder, og slik var det heilt fram til slutten av 1700-talet. Trafikken som oppstod mellom folk på Agder og dei utanlandske kjøpmennene var ganske enorm. Heile Agder-kysten frå Risør til Flekkefjord var med på dette handelseventyret. Vest på Agder var det samkvæmet med *hollendarane* som blei særleg viktig. Hollendarane kom til landsdelen for å kjøpa trelast og tømmer. Særleg var den sørlandske eika ettertrakta på grunn av den verdien ho hadde som byggjemateriale for hus og for den store hollandske flåten. Men hollendarane kjøpte òg ferdigbygde båtar. Mykje eikemateriale blei også bruka som pålar under hus i Amsterdam. (Rådhuset i Amsterdam står på 15000 pålar frå Noreg.) I tillegg til trelast og ferdigbygde båtar var hollendarane interesserte i å kjøpa fleire varer nordmennene hadde å selja, slikt som huder, skinn, mose til farging, smør, talg, ull og fisk. Til og med levande kyr og hestar blei førte over havet frå sørlandsbygdene. Ved til brennstoff blei tatt med i båten så sant denne kunne staplast på toppen av den andre lasta (jf. Sogner 1996:188).

Dei hollandske skipa som kom til landet, hadde også med seg mange varer som nordmennene kjøpte eller bytta til seg. I særleg grad gjaldt dette matvarer: Korn, salt, sild, ost og krydder. Fleire nytteprodukt, som kork og jernvarer, hadde sjøfararane òg med seg til Agder-kysten, seinare òg mur- og teglstein. Av meir luksusprega ting som hollendarane kom med, var leirkrukker, steintøy og all slags fajanse; vidare sølv, tekstilar av ulike slag og farge, gjerne raude tøystoff og blå hattar. Hollandske møblar (m.a. fine skatoll og kister) var også populære. Tobakk og kritt Piper kom til landsdelen vår gjennom hollendarane. I tillegg må

nemnast “det som var flytande”, det var mykje drikkevarer som blei løfta over skipssida i desse åra, både vin, brennevin og hollandsk øl (jf. Lillehammer 2001:20 og Sogner 1996:188).

For det meste føregjekk fraktefarten mellom Noreg og Holland med hollandske skip, men det var òg nokre bønder i området som eigde skip og “drev ikke så lidet skipsfart” (Bugge 1925 II:136).

At det var mykje liv og røre på 1500-talet i denne landsdelen, skjønar me av at det i 1549 kom opp framlegg om å få til ein kjøpstad i Lister len (Bugge 1925 II:135).

I bygdeboka for Kvinesdal har Ånen Årli eit eige kapittel om “Utvandringa til Holland” (Årli 1972:137 f.). Han fortel at mange folk frå Agder-kysten reiste til Holland for å ta seg arbeid. Han viser til gamle skiftebrev der det kjem fram namn på 26 gardar i Kvinesdal som folk har reist frå til Holland. Nokre kom aldri tilbake til heimbygda, medan andre vende heim etter å ha tent seg opp pengar i det framande landet. Årli seier at “dette sette djupe merke etter seg i Kvinesdal. [...] Hollandsfarten må ha sett merke på bygdelivet i mest 200 år” (s. 137–139). Njål Vere skriv i *Listamålet* (bd. I, 1993:26) om “Hollendertida”: “Ein kan mest samanlikna det med seinare Amerika-fart. [...] Det heitte seg på Lista at mest i kvart hus hadde dei familie i Holland.”

Eit særtrekk ved utvandringa til Holland må nemnast spesielt. Det gjeld dei svært mange unge mennene frå Agder som tok seg hyre på hollandske skip. Eventyrlysten drog nok mange ut. Men for dei fleste var motivet for utreisa å tena seg opp pengar slik at dei kunne ta over gardsbruket heime. Hyrene på dei hollandske båtane var jamt over mykje høgare enn på dei norske. Ein annan faktor som også verka med når det gjaldt interessa for å gå inn i sjømannsyirket, var frykta for å bli innkalla til teneste i den felles dansk-norske marinen. Mange unge menn blei etter kvart kommanderte til slik teneste. Dette var ein plikt dei innrullerte såg på med stor ulyst. Gamle dokument fortel at mange innrullerte frå Lister og Mandal rømde til Holland under Den store nordiske krigen (Sætra 2001:77). Dei unge jentene som reiste ut, tok seg for det meste arbeid som hushjelper i Holland, jf. nedafor.

Eg vil litt tilbake til handelen med hollendarane (og frisarane). Denne handelen var så utbreidd og livleg at denne tida har fått ei særstilling i egdene sitt medvit. “Hollendartida” har blitt eit omgrep for folk på Agder. Jamfør også at Flekkefjord den dag i dag har sin “Hollendarby”, ein bydel med smale gater og særprega byggjestil på husa.

Arnvid Lillehammer er mellom dei som har skrive om samhandelen mellom Agder og Danmark, Holland og Skottland (Lillehammer 1997 og 2001). Noko av det mest interessante hjå Lillehammer er drøftinga hans av kva *form* kontakten mellom kjøpar og seljar hadde. Han meiner at omgrep som “Hollendartida” på Agder og “Skottetida” i Ryfylke knapt dukkar opp bare fordi desse nasjonane var dominerande innafor kvart sitt område. “Eg vil tru at dei også dukka opp på

grunn av den forma handelen lenge hadde, fordi dei framande skipperane i lang tid kom i direkte kontakt med bøndene, med ålmugen” (Lillehammer 1997:25). I følgje Lillehammer oppstod det nær kontakt mellom oppkjøpar og seljar, mellom hollandske, skotske og danske skipperar og norske bønder. Ein slik direktekontakt er viktig å ha i minne når ein skal forklara den språklege påverknaden frå hollendarane. Samkvæmet mellom Holland og Agder heldt fram til langt ut på 1700-talet.

### 2.1 Vest-Agder skil seg ut

Sølvi Sogner fortel i artikkelen “Tjenestejenter til Holland” (1998) om korleis jenter frå Sørlandet fekk seg arbeid som tenestejenter og hushjelper i Holland.

Sogner har studert kjeldemateriale i Amsterdam, m.a. lysingsprotokollar for ekteskap og medlemslister for lutherske trussamfunn i åra 1601–1800. Ho fann at det i Amsterdam i dette tidsrommet blei registrert i alt 11557 bruder og brudgommar frå Noreg. Av desse utgjorde dei frå Lister og Mandal amt totalt 4946, dvs. 43 % av alle registrerte! Heile 55% av alle registrerte norske kyrkjelydsmedlemmer i Amsterdam i denne perioden var frå Lister og Mandal. Sølvi Sogner seier: “Emigrasjonen fra Vest-Agder, eller Lister og Mandals amt, som det het den gang, var helst spesiell. De fleste migrantene til Holland kom derfra” (Sogner 1998:25). Vestlege delar av Vest-Agder skil seg også ut på den måten at kvinneandelen av utvandrarane var særleg stor i dette området (jf. òg Hodne 2001:38, 45).

Hollandsfararane tilførte kystbygdene på Agder kjærkomne midlar, slik amerikafararane seinare skulle gjera det. I mest to hundre år var gardbrukarane og ei stadig veksande befolkning i desse bygdene avhengig av inntekter frå Holland.

## 3 Hollandsk språk til Noreg

Det nære sambandet mellom egdene og hollendarane sette spor etter seg i språket på Agder, slik det nedertyske hansaspråket påviseleg gjorde det på 1300- og 1400-talet. Sjølv meiner eg at hopehavet mellom Sørlandet og Holland har vore med på å fasthalda dei tidlegare innlånte nedertyske orda frå hansatida (sidan det hollandske språket var så likt det språket hanseatane brukte). I tillegg har mange hollandske ord kome til i denne perioden fram mot slutten av 1700-talet. Dette er grunnen til at det i talemålet på Agder, særleg langs kysten, synest vera fleire nedertyske lånord i bruk enn i mange andre stader i landet.

### 3.1 Interessante lånord kjem inn i dialekten

Mykje av Agder høyrer til det sørlandske dialektområdet som har ”blaute konsonantar”, dvs. overgang frå *p, t, k* til *b, d, g* i stilling etter vokal, jf. døme som: *ei piba*, *å eda* (eta), *å бага* osb.

Ein del ord skil seg likevel ut ved ikkje å ha blaute konsonantar, dei har halde

på *p*, *t*, *k* der me skulle ha venta overgang til *b*, *d*, *g*. Nokre av desse orda kan berre forklarast som lån frå nedertysk eller hollandsk. Orda må vera innlånte etter at den sørlandske prosessen med overgang til "blaute konsonantar" var avslutta. Døme på dette er: *eit prat*, *å prata* 'prata, snakka' (frå nedty.), *å pruta* (frå nedty.), *ei fauta* 'ei fuge', "*ikkje i fautan*" 'ikkje i form' (frå nedty.), *ein døyt* (frå holl.), *kaut* adj. 'kry' (tyder 'triveleg' i holl.), *ein kaper* (frå holl.) (væpna privatfarty), *staut* adj. (frå holl.), *ei sjøyta* 'ei skøyte' (frå holl.), *ein sprøyt*, *å sprøyta* (frå holl.) (jf. elles Frestad 1991:178f.).

### 3.2 Andre ord frå hollandsk

Bram de Boer (2002) nemner desse orda i norsk frå nederlandsk: *akterut* (holl. "acheruit", "achteruitrijden" = 'bakoverkøyre, rygge'), *baugspryd* (holl. "boegspriet"; "spriet" = stake), *spant* (= 'innvendig ribbe i båt'), *vant* (holl. "want" = tauverk), *bramseil* (holl. "bramzeil"), *malstrøm* (holl. "maalstroom". Hollandsk *malen* svarer til norsk *male* = knuse), *klinker* om murstein (= klingande murstein; når steinen blir delt i to), *papp* (holl. "pap" = grøt), papp blir laga av grautaktig pappirmasse. De Boer nemner også  *fartøy*, *kahytt*, *hyre*, *vindås* og *kyst*. Her nedafor følgjer elles eit utval av andre ord som eg har plukka ut frå databasen til *Nynorskordboka*. Det gjeld ord som alle er oppførte med opphav i hollandsk. Ein del av orda er omlaga i form og tyding. Nokre av orda har opphav i fjernare språk, men har kome til oss gjennom hollandsk:

*appelsin* (eig. flamsk "appelsien", holl. i dag: "sinaasappel"), *å balera* ('trampa, ståka'), *bambus* (eig frå malayisk "bambu"), *bas*, *bederva*, *bengel*, *bestefar* og *bestemor* (holl. i dag: "grootvader" og "grootmoeder" i skrift, "opa" og "oma" i talemålet), *bestikk*, *bever*, *å brasa*, *bus*, *børs*, *drill*, *døyt*, *edamer*, *eddik*, *filibuster* (ei spansk-porugisisk omlaging av holl. "vrijbuiiter", 'fribyttar'), *forsuffa* ('forfjamsa'), *å fortona* (*seg*), *frakt*, *gauda* (frå Gouda = bynamn), *gut*, *å heisa*, *kalkun*, *kaut*, *å knabba*, *komfyr*, *kvakksalvar*, *kvalross*, *å kvesta*, *lubben* (holl. "lobbig". 'romsleg'), *lurendreier*, *løyert*, *å mangla*, *å mankera*, *nickers* (frå "knickerbockers", amerikansk nemning på nederlendarane i USA på 1600-talet, eig. hollandsk familienamn: Knikkebakker), *nøkkelost* (etter nøklane i Leidens byvåpen, prenta på osten), *palme*, *passiar* (eig. malayisk "bicara", holl. "pitsjaren"), *å pelma*, *plattung*, *å rappa*, *singel* (småstein), *sjaber*, *sjenever*, *å skaffa*, *å skeisa*, *å skofta*, *å skolera*, *skøyar*, *å skøya*, *slitasje*, *å sprøyta*, *å spyla*, *stillas*, *å svaia*, *tralt*, *vaterpass* og *vatt* m. fl.

### 3.3 Ord frå sjømannsyrket

Eg nemnde ovanfor korleis dei unge egdene i stor grad lærte det "moderne" sjømannsyrket av hollendarane. På bakgrunn av dette er det ikkje merkeleg at mange av dei orda våre som er knytte til sjømannslivet, har opphavet sitt i hollandsk. Eg har alt nemnt ein del slike ord. Her er fleire:

*akter, anker, bakk, baugspryd, bedding, bramrigg, bris, å buksera, båtsmann, certeparti, cruise, dekk, dis, dokk, dregg, dørk, fallreip, farty, farvatn, fokk, å forhala, forlis, å forlisa, å fortøya, frakt, harpun, å heisa, ein hyling (ugreie på fiskegarn), jagar, jakt, å jibba, kai, å kantra, kaper, kjetting, å kjølhala, klyver (trekanta stagsesgl), korvett, kryssar, kuling, køy, laber, lekkasje, å lensa, matros, orkan, orlog, å overhala, passat, å peila, å praia, reiar, reling, ropert, ruff, sjau, å skalka, skøyte, å slingra, slupp, spant, å spleisa, stropp, å svaia, talje, tokt, trinse, utkikk, vasspegel, vaterbord (planke ytst i skipsdekket), vimpel, vrak og yacht m. fl.*

### 3.4 Sørendingane kjenner hollandske ord

I samtalar med egder om språk og kultur har det ofte kome opp ord som informanten meiner har opphav i Holland. Det gjeld ord som *allgodt* adj. ('sikkert'), *ankerasje* subst. ('ankerplass'), *beje* v. ('gje opp', "vil du beje?"), *becomst* subst. ('det ein fortener'), *bomse* subst. ('stor kvinne', og nemning for ein bestemt båttype i hollandsk), *dandera* v. ('stella til, pynta'), *flensje* subst. (i hollandsk: 'ei lita kake av pannekakerøre'), *nå forbaser det* ('det går for langt'), *hautenhaler* subst. ('energisk person, våghals'), *holk* subst. ('dårleg båt'), *kapseisa* v. ('kanta'), *lappsalva* v. ('setja inn tauverk med tjøre'), *maskepi* subst. ('intrige'), *pøs* subst. ('bøtte'), *remedie* subst. ('saker og ting', frå latin, gjennom hollandsk), *spons* subst. ('propp til ei tønne'), *søsterkake* subst. (bestemt kake frå Holland).

Interessant hadde det vore å leita opp alle ord i norsk som har kome frå hollandsk. I ei slik gransking ville det truleg vore fruktbart å dela inn ordmassen i semantiske grupper, dvs. sakgrupper som t.d. "mat og drikke", "utstyr i huset", "kunst og kultur", "sjømannsskap" osv., slik eg ser det er gjort av Torben Arboe Andersen i arbeidet med lån frå lågtysk til jysk (jf. Andersen, T.A. 1995).

### 3.5 Lån av faste uttrykk

Mange faste uttrykk i språket har kome frå nedertysk og hollandsk. Døme er: *akta og æra* (= nedty. *achten, eren*), *gå over styr* (= holl. *overstuur gaan*), *stikka i brann* (holl. *in brand steken*), *spott og spe* (= nedty. *spott un spee*), *rubb og stubb* (= nedty. *rump un stump*), *ditt og datt* (= nedty. *dit un dat*), *hulter te bulter* (= nedty. *huller de buller*), *reie pengar* (= nedty. *rêden penningen*), *på lykke og fromme* (= nedty. *vrômen*), *til punkt og prikke* (= nedty. *tô pricke*), *pikk og pakk* (= nedty. *pak*), *basta og binda* (= nedty. *basten*), *velja og vraka* (= nedty. *wraken*), *koma i beknip* (= nedty. *in beknip komen*), *koma i stand* (= holl. *tot stand komen*), *ta til ekte* (= nedty. *tô êchte nemen*) m.fl. (jf. Naumann 2000: 95 f.).

## 4 Konstruksjonen "han sin" og "henna sin" i dialekten

Eit svært sentralt grammatisk fenomen i dialektane på Agder er genitivsomskriving med reflektivt eigedomspronomen. Det gjeld typen: "dæ æ *Ola sin* hest"; "æ

dæ *Kari sin* kjole?"; "dæ æ *hans sin* båt"; "dæ æ *henna sin* katt" osv. Det er semje mellom språkforskarane om at denne konstruksjonen har opphav i det nedertyske og hollandske språket (jf. Torp 1973:131–133). *Sin*-genitiv blir i dag ikkje brukt i skriftleg og formell nederlandsk, men er vanleg i talemålet. Ordet "zijn" på hollandsk betyr 'hans', så det har skjedd ei omtolking i norsk pga. lydlikskap.

Ein legg merke til at det mest utprega norske "sin"-området finn ein i dialektane vest på Agder. I eit område i vestre Vest-Agder og sørlege Rogaland har konstruksjonen også trengt inn som genitivsomskriving ved personlege pronomener, typen: *han sin, henna(r) sin, ho sin* osv. Det geografiske kjerneområdet for "sin"-konstruksjonen på Agder fell saman med det området der samkvæmet med hollendarane var særleg sterkt på 1500- og 1600-talet. Dette underbyggjer min teori om at hollendarane, forutan å bidra med sine eigne ord og uttrykk til dialektane, også styrkte den tidlegare nedertyske språkpåverknaden i hansatida i dei områda der hollendarane etablerte seg med handel. Dei mange hollandske lånorda og mykje bruk av genitivsomskriving med *sin* er etter mitt syn eit resultat av dette.

## 5 Namneskikk frå Holland

Frå 1600-talet og utover merkar me at det kjem inn ein del hollandske og tyske personnamn i egdemåla. Tidleg på 1600-talet heitte dei fleste mennene her *Anders, Ola, Arne, Halvor, Ånun, Torgeir, Jon, Tobias, Tallak* osv. På 1600- og 1700-talet kjem det inn namn som *Arnt, Adrian, Bertel, Bernt, Cornelius, Claus, Didrik, Evert, Hans, Hermann, Jan, Laurits, Reinert, Vollert, Villum, Valter, Verner, Teis* m.fl. Kvinnenamna endrar seg også: *Anna, Sigrid, Guri, Ingrid* og *Åse* får nå konkurrentar i namn som *Alida, Amalie, Berte, Bertine, Alette, Emma, Gesine, Gunda, Henrikke, Hansine, Juliane, Jørgine, Katrine, Susanne* m.fl.

## 6 Hollandske uttrykk som skildrar "nordmannen"

I det hollandske språket har det funnest ein del uttrykk der den norske "veremåten" blir trekt fram for å karakterisera ein person. Sølvi Sogner (1996:194) nemner ein del slike hollandske uttrykk: "Hij kijkt als een bok uit Noorwegen" ('han glaner som ein geitebukk frå Noreg'), "De duivel voere hem naar Noorwegen" ('måtte djevelen ta han til Noreg'), "Hij slacht de Noorman: hij kan niet beloven en houden" (eldre språk, tyder: 'han slektar på nordmannen, han kan ikkje halda ein lovnad.' Verbet "slachten" tyder i dagens hollandsk 'å slakta', og nemninga "Noormann" blir i dag bare bruka om vikingane; moderne nordmenn heiter "Noren", eintal "Noor".), "Dat is een zoet lief beestje, zei Pieter de Noorman, en hij zag een platluis" ('det var eit søtt lite dyr, sa nordmannen då han såg ei flatlus'), "Hij bromt als een Noorse beer" ('han knurrar/brummar som ein norsk bjørn'), "Men zou eerder met een leuke schuit naar Noorwegen varen" ('eg ville heller segla til Noreg med ein båt som var lekk').

Dersom ein skal ta desse uttrykka på alvor, så stod ikkje nordmennene så “høgt i kurs” hjå hollendarane. Men uttrykka fortel i alle fall om nær kontakt mellom dei to landa.

## 7 Etterord

Eg har i denne artikkelen peika på nokre særdrag i Agder si historie, og då særleg den historia som gjeld dei vestlege bygdene i landsdelen vår. Eg har prøvd å få fram nokre av dei kulturelle impulsane egdene tok med seg frå framande land. I hovudsak har eg teke for meg sambandet med Holland på 1600- og 1700-talet. Men det er to andre viktige utferdperiodar eg ikkje har nemnt. Det gjeld for det første stordomstida for seglflåten i dei to fylka utover på 1800-talet. Kontakten med andre land og kulturar har såleis Agder hatt i lange tider. Fleire generasjonar mottok formande inntrykk frå Holland frå 1500-talet og framover (og seinare frå USA).

Egdene har såleis alltid vore eit folkeferd med andletet vendt mot havet. Historikaren Bjørn Slettan (1998: 255) viser til Vilhelm Krag, som meiner at vår kontakt med andre land og folkegrupper har blitt altfor lite påakta som ei viktig kulturell ballast. Krag sa det slik (i *Tidens Tegn* 1917): “Vi sørlændinger har været fattige og forurettede og forglæmte altfor længe (...) enda vi har havet udenfor og derfor hele verden til nærmeste nabo” (her sitert etter Andreassen 1994: 52).

## 8 Litteratur

### Ordbøker:

*Nynorskordboka* 1993 = *Nynorskordboka. Definisjons- og rettskrivingsordbok*, redigert av Hovdenak, M., Killingbergtrø, L., Lauvhjell, A., Nordlie, S., Rommetveit og M., Warren, D. Oslo: Det norske samlaget.

### Annan litteratur:

- Andreassen, Jostein 1994: *Begrepet Sørlandet : en historisk oversikt*. Søgne: J. Andreassen.
- Andersen, Kåre A. m.fl. 2001: *Kontakten mellom Agder og Holland på 1600- og 1700-tallet*. (= Seminarrapport, Flekkefjord Museum 30. juni – 1. juli 2000.) Flekkefjord: Flekkefjord Museum. 2001.
- Andersen, Torben Arboe (1995): *Niederdeutsche Lehnwörter in jütischen Dialekten*. I: Braunmüller, Kurt (red.) 1995.
- Braunmüller, Kurt (red.) 1995: *Niederdeutsch und die skandinavischen Sprachen II*. Heidelberg: C. Winter.
- Bugge, Alexander 1925: *Den norske trælhandels historie*, bd. I-II. Skien : Fremskridts boktrykkeri.
- de Boer, Bram 2002: Norsk-nederlandske lånord og faslke venner . I: *Språklig samling* 2/2002. 3–7.



- Frestad, Randi Lohndal 1991: *Lydverket og formverket i Lista-målet. Utviklinga i dette hundreåret* (uprenta hovudoppgåve). Universitetet i Bergen.
- Hodne, Kåre Oddleif 2001: Utvandringen til Nederland. Hovedlinjer og særtrekk. I: Andersen, Kåre A. m.fl. 2001. 37–59.
- Indrebø, Gustav 1932: *Nynorsk og bokmål*. Bergen: Lunde.
- Indrebø, Gustav 1951: *Norsk målsoga*. Utgjeve av Per Hovda og Per Thorson. Bergen.
- Jahr, Ernst Håkon (red.) 2000: *Språkkontakt – Innverknaden frå nedertysk på andre nord-europeiske språk*. København: Nordisk Ministerråd. [Findes ikke i teksten] [Sjå Naumann nedanfor]
- Lillehammer, Arnvid 1997: Trelastutførsel til Holland, Skottland og Danmark. I: Slettan, B. (red.): *Skogbrukspolitik og trelasthandel. Ei artikkelsamling*. (= Skriftserien ved Høgskolen i Agder 20/1997.) Kristiansand: Høgskolen i Agder.
- Lillehammer, Arnvid 2001: Skottar og hollendarar på Agder- og Rogalandskysten. I: Andersen, Kåre A. m.fl. 2001. Flekkefjord: Flekkefjord Museum. 4–24.
- Naumann, Hans-Peter 2000: Lågtysk fraseologitransfer: Strukturer och modeller. I: Jahr, E.H. (red.) 2000. 95–110.
- Sandøy, Helge 2000: *Lånte fjører eller bunad? Om importord i norsk*. Oslo: Landslaget for norskundervisning LNU / J.W. Cappelens Forlag A/S.
- Skjekkeland, Martin 1999: *Tysk-danske lånord i nynorsk og i bygdemåla*. (= Forskingsserien ved Høgskolen i Agder 16/1999.) Kristiansand: Høgskolen i Agder.
- Slettan, Bjørn 1998: *Agders historie 1840-1920*. Kristiansand: Agder historielag.
- Sogner, Sølvi 1996: Popular contacts between Norway and the Netherlands in the early Modern Period. I: Juliette Roding and Lex Heerma van Voss (red.) (1996): *The North Sea and Culture (1550-1800)*. Proceedings of the International Conference held at Leiden 21-22 April 1995. Hilversum/Verloren 1996. 185–198.
- Sogner, Sølvi 1998: Tjenestejenter til Holland. I: *Vår barndoms have*. (= Årbok for Vest-Agder fylkesmuseum.) Kristiansand: Vest-Agder fylkesmuseum. 22–35.
- Sætra, Gustav 2001: Norske sjøfolk i hollandsk tjeneste. I: Andersen, Kåre A. m. fl. 2001. 74–82.
- Tidens Tegn* (dagsavis). 1910 til 1941. Redigert av [Olaf Anton Thommessen](#). Oslo.
- Torp, Arne 1973: Om genitivsomskrivninger og -s-genitiv i norsk. I: *Maal og Minne* 1973. 124–150.
- Torp, Arne 1982: *Norsk og nordisk før og nå*. Oslo: Universitetsforlaget.
- Torp, Arne 1998: Praat je Nederlands? I: *Språknytt* 2/1998. 10-13.
- Vere, Njål 1993: *Listamålet*, bd. I. Lista: Eige forlag.
- Årli, Ånen 1972: *Kvinesdal. Ei bygdebok*, bd. III. Kvinesdal: Kvinesdal kommune.

Martin Skjekkeland  
 professor i nordisk språkvitenskap, f. 1943  
 Universitetet i Agder (University of Agder)  
 Institutt for nordisk og mediefag  
 Serviceboks 422  
 NO-4604 Kristiansand  
 martin.skjekkeland@hia.no



## Fire tematiske ordbøger fra renæssancens Danmark

This article offers a short presentation of four thematic dictionaries from the Danish Renaissance, viz. Jon Tursten: *Vocabularius rerum* (1561), Henrik Smith: *Libellus vocum Latinarum* (1563), P[oul] N[ielsen] H[ingelberg]: *Vocabulorum variorum expositio* (1576), and Mads Pors: *De nomenclaturis Romanis* (1594). As the titles imply the main purpose of these dictionaries was the teaching of Latin; however, I am more interested in their representation of the Danish vocabulary. The article deals with various aspects of the dictionaries, mainly their structure and their sources.

### 1 Indledning

Jeg skal her præsentere de fire tematisk ordnede latinsk-danske ordbøger fra sidste halvdel af 1500-tallet der udgør grundlaget for mit i februar 2007 påbegyndte ph.d.-projekt. Desuden er det tanken kort at komme ind på fra hvilke vinkler jeg vil angribe dem i dette projekt.

### 2 De fire ordbøger

De fire ordbøger er i moderne tid udgivet som faksimiler i serien *Det 16. århundredes danske vokabularier*. I samme serie er udgivet to ældre alfabetiske ordbøger: I 1510 udkom nemlig i Paris den kendte humanist Christiern Pedersens *Vocabularium ad usum dacorum*, og et tiår efter så Henrik Smiths *Hortulus synonymorum* dagens lys. Den sidste udmærker sig ved at have dansk som kildesprog.<sup>1</sup> Yngre end disse seks 1500-talsordbøger er Poul Jensen Coldings meget omfattende *Etymologicum Latinum* fra 1622 og hans mindre dansk-latinske *Dictionary Herloviaenum* (1626). Også disse to er alfabetiske. Fælles for de nævnte otte værker er at de indgår i det site som projektet *Renæssancens Sprog i Danmark* er ved at etablere. Jeg skal vende tilbage til dette projekt til sidst.

De fire tematiske ordbøger er:

1. Jon Tursten: *Vocabularius rerum* (1561)
2. Henrik Smith: *Libellus vocum Latinarum* (1563)
3. P[oul] N[ielsen] H[ingelberg]: *Vocabulorum variorum expositio* (1576)
4. Mads Pors: *De nomenclaturis Romanis* (1594).

<sup>1</sup> De seks vokabularier er tidligere blevet præsenteret i Anderson, Hjorth og Jørgensen 1997 og i Jørgensen 2000.

Der knytter sig forskellige problemer til hver af disse, og jeg vil derfor slå ned på forskellige enkeltheder i det følgende.

### 2.1 Jon Tursen

Jon Tursen blev født omkring 1512 og døde 1577. Han var beskæftiget som kanik og rektor for domskolen i Lund. Han samlede historisk materiale og oversatte bl.a. Johan Carions udbredte historieværk *Chronica*.<sup>2</sup>

Tursens ordbog er meget simpelt opbygget med en nøgen artikelstruktur bestående af latinske og danske modsvarigheder i hver deres spalte. Makrostrukturen er bestemt af den saglige orden. Bogen begynder med et kapitel om Gud og de himmelske ting hvorpå følger kapitler om tiden, menneskets lemmer, klæder, om husets indretning, om træer, frugt og meget mere. I alt 58 kapitler.

Hovedparten af opslagsordene er substantiver og adjektiver fordi Tursen bagerst har sat kapitler med verber opdelt efter de latinske konjugationer. Den kapitelinterne ordning af disse er i høj grad associativ, og af pædagogiske grunde er der sammen med verberne anført beslægtede nominer: "*Canto, Decanto* ieg siunger / *Recanto* ieg igen siunger / *Cantus* sang". Den associative ordning kendetegner i det hele taget Tursens ordbog (og genren i det hele taget). Gennemgangen af den menneskelige anatomi kan tjene som eksempel. Den findes i kapitlet *De homine & eius partibus* der indledes med generelle betegnelser på mennesker og dets aldre fra "*Homo* Menniske" til "*Pubertas* [Suendom]",<sup>3</sup> og efter en række ord for menneskets sjæl og sanser følger kapitlets hoveddel startende med "*Corpus* Legomme". Beskrivelsen af kroppen følger et udbredt mønster fra hoved til fod (jf. fx Hüllen 1999: 223), det associative kommer til udtryk i at ord der i egentlig forstand ikke hører til anatomen anføres her. Egentlig anatomiske er "*Mucus* snøt i Næsen" og "*Lotium* piss" vel ikke for en moderne betragtning, men interessantere er det at kapitlet opregner ord som "*Vox* røst", "*Osculum* kyss" (efter "*Os oris* mund"), "*Palmus minor* tuer haand bred" og "*Vestigium* fodspor".<sup>4</sup>

Tursens ordbog er den første af sin art i Danmark, men den skriver sig ind i en lang tradition af tematiske ordbøger (jf. Hüllen 1999 og Müller 2001: 285ff.) der oplever sin storhedstid i perioden fra bogtrykkerkunstens fremkomst til omkring 1700. Det er sigende at af de seks danske ordbøger der kendes fra 1500-tallet er de fire tematiske.

Denne tradition betyder også at Tursen næppe har sammenstykket sin ordbog uden væsentlig inspiration fra andre kilder, og snarest fra tematiske ordbøger. Alene for det tyske område opremser Müller (2001: 286ff.) 28 tematiske ordbøger (ofte i adskillige oplag) hvis datering muliggør at Tursen kan have kendt til dem. Materialet er altså omfattende, og jeg er ikke kommet frem til at kunne pege på

2 Tursen er bl.a. biograferet i *Dansk Biografisk Leksikon*, 3. udg. (DBL<sup>3</sup>), 15: 75f.

3 I skarpe parenteser sættes her det ord originalens *idem*. henviser til.

4 Der er ikke spor af at danskens "fodspor" har været brugt metonymisk om fodsålen eller hele foden, således som latinens *vestigium* (jf. fx Jensen & Goldschmidt).

hans hovedkilde. Påfaldende er det at Tursens bog bærer samme titel som en af de tidligste i genren, den senmiddelalderlige latin-tyske ordbog *Vocabularius rerum* af Wenzeslaus Brack. Umiddelbart synes der ikke at være nogen forbindelse mellem Bracks og Tursens bøger. Brack begynder fx meget uligt det tidstypiske med “*Celum* .. Der himel”, mens Tursen som det er almindeligt er en *De Deo*-ordbog og som første indgang har “*Devs* Gud”. Tursen kunne jo her have redigeret i forhold til sit forlæg, og hans første kapitel indeholder – som det er sædvanligt – også de himmelske ting. Bortset fra denne uoverensstemmelse er også Bracks ordbog struktureret over hvad Hüllen har kaldt “a movement *from above to below*” (1999:179) og tilhører dermed en bred strøm af tematiske ordbøger.

Bracks ordbog indeholder 105 kapitler, altså 47 flere end Tursen, men denne sammenligning overdriver uligheden mellem de to bøger. Brack opsplitter ofte et emne på flere kapitler, mens Tursen har en mindre detaljeret makrostruktur.

Hvordan forholder det sig så med indholdet? En undersøgelse af Bracks to kapitler om skibsfart viser at de indeholder lidt færre (40+35) indgange end Tursens tilsvarende kapitel (med 84 indgange). Til gengæld er informationsmængden større hos Brack der ofte har latinske forklaringer, men stedvis udelader tyske modsvarigheder. Det udelukker selvfølgelig ikke at Tursen kan have skovet hos Brack, men sandsynligt er det ikke. Under de åbenlyse uoverensstemmelser synes der dog at være en vis enighed i lemmerækkefølgen. Således er de to stort set enige om rækkefølgen i afsnittet fra *Carina* til *Anchora*: *Carina* – [*Sentina* Tursen] – *Fori* – [*Agiania* – *Columbaria* Brack] – *Transtra* – *Remus* – *Palmula* – *Remex* – [*Remigium* Brack] – *Antenne* – *Malus* – *Clauus* – [*Gubernaculum* Tursen] – *Anchora*. Denne overensstemmelse er påfaldende, men plusstoffet gør det usandsynligt at Brack er kilde til Tursen. Om Bracks to skibsfartskapitler skriver Pleuger i sit afsnit om ordbogens hovedkilde, Isidor af Seville *Etymologiae*:

Zu Kapitel 32 und 33 über Schiffahrt finden sich ebenfalls im XIX. Buch Entsprechungen mancher Artikel sowie einiger Etymologien, aber deren Abfolge ist wiederum aufgrund einer anderen thematischen Ordnung anders. (Pleuger 2005:93).

Det netop anførte afsnit svarer dog nogenlunde til rækkefølgen hos Isidor XIX.2.

At Tursens ordbog skulle være direkte afhængig af Bracks, kan afvises. I stedet kan de sammenfald der er, forklares på en af følgende måder: 1) Tursens ordbog bygger på et værk der har udnyttet Brack, 2) Tursen og Brack har fælles direkte forlæg eller 3) Brack og Tursens forlæg har samme forlæg (sandsynligvis Isidors *Etymologiae*).

## 2.2 Henrik Smith

Henrik Smith (ca. 1495–1563) er en kendt person i dansk litteratur- og sproghistorie. Han er også en meget fascinerende person. En praktisk og mangestedsvirkende folkeoplyser. Især kendt for sine læge- og urtebøger. Men han har også virket ved 1524-oversættelsen af Det Ny Testamente, ligesom han har filologiseret i Skånske Lov.<sup>5</sup>

Som sagt udgav Smith i 1520 den første ordbog med dansk som kildesprog. Men allerede to år tidligere var han på banen med en ny udgave af Christiern Pedersens pionerværk (fra 1510). Til denne tredjeudgave (der kom i Leipzig) føjede Smith nogle tillæg der hidtil har været ignorerede, men som er spændende fordi de er de første danske forsøg på en saglig opstilling af dele af ordforrådet. Smith – der for eftertiden særligt er kendt for sine lægebøger – har i disse tillæg bl.a. opremset planter, urter og sten, men også slægtskabsbetegnelser. Latin dominerer fuldstændigt, og adskillige artikler mangler danske modsvarigheder.

Disse lister er ikke en anderledes præsentation af Christiern Pedersens alfabetisk ordnede materiale. Dels omfatter tillæggene kun en meget lille del af Christiern Pedersens omfattende ordforråd, dels finder man i tillæggene ord der ikke kendes fra Christiern Pedersens ordbog, fx "*Amarus[c]a* hønseydyl" 'stinkende kamille' og "*Arenarius* sondsten".<sup>6</sup> Med Smiths egen tematiske ordbog er der heller ingen ligheder. De tilsvarende kapitler dér har en anden struktur end i tillæggene, og hverken lemmaerne eller de danske ækvivalenter er de samme. Som eksempel kan behandlingen af mad i de to kilder bruges.

Mens tillægget går direkte til – ikke makronerne, men – "*Azimus*. brød til giort foruden surdeg", "*Placenta*. panekage" og "*Lebeta*. hunig kage", lægger 1563 ud med overbegreber som "*Cibus* .. Mad", "*Cibaria* .. Alt det som ædes/ enten aff folck eller queg" og "*Dapes* .. Kresen mad".<sup>7</sup>

1563 mangler den sidste af tillæggets citerede artikler, mens de danske modsvarigheder til de to andre er: "Vsyret brød" og "Kage aff mel oc melck med Honning beblendt". Tillægget har ingen af 1563's overbegreber, men går hurtigt over til drikkevarerne hvor der bliver længere mellem de danske modsvarigheder. De dukker dog op, og vi møder heldigvis de to fraser "det geller dig. eller du for eth got tor" som oversættelse af latinens *propino tibi*.<sup>8</sup>

Smiths ordbog fra 1563 er den største af de fire, og han har et meget stort ordforråd. Hans ordbog anfører de latinske ords bøjningsoplysninger – i form

5 DBL<sup>3</sup> 13:527f., Smith er udførligere biograferet af Lauritz Weibull i 1897.

6 Sammenligningen bygger på faksimilen af 1510-udgaven, at der heller ikke er overensstemmelse med Smiths revision antydes af at denne (som førsteudgaven) har "*Placenta te*. kage" (planche VII i Skautrup 1947:147), mens tillægget har "*Placenta*. panekage".

7 Udeladt i citaterne her er 1563's bøjningsoplysninger og paradigmehenvisninger.

8 Det har vist sig at Smiths tillæg i udstrakt grad bygger på en latin-tysk ordbog af Balthasar Trochus fra 1517. Derfra er et meget stort antal artikler overtaget og tysken erstattet med dansk ordstof.

af paradigmehenvisninger – efter den danske modsvarighed. Som Tursten tager Smiths kapitelstrukturering udgangspunkt i Gud, men han dvæler længe ved mennesket fra dets lemmer over dets sygdomme til dets familierelationer. Det er nok Smiths medicinske interesser der slår igennem her, ligesom i de store kapitler om urter, blomster og vellugtende urter og rødder og ”om den fruct som er i beller oc icke affskeres, men opdrages oc affpluckles met henderne” (som hans kapitel 52 hedder).

### 2.3 Poul Nielsen Hingelberg

Forfatterangivelsen P: N: H: i det ellers anonyme *Vocabulorum variorum expositio* (1576) er blevet udlagt som Poul Nielsen Hingelberg om hvem vi intet ved. Derimod ved vi en hel del om hans kilde og hans arbejdsmetode, for i efterskriften vedgår han arv og gæld:

Fromme Læsere her haffuer du denne føye Baag Vdsat oc Fordanskedt aff Suenske fattige simpel Børn til hielp/ findis her nogid som icke saa ræt er Fordanskedt aff Suensken da anamme det dog for gode/ Oc forbædre det huor du kandt.

Det forlæg Hingelberg har brugt er den latin-svenske tematiske ordbog *Variarum rerum vocabula* hvis førsteudgave udkom 1538. Disse bøger anvendelse i skolesammenhænge, og det slid der følger deraf gør det usandsynligt at Hingelberg skulle have brugt en knap 40 år gammel bog. Og når man ser nærmere på materialet eller blot på Isak Collijns bibliografiske beskrivelser af det (1925 og 1928), bliver det klart at Hingelberg har brugt en senere udgave, en af de overleverede 1545, 1559 eller 1574 – eller måske en fjerde, nu tabt. Jeg har sammenholdt Hingelbergs tolv første kapitler med de tilsvarende i 1559-udgaven, og selvom det næppe er den han har benyttet kan kollationeringen sige mere om hans arbejdsmetode end en tilsvarende med 1538-udgaven.<sup>9</sup> I sin bibliografiske beskrivelse af det af ham fundne unikum af 1559-udgaven opremser Collijn nyindkomne trykfejl i forhold til førstetrykket. Af dem han nævner, genfindes én i den danske ordbog: ”blod i hoffuedet” (*Heterocrania*, for ”bold”). Ligesom samtlige de nye artikler Collijn anfører fra 1559, genfindes i den danske ordbog. At det netop er 1559 – eller en udgave med samme layout – Hingelberg har haft foran sig forklarer usøgt visse fejl i hans tekst. I dette eksempel gengives de to ordbøgers artikelopsætning:

<sup>9</sup> Kun udgaverne 1538 og (den i denne sammenhæng uvæsentlige) 1579 er genudgivet i moderne tid (af Andersson 1890 hhv. Collijn 1925). Udgaven 1545 er kun overleveret i fragment, jf. Lindberg 1960.

1559:	Hingelberg:
Pronuba brudsäta / brwd	<i>Pronuba</i> Brudsæde
fremma / brwdtämna	<i>Fremma</i> Brudtammer

Her har han simpelthen opfattet et linjeskift som artikelskift, med det ulatinske *Fremma*, og udanske "Brudtammer" som resultat. Hans misforståelse er blevet lettet af at 1559 – i modsætning til tidens almindelige praksis – ikke skelner typografisk mellem latinske og svenske modsvarigheder, men sætter al tekst i fraktur.<sup>10</sup>

En sammenligning af Hingelbergs bog med udgaven 1559 viser at Hingelberg har redigeret langt mere end hvad Ehrlings udtalelser om hans værks svecismer giver udtryk for. Der er langt fra tale om at Hingelberg gennemgående har overtaget 1559's ordlyd ukritisk (jf. Ehrling 1929–30: 27). Der er fx således både eksempler på at han tilføjer og fjerner modsvarigheder. I de følgende eksempler angives først 1559's svenske tekst og efter kolonet Hingelbergs danske, i parentes sættes den latinske modsvarighed de er fælles om.

lekamen/ krop : Legomme (*Corpus*)  
 framdelen på halsen/ förhals : Fremdelen aff hals (*Iugulus*)  
 brödrabarn bröllinger/ syskenebarn : Brødrebørn eller syskenebørn (*Patruelis*)  
 frijare/ bedhel : Friere (*Procus*)  
 spott : Rege eller spöt (*Saliua*)  
 term : Tarme eller induol (*Intestinum*)  
 lijktorn : Lygtorn eller naffuer bor vnder foden (*Clauus*).

Der er også eksempler på at Hingelbergs hele ordlyd adskiller sig fra 1559's. Nogle få eksempler blandt mange andre er

smoorder : Salffuede (*Christus*)  
 thet himmel tekn som flere än een stierna är vthi : Tegen paa himmelen (*Sydus*)  
 Moln : Sky (*Nubes*)  
 ijs drope : Huss egle (*Stiria*)  
 änkling/ fråskild : Æncke mand (*Viduus*)  
 krögare then ther äta och dricka haffuer til köps : Herbergærer den som Øl oc Mad haffuer falt (*Caupo*)  
 hettotimp eller stola : Hette Strud (*Orarium*)  
 likmadk : Ligorm eller krafft (*Cancer*).

En del af dette materiale udgøres af svensk ordstof som Hingelbergs begrænsede

<sup>10</sup> Udgaven 1574 – der er trykt i Rostock – har et layout der afviger fra 1559. Fragmentet af 1545 har jeg endnu ikke haft lejlighed til at undersøge.



sprogforførmelse har set som for svensk. Det skulle altså være muligt at sige noget interessant om forholdet mellem svensk og dansk på denne baggrund.

Hingelbergs ordbog er den mindste af de fire. Og i modsætning til de andre er den også meget sparsom med sine udenværker, men har dog til sidst et latinsk kapitelregister og den omtalte efterskrift. Artikelstrukturen er meget nøgen: et latinsk lemma og en dansk ækvivalent. Af og til dog også en udtaleangivelse i form af et "p.p." eller "p.c." for "penultima producta" og "penultima correpta". Disse forkortelser opløses allerforrest i ordbogen.<sup>11</sup>

#### 2.4 Mads Pors

Mads Pors (d. 1618) var efter uddannelse i Tyskland ansat som rektor i Ribe. Og hans ordbog, *De nomenclaturis Romanis*, har i den moderne udgave fået genremærketat "skolebog" i undertitlen. Af *Dansk Biografisk Leksikon* fremgår det at han har været en lidenskabelig mand der til stadighed lå i proces med sin slægt, blev afskediget på grund af usømmelig optræden m.m. (DBL<sup>3</sup> 11: 453).

Bogen adskiller sig på flere måder fra resten. For det første har den en helt anderledes makrostruktur; den er opdelt i fire bøger, og den sidste af dem er igen opdelt efter de ni musen. Denne sidste strukturingsvis er jeg endnu ikke stødt på andre steder. Den information der gives, viser at også med hensyn til mikrostrukturen er Pors' ordbog af en helt anden type end de tre andre. Alene de hyppige henvisninger til klassiske autoriteter viser at artiklerne i Pors' ordbog har et andet sigte end de tre mere elementære skolebøger. Bogen der er trykt i Frankfurt am Main i det feyerabendiske officin, er en encyklopædi over den romerske oldtid iblandet en god del kristendom. Et andet påfaldende træk ved Pors' ordbog er at den stort set kun indeholder substantiver og var tænkt suppleret af en ordbog over verberne (hvoraf ingen spor dog kendes).

Som i de andre ordbøger er også Pors' første indgang "*Deus* .. gud". Og hele hans første bog (ca. 10 store, tospaltede sider, 150 artikler) omhandler religiøse sager: forskellige betegnelser for Gud, for onde ånder (heriblandt de antikke guder), om religion, om kirkelige embeder, kirker og om helligdage. Det både minder om og adskiller sig fra de øvrige ordbøger. Også de begynder med *Deus*, men deres behandling er meget mindre omfattende. Tursens første kapitel er på seks små sider (136 artikler), Smiths på to (26 artikler), Hingelbergs på ca. 9 små sider (157 artikler). Disse tal viser at Pors giver en langt mere tilbundsående behandling i sine artikler. Og hans artikler rummer da også flere elementer end hvad vi finder i de andre ordbøger, de latinske opslagsord suppleres med latinske forklaringer, en eller flere danske modsvarigheder, og i omtrent hver artikel med græske. Klassikercitater og henvisninger vha. forkortelser som *Cic.* og *Plavt.* møder vi flere gange pr. side.

En anden væsentlig forskel mellem ordbøgernes indledende kapitler er at

<sup>11</sup> Også i Tursens og Smiths ordbøger finder vi disse prosodiske oplysninger, men mindre udbredt.

Tursen, Hingelberg og Smith (i sit kapitel 2), omhandler ikke blot gudelige sager, men også "himmelske" i bred forstand. Dvs. at Gud og månen, stjernebilleder og vejrfænomener som hagl og vindretninger behandles sammen. Helt anderledes er det grebet an hos Pors. Her er astronomien og meteorologien anført i den sidste bog (under muserne Urania og Polymnia). Og ikke nok med det – disse to emner omtales som "Meteorologicis" og "Astronomicis". Det er ganske vist på latin og ikke i funktion som fremmed- eller låneord i dansken. Men det viser en helt anden tilgang til stoffet end i de tidligere bøger. Denne forskel er nok snarere udtryk for forskellige aftagergrupper, end for en tidsmæssig bevægelse væk fra den middelalderlige verdensopfattelse henimod en mere moderne, empirisk funderet. Pors' bog er med hensyn til udstyr og indhold af en helt anden kaliber end de tre andre bøger.

Flere af ordbøgerne indeholder grammatiske oplysninger og udtalehjælp til de latinske opslagsord. Men Pors giver stedvis sproglige oplysninger også om de danske ækvivalenter. Under opslagsordet *Gallus* skriver han fx "Hane. *Cimbris* Kock". Det vidner om en interesse for modersmålet og ikke blot for antikviteterne. Både Smith og Pors har latinske registre bagerst, men Pors er ene om at have et tilsvarende dansk. Også det vidner jo om interesse for det danske sprog.

### 3 Hvad vil jeg gøre med ordbøgerne?

Den foreløbige titel på mit projekt er "Begrebssystemerne i renæssancens danske ordbøger", så jeg bliver nødt til at undersøge hvilke begrebssystemer disse ordbøger har, hvordan de præsenterer verden. Her er jeg navnlig inspireret af den tyske anglist Werner Hüllen der forsøger at læse ordbøger som sammenhængende tekster (især Hüllen 1999). Men denne del af mit projekt er jeg ikke kommet langt nok med til at kunne meddele andet end hvad jeg netop har sagt om Pors' anderledes strukturering.

En anden interesse drejer sig om ordforrådet. Det vil dels sige en overordnet undersøgelse af hvilke ord de enkelte ordbøger indeholder, dels af hvordan de forholder sig til latinen. Hvad stiller de fx op med latinske synonymer? Bliver de sammenholdt med *idem*, eller differentieres der også mellem synonyme danske modsvarigheder? Dette og andre spørgsmål lader sig imidlertid bedre belyse når det omtalte projekt *Renæssancens Sprog i Danmark* har færdiggjort sin ordbogsbase (jf. ndf.). Som en smagsprøve har jeg derfor foreløbig udvalgt at sammenligne ordstoffet i et kapitel der findes i alle bøgerne, nemlig det om fugle. Jeg har undersøgt hvilke fugle der optræder hvor, og hvordan de samme latinske ord oversættes i de enkelte ordbøger. Det er fælles for bøgerne at fuglekapitlet er placeret relativt sent. Det tyder på at dette område har stået mindre centralt end fx mennesket og byen. Hovedparten af kapitlerne består af en opremsning af fuglenavne, men ofte indeholder de også anatomiske betegnelser el. lign. Smith har udskilt et særligt kapitel til det: "Om fuglenis lemmer". Hos

Pors er der også knyttet et ekstra afsnit til fuglekapitlet, men det er om flyvende insekter.

Tre af de mest interessante detailforskelle er følgende:

*Fenix*

Tursen: “den eniste Fugel aff det slau i Arabia”

Smith: “Er en eniste Ful i verden aff det slau”

Pors: “Phenix”.

Hingelberg kender ikke ordet, og hos Pors er de encyklopædiske oplysninger anbragt i latinen. Tursens modsvarighed til *Fenix*, “den eniste Fugel aff det slau i Arabia”, findes allerede hos Brack: “*Fenix*. eyniger vogel. in arabia” (jf. Pleuger 2005: 274), men Bracks interpunktion kan tilmed forklare Tursens ordlyd (hvis man forestiller sig at Tursen har læst henover Bracks punktum). Oplysningen om føniksens arabiske habitat findes dog allerede hos Isidor (XII.7.22): *phoenix Arabiae avis*.

*Onocrotalus*

Smith: “En ful saa stor som en Suan/ haffuer en Pose vnder hagen”

Pors: “Rhørdrum”

Heller ikke dette ord findes hos Hingelberg. Tursen og Pors har “*Pillicanus Pellikan*”, men Smith kender tilsyneladende ikke denne fugl. Pors’ identifikation af *Onocrotalus* med rørdrummen er en almindelig sammenblanding der skriver sig fra middelalderen og bl.a. kommer til udtryk i tidlige bibeloversættelser, se KLNLM under *rørdrom*.

*Vpupa*

Tursen: “Vibe”

Smith: “Herful”

Pors: “Vibe”

At sammenholde den nordiske betegnelse “vibe” med lat. *upupa* er udbredt i middelalderen og renæssancen, men forkert, idet latinen betegner hærfulgen (jf. KLNLM under *Vipa*).<sup>12</sup>

Endelig er jeg særligt interesseret i forholdet mellem dansk og svensk ordstof sådan som det gestalttes i Hingelbergs ordbog.

#### 4 Renæssancens Sprog i Danmark – en kort præsentation

Titlen på projektet *Renæssancens Sprog i Danmark* udnytter at “sprog” har umark-eret pluralis. Det er nemlig ikke dansk eller latin der er fokus på, men forholdet

<sup>12</sup> At den eksotiske hærfulg ikke har været ukendt i Danmark i oldtiden vidner den gængse tolkning af Herfølge som en ia-afledning, opr. \*Hærfyghli, om.

mellem de to, de to vigtigste sprog i Danmark på det tidspunkt.<sup>13</sup> Projektet der er støttet af Forskningsrådet for Kultur og Kommunikation, huses af Det Danske Sprog- og Litteraturselskab. De to forskere der er knyttet til projektet, Peter Zeeberg og Jonathan Adams, udforsker hhv. det dansk-latinske blandingssprog og latinens indflydelse på periodens danske orddannelse. Grundlaget for dette arbejde er to baser der er under opbygning. Den ene er et opslagsværk, en ordbogsbase. Den indeholder i opmærket form de fire omtalte ordbøger plus de to ældre: Christiern Pedersens fra 1510 og Henrik Smiths fra 1520, samt Poul Jensen Colding fra 1626. Den anden base er en tekstbase der skal indeholde forskellige danske og latinske tekster fra tiden. Begge baser bliver tilgængelige på nettet, primo 2009. På projektets hjemmeside kan man læse meget mere om baserne og delprojekterne: [dsl.dk](http://dsl.dk) og [renaessancesprog.dk](http://renaessancesprog.dk)

## Litteratur

### Ordbøger

- Brack, Wenzeslaus: *Vocabularius rerum*, se Pleuger 2005.
- Colding, Poul Jensen: *Etymologicum Latinum* (1622).
- Colding, Poul Jensen: *Dictionarium Herlovianum* (1626).
- H[ingelberg], P[oul] N[ielsen]: *Vocabulorum variorum expositio* (1576) [= Det 16. århundredes danske vokabularier, V. København: UJDS/C.A. Reitzel 1995].
- Jensen & Goldschmidt = Jensen, J.Th. & M.J. Goldschmidt 1920: *Latinsk-dansk Ordbog*. 2. udg. København: Gyldendalske Boghandel/Nordisk Forlag.
- Pedersen, Christiern: *Vocabularium ad usum dacorum* (1510) [= Det 16. århundredes danske vokabularier, I. København: UJDS/C.A. Reitzel 1973].
- Pedersen, Christiern: *Vocabularium ad usum dacorum* (1518).
- Pors, Mads: *De nomenclaturis Romanis* (1594) [= Det 16. århundredes danske vokabularier, VI. København: UJDS/C.A. Reitzel 1995].
- Smith, Henrik: *Hortulus synonymorum* (1520) [= Det 16. århundredes danske vokabularier, II. København: UJDS/C.A. Reitzel 1974].
- Smith, Henrik: *Libellus vocum Latinarum* (1563) [= Det 16. århundredes danske vokabularier, IV. København: UJDS/C.A. Reitzel 1991].
- Tursen, Jon: *Vocabularius rerum* (1561) [= Det 16. århundredes danske vokabularier, III. København: UJDS/C.A. Reitzel 1975].
- Variarum rerum vocabula* (1538), se Andersson 1890.
- Variarum rerum vocabula* (1579), se Collijn 1925.
- Variarum rerum vocabula* (1559), kopi af unikum i Universitetsbiblioteket i Uppsala.

13 Det tredje hovedsprog i renæssancens Danmark, nemlig tysk, inddrages derimod ikke.

*Anden litteratur*

- Andersson, Aksel 1890 (udg.): *Variarum rerum vocabula cum sueca interpretatione*. Uppsala.
- Andersson, Henrik; Ebba Hjorth og Merete K. Jørgensen 1997: Man skal ikke give bagerbørn hvedebrød. En artikel om leksikografisk brødarbejde, i: *Ord, Sprog og artige Dict. Et overblik og 28 indblik 1500–1700. Festskrift til Poul Lindegård Hjorth*. København: UJDS/C.A. Reitzel, s. 101–114.
- Collijn, Isak 1925 (udg.): *Variarum rerum vocabula cum sueca interpretatione*. Stockholm [= Faksimileopplagor av äldre svenska tryck, VI].
- Collijn, Isak 1928: En hittills okänd upplaga av *Variarum rerum vocabula med Formulae puerilium colloquiorum* Stockholm 1559, i: *Nordisk tidskrift för bok- och biblioteksväsen* XV, s. 1–39.
- DBL<sup>3</sup> = *Dansk Biografisk Leksikon*, 3. udg. København: Gyldendal. 1979–1984.
- Ehrling, Sigfrid 1929–30: Anmärkningar till Kalkars Ordbog til det ældre danske Sprog, i: *Acta Philologica Scandinavica* IV, s. 26–39.
- Hüllen, Werner 1999: *English Dictionaries 800-1700. The Topical Tradition*. Oxford: Clarendon Press.
- Jørgensen, Merete K. 2000: Vocabularia ad usum dacorum – og mange andre, i: *Lexico-Nordica* 7, s. 91–100.
- KLNM = *Kulturhistorisk Leksikon for Nordisk Middelalder I-XII*. København 1956–1978.
- Lindberg, Sten G. 1960: Om prinsuppföstran och tryckeriteknik i Stockholm 1543-1545, i: *Lychnos* 1959, s. 209-217.
- Müller, Peter O. 2001: *Deutsche Lexiographie des 16. Jahrhunderts. Konzeptionen und Funktionen frühneuzeitlicher Wörterbücher*. Tübingen: Niemeyer.
- Pleuger, Nina 2005: *Der Vocabularius rerum von Wenzeslaus Brack. Untersuchung und Edition eines spätmittelalterlichen Kompendiums*. Berlin/New York: Walter de Gruyter.
- Skautrup, Peter 1947: *Det Danske Sprogs Historie*. Bd. 2. København: DSL/Gyldendalske Boghandel/Nordisk Forlag.
- Weibull, Lauritz 1897: Henrik Smith, i: *Skånska samlingar* 1894–1895. IV,1.

Simon Skovgaard Boeck  
 ph.d.-stipendiat, mag.art., f. 1977  
 Institut for Nordiske Studier og Sprogvidenskab  
 Københavns Universitet  
 Njalsgade 120  
 DK-2300 København S  
 ss@dsl.dk



EMMA SKÖLDBERG

## Från *vrångstrupen* till *fotabjället*

### Om presentationen av unika konstituenten i Svensk ordbok utgiven av Svenska Akademien

The paper presents a study of how unique constituents, by which I mean independent lexemes that occur primarily in a specific phrase, are treated in *Nationalencyklopedins ordbok* (1995–96). I explore aspects of unique constituents including their number, grammatical form when they constitute lemmas, and whether they are possible to define. I also compare how similar words are presented in *Den Danske Ordbog* (2003–2005). The results will be used to achieve a more stringent presentation of the relevant lexemes in the forthcoming *Svensk ordbok utgiven av Svenska Akademien*, which is scheduled for release in 2009.

#### 1. Inledning

I svenskan finns det fristående lexikala enheter som företrädesvis förekommer i en viss fras. Ett exempel är substantivet *vrångstrupe*. I Språkbanken, på cirka 100 miljoner ord, används ordet i princip alltid i bestämd form singularis och i uttrycket *sätta ngt i vrångstrupen* ('sätta ngt i halsen'). Ett annat exempel är substantivet *fotabjälle* som nästan uteslutande uppträder inom frasen *från hjässan till fotabjället* ('från det översta av huvudet till fötterna') i ovan nämnda material. Men den här typen av konstruktionsbundna lexikala enheter är självfallet inte unik för svenskan. Ett exempel från danskan är *hornlygte* som är starkt förknippat med uttrycket *en fis i en hornlygte* (jfr *Den Danske Ordbog* 2003–2005). För exempel på sådana ord i bl.a. engelska, tyska, nederländska, se t.ex. Dobrovolskij (1988), Fleischer (1997) och Moon (1998).

I den här artikeln resonerar jag kring den här typen av lexikala enheter och hur vi inom projektet Svenska Akademiens samtidsordböcker arbetar mot att ge dem en mer stringent presentation i kommande *Svensk ordbok utgiven av Svenska Akademien* (f.d. *Nationalencyklopedins ordbok* 1995–96). Jag jämför också med hur liknande ord behandlas i *Den Danske Ordbok* (2003–2005). Bland annat diskuterar jag deras antal i den kommande ordboken och huruvida de utgör lemmen eller lexem (dvs. betydelsemoment). Jag resonerar också kring vilken grammatisk form de ska ha när de utgör lemmen och hur det i sin tur påverkar angivelser om aktuella böjningsformer inne i ordboksartiklarna. I viss mån behandlar jag också hur den här typen av lexikala enheter kan definieras. Artikeln avslutas med en kortfattad diskussion.

## 2. Kortfattad forskningsöversikt

Självständiga lexem som endast uppträder i ett eller några få flerledade uttryck är föga uppmärksammade inom svensk språkvetenskap och mig veterligen finns det ingen svensk etablerad benämning på den typen av ord. Emellertid kallar Fillmore et. al (1988: 506) dem för *unfamiliar pieces* och Nenonen och Niemi (1999) för *idiomatic isolates*. I tysk litteratur benämns de bl.a. *unikale Komponenten* (Burger 1998: 12 m.fl.) och i spanska texter *palabras diacríticas* (Corpas Pastor 1996: 115). (Jfr även termen *unikal* i förteckningen över fraseologisk terminologi i Brink 2006: 50.)<sup>1</sup> Själv har jag valt arbetsnamnet *unika konstituent*er som beteckning för den här typen av fristående och konstruktionsbundna lexikala enheter.

I vissa arbeten, t.ex. Dobrovolskij & Piirainen (1994) samt Fleischer (1997: 40–42) diskuteras de problem som uppträder när man ska avgränsa kategorin unika konstituent. Ruiz Gurillo (1998: 12, 19) m.fl. koncentrerar sig på vilka typer av unika konstituent det finns. Till denna litteratur har också jag själv bidragit (Sköldberg 2007). En typ är ord med suffix som återgår på äldre kasusformer (jfr Teleman et al. 1999, del 3: 649–650). Till dessa hör bl.a. *andanom* i uttrycket *i andanom* 'för sitt inre'. Dessa konstituent är speciella eftersom de utgörs av en viss ordform hos ord som i sig inte är unika. Formen *andanom* är bestämd dativform av substantivet *ande*, som används i en rad olika kontexter och sålunda inte kan klassificeras som unikt. Kopplingen mellan *andanom* och *ande* torde dock vara okänd för många språkbrukare av idag vilket motiverar att *andanom* utgör ett eget lemma i en samtidsordbok. En annan typ av unika konstituent är lånord från klassiska språk som latin. Ett sådant är adjektivet *magnificus* som uppträder i uttrycket *rector magnificus*, en titel för universitetsrektor. En tredje typ är sådana som uppkommit genom rim och allitteration. Här kan adverbet *huller* i uttrycket *huller om buller* 'utan någon ordning' anföras.

Slutligen resonerar t.ex. Moon (1998: 78–80) kring de fraser som unika konstituent uppträder i. I sin typologi över engelska fasta uttryck har hon avgränsat en särskild grupp, *cranberry collocations*, som innehåller den här typen av lexikala enheter. Svenska unika konstituent återfinns i olika typer av mer eller mindre lexikaliserade fraser. Ett antal sådana utgör ett eller båda led i s.k. *ordpar* (Bendz 1965: 17). Det gäller t.ex. *dus* i prepositionsfrasen *i sus och dus* 'under ständigt festande'. Andra svenska unika konstituent återfinns i mer typiska idiom, t.ex. *hårdhandskarna* i *ta i med hårdhandskarna* 'tillgripa stränga åtgärder'. Vidare kan unika konstituent uppträda i liknelser, som exempelvis *korvspad* i *vara klart som korvspad* 'vara helt klart'. De återfinns också i ordspråk. Exempel på den sista typen är såväl *hjärterum* som *stjärterum* i *finns det hjärterum så finns det stjärterum* med betydelsen 'är man goda vänner så räcker platsen till även om det är trångt'.

1 Jfr också beskrivningen av svenskans *restmorfem*, dvs. bundna rotmorfem som "endast återfinns i ett eller några få avledda eller sammansatta ord" i Teleman et al. (1999, del 1: 218). Ett exempel på ett sådant morfem är *ling-* i *lingon*.



(Se vidare om svenska unika konstituenters ordklassstillhörighet, morfologiska egenskaper och ursprung i Sköldberg 2007.)

### 3. Material

Genom mitt arbete inom det lexikografiska projektet Svenska Akademiens samtidsordböcker vid Institutionen för svenska språket, Göteborgs universitet, har jag tillgång till en databas som ursprungligen är baserad på innehållet i *Nationencyklopedins ordbok* (1995–96; hädanefter NEO) och som nu bearbetas inför den kommande *Svensk ordbok utgiven av Svenska Akademien* (hädanefter SO). I NEO finns drygt 600 lemman som anges förekomma i ett speciellt uttryck och därigenom torde kunna klassificeras som unika konstituent. Den kombinatoriska begränsningen hos de aktuella lemmarna markeras genom olika formella kommentarer, framför allt formuleringen ”i frasen”. Men i NEO finns också andra kommentarer med liknande funktion, t.ex. ”nästan enbart i frasen”, ”endast i ett fåtal uttryck” och ”mest i ett uttryck”. Samtliga sådana uppslagsord har jag excerperat. Bland de excerperade lemmarna finns bland annat **vrångstrupen** (i artikeln anges samtliga lemman med fet stil). Lemmat presenteras på följande vis:

- (1) **vrång`strupen** subst., best. f.  
 ORDLED: *vrång-strup-en*  
 • i frasen *sätta/få ngt i ~* av misstag få ned ngt i struphuvudet eller luftstrupen med åtföljande host- el. kvävningssattack  
 HIST.: sedan 1538; till **vrång** i bet. ’oriktig’

Vidare har jag en förteckning över de närmare 500 ord i *Den Danske Ordbog* (2003–2005; hädanefter DDO) som kombineras med kommentarer som ”kun i uttrycket”, ”kun i uttrykkene”, ”kun i förbindelsen” och ”kun i forbindelserne”.<sup>2</sup> Ett lemma som följs av en sådan kommentar är **bollemælk**:

- (2) **bollemælk** sb. fk.  
*kun i udtrykket*  
**ikke stikke op for bollemælk** ikke være bange for noget som helst □  
*De skulle nok få se at, de der håndværkmestre, han stak ikke op for bollemælk*  
 JaAamu89.

Många av de svenska lexikala enheter som står i fokus här har direkta motsvarigheter i danskan, vilket gör en jämförelse mellan NEO/SO och DDO relevant. Nedan återges exempel på svenska unika konstituent som har uppenbara motsvarigheter i danskan. Först anges lemmarna och sedan de aktuella fraserna.

- (3) **datt** (*ditt och datt*)                      **dat** (*dit og dat*)  
**fåraklæder** (*en ulv i fåraklæder*)      **fåreklæder** (*ulv i fåreklæder*)

<sup>2</sup> Ett varmt tack till Henrik Lorentzen på Det Danske Sprog- og Litteraturselskab som har låtit mig få ta del av detta material.

<b>giftastankar</b> ( <i>gå i giftastankar</i> )	<b>giftetanke</b> ( <i>gå i giftetanker</i> )
<b>gordisk</b> ( <i>gordisk knut</i> )	<b>gordisk</b> ( <i>gordisk knude</i> )
<b>tåls</b> ( <i>ge sig till tåls</i> )	<b>tåls</b> ( <i>slå sig til tåls</i> )

I artikeln hänvisar jag också till exempel ur olika textkorporusar. För svenskans del har jag använt den redan nämnda Språkbanken vid Institutionen för svenska språket, Göteborgs universitet (<<http://spraakbanken.gu.se/>>) samt Presstext, som är Nordens största pressarkiv (<<http://www.presstext.se/>>). Jag har även granskat andra texter på Internet som jag funnit genom sökmotorn Google (<<http://www.google.se/>>). För att studera danskt språkbruk har Korpus 2000 utnyttjats (<<http://korpus.dsl.dk/korpus2000/indgang.php>>).

#### 4. Unika konstituentier – finns de?

En viktig fråga är om det över huvud taget existerar fristående lexikala enheter som är så konstruktionsbundna att de kan betraktas som unika konstituentier i snäv bemärkelse. Vissa ord, som t.ex. det redan nämnda *vrångstrupe*, används som sagt på ett mycket enhetligt sätt i Språkbanken, dvs. alltid i bestämd form singularis och i frasen *i vrångstrupen*. Nästan alltid föregås frasen av verbet *sätta*. Om man endast tar hänsyn till bruket i Språkbanken torde det sålunda vara berättigat med en kommentar av slaget ”i frasen” i den kommande ordboken. Nedan återges några exempel ur korpusen (samtliga kursiveringar i exemplen är mina egna).

- (4) Falska nyheter i tidningarna lockar fram både skratt och vrede hos läsarna. Det var säkert ett antal läsare i Göteborg som häromåret *satte* frukostkaffet *i vrångstrupen* när de läste sin morgontidning. På bästa nyhetsplats fanns en artikel om att den som har utträtt ur Svenska kyrkan inte kan räkna med att få ledigt från jobbet på söndagar och helgdagar. (GP04)
- (5) Åtskilliga *satte* nog middagskaffet *i vrångstrupen* då de slog på radions luncheko onsdagen den 28 oktober 1981. En ubåt hade snirklat sig långt in i Karlskrona skärgård och satt ohjälpligt på grund mitt i ett svenskt militärt skyddsområde. En ubåt från den sovjetiska krigsflottan! (p06)

Ofta är det en person som sätter kaffet i vrångstrupen när han eller hon tar del av en (för honom eller henne) chockerande nyhet. I det mer omfattande textmaterialet PressText används uttrycket på liknande sätt, men det uppträder också i andra fraser och i obestämd form:

- (6) Sellers har tre olika roller i Kubricks oerhört svarta komedi om det kalla krigets paranoia. Bäst är han som titelfiguren, en rullstolsburen före detta nazistisk forskare vars armprotes hakar upp sig i avslöjande poser. Möjligheten att det skulle kunna ha hänt på riktigt gör att skrattet *fastnar i vrångstrupen*. (PressText)

- (7) En just avsmakad bildtext om Aretha Franklin *sitter* fortfarande *i min vrångstrupe*: ”Hon sjunger med makalös auktoritet, trots att hon fyller femtio år nästa år.” Trots? Varför inte lika gärna ”eftersom”? Eller rentav ”tack vare”? (PressText)

Föga förvånande är det också lätt att finna mindre typiska användningar av substantivet *vrångstrupe* med hjälp av Google. Följande citat (hämtade den 15 augusti 2007) har det gemensamma innehållet att kaffe (förutsätts) sättas i vrångstrupen på grund av någon chockerande nyhet. Men vad gäller formen kan man se att det aktuella ordet ingår i en rad olika konstruktioner. Som synes används det i singularis, pluralis, bestämd och obestämd form:

- (8) ...chocken gjorde att kaffet *blockerade min vrångstrupe*. <<http://namedrop.se/tankefrihet/>>
- (9) Det var nog inte bara på DIs redaktion som kaffetårar *fastnade i vrångstrupar* när Svea hovrätts dom [...] kom på torsdagen. <<http://di.se/Nyheter/?page=%2fAvdelningar%2fArtikel.aspx%3fO%3dIndex%26ArticleID%3d2002%5c04%5c12%5c46811>>
- (10) Varning för kaffe *i vrångstrupen!* <<http://www.canit.se/~per/scen.htm>>
- (11) ” Vilket ramaskri det skulle bli!!!! vad mycke kaffe, som skulle *sättas i vrångstruparna*.” <<http://www.debatten.fi/index.php?topic=1306.60>>

Ordet ingår sålunda i konstruktioner av mycket varierande slag.

Även andra ord, t.ex. *fårakläder* och *kålsupare*, som starkt förknippas med en viss form och en viss kontext, används på ett enhetligt sätt i Språkbanken. Däremot varierar de mer i PressText och i andra texter som man finner på Internet. Det är emellertid komplicerat med språkbruket i dessa senare texter eftersom de är av så olika kvalitet.

Jag har dock kommit fram till att endast ett fåtal av de lemman som följs av kommentaren ”i frasen” i NEO är helt konstruktionsbundna i större svenska textsamlingar. Givetvis kan varken NEO eller den kommande SO återge alla möjliga användningar av ett ord utan de strävar efter att beskriva ett mer typiskt språkbruk. Det är dock vanskligt att i ordboken använda så precisa kommentarer som ”i frasen”, vilken lätt tolkas som att ordet ifråga endast används i en viss fras. Det kan vara lämpligt att uttrycka sig på ett mer försiktigt sätt och i den kommande SO överväger vi därför att mer genomgående använda kommentaren ”främst i uttrycket”.

Förutsättningarna för DDO är annorlunda eftersom den ordboken är baserad på en sluten korpus (se DDO 2003: 53ff.). Det finns lexikala enheter i Korpus 2000, t.ex. *morgenkvisten*, som endast förekommer i en viss form och i en viss konstruktion. När enheter som dessa beskrivs i ordboken kan man givetvis in-

foga kommentarer som ”kun i uttrycket” etc. Men liksom i NEO används kommentarer av detta slag kanske väl frikostigt. Som exempel kan lemmat **krigsstien** nämnas.

- (12) **krigsstien** sb. fk. bf.

[-,sdɪʔən]

*kun i uttrycket*

**være på krigsstien** gå el. være i krig • om nordamerikanske indianere □ *apacherne er gået på krigsstien som følge af den amerikanske regerings brudte fredsløfter* Pol98 • (ofø) opføre sig aggressivt og søge konfrontation i en bestemt sag □ (varierer i form) *Så sent som sidste sommer gik kansler Helmut Kohl på krigsstien, fordi han mente, at Kommissionen blandede sig i for meget* EU-kommissionens hj.side 1999.

Artikeln är mycket utförlig, och informationen om uttryckets semantik är klar. Men jag ställer mig frågande till de formangivelser som finns i textavsnittet. Som utgångspunkt tar jag det språkbruk som är belagt i Korpus 2000. När man söker på strängen ”krigsti\*” i korpusen, visas sammanlagt nio belägg. I fem av dessa föregås *på krigsstien* av verbet *være*, men i övriga fall föregås prepositionsfrasen av andra verb, t.ex. *bringe* och *drage*. Substantivet förekommer också i obestämd form (i ”den moderne køns-debats krigssti”). För övrigt kan man se att ordboksartikeln innehåller två exempel med formerna ”er gået på krigsstien och ”gick [...] på krigsstien”. I verket konstateras också att när *være på krigsstien* används överfört varierar uttrycket till formen. Ordboksartikeln ger sålunda till viss del motsägande information. Sammanfattningsvis anser jag att den semantiska analys som presenteras är god och att exemplen är väl valda, men att kommentaren ”kun i uttrycket” inte är riktigt rättvisande.

#### 4. Antalet unika konstituenten i SO

Vi har sålunda gått igenom alla unika konstituenten i NEO och det har resulterat i att deras antal i kommande SO är mindre än i föregångaren. För att visa hur vi har arbetat återger jag hur lemmat **hästvæg** presenteras i NEO respektive SO:

- (13) **häs`tvæg** subst., ingen böjning

ORDLED: häst–væg

• i frasen *något/någotning i* ~ ngt som gör ett mycket starkt intryck t.ex. genom att vara mycket stort el. kraftigt; ibl. med negativ biton

HIST.: åtm. sedan 1930-talet; till **væg 4**

- (14) **häs`tvæg** subst., ~en, ~ar [...]

• væg för ridning och hästransporter *gamla ~ar från 1800-talet har iordningställt och används nu som vandringsleder*

\* **något i hästvæg** ngt som gör ett mycket starkt intryck t.ex. genom att

vara mycket stort el. kraftigt: *årets nyårsfyrverkeri förväntas bli något i hästväg [...]*

Sammansättningen *hästväg* är i hög grad genomskinlig (jfr Svensén 2004: 81) och det kan tyckas att lexemet inte hör hemma i NEO eftersom ordboksförfattarna, för att begränsa omfånget, har fått vara restriktiva vid urvalet av lemman. Eftersom lexemet ingår i ett idiom, *något/någonting i hästväg*, behandlas det ändå i ordboken. I SO väljer vi att också ange betydelsen 'väg för ridning och hästtransporter' och kompletterar denna med ett syntaktiskt exempel, eftersom lexemet används både bokstavligt och bildligt i korpusarna, även om det huvudsakligen uppträder som en del av idiomet. Därmed blir den här typen av lexikala enheter mer konsekvent behandlade i den kommande ordboken. Som en jämförelse kan presentationen av olika typer av löv i NEO/SO anföras. Det finns många sammansättningar som betecknar olika slags löv (*björklöv, eklöv, lindlöv* etc.) och alla dessa sammansättningar är inte med i ordböckerna eftersom de betraktas som genomskinliga. Ett undantag är *asplöv* som behandlas eftersom substantivet ingår i liknelsen *darra som ett asplöv* 'darra häftigt vanligen på grund av rädsla'. Först anges sålunda den kompositionella betydelsen 'blad av asp' och därefter det fasta uttrycket.

Man kan invända att ordboksanvändaren inte är särskilt betjänt av de bokstavliga och något triviala betydelseerna hos sammansättningar som *hästväg* och *asplöv*. De är som sagt föga belagda – åtminstone i skrift. De bildliga betydelseerna hos orden kan dock lyftas fram i ordboksartiklarna med hjälp av kommentarer om typisk användning.

I några fall har vi dock bortsett från den kompositionella och genomskinliga betydelsen hos sammansättningen, men angivit en bildlig sådan. Som exempel väljer jag lemmat **guldkalv**. Detta definieras inte i NEO utan i ordboken står det bara att lemmat förekommer i en viss fras, nämligen *dansen kring guldkalven*:

- (15) **gul'dkalv** subst. ~en ~ar  
 ORDLED: *guld-kalv-en*  
 • i frasen *dansen kring ~en* dyrkan av och begär efter pengar  
 HIST.: före 1520; fornsv. *guldkalver*

I de utnyttjade textmaterialen används dock substantivet i olika sammanhang, men det rör sig ändå om den metaforiska betydelsen hos ordet och inte om en konkret kalv tillverkad av guld. I SO väljer vi sålunda att först ge en definition, dvs. 'ngn eller ngt som ger gott ekonomiskt utbyte' följt av ett exempel som illustrerar denna betydelse. Först därefter presenteras idiomet (i form av en hänvisning):

- (16) **gul'dkalv** subst. ~en ~ar  
 ORDLED: *guld-kalv-en*

- ngn eller ngt som ger gott ekonomiskt utbyte *den nya spelkonsolen har blivit Nintendos verkliga ~*
- \* **dansen kring ~en** se under **dans**
- HIST.: sedan 1400–25 (Heliga Birgittas uppenbarelser): fornsv. *guld-kalver*

Av exemplet framgår också att vi inom projektet har arbetat vidare med datering av ord som är belagda före år 1520 och att SO även kommer att ge information om i vilken text som förstabelägget återfinns.

## 5. Unika konstituenten som lemman och lexem

Ett stort antal unika konstituenten är egna lemman i NEO/SO (se vidare Svensén 2004: 116ff. om lemmaansättning). Detta gäller t.ex. substantivet *jota* i uttrycket *inte ett jota* 'inte någonting' och adjektivet *lidnersk* i uttrycket *få en lidnersk knäpp* 'drabbas av plötslig klarsyn'. Andra exempel är adverbet *barbacka* (i *rida barbacka* 'rida utan sadel') och verbet *tråda* (i *tråda dansen* '(börja) dansa'). Några av dessa konstituenten är homografa med en eller flera lexikala enheter ur samma eller andra ordklasser. Det gäller t.ex. adverbet *splitter* som återfinns i frasen *splitter ny* ('aldeles ny') och substantivet *splitter* med betydelsen 'små vassa stycken av ngt hårt som slagits sönder'.

Antalet unika konstituenten som utgör egna lexem är mindre i den kommande ordboken än i föregångaren. För att förtydliga återger jag hur valda delar av lemmat *nickel* presenteras i NEO respektive SO:

- (17) **nickel** [...] subst. ~n, äv. nicklet  
 ORDLED: *nickl-et*  
 1 en silvervit, starkt glänsande, tånj- och smidbar metall som är motståndskraftig mot luft och vatten [...] *nickelallergi; nickellegering* HIST.: sedan 1754; kortform av *kopparnickel*, av ty. *Kupfernickel* [...]  
 2 i frasen (*jag har*) *inte ett/en nickel* (*jag har*) inga pengar alls  
 HIST.: sedan 1929; se **nickel 1**
- (18) **nickel** [...] subst. ~n, äv. nicklet [...]  
 • en silvervit, starkt glänsande, tånj- och smidbar metall som är motståndskraftig mot luft och vatten [...] *nickelallergi; nickellegering*  
 \* **inte ett nickel** inga pengar alls [...]  
 HIST.: sedan 1754; kortform av \**kopparnickel*, av ty. \**Kupfernickel* [...]

I SO har vi, med tanke på uttryckets semantik, placerat frasen *inte ett nickel* som idiom under lexemet/betydelsemomentet *nickel1*. Man kan emellertid ifrågasätta hur detta förfarande förhåller sig till den följda lemma-lexemmodellen, som innebär att lexikala enheter med olika stavning, uttal och böjning ska utgöra egna

artiklar (jfr NEO 1995: Inledning). Lexemet *nickel* i *inte ett nickel* är neutralt och följer sålunda inte helt det böjningsmönster som anges i artikelns huvud. Vi anser dock att ordboksanvändarna är betjänta av att finna idiomet *inte ett nickel* under lemmat **nickel** med tanke på idiomets metonymiska koppling till uppslagsordet. Eftersom neutrum anges som alternativform hos lemmat är lösningen inte helt regelvidrig. Man kan också invända att *inte ett nickel* borde placeras under ett eget lemma med tanke på att ämnesordet *nickel*, som behandlas i **nickel** knappast föregås av obestämd artikel. Enligt lemma-lexemmodellen är dock denna skillnad inte lemmaskiljande (jfr t.ex. lemmat **guld** som inkluderar användningar av slaget *hon har vunnit ett guld och två silver*).

## 6. Lemmats form och böjningsangivelser i artikeln

En annan fråga gäller den grammatiska form som en unik konstituent har när den utgör ett lemma. Om ett substantiv (nästan) bara används i plural, bör pluralformen fungera som lemmaform (Svensén 2004: 128). Om ett substantiv däremot (nästan) bara används i bestämd form, bör istället den bestämda formen vara lemmaform. Svensén tar också upp ord som enbart förekommer i idiom och där alltid står i bestämd form. Som exempel anförs uttrycket *vara i faggorna*. Enligt Svensén brukas inte obestämd form singularis *fagga* (ung. 'klädtrasa') längre och den hör i huvudsak hemma i en historisk ordbok.

Såväl NEO och SO som DDO följer i huvudsak den princip som företräds av Svensén och som beskrivs strax ovan. Nedan återges några substantiviska lemmen i NEO/SO och den grammatiska form de återfinns i. Utöver detta anges hänvisningar till aktuella lemmen i DDO.

- |      |   |                                   |
|------|---|-----------------------------------|
| (19) | <b>snålskjuts</b> (obestämd form singularis)  | (jfr DDO:s <b>bram</b> )          |
|      | <b>dövörat</b> (bestämd form singularis)      | (jfr DDO:s <b>firetoget</b> )     |
|      | <b>rövarhänder</b> (obestämd form pluralis)   | (jfr DDO:s <b>listefødder</b> )   |
|      | <b>hårdhandskarna</b> (bestämd form pluralis) | (jfr DDO:s <b>spenderbukser</b> ) |

Men NEO är inte helt konsekvent på denna punkt och därför har vi ändrat den grammatiska formen hos vissa lemmen inför SO. Detta gäller t.ex. lemmat **fotabjälle** (i frasen *från hjässan till fotabjället* 'från det översta av huvudet till fötterna') som i SO får bestämd form, dvs. **fotabjället**. På liknande har vi ersatt **målro** (i *hålla målron vid makt* 'hålla samtalet i gång') med **målron** och **geniknöl** (i *gnugga geniknölarna* 'göra en intellektuell ansträngning') med **geniknölarna**.

För danskans del råder stor överensstämmelse mellan bruket i Korpus 2000 och lemmats grammatiska form i DDO. Vad jag kan se finns det bara några få undantag och det är t.ex. lemmat **spenderbukser** med kommentaren "kun i uttrycket" *have spenderbukserne på*. Substantivet ifråga uppträder företrädesvis i bestämd form pluralis i korpusen. Vidare anger ordboken att det endast uppträder i en viss

konstruktion och om man hade varit helt konsekvent i verket hade det framgått av lemmats form.

Den lemmaform som väljs påverkar i sin tur böjningsangivelserna i artikeln. När ett visst ord (nästan) enbart förekommer i grundform och man valt denna form som lemmaform, följs den i NEO/SO av kommentaren ”ingen böjning”. Det gäller t.ex. lemmat som **kreti** (i *kreti och pleti* ’personer av blandad, huvudsakligen lägre, social ställning’) och **pin** (i *på pin kiv* ’bara för att retas’). I DDO har man valt en annan lösning. När lemmat står i grundform och ingen böjning är aktuell, ges ingen information om böjning alls i artikeln (se t.ex. **baglås** och **talefod**). Däremot ser det ut på liknande sätt NEO/SO respektive DDO när lemmat står i bestämd form och/eller i pluralis. Lemmat **bättringsvägen** har böjningsinformationen ”bestämd form”. Lemmat **låtanden** följs av kommentaren ”pluralis” och **arbetstagen** av ”bestämd form pluralis”. (Se även t.ex. DDO:s lemmat **baduljen** och **fåreklæder**.)

Men som jag redan visat varierar språkbruket, framför allt i de större korpusarna, och det påverkar givetvis också den böjningsinformation som ges. Ibland har vi för den kommande SO valt bestämd form som lemmaform, men det betyder inte att obestämd form inte används. Låt mig åter nämna substantivet *geniknölarna*. I SO har vi som sagt tagit fasta på lemmaformen **geniknölarna** och den följs av frasen *gnugga geniknölarna*. Men i en sjundedel av beläggen i Språkbanken används substantivet i obestämd form pluralis, t.ex. i ”nu gnuggar gubbarna sina geniknölar i styrelserummet” (DN 1987). Ett pronomen har infogats i uttrycket och därigenom är den obestämda formen på substantivet oundviklig. Återigen måste vi välja mellan att återge prototypisk användning och vara helt sanningsenliga vad gäller faktiskt bruk och dessa val vållar ofta en hel del huvudbry.

## 9. Unika konstituentier – kan de definieras?

Avslutningsvis vill jag också säga något om arbetet med att definiera unika konstituentier. Det är uppenbart att många lexikala enheter av detta slag är svåra att definiera, eftersom de ofta har en mycket vag betydelse (Bendz 1965: 19; Corpas Pastor 1996: 116). Detta medför att många av dem inte definieras i NEO eller SO. Det gäller t.ex. *bång* (i *med buller och bång* ’under uppseendeväckande former’), *pickalurven* (i *vara på pickalurven* ’vara berusad’), *gordisk* (i *gordisk knut* ’olöslig uppgift’) och *svejs* (i hälsningsfrasen *hej svejs!*). I dessa fall är det endast fraserna som enheterna ingår i som får en betydelsebeskrivning. I några fall är betydelsen möjlig att härleda men inte särskilt relevant att ange i ordboken. Det gäller t.ex. svenskans *dövörat* (i *slå dövörat till*) som skulle kunna få definitionen ’det öra som har kraftig nedsatt hörsel’. I DDO har man arbetat efter samma principer och exempelvis följande unika konstituentier saknar en definition: *fåraklæder*, *pjåketarm* och *halvkvædet*.

Det finns dock unika konstituentier som har en mer urskiljbar betydelse och



som därför definieras i NEO/SO. Två exempel är *söcken* ('vardag') i frasen *helg och söcken* och *inomhusbruk* ('användning inomhus') i frasen *för inomhusbruk*. I NEO har unika konstituenten med kommentaren "i frasen" inte någon betydelsebeskrivning. Unika konstituenten som följs av formuleringen "i ett uttr." etc. definieras emellertid. I SO gör vi ingen sådan åtskillnad, dvs. förekomsten av definition ska inte styra kommentarens utseende.

## 10. Slutord

I den här artikeln presenterar jag hur unika konstituenten, dvs. fristående lexikala enheter som framför allt används i en viss fras, behandlas i *Nationalencyklopedins ordbok*. Vidare diskuterar jag hur vi inom projektet Svenska Akademiens samtidsordböcker vid Institutionen för svenska språket, Göteborgs universitet, arbetar med den här typen av lexikala enheter i kommande *Svensk ordbok utgiven av Svenska Akademien*. Som en jämförelse tar jag också upp hur liknande ord beskrivs i *Den Danske Ordbok*.

En slutsats är att det torde vara få lexikala enheter som kan betraktas som unika konstituenten i snäv bemärkelse, dvs. som bara uppträder i en enda konstruktion i modern svenska. Däremot finns det ett relativt stort antal lexem som är starkt konstruktionsbundna och det måste givetvis framgå när dessa behandlas i en ordbok. Förhoppningsvis ger artikeln en inblick i lexikala egenskaper hos den aktuella typen av ord. En mer uttömmande beskrivning av dessa lexikala egenskaper är också en förutsättning för att unika konstituenten ska bli mer stringent beskrivna i kommande ordböcker, såväl svenska som beträffande andra språk.

## Litteratur

### Ordböcker

DDO = *Den Danske Ordbog* 2003–2005: Del 1–6. København: Gyldendal.

NEO = *Nationalencyklopedins ordbok* 1995–9: Del 1–3. Höganäs: Bra Böcker.

### Övrig litteratur

Bendz, Gerhard 1965: *Ordpar*. (Svenska Humanistiska Förbundet 74.) Stockholm: P. A. Norstedt & Söners Förlag.

Brink, Lars 2006: Den fraseologiske terminologi. I: Lorentzen, Henrik & Lars Trap-Jensen (red.), *Nordiske Studier i Leksikografi 8. Rapport fra Konference om Leksikografi i Norden, Sønderborg 24.–28. maj 2005*. (Skrifter udgivet af Nordisk Forening for Leksikografi 9.) København, 39–51.

Burger, Harald 1998: *Phraseologie: eine Einführung am Beispiel des Deutschen*. (Grundlagen der Germanistik 36.) Berlin: Erich Schmidt.

Corpas Pastor, Gloria 1996: *Manual de la fraseología española*. Madrid: Gredos.

- Dobrovol'skij, Dmitrij 1988: *Phraseologie als Objekt der Universalienlinguistik*. Leipzig: Verlag Enzyklopädie.
- Dobrovol'skij, Dmitrij & Elisabeth Piirainen 1994: Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. I: *Folia Linguistica* XXVI-II/3–4, 449–473.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor 1988: Regularity and idiomatcity in grammatical constructions: the case of *let alone*. I: *Language* 64, 501–538.
- Fleischer, Wolfgang 1997: *Phraseologie der deutschen Gegenwartssprache* (2. durchgesehene und ergänzte Auflage.) Tübingen: Niemeyer.
- Moon, Rosamund 1998: *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.
- Nenonen, Marja & Jussi Niemi 1999: Morphological Isolates in Idioms: Cranberries or Real Words? I: *Brain and Language* 68, 158–164.
- Ruiz Gurillo, Leonor 1998: *La fraseología del español coloquial*. Barcelona: Ariel.
- Sköldberg, Emma 2007: Birds of a feather and sheep's clothing: On unique constituents in Swedish. I: Nenonen, Marja & Sinikka Niemi (eds.), *Collocations and Idioms 1. Papers from the First Nordic Conference on Syntactic Freezes*. Joensuu, Finland, May 19–20, 2006. (Studies in Languages 41.) Joensuu: University of Joensuu, 297–305.
- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. (2 uppl. [1987]). Stockholm: Norstedts Akademiska Förlag.
- Teleman, Ulf, Staffan Hellberg & Erik Andersson 1999: *Svenska Akademiens grammatik* Del 1–4. Stockholm: Norstedts Ordbok.

#### Internet

- Google: <<http://www.google.se>>, hämtat 2007-08-21.
- Korpus 2000: <<http://korpus.dsl.dk/korpus2000/indgang.php>>, hämtat 2007-08-21.
- Presstext: <<http://www.presstext.se/>>, hämtat 2007-08-21.
- Språkbanken: <<http://spraakbanken.gu.se/>>, hämtat 2007-08-21.

Emma Sköldberg  
 FD, f. 1968  
 Institutionen för svenska språket  
 Göteborgs universitet  
 Box 200  
 SE-405 30 Göteborg  
 emma.skoeldberg@svenska.gu.se

## Modell för beskrivning av kollokationer i ett medicinskt lexikon (MedLex)

The paper presents a pilot study of the treatment of collocations in an electronic medical lexicon, MedLex, currently under construction at the University of Gothenburg, Department of Swedish. The lexicon is descriptive and intended to support both reception and production. The study focuses on noun collocations, especially those which describe diseases and medical conditions. We present an analysis model based on a combination of semasiological and onomasiological premises. The model results in a classification into lexically conditioned collocations and type-related collocations. We also discuss how medically related collocations may be described in the MedLex lexicon.

### 1. Inledning

Dagens samhälle kännetecknas av en allt större konsumtion av hälso- och sjukvård. Människor förväntas vara aktiva hälsokonsumenter som tar del av all den medicinska information som är tillgänglig på Internet och i massmedierna. Dessutom ges patienter numera möjlighet att läsa sina egna patientjournaler. För att ta del av all denna hälsoinformation krävs olika grad av förkunskaper. Även i en konkret vårdssituation måste läkare, patienter och anhöriga förstå vad som sägs men också kunna uttrycka sig på ett nyanserat sätt för att undvika missförstånd. Kollokationer utgör en viktig del i en sådan kommunikation, men deras reception och produktion kan vålla problem för dem som är mindre bevandrade i det svenska språket. Av den anledningen utgör medicinska kollokationer ett intressant studieobjekt som bör uppmärksammas i högre grad i det lexikografiska sammanhanget.

Syftet med den här artikeln är att presentera en pilotstudie om hantering av kollokationer i ett framtida deskriptivt medicinskt elektroniskt lexikon, Medlex. Det är tänkt att vara polyfunktionellt och avsett för reception men i kanske ännu högre grad för produktion. Det ska kunna vara till nytta för både läkare och patienter med eller utan svenska som förstaspråk (jfr Bergenholtz & Tarp 1994: 16–25). Fokus i studien ligger på kollokationer med substantiv och de undersöks ur ett synkront och funktionellt perspektiv. Genom studien hoppas vi kunna bidra till en mer systematisk beskrivning av svenska medicinska kollokationer ur både ett lexikologiskt och lexikografiskt perspektiv. Tanken är att det blivande lexikonet ska vara ett slags brygga – och fylla en lucka – mellan allmänspråkliga

ordböcker och mer renodlade fackordböcker. Innehållet i lexikonet kan också komma till användning inom språkteknologi för bl.a. semi-automatisk igenkänning av kollokationer i medicinska texter.

Artikeln är disponerad på följande vis: inledningsvis redogör vi för vår syn på termen *kollokation* och därefter presenterar vi vårt material. Därpå diskuterar vi vår analysmodell och innehållet i det elektroniska lexikonet. Artikeln avslutas med en sammanfattning och några slutsatser.

## 2. Om termen *kollokation*

Litteraturen om kollokationer är omfattande och termen *kollokation* används med skiftande innehåll (se t.ex. Hausmann 2004, Svensén 2004 och Siepmann 2005 med anförd litteratur). De nämnda forskarna diskuterar bl.a. kollokationers grundläggande egenskaper inklusive den gradvisa övergången mellan kollokationer och angränsande kategorier som idiom och fria kombinationer. För att undvika oklarheter redogör vi nedan för vår syn på formella och semantiska drag hos prototypiska kollokationer. Vår syn är inspirerad av framför allt den strukturella och funktionella beskrivning av engelska kollokationer som presenteras i Bartsch (2004). Andra viktiga inspirationskällor är Hausmann (1985), Malmgren (2003), Heid (1994, 2004) och Martin (2006).

- En prototypisk kollokation är en återkommande kombination av två eller flera ord som har en intern semantisk och syntaktisk struktur och som uppvisar lexikalisk bindning, t.ex. *ta tempen* och *falla i koma*. Med det ovan sagda vill vi markera vårt väsentligen systemorienterade synsätt på kollokationer. Det står i kontrast till ett korpusinriktat perspektiv enligt vilket kollokationer betraktas som ett statistiskt fenomen, beskrivna i termer av bigram och trigram, t.ex. *läkaren och* och *patienter som lider* (jfr Sinclair 1991; Kjellmer 1994).
- En prototypisk kollokation kan delas upp i en bas och en kollokator. Basen betraktas som ett slags semantiskt huvudord och kollokatorn som ett biord. I kollokationen *ta tempen* är *tempen* bas och det delexikaliserade verbet *ta* kollokator. Man kan här tala om en form av direktionalitet inom ordförbindelsen i och med att basen *tempen* styr valet av kollokatorn *ta* och inte tvärtom. Det finns dock kollokationer där direktionaliteten är svårare att fastställa och det är t.ex. när verbet har en mer pregnant betydelse. Exempel är *ordinera vila* och *ordinera läkemedel*.
- En prototypisk kollokation är vanligen genomskinlig. Det medför att den ofta är relativt lätt att tolka. Men graden av genomskinlighet kan variera, och den påverkas av flera olika faktorer, t.ex. de ingående ordens eventuella polysemi. I många fall disambigueras polysema ord när de ingår i en kollokation, t.ex. *förband* i *lägga förband* (jfr användning av substantivet

i medicinsk respektive militär domän). Något som också påverkar genomskinligheten är den eventuella metaforiken hos kollokationens led, t.ex. *tung i tung andning* och *djup i djup medvetlöshet*.

- Många prototypiska kollokationer uppvisar samma typ av relation mellan bas och kollokator. De oftast förekommande relationerna har beskrivits i termer av lexikala funktioner (Mel'čuk 1998). Exempelvis uttrycker den lexikala funktionen **Magn** ett gradförhållande, och den kommer till uttryck i kollokationer som *hög feber*, *djup depression* och *svår huvudvärk*.
- En prototypisk kollokation rymmer ofta implicit kulturell eller encyklopedisk information. Tillgång till den implicita informationen är en förutsättning för att man ska kunna tolka och använda kollokationerna på ett korrekt sätt. Till exempel har kollokationen *fri abort* en innebörd som är samhällsrelaterad. För att tolka *medicinsk* respektive *kirurgisk abort* krävs det däremot fackkunskaper. Andra exempel, som kan vålla både receptions- och produktionsproblem, är medicinska kollokationer med attributen *positiv* eller *negativ*, t.ex. *positivt/negativt graviditetstest*.

Vidare brukar kollokationer klassificeras utifrån deras inre syntaktiska struktur och klassificeringar av svenska, engelska, tyska respektive spanska kollokationer görs av bl.a. Svensén (2004), Bartsch (2004), Heid & Gouws (2006) samt Koike (2001). Nedan återges några återkommande syntaktiska klasser inom det medicinska området:

ADJ + SUBST	<i>öppet sår, molande värk</i>
V + SUBST	<i>ställa diagnos, lägga om ett sår</i>
V + PP	<i>hoppa ur led, svara på medicinen</i>
V + ADV	<i>andas djupt, må illa</i>
ADV + ADJ	<i>djupt medvetlös, gravt synskadad</i>
SUBST + V	<i>febern stiger, blodtrycket sjunker</i>

Ett tydligt drag är också att det bland kollokationerna förekommer parallella lexikala strukturer. I princip samma semantiska innehåll kan, genom morfologisk variation, realiseras med olika syntaktiska strukturer. Ett exempel är den redan nämnda kollokationen *djupt medvetlös* (ADV + ADJ), som är starkt besläktad med *djup medvetlöshet* (ADJ + SUBST). Ett annat fall är *blodtrycket sjunker* (SUBST + V) och *sjunkande blodtryck* (ADJ + SUBST). Vad beträffar morfologin uppträder vissa kollokationer ofta i en särskild grammatisk form. Det gäller t.ex. *tillstöta komplikationer*, vars bas främst förekommer i pluralis.

Ett annat fenomen, som uppmärksammats av bl.a. Heid (2004: 732), är att kollokationer kan kombineras med varandra och bilda mer komplexa flerordsenheter. Exempelvis kan kollokationerna *starkt beroendeframkallande* och *beroendeframkallande medel* fogas samman till *starkt beroendeframkallande medel*.

### 3. Material

Som utgångspunkt för vår studie har vi valt cirka 50 allmänspråkliga ettordsenheter som är representativa för ämnesområdet medicin. Det rör sig t.ex. om beteckningar på sjukdomar, ingrepp, behandlingar och symtom. Med tanke på det begränsade utrymmet här koncentrerar vi oss främst på ord som betecknar sjukdomar och sjukdomstillstånd. De aktuella orden upptas i allmänspråkliga svenska ordböcker, t.ex. NEO (1995–96), och fackspråkliga medicinska lexikon, exempelvis Lindskogs *Medicinsk terminologi* (2004).

För att finna relevanta kollokationer med de utvalda orden har vi använt både allmänspråkligt och fackspråkligt korpusmaterial. De allmänspråkliga texterna ingår i Språkbanken (på cirka 100 miljoner ord) vid Institutionen för svenska språket, Göteborgs universitet (<<http://spraakbanken.gu.se/>>). De medicinska texterna är hämtade ur Medlex korpus vid samma institution. Medlex korpus, som ännu inte är tillgänglig för allmänheten, innehåller drygt 14 miljoner ord och består av medicinska texter hämtade från bl.a. Läkartidningen och Socialstyrelsens hemsida. Våra data har också jämförts med motsvarande sökningar på Google (<<http://www.google.se>>). Som verktyg har vi dels använt allmänt tillgängliga konkordansprogram, dels ett kollokationsprogram, utarbetat av Christian Sjögreen vid Institutionen för svenska språket, Göteborgs universitet.

Fokus i vår studie ligger således på kollokationer. Däremot utesluter vi för tillfället etablerade medicinska flerordstermer och eponymer som t.ex. *gula fläck-en* och *Downs syndrom*. Det kan dock noteras att det enligt Manning & Schütze (1999: 141) föreligger en överlappning mellan begreppet kollokation och fackspråkliga flerordstermer.

### 4. Analysmodellen

Vår analysmodell bygger på en kombination av ett semasiologiskt och ett onomasiologiskt synsätt. Det semasiologiska perspektivet har diskuterats mer utförligt i avsnitt 2 ovan. Hädanefter uppmärksammas det onomasiologiska perspektivet och mer onomasiologiskt grundade kategorier. Inom det medicinska området finns det flera subdomäner: ANATOMI, SJUKDOM, BEHANDLING etc. Varje subdomän förutsätter en uppsättning av specifika, och sinsemellan ganska olika, scenarier som kan beskrivas med vissa onomasiologiska parametrar. Till exempel kan subdomänen SJUKDOM analyseras utifrån parametrar som Insjuknande, Tillstånd, Grad, Förlopp, Typ och Tillfrisknande. För att illustrera tankegången anges här några typiska kollokationer med *depression* som bas:

- Insjuknande: *drabbas av (en) depression, gå in i en depression*
- Tillstånd: *ha en depression, lida av depression*
- Grad: *lätt(are) depression, djup depression,*
- Förlopp: *akut depression, kronisk depression*

- Typ: *årstidsbunden depression, bipolär depression*
- Tillfrisknande: *komma ur en depression, ta sig ur en depression*

För att kunna upptäcka generella mönster har vi valt att jämföra *depression* med tre andra sjukdomar och sjukdomstillstånd, närmare bestämt *cancer*, *diabetes* och *förkylning*, som inbördes är mycket olika, vad gäller t.ex. symtom och allvar. I samband med beteckningarna aktualiseras ett antal typiska parametrar, men listan är inte uttömmande. I samtliga exempel nedan utgör sjukdomarna bas i de enskilda kollokationerna.

Parameter	<i>depression</i>	<i>cancer</i>	<i>diabetes</i>	<i>förkylning</i>
<b>Insjuknande</b>	<i>få en d.</i> <i>drabbas av (en) d.</i> <i>gå in i en d.</i>	<i>få c.</i> <i>drabbas av c.</i> <i>utveckla c.</i>	<i>få d.</i> <i>drabbas av d.</i> <i>utveckla d.</i>	<i>få en f.</i> <i>drabbas av en f.</i> <i>dra på sig en f.</i> <i>ådra sig en f.</i>
<b>Tillstånd</b>	<i>ha d.</i> <i>lida av (en) d.</i>	<i>ha c.</i> <i>lida av c</i>	<i>ha d.</i> <i>lida av d.</i>	<i>ha en f.</i>
<b>Grad</b>	<i>djup d.</i> <i>svår d.</i> <i>lätt(are) d.</i> <i>mild d.</i>	<i>långt framskriden c.</i>	<i>svår d.</i>	<i>elak f.</i> <i>kraftig f.</i> <i>lätt f.</i>
<b>Förlopp</b>	<i>kronisk d.</i> <i>långdragen d.</i> <i>akut d.</i>	<i>kronisk c.</i>	<i>långvarig d.</i>	<i>långvarig f.</i> <i>långdragen f.</i> <i>envis f.</i>
<b>Typ</b>	<i>årstidsbunden d.</i> <i>bipolär d.</i>	<i>ärfilig c.</i> <i>obotlig c.</i> <i>aggressiv c.</i>	<i>ärfilig d.</i>	
<b>Tillfrisknande</b>	<i>komma ur en d.</i> <i>ta sig ur en d.</i>	<i>botad c.</i>		

Tabell 1: Parameterbaserad klassificering av kollokationer med substantiven *depression*, *cancer*, *diabetes* och *förkylning*

Av exemplen i tabell 1 framgår att det finns kollokationer som är typiska för endast en av sjukdomarna men som inte används i samband med de andra. Man kan t.ex. *gå in i en depression*, men inte *\*gå in i en cancer/diabetes/förkylning*. Man talar om en *djup depression* men inte om *\*djup cancer/diabetes/förkylning*. Andra kollokationer med något begränsad användning är *dra på sig i dra på sig en förkylning* och *obotlig i obotlig cancer*.

Vidare finns det kollokationer som uppträder med några av de aktuella substantiven. Till exempel finns det flera förekomster av *lida av (en) depression/cancer/diabetes* i materialet, men däremot inte *lida av (en) förkylning*. Detta faktum torde kunna förklaras med att den senare åkomman inte är fullt så allvarlig som de andra. Med andra ord är det språkanvändarnas kunskaper om faktiska förhållanden som spelar en viktig roll för deras val av adekvata kollokationer.

Slutligen finns det kollokatorer som är aktuella i samband med samtliga sjukdomar ovan. Det gäller t.ex. *få* i *få (en) depression/cancer/diabetes/förkylning* och *drabbas av* i *drabbas av (en) depression/cancer/diabetes/förkylning*. En annan återkommande kollokator är *ha*. Dessa kollokatorer förekommer troligtvis med alla typer av sjukdomar, och utgör därför underlag för ett återkommande mönster med begreppet SJUKDOM i basen.

De förhållanden som illustreras i tabell 1 kan relateras till ett arbete av Willy Martin (2006). I sin artikel argumenterar han för en ramsemantisk modell för beskrivning av lexikonenheter i elektroniska ordböcker. Som illustration använder Martin exempel hämtade från det semantiska fältet Musikinstrument. Modellen utgår från en uppsättning parametrar, som är relevanta för det aktuella semantiska fältet, samt deras språkliga realisationer. De postulerade parametrarna utgör också ett viktigt stöd för en systematisk och uttömmande beskrivning av kollokationer.

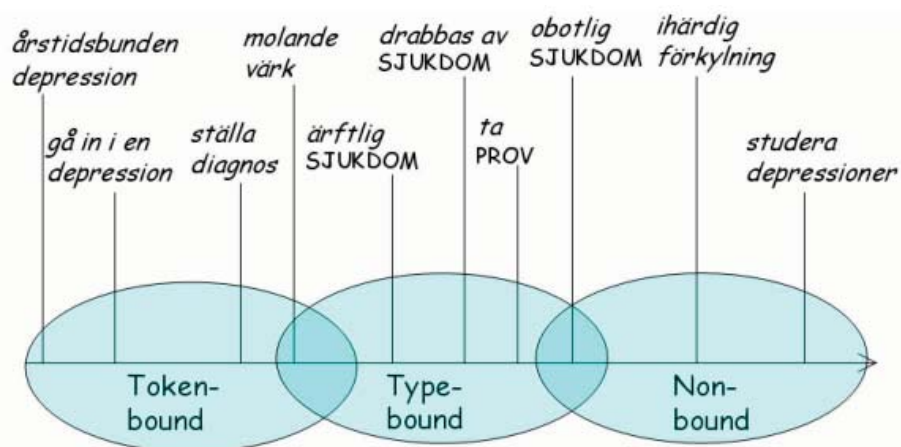
Martin (2006: 289) förespråkar en uppdelning i *token-bound* och *type-bound collocations* och relaterar dem till *non-bound/free combinations*. Skillnaden mellan kollokationstyperna och de fria kombinationerna ligger i en fallande grad av bindning mellan de ingående leden. För fria kombinationer är bindningen svagast eftersom dessa ordförbindelser uppvisar minst begränsningar beträffande konstituenternas semantiska kombinatorik. Av Martins modell framgår det dock tydligt att gränserna mellan kategorierna inte är särskilt skarpa och att de går in i varandra.

Med utgångspunkt i Martin (2006) skiljer vi mellan *lexikalt betingade kollokationer* (token-bound) och *typrelaterade kollokationer* (type-bound). De lexikalt betingade kollokationerna uppvisar ett starkt beroende mellan leden och tillåter endast begränsade synonymiska byten vad gäller ingående konstituent, t.ex. *årstidsbunden/årstidsrelaterad depression*. Kännetecknande för typrelaterade kollokationer är att deras kollokatorer återkommer i ett antal olika kollokationer och att deras bas i de flesta fall kan generaliseras och uttryckas med hjälp av en kategori, t.ex. SJUKDOM. Möjligheten att byta ut basen mot en hypo- eller hyperonym tyder på att en typbunden relation finns inbäddad i kollokationen, t.ex. *ärfilig SJUKDOM* och *drabbas av SJUKDOM*. De lexikalt betingade kollokationerna uppvisar således högre grad av bindning mellan leden än de typrelaterade.

Förhållandet mellan lexikalt betingade och typrelaterade kollokationer samt fria ordkombinationer kan illustreras med hjälp av figur 1.

Hittills har vi diskuterat den modell som vi använder för att analysera de medicinska kollokationerna. Nedan följer en mer ingående presentation av hur de beskrivs i själva lexikonet.





Figur 1: Ordkombinationer ordnade efter grad av bindning (efter Martin 2006)

## 5. Kollokationer i MedLex ordbok

Den elektroniska MedLex-ordboken kommer att bestå av två länkade moduler. Den ena är en allmänspråklig lexikonmodul och den andra är en kollokationsmodul. Den senare ska ses som ett komplement till den första, men tanken är att modulerna även ska kunna användas oberoende av varandra.

Artiklarna kommer att vara uppbyggda kring ord som utgör bas i kollokationerna. Dessa ord betraktas som primära ingångar i Medlex. Som sekundära ingångar fungerar kollokatorer inklusive relevanta kollokationer. Som exempel kan *ta pulsen* anföras. Substantivet *puls* är primär ingång till båda modulerna. Däremot är *ta* och *ta pulsen* sekundära ingångar. Man kan undra varför vi kommer att lista kollokatorer som t.ex. *ta* som enskild ingång. Skälet är att man ska kunna navigera i lexikonet utifrån alla ord som förekommer som bas eller kollokator i det medicinska ordförrådet utan att på förhand behöva veta kollokationens exakta form. Om man söker på hela kollokationen, får man direkt information om hur just den används (se vidare avsnitt 5.2 nedan).

I det följande ger vi en översikt över typer av information som ska inkluderas i lexikonet. Däremot gör vi inte anspråk på att visa hur informationen ska visualiseras i lexikonet.

### 5.1. Allmänspråklig lexikonmodul

I artiklarna i den allmänspråkliga lexikonmodulen ges information om uttal (i hörbart format), ordklassinformation och uppgifter om böjning samt betydelse i form av betydelseindikator, definition och definitionstillägg. Med en betydelseindikator menas en koncis formulering som speglar det huvudsakliga innehållet i ett visst betydelsemoment.

Det finns flera syften med att inkludera betydelseindikatorer. För det första

blir ordets övergripande begreppskategori synlig. För det andra vägleds användaren snabbt i valet av betydelsemoment. Detta gäller särskilt för polysema medicinska ord, t.ex. *cancer*, som får två betydelseindikatorer: SJUKDOM respektive TUMÖR. För det tredje möjliggör betydelseindikatorerna att man kan göra tesauristiska sökningar. Ordboksanvändaren kan exempelvis få fram en lista över alla ingångar som har betydelseindikatorn SJUKDOM.

Artiklarna ska också innehålla semantiskt relaterade ord (som synonymer, antonymer, hyperonymer och hyponymer) samt information om kollokationer. I artiklarna presenteras även några aktuella sammansättningar, enklare redigerade språkexempel och information om morfologiskt relaterade ord som tillhör samma eller andra ordklasser. Dessutom ges en länk till motsvarande ingång i kollokationsmodulen. Mer konkret kan det, utifrån substantivet *depression*, se ut på följande sätt:

**uppslagsord:** *depression*

**uttal:** (ljud)

**ordklass:** substantiv

**böjning:** *depression, depressionen, depressioner, depressionerna* m.fl.

**betydelseindikator:** SJUKDOM

**definition:** sjukligt tillstånd med svår (långvarig) nedstämdhet

**definitionstillägg:** ofta omfattande ångest, handlingsförlamning och livströtthet

**sammansättningar:** *utmattningsdepression, förlossningsdepression* m.fl.

**semantiskt relaterade ord:** motsats *mani*

**kollokationer:**

ADJ + *depression*: *djup, svår; lätt(are) | akut, långdragen, kronisk | årstidsbunden*

V + *depression*: *drabbas av (en), gå in i en | ha, lida av | ta sig ur en*

**exempel:** *hon lider av en djup depression som orsakar sömnlöshet*

**morfologiskt relaterade ord:** v. *deppa*; adj. *deprimerad, deprimerande, deppig, depressiv*

**länk:** till kollokationsmodul

Som synes listas kollokationerna först utifrån grammatisk struktur. I nästa steg indelas de med lodstreck efter aktuella parametrar: Insjuknande, Tillstånd, Grad etc. (jfr OCDSE 2002). Semikolon avgränsar semantiska undergrupper inom respektive parameter. Denna parameterbaserade indelning förefaller i många fall sammanfalla med en klassificering utifrån typer av lexikala funktioner, presenterade av Mel'čuk (se avsnitt 2 ovan).

### 5.2 Kollokationsmodul

Den andra modulen, kollokationsmodulen, kan som sagt användas antingen självständigt eller som ett komplement till den allmänspråkliga modulen. I kollokationsmodulen ligger tyngdpunkten på en mer uttömmande beskrivning av kollokationernas strukturella, typologiska och pragmatiska egenskaper. Denna

information kommer att göras åtkomlig för användaren dels i form av löpande artikeltext, dels genom ett utökat utbud av sökmöjligheter. I den löpande artikeltexten är den alfabetiska ordningen default. Det utökade utbudet av sökmöjligheter innebär främst tesauristiska sökningar utifrån betydelseindikatorer samt parametrar. Tack vare dessa sökmöjligheter skapas nya genvägar till innehållet i lexikonet, och mycket av den implicita informationen i den allmänspråkliga modulen blir explicit. Nedan ges en översikt över mikrostrukturen i artiklarna och sökmöjligheterna i kollokationsmodulen.

I artikelns mikrostruktur inkluderas följande delmoment: uppslagsord, ordklass, betydelseindikator (enbart för kollokationers basord) och definition. Informationen om uppslagsordets ordklass och betydelseindikator fyller främst en disambiguerande funktion och signalerar vilken betydelse som aktualiseras i artikeln. (Det kan tilläggas att en sådan hantering av ingångar förespråkas i många elektroniska ordböcker, exempelvis CALD 2003, eftersom hanterandet effektiviserar åtkomsten av information.) Efter den inledande delen följer en mer uttömmande lista över kollokationer. Vid listningen av kollokationerna följer vi samma format som används i den allmänspråkliga modulen, dvs. kollokationerna sorteras efter grammatisk struktur och relevanta parametrar. Det ska också finnas en länk till motsvarande ingång i den allmänspråkliga modulen. De ovan nämnda momenten i mikrostrukturen kan exemplifieras på följande sätt utifrån substantivet *depression*:

**uppslagsord:** depression

**ordklass:** substantiv

**betydelseindikator:** SJUKDOM

**definition:** sjukligt tillstånd med svår (långvarig) nedstämdhet

**kollokationer:**

ADJ + *depression*: *djup, svår, allvarlig; mild, lätt(are) | akut, långdragen, kronisk | årstidsbunden, årstidsrelaterad | bipolär | postnatal | egentlig*

V + *depression*: *få, drabbas av, gå in i en | ha, lida av | leva med | behandla, medicinera mot | ta sig ur en, komma ur en | bota*

**länk:** till allmänspråklig modul

I kollokationsmodulen är varje enskild kollokation försedd med en länk till mer specifik information om just den aktuella kollokationen. Om man klickar på länken, får man explicit information om bakomliggande parameter, kollokationens relativa frekvens och bruklighet (med hänsyn till ämnesområde och stilnivå). Dessutom ges autentiska exempel hämtade ur Medlex korpus samt ett kommentarfält med annan relevant information om kollokationen ifråga. Där förtydligas kollokationens innebörd och kombinatoriska egenskaper. Kommentarfältet kan också innehålla uppgifter av encyklopedisk eller pragmatisk karaktär som kan vara nödvändiga vid reception och produktion av ordförbindelsen. Om man t.ex.

väljer att aktivera länkarna till kollokationerna *svår depression* respektive *årstidsbunden depression*, får man ta del av följande uppgifter:

*svår depression*

**parameter:** Grad

**frekvens:** \*\*\*\* (= mycket frekvent)

**bruklighet:** allmänspråk

**exempel:** *Vid en medelsvår eller svår depression väljer läkaren ofta att kombinera samtalsterapi med en medicinsk behandling.*

**kommentar:** Indelning i *lätt*, *måttlig* och *svår depression* bygger på en skattningsskala utifrån förekommande symtom. Det är betydelsefullt att gradera depressionen eftersom patientens förutsättningar att bli fullt frisk ökar om behandlingen anpassas efter depressionens svårighetsgrad.

*årstidsbunden depression*

**parameter:** Typ

**frekvens:** \*\* (= vanligt)

**bruklighet:** allmänspråk

**exempel:** *Ljusbehandling är ett behandlingsalternativ vid årstidsbunden depression*

**kommentar:** De flesta patienter med årstidsrelaterad depression har en s.k. vinterdepression, dvs. de blir dåliga på hösten och tillfrisknar när våren kommer. Sommardepression verkar inte vara så vanligt i Sverige.

Som redan antytts kompletteras den deskriptiva delen i den löpande artikeltexten med ett utökat utbud av sökmöjligheter. Följande tillval vore önskvärda:

- **Enkel tesauristisk sökning** med betydelseindikator eller parameter som sökkriterium. I lexikonet kan man t.ex. söka på alla ingångar som har betydelseindikatorn SJUKDOM och få fram en lista med *cancer, depression, diabetes* etc. Söker man på parametern Grad, får man träffarna *svår, lätt, framskriden* etc.
- **Kombinerad tesauristisk sökning** där en betydelseindikator och/eller parameter kombineras med ett konkret ord. Man kan t.ex. vilja veta vilka sjukdomar som tar *akut* som attribut. En lista med t.ex. följande kollokationer skulle presenteras: *akut depression, akut pankreatit* och *akut lymfatisk leukemi*. Vidare kan man vilja veta hur en specifik parameter, exempelvis Grad, kan uttryckas när det gäller ett visst ord, t.ex. *depression*. Sökningen skulle ge följande resultat: *djup depression, svår depression, lätt(are) depression* etc.
- **Mönstersökning** efter typer av semantiska mönster som finns i lexikonet. En sådan sökning lämpar sig om man t.ex. vill veta vilka parametrar som är aktuella vid SJUKDOM. Här följer några exempel på sådana mönster: Förlopp + SJUKDOM, Grad + SJUKDOM, Insjuknande + SJUKDOM och Tillfrisknande + SJUKDOM.

- **Mönsterbaserad sökning** syftar till att finna alla belägg i hela lexikonet som har ett visst mönster. (För att hitta befintliga mönster används den mönstersökning som beskrivs strax ovan.) Om man väljer mönstret *Insjuknande + SJUKDOM*, får man resultatet *få en depression, drabbas av en depression, gå in i en depression, få cancer, drabbas av cancer, utveckla cancer* etc.

Med hjälp av dessa tillvalsökningar fås såväl lexikal som metaspråklig information om kollokationer, deras strukturer och typer. Denna information kan förhoppningsvis stödja lexikonanvändaren i jakten på en lämplig kollokator till en viss bas. Den kan också tjäna som ett verktyg för lexikografer som vill kontrollera systematiken och konsekvensen i material som de arbetar med.

## 6. Sammanfattning och slutsatser

I den här artikeln presenterar vi en pilotstudie om hantering av kollokationer i ett medicinskt elektroniskt lexikon, MedLex. Lexikonet är under uppbyggnad vid Institutionen för svenska språket, Göteborgs universitet. Det är deskriptivt och avsett att stödja såväl reception som produktion. Fokus i studien ligger på kollokationer med substantiv, framför allt sådana som beskriver sjukdomar och sjukdomstillstånd. Den analysmodell som presenteras ovan bygger på en kombination av semasiologiska och onomasiologiska utgångspunkter. Modellen resulterar i en indelning i dels lexikalt betingade, dels typrelaterade kollokationer. I artikeln diskuteras också hur de medicinskt relaterade kollokationerna kan beskrivas i det konkreta elektroniska lexikonet.

Vår pilotstudie har bekräftat vikten av att anlägga ett helhetsperspektiv, dvs. att inkludera såväl en semasiologisk som onomasiologisk dimension vid analys- och struktureringsarbetet. Vi hoppas att våra iakttagelser vad gäller kollokationer kring sjukdomar och sjukdomstillstånd också är relevanta vid beskrivning av kollokationer i andra medicinska sub-domäner. Studien visar även att det föreligger långtgående korrelationer mellan metodologin och det elektroniska formatet som skapar nya och mycket raffinerade sök- och åtkomstmöjligheter. Med andra ord utgör en nyanserad lexikologisk och encyklopedisk beskrivning av dessa enheter en nödvändig förutsättning för att skapa ett nytt elektroniskt lexikon över medicinska kollokationer. I dessa framtida elektroniska lexikon bör det också finnas möjligheter att på ett smidigt sätt komplettera det lexikala perspektivet med ett encyklopediskt genom att förse användaren med ett antal relevanta länkar till andra elektroniska källor.

## Referenser

### Ordböcker

- CALD = *Cambridge Advanced Learner's Dictionary* 2003. Cambridge: Cambridge University Press.
- Medicinsk terminologi* 2004. Utarbetad av Bengt Lindskog. Stockholm: Norstedts Akademiska Förlag.
- NEO = *Nationalencyklopedins ordbok* 1995–96. Utarbetad vid Språkdata, Göteborgs universitet. Höganäs: Bra Böcker.
- OCDSE = *Oxford Collocations dictionary for students of English* 2002. Oxford: Oxford University Press.

### Övrig litteratur

- Bartsch, Sabine 2004: *Structural and Functional Properties of Collocations in English. A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr Verlag.
- Bergenholtz, Henning & Sven Tarp 1994: *Manual i fagleksikografi. Udarbejdelse af fagordbøger – Problemer og løsningsforslag*. Herning: systime.
- Hausmann, Franz Josef 1985: Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. I: Bergenholtz, H. & J. Mugdan (eds.), *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28. –30.6. 1984*. Tübingen: Niemeyer, 226–257.
- Hausmann, Franz Josef 2004: Was sind eigentlich Kollokationen? I: Steyer, K. (Hrsg.), *Wortverbindungen – mehr oder weniger fest*. (Institut für Deutsche Sprache, Jahrbuch 2003.) Berlin/New York: Walter de Gruyter, 309–334.
- Heid, Ulrich 1994: On Ways Words Work Together. Topics in Lexical Combinatorics. I: Martin, W., W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (eds.), *Euralex 1994 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*. Amsterdam, 226–257.
- Heid, Ulrich 2004: On the presentation of collocations in monolingual dictionaries. I: Williams, G. & S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress, Lorient, France July 6–10, 2004*. Lorient: UBS, 729–738.
- Heid, Ulrich & Rufus H. Gouws 2006: A Model for a Multifunctional Dictionary of Collocations. I: Corino, E., Marelllo C. & C. Onesti (eds.), *Proceedings of the XII Euralex International Congress, Torino, Italia, September 6–9, 2006*. Alessandria, 979–988.
- Kjellmer, Göran 1994: Introduction. I: *A Dictionary of English Collocations. Based on the Brown Corpus*. Utarbetad av Göran Kjellmer. Oxford: Clarendon Press, ix–xliii.
- Koike, Kazumi 2001: *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá.
- Malmgren, Sven-Göran 2003: Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet. I: *Språk & Stil. Tidskrift för svensk språkforskning* 13 Ny följd 2003, 123–168.
- Manning, Christopher & Hinrich Schütze 1999: *Foundations of Statistical Natural Language Processing*. MIT Press.

- Martin, Willy 2006: Frame-based Lexicons and Making of Dictionaries. I: Corino, E., Marengo C. & C. Onesti (eds.), *Proceedings of the XII Euralex International Congress, Torino, Italia, September 6–9, 2006*. Alessandria, 281–293.
- Mel'čuk, Igor 1998: Collocations and Lexical Functions. I: Cowie, A.P., (ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press, 23–53.
- Siepmann, Dirk 2005: Collocation, Colligation and Encoding Dictionaries. Part I. Lexicological Aspects. I: *International Journal of Lexicography* 18:4, 2005, 409–443.
- Sinclair, John 1991: *Corpus, Concordance, Collocations*. Oxford: Oxford University Press.
- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. (2 uppl.). Stockholm: Norstedts Akademiska Förlag.

#### Internet

- Google <<http://www.google.se>>, hämtat 2007-08-28.
- Språkbanken <<http://spraakbanken.gu.se/>>, 2007-08-28.

Emma Sköldberg  
FD, f. 1968  
Institutionen för svenska språket  
Göteborgs universitet  
Box 200  
SE-405 30 Göteborg  
emma.skoeldberg@svenska.gu.se

Maria Toporowska Gronostaj  
FD, f. 1945  
Institutionen för svenska språket  
Göteborgs universitet  
Box 200  
SE-405 30 Göteborg  
maria.gronostaj@svenska.gu.se





LARS TRAP-JENSEN

## Tilgangs- og henvisningsstruktur i digitale ordbøger Overvejelser baseret på ordnet.dk

The concept of structure cannot be transferred from the printed book to the digital medium: the alphabetical ordering principle does not apply to a digital dictionary, hyperlinks are possible in the digital medium only, and conditions of spatial orientation, overview and browsing are different in the two media. Therefore, a digital dictionary must seek new solutions to provide user-friendly access to its contents. In this article, the task is viewed from two angles: Various solutions employed by existing online dictionaries are analyzed and evaluated as viewed by the user, and from the perspective of the dictionary-maker, special attention is devoted to the possible need to rearrange the (print-based) data structure underlying the presentation displayed on the screen.

### 1. Indledning

De fleste er vist enige om at traditionelle definitioner af begreberne tilgangsstruktur og henvisningsstruktur – eller mediostruktur som det også benævnes nogle steder – ikke dækker digitale ordbøger. I *Nordisk leksikografisk ordbok* lyder de korte definitioner:

**Tilgangsstruktur:** strukturen av de indikatorene som fører ordboksbrukeren til de opplysningene han/hun ønsker.

**Henvisningsstruktur:** strukturen av de eksplisitte og implisitte indikatorer som leder brukeren rundt i ordboken for å supplere eller utdype de opplysningene en har funnet.

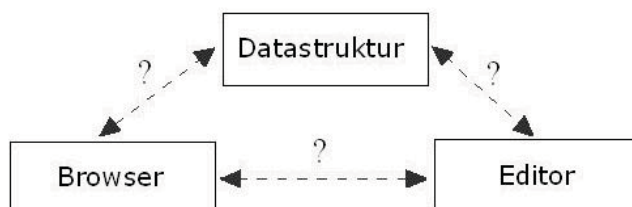
Problemet er, tror jeg, ordet struktur. Struktur handler om relationer og ordning af enheder, og det er ikke mærkeligt at en strukturel beskrivelse af den trykte ordbogs artikler og ordbogsoplysninger, ordnet sekventielt i de to dimensioner som bogsiden har, ikke kan overføres til elektroniske medier, især fordi muligheden for hyperlinks føjer en ny dimension til som gængse beskrivelser af gode grunde ikke har taget højde for. Den samme indvending kan bruges om centrale leksikografiske begreber som mikrostruktur og makrostruktur. Spørgsmålet er om begreberne kan udvides til at omfatte såvel traditionelle trykte ordbøger som digitale ordbøger. Denne artikel giver ikke det endelige og udtømmende svar på spørgsmålet, men tager fat på problemet med udgangspunkt i en række praktiske valg der har skullet træffes i forbindelse med en netordbog baseret på data

fra *Den Danske Ordbog*. I det følgende vil jeg gå ud fra et andet og meget løsere grundlag end de ovennævnte definitioner og overveje *hvad som helst der tjener til at hjælpe brugeren med at finde det han/hun ønsker*.

Jeg interesserer mig for hvordan man kan søge i en online-ordbog, hvordan resultatet af søgningen præsenteres, hvilken hjælp brugeren får, og jeg vil komme ind på de ekstra muligheder man har i en digital ordbog. På artikelniveau gælder det navigering og overblik over en artikels opbygning, henvisninger i almindelighed og derudover flerordsforbindelser og deres håndtering i den digitale ordbog. Afslutningsvis vil jeg se på hvilke leksikografiske konsekvenser det har, og hvordan det påvirker leksikografens arbejdsbetingelser.

## 2. Data, editor, browser

Udgangspunktet for artiklen har været *Den Danske Ordbog* (DDO) som papirordbog og brugen af data derfra til en netordbog. Interessen har især koncentreret sig om organiseringen af data og hvordan de valg der træffes, både bestemmes af og har konsekvenser for datastrukturen. Med en given struktur af data i basen kan man vælge en bestemt præsentation af dem, enten i den grænseflade som redaktørerne sidder med, i form af en editor eller et særligt redigeringssystem, eller i den brugergrænseflade som slutbrugeren anvender til at slå op i ordbogen via sin browser. Det vil sige at det er samspillet mellem database, editor og browser der er i fokus. Enhederne kan skematisk fremstilles som i figur 1, mens relationerne mellem dem indtil videre ikke er klarlagt.



Figur 1. Samspillet mellem enhederne er det der ligger bag meget af det følgende, og målet er at blive klogere på deres indbyrdes sammenhæng.

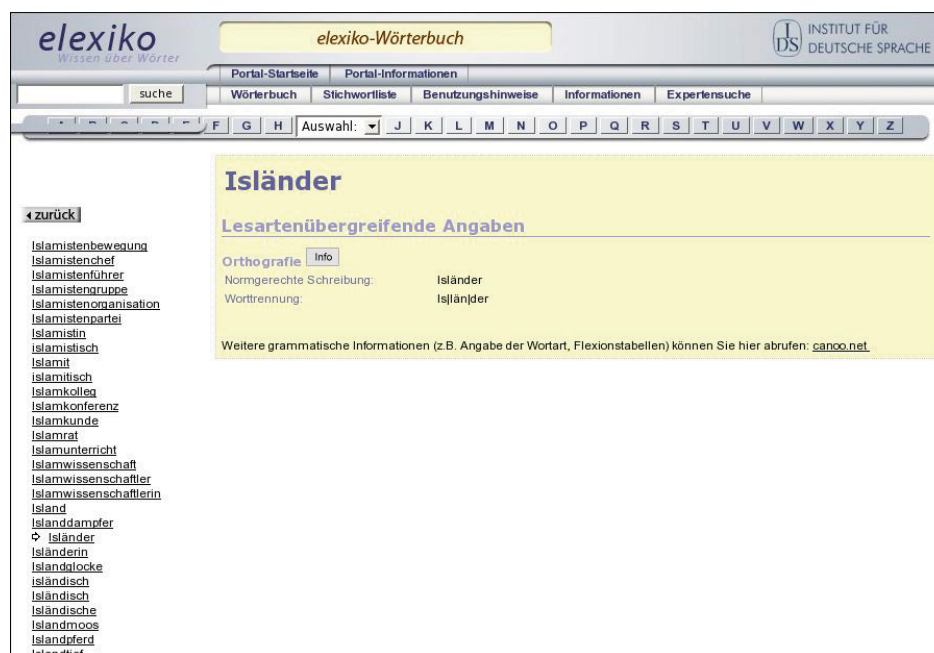
## 3. Opslag

### 3.1 Søgemåder

Den mest udbredte opslagsmulighed i online-opslagsværker er et søgefelt hvori brugeren indtaster det ønskede opslagsord. Ofte tilbydes også en trunkeret søgning, enten med jokertegn eller med valg fra en rullemenu mellem forskellige muligheder. Opslag ved hjælp af søgefelt er dog ikke den eneste mulighed. Man kan også med mus og taster navigere sig frem til det ønskede lemma gennem en systematisk eller alfabetisk ordnet opstilling og derfra klikke sig til den ønskede

artikel. Det er en løsning man især benytter i mindre værker hvor det er forholdsvis nemt at orientere sig i ordstoffet, fx privatpersoners offentligt tilgængelige ordsamlinger, eller i faglige værker hvor det er vigtigt at visualisere den systematiske opbygning, og i indscannede ældre ordbøger der kun findes som billedfiler uden indeksering på lemmaniveau. Der er sådan set ikke noget til hinder for at man kan kombinere de to søgemåder, og det finder man også i nogle netordbøger, fx det tyske projekt *elexiko*, jf. figur 2. Fordelen ved navigation fra en oversigt er at brugeren får et samlet overblik over stoffet – på samme måde som når man læser en trykt bogs indholdsfortegnelse. Dog mindskes overblikket hurtigt, jo større antallet af indgange i basen er. Hvis man skal scrolle i lang tid for at bladre forbi et stort antal poster eller klikke mange gange for at nå frem til en bestemt artikel, vendes fordelene ved overblikket nemt til en ulempe i den tid det tager at finde frem til den ønskede artikel.

For DDO's vedkommende er antallet af artikler så stort at navigationsløsningen aldrig har været overvejet seriøst – i hvert fald ikke som en primær løsning. Dog er vi opmærksomme på det tilfredsstillende ved den trykte bogs mulighed for overblik og bladring til andre artikler og vil gerne imødekomme det i et vist omfang.



Figur 2. Elexiko tillader både alfabetisk liste og søgefelt som tilgangsmulighed

Det gøres fx i netversionen af *Ordbog over det danske Sprog* (ODS) ved at tilbyde et udsnit af lemmalisten i skærmens venstreside. Her har brugeren mulighed for at bladre

frem og tilbage til artikler i alfabetisk nærhed af det søgte. En anden mulighed der minder om denne, er at vise artiklerne i rækkefølge efter hinanden på selve hovedsiden. Det bruges fx i *Macmillan English Dictionary Online* (MEDO), jf. figur 3.



Figur 3. I MEDO vises artiklerne fortløbende på skærmen med mulighed for at klikke sig videre videre frem og tilbage i den alfabetiske liste

Begge dele er selvfølgelig en illusion, nemlig om en struktur der ligner den trykte sides to dimensioner hvor artiklerne er ordnet sekventielt efter alfabetisk princip. Det har til gengæld den fordel at brugeren kan bladere rundt og læse artikler for sin fornøjelses skyld – også artikler hun ikke havde søgt efter i første omgang. Løsningen med fortløbende artikler er dog ikke så attraktiv i vores tilfælde – udelukkende af praktiske grunde. Den fungerer bedst hvis hver enkelt artikel kun er på nogle få linjer. Og løsningen med en ordliste har den fordel at man ret enkelt kan tilbyde ordbogen som baglænsordbog ved at tilføje en knap “alfabetiser baglæns” hhv. “alfabetiser forlæns”.

### 3.2 Søgning på hovedopslagsord

Lad os herefter se på de beslutninger der skal træffes før man får et resultat af sin søgning. Når man går ind på ordbogssiden, bør det være sådan at cursoren står i søgefeltet klar til indtastning (det er ikke altid en selvfølge). Når søgeordet er tastet ind, bør der være mulighed for enten at klikke på en søgeknop eller trykke <retur> – eller helst begge dele. I nogle ordbøger (et eksempel er MEDO) er muligheden for at trykke <retur> ikke til stede, og det er upraktisk. Når ordet er tastet ind og der er trykket <retur>, går søgningen i gang. Men hvor søges der, og hvordan vises resultatet? Hvis vi antager at det i første omgang kun drejer sig om søgning på hovedopslagsord, kan man tænke sig tre logiske muligheder.

### 3.2.1 Søgeordet matcher én og kun én post i basen

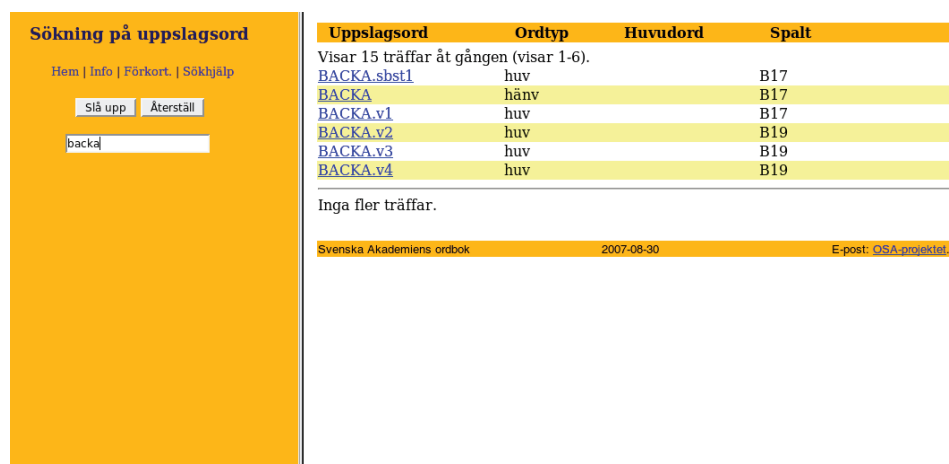
Den situation er som regel uproblematisk. Når man har trykket <retur>, kommer man straks til den pågældende artikel. Dog skal man være opmærksom på det tilfælde hvor der er et match, men bare ikke det rigtige. Et eksempel kunne være søgning efter det engelske adjektiv *savory* 'krydret, velsmagende'. Det har varianten *savory*, som er den amerikanske stavemåde, og hvis man søger efter denne form i *Oxford English Dictionary* (OED), får man ganske vist et resultat; det er bare ikke det ønskede. Det resultat der vises, er artiklen for substantivet *savory* 'sar' (plante af slægten *Satureia*), hvorimod man selv må regne ud at adjektivet skal søges under den britisk engelske stavemåde *savory*. Hvis man vil hjælpe brugeren i et tilfælde som dette, er svaret at tilbyde en udvidet søgning hvor der ikke kun søges efter opslagsordet, men også i andre elementer. Dette må man gøre som en generel mulighed da man naturligvis ikke kan vide om den viste artikel er den brugeren søgte.

### 3.2.2 Søgeordet matcher flere poster i basen

Her kan man forestille sig (mindst) tre muligheder:

1. Der vises en søgeresultatside hvorfra brugeren kan klikke sig videre blandt de opstillede muligheder
2. Mulighederne fremgår af en sidemenu, mens den første mulighed vises centralt
3. De matchende poster vises direkte på skærmen i rækkefølge efter hinanden

Mulighed 1 er vist den mest almindelige. Den bruges i fx OED, *Svenska Akademiens Ordbog* (SAOB) og *Longman's Dictionary of Contemporary English* (LDOCE), jf. figur 4.



**Sökning på uppslagsord**

Hem | Info | Förkort. | Sökhjälp

Slå upp Återställ

backa

Uppslagsord	Ordtyp	Huvudord	Spalt
Visar 15 träffar åt gången (visar 1-6).			
<a href="#">BACKA.sbst1</a>	huv		B17
<a href="#">BACKA</a>	häv		B17
<a href="#">BACKA.v1</a>	huv		B17
<a href="#">BACKA.v2</a>	huv		B19
<a href="#">BACKA.v3</a>	huv		B19
<a href="#">BACKA.v4</a>	huv		B19

Inga fler träffar.

Svenska Akademiens ordbok 2007-08-30 E-post: [OSA-projektet](#)

Figur 4. I SAOB vises søgeresultater i hovedvinduet som en menuliste hvorfra brugeren kan klikke sig videre til den ønskede artikel

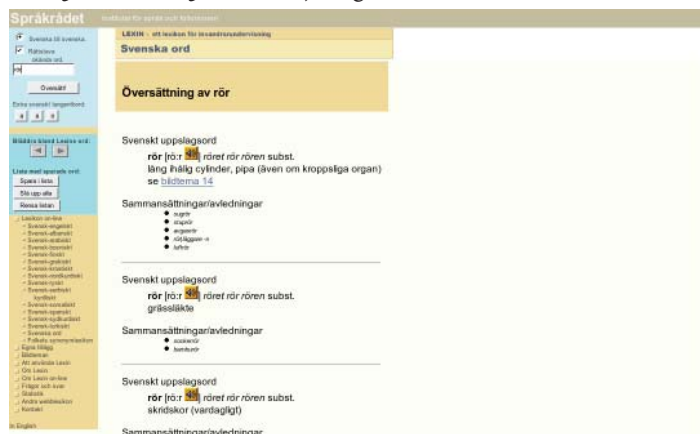
Mulighed 2 kræver at der er plads i sidemenuen til denne funktion. Til gengæld sparer den brugeren for et klik hvis den viste artikel er den søgte. I ODS er denne løsning valgt, og samtidig kombineres søgeresultatoversigten med den alfabetiske liste vha. faneblade, jf. figur 5.



Figur 5. I ODS fremgår søgeresultatet af oversigten til venstre, og det første resultat vises i hovedvinduet. Man kan veksle mellem søgeresultat og alfabetisk liste ved at vælge det ønskede faneblad.

Den løsning er ganske velfungerende og vil sandsynligvis blive videreført i den endelige grænseflade. Derimod er det muligt at listen vil blive flyttet til højremenuen fordi venstremenuen reserveres til andre formål.

Den tredje mulighed bringer brugeren direkte til de resultater der findes, men man må selv scrolle op og ned for at finde den ønskede artikel. Den mulighed finder man fx i *Lexin*-ordbøgerne, den danske *Retskrivningsordbogen* og i *Chambers 21<sup>st</sup> Century Dictionary Online*, jf. figur 6.



Figur 6. Lexin viser resultatet af en søgning på rör direkte på skærmen, og brugeren kan scrolle ned og vælge hvilken af de otte artikler der er den ønskede

### 3.2.3 Søgeordet matcher ingen post i basen

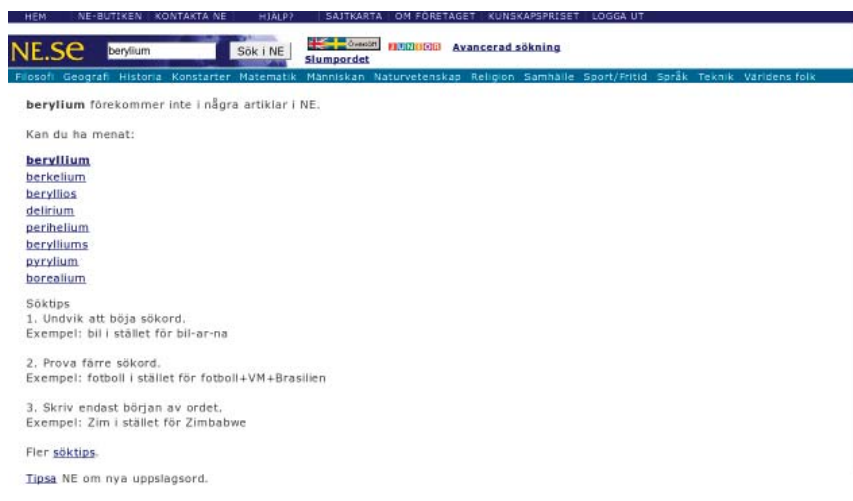
Hvis der ikke er noget match for søgningen, kan det skyldes at det pågældende ord ikke findes i basen. Der bør da komme en standardmeddelelse om det, og evt. kan der samtidig vises en relevant artikel på skærmen, fx nærmeste alfabetiske nabo. Brugeren kan også gives mulighed for at kontakte redaktionen og foreslå ordet optaget.

Den situation kan i praksis ikke skelnes fra den situation hvor det søgte ord faktisk findes i basen, men i en anden form end den der er tastet ind, hvad enten det skyldes en tastefejl eller at brugeren ikke kender den rette stavemåde.

Her bør der også komme en fejlmeddelelse der hjælper brugeren videre med søgningen. I meddelelsen kan der gives en oversigt med forslag til hvilke ord der minder om det søgte (resultatet af en såkaldt *fuzzy search*, jf. Svensén 2004: 521). I eksisterende netordbøger er det forskelligt hvor meget hjælp der tilbydes. God hjælp får man fx i online-versionen af *Nationalencyklopedins Ordbok* (NEO), hvor der ofte tilbydes brugbare alternative forslag, jf. figur 7.

Man kan se på forslagene at de er fremkommet som resultat af en algoritmisk beregning, og det er selvfølgelig ikke alle algoritmer der er lige gode.

Hvis brugeren tror at søgeordet staves anderledes, er det en fordel hvis ordbogen i forvejen kender til de mest almindelige både officielle og uofficielle varianter og indekserer dem så man kan hjælpe brugeren til at komme til det rigtige sted. I trykte ordbøger er det almindeligt at gøre det i form af henvisningsartikler: Fra officielle eller registrerede varianter henviser man til hovedopslagsordet. Det kan man også godt gøre i en netordbog, sådan som man kan se det i fx ODS eller *Grimms Deutsches Wörterbuch*.



Figur 7. En søgning i NEO på *beryllium* giver ikke noget resultat. I stedet foreslås følgende muligheder: *beryllium*, *berkelium*, *beryllios*, *delirium*, *perihelium*, *berylliums*, *pyrylium* og *borealium*.

Som regel vælges denne løsning dog fordi man af resurse-mæssige grunde har været nødt til at følge den løsning der er valgt i den trykte ordbog. Mere brugervenligt er det at spare brugeren for et ekstra opslag ved at bringe ham/hende direkte til den søgte artikel. Et eksempel på det ses i NEO, som vist stort set ikke rummer henvisningsartikler. I nogle learnerordbøger foretrækkes henvisningsløsningen dog bevidst – af pædagogiske grunde (fx i MEDO). Af samme grund foretrækker vi kun at springe henvisningen over ved dobbeltformer og acceptable variantformer; derimod ikke ved stavefejl og uofficielle former. Det kan desuden blive nødvendigt at fremhæve det søgte ord i den artikel der vises som søgeresultat, så brugeren straks kan se hvorfor artiklen er et relevant resultat af den pågældende søgning.

For datastrukturen betyder denne løsning ikke andet end at henvisningsartiklerne slettes fra basen. Indekseringen af variantformerne sikrer at brugeren kan komme til den rigtige artikel, og hvis man sørger for at registrere varianterne i separate elementer, kan man også i HTML-visningen tildele dem forskellige ledsagetekster: Ved dobbeltformer og sideformer kan brugeren bringes direkte til artiklen, ved fejl og uofficielle varianter først via en meddelelsside hvorfra de selv må klikke sig videre. I DDO's tilfælde er lavfrekvente varianter og fejl de eneste tilfælde hvor det er nødvendigt ikke bare at slette, men også at ændre i datastrukturen. I DDO-strukturen findes lavfrekvente varianter og fejl nemlig *kun* som henvisningsartikler og er ikke beskrevet i hovedartiklen. Her er det derfor nødvendigt at oprette fejlformen i det rigtige element i hovedartiklen. For brugeren gør ændringen selvfølgelig ingen forskel; det kan godt vises i form af en meddelelse på skærmen med mulighed for at klikke sig til den rette artikel. Derimod vil fejlformerne ikke længere optræde på den alfabetiske liste i siden af skærmen.

Endelig kan man tænke sig at søgeordet ikke findes som hverken hovedlemma, variant eller bøjningsform, men forekommer i et relevant element andetsteds i artiklen (fx som synonym, konstruktionseksempel eller orddannelsesmulighed). I det tilfælde er det ligeledes oplagt at vælge en løsning hvor der kommer en meddelelse om at der ikke er et match med det pågældende søgeord, men at ordet optræder i en anden artikel som brugeren så kan klikke sig hen til. Evt. kan man lade rækkefølgen af forslag være bestemt af en prioritering af artikelelementerne.

### 3.3 Udvidede søgemuligheder

I en netordbog har man flere muligheder end i en papirordbog. Hyperlinks gør det muligt at tilbyde opslag andre steder. Det kan være i eksterne resurser, fx Google eller *Retskrivningsordbogen*. Hvis det tilbydes som en generel mulighed, adskiller netordbogen sig ikke fra andre netresurser hvor der kan være integreret en søgemaskine i en bestemt applikation. Men muligheden kan også tilbydes i redigeret form, altså hvor en redaktør har taget stilling til at det er relevant at henvise til et websted uden for ordbogen selv. Det kan være en henvisning fra et



ord eller en bestemt betydning til en mere udførlig encyklopædisk beskrivelse, en mulighed fx *Wikipedia* benytter sig meget af. Blot skal man huske at koble en automatisk overvågning af linkene til så der ikke opstår “døde links” når en udbyder lukker sin tjeneste eller ændrer på opbygningen af sitet.

Det kan også være i form af semi-eksterne links, dvs. links til resurser som ligger uden for den aktuelle ordbog, men inden for ens egen kontrol. I vores tilfælde gælder det links til DSL’s korpusresurser og til ODS, og det er en mulighed vi ønsker at udvikle (se nærmere i fx Lorentzen og Trap-Jensen 2006, Trap-Jensen 2005). Det er fx oplagt at der samtidig med søgningen i netversionen af DDO sker en søgning i baggrunden i de andre resurser. Resultatet af baggrundssøgningen kan så tilbydes hvis en søgning i DDO ikke giver resultat, men findes i ODS (“det søgte ord findes ikke i DDO, men er registreret i ODS. Vil du se artiklen i ODS?”). Det kan også tilbydes som en generel mulighed når man står i en artikel (“Der findes også en artikel i ODS med dette ord. Vil du se artiklen i ODS?” eller “Se forekomster af dette ord i korpus”). Endelig er der de udvidede søgemuligheder som kan tilbydes internt i ordbogen. Indtil nu er det kun varianter af opslagsordet der har været behandlet, og hertil kan man også regne bøjningsformer. Disse søgemuligheder gælder de forskellige former som opslagsordet kan optræde i, men derudover er en ordbogsartikel jo opmærket i en række delelementer i mikrostrukturen, og det er oplagt at tilbyde at man også kan søge på dem, enten dem alle eller et relevant udsnit af dem. Det gør man formentlig bedst ved at tilbyde søgemuligheden “udvidet søgning” sådan som man har gjort det i fx OED, jf. figur 8.

Figur 8. I OED tilbydes udvidede søgemuligheder som valg fra et sæt af på forhånd opstillede muligheder

Her kan man vælge mellem fritekstsøgning eller søgning på bestemte elementer. Man kan se at det giver nogle helt nye søgemuligheder som ikke er til stede i papirordbogen. Fx kan man søge målrettet efter citater af bestemte forfattere eller få eksempler på ord der er lånt fra et bestemt sprog. Og man får en rimordbog forærende hvis det er muligt at søge i udtaleelementet hvor søgning på tekststregen alene kun giver meget upræcise rim (jf. Burke 2003: 244). Denne mulighed kan naturligvis også tilbydes som en selvstændig mulighed (“find ord der rimer på X”).

### 3.4 Artikelnavigering

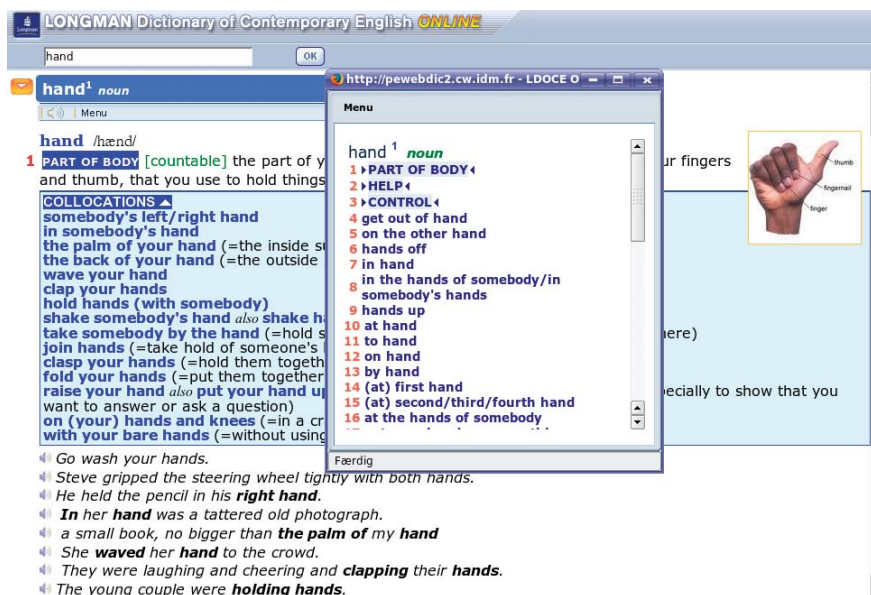
Når man er nået frem til den ønskede artikel, er spørgsmålet hvordan den ønskede information findes inden for artiklen. Det gælder at et skærmbillede på én gang er mere begrænset og rummer flere muligheder end papiret. Skærmen er begrænset i forhold til bogens sider ved at der står betydelig mindre på et skærmbillede end der gør på end papirsider. Og da undersøgelser samstemmende viser at læsbarheden nedsættes på skærmen, er det forståeligt nok sværere at skaffe sig et overblik over den samlede artikel på en skærm. En løsning på dette problem kan være at vise et resumé af artiklens hovedinddeling foroven. Det skaber overblik, og hvis oversigten er klikbar, har brugeren mulighed for straks at klikke sig frem til det ønskede sted frem for at skulle scrolle sig gennem et måske betragteligt antal skærmbilleder. For den der ikke vil læse artiklen i dens helhed, er det en god måde at finde den ønskede information hurtigt og præcist. Det er mærkelig nok ikke så udbredt som man kunne tro, men kendes dog i fx *Grimms* ordbog, i LDOCE og i *Wikipedia*. Alene af denne grund er *Grimms* ordbog væsentlig nemmere at bruge i netversionen end papirordbogen. Se figur 9.

Af figur 9 kan man også se at betydningsresuméet i *Grimm* er fremkommet ved simpelthen at tage de første ca. 100 karakterer af definitionen. Der er ikke foretaget nogen manuel redigering fra redaktørernes side. Det er gjort anderledes i LDOCE, hvor resuméet er holdt i korte, knappe overskrifter der dækker definitionen. Denne facilitet er med andre ord tænkt ind i konceptualiseringen af ordbogen (også af den trykte ordbog), jf. figur 10.

Det er klart at det gør en stor forskel om man vælger den ene eller den anden løsning. Hvis hver enkelt betydning skal forsynes med et særligt element der indeholder en glosse eller en sammentrængt definition til brug i resuméet, kræver det en betydelig redaktionel indsats, og det ligger uden for mulighederne i dette projekt. Der kan dog godt tænkes en automatisk eller semi-automatisk løsning som er anderledes end *Grimms* simple løsning. Fx fungerer fag- og stilmarkører ofte som måder at orientere sig i et ords betydninger på, og redaktionsinterne eksperimenter tyder på at en kombination af markører og uddrag af definitioner kan føre til et brugbart resultat (jf. Lorentzen og Trap-Jensen 2006, Trap-Jensen 2005).



Figur 9. Grimms Deutsches Wörterbuch. Bemærk at hele skærmens højrespalte er forbeholdt artikelnavigation.



Figur 10. Betydningsresumé i LDOCE. Et klik på "Menu" i øverste venstre hjørne kalder det centrale vindue frem.

I OED, som ellers har gjort meget for at fremme overskueligheden, gives der ikke artikelresuméer. En artikel som *go* har fx 96 nummererede hovedbetydninger og endnu flere underbetydninger (*go out* har alene underbetydninger op til bogstav u, dvs. 21 betydninger). Til gengæld er det en stor hjælp at man har mulighed for

at slå visse oplysningstyper til og fra i en såkaldt teleskopvisning. Navnlig citater fylder meget i en artikel, og det giver et væsentlig bedre overblik over artiklen hvis de ikke vises på skærmen. Af samme grund bør man overveje om artikelresuméet også skal være klikbart så brugeren selv kan slå det til og fra. Navnlig ved korte, overskuelige artikler vil mange nok finde det generende.

Den mulighed at brugeren selv bestemmer om oplysninger skal vises eller skjules, vil formentlig kunne udnyttes endnu bedre og derved få større betydning i fremtidens netordbøger. På den måde vil nøgleord som brugergrupper og ordbogsfunktioner komme til at få en anden rolle, idet den samme resurse kan rumme flere forskellige funktioner og henvende sig til forskellige brugergrupper. Det kommer nemlig helt an på hvilke oplysningstyper brugeren har slået til i sin opsætning, og tilsvarende kan den samme type oplysning formidles på forskellige måder. Man kan forestille sig at brugeren definerer en brugerprofil som svarer til netop de forudsætninger og behov han eller hun har, og i overensstemmelse hermed bliver de relevante oplysningstyper vist på skærmen. Idéen med at have fx både en kort og præcis ordbogsdefinition og en længere uddybende forklaring er ikke ny, men kendes også fra papirordbøger, ikke mindst fagordbøger (et eksempel er *Nordisk leksikografisk ordbok*). Der er ikke langt fra denne idé til at overlade det til brugeren at vælge hvilken udgave der skal vises i en netordbog. Et eksempel er *Musikordbogen*, der giver mulighed for at vælge kort eller lang visning.

## 4. Henvisninger

### 4.1 Henvisninger i trykte ordbøger

Hvis man tager udgangspunkt i den trykte ordbogs mere klart definerede henvisningsstruktur, er der i hvert fald tre forskellige argumenter for de henvisninger der optræder i en trykt ordbog som DDO. For det første et pladsøkonomisk argument. Det tager for meget plads op at gentage en oplysning to steder. Derfor henvises der fra ét sted til et andet. Eksempler er synonyme betydninger og dobbeltformer, altså officielt godkendte alternative skrivemåder. Ved redigeringen af DDO blev der i samarbejde med Dansk Sprognævn og Den Store Danske Encyklopædi udarbejdet en liste over hvilke dobbeltformer der skulle foretrækkes som hovedopslagsord og i redaktionel tekst.

For det andet er der et pædagogisk argument. Hvis man ikke ved hvordan et ord staves, kan man jo ikke slå det op. Derfor er henvisninger nødvendige hvis brugeren skal have reel hjælp til at komme fra en fejlagtig form til den rigtige. I tillæg er der en, omend indirekte, vejledende funktion i at henvise fra en lavfrekvent til den mere almindelige form eller betydning.

For det tredje bruges henvisninger til at gøre brugeren opmærksom på yderligere oplysninger i artikler andetsteds i ordbogen som kan være relevant for det ord eller den betydning man har slået op, og som man måske slet ikke var op-

mærksom på i udgangssituationen. Den slags henvisninger kaldes her plushenvisninger.

#### 4.2 Henvisninger i digitale ordbøger

Set fra netordbogens synspunkt ser de samme argumenter – ikke overraskende – lidt anderledes ud: Det pladsøkonomiske argument gælder ikke som det efterhånden nok vil være mange bekendt. Og det er en grund til at nedlægge de variantartikler der har været nævnt. Ud fra samme argument kan man i princippet også nedlægge alle henvisninger mellem betydninger. Men inden den beslutning tages, må de øvrige argumenter lige overvejes.

Det pædagogiske argument falder i to dele. Teknisk set er der ikke noget til hinder for at lede brugeren direkte til det rette opslagsord uden et ekstra klik. Men der kan godt være en pædagogisk pointe i at vælge omvejen alligevel: Hvis man indtaster en anden form end opslagsformen og derefter kommer til en resultatside hvor det oplyses at det indtastede ikke er den rette form, er det muligt at brugerens opmærksomhed rettes mod forholdet og at han/hun måske lærer af det. Hvis man derimod kommer til den rigtige artikel med det samme, er der risiko for at den pædagogiske effekt udebliver. Derfor ser man i mange ordbøger at der stadig opereres med henvisninger, ikke mindst i learnerordbøger, fx i LDOCE og MEDO hvor der henvises fra fx *children* til *child*. Andre ordbøger vælger at bringe brugeren direkte til den rette opslagsform, fx *Oxford Advanced Learner's Dictionary Online* i eksemplet *children*.

Personlig finder jeg at man bør skelne mellem officielle og ikke-officielle varianter sådan at man kun føres direkte til opslagsordet hvis der er tale om en autoriseret form. Men man skal bemærke at det standpunkt er pædagogisk begrundet, og valget – uanset hvilken løsning man vælger – påvirker ikke beslutningen om at nedlægge henvisningsartikler som selvstændige poster i basen. Så længe tilladte, uofficielle og fejlagtige varianter er opmærket med separate tags i hver deres felt, kan man knytte forskellige visninger til dem i HTML-versionen. Med andre ord gør det ikke nogen forskel for redaktøren i det daglige arbejde. Men præsentationen af datastrukturen for brugeren bliver forskellig. Hvis der er gode argumenter for det, kan beslutningen nemt gøres om uden at det kræver anden redaktionel indsats end at ændre HTML-visningen af de involverede elementer globalt.

Det andet tilfælde, hvor der henvises fra en sjælden til en hyppig betydning, er lidt anderledes. I langt de fleste tilfælde vil brugeren være bedst tjent med at få besked på stedet, frem for at blive sendt til en anden artikel. Men undertiden kan man af pædagogiske grunde overveje at bevare henvisningen, mest oplagt hvis der er tale om en ekstrem fordeling, eller hvis betydningen er forkættet, fx ved kontamination af to betydninger.

Man bør dog være opmærksom på at det er nødvendigt at ændre datastrukturen en del for at den fungerer hensigtsmæssigt. Hvis man ønsker at vise den

samme oplysning to gange i forskellige artikler, bør det ikke ske ved at gentage samme tekst i identiske elementer i flere artikler. Det fører til uønsket redundans, som kan skabe problemer hvis man foretager en revision i den ene artikel, men glemmer den anden. Derfor bør betydningsbeskrivelsen kun stå i én artikel, mens der i eventuelle synonymartikler blot optræder en "pegepind" der viser videre til den fulde beskrivelse. I HTML-visningen ser brugeren selvfølgelig kun målartiklens fulde ordlyd. Derved sikres at en ændring af definitionen automatisk afspejler sig i artikler med synonyme betydninger.

En forudsætning for dette princip er at det er muligt at pege entydigt på en bestemt betydning. Det var ikke muligt under redigeringen af den trykte DDO fordi lemmaet som grundenhed var det eneste element der kunne bestemmes entydigt ved hjælp af en ID-kode. I forbindelse med projektet DanNet (se mere herom i fx Nimb, Pedersen og Trap-Jensen i denne publikation) er der imidlertid oprettet ID-numre på betydningsniveau, og det kan nu udnyttes til dette formål. Det indebærer en vis redaktionel indsats at ændre datastrukturen for eksisterende artikler så den konsekvent følger det beskrevne princip. For redaktørerne betyder det en ændret praksis fremover, på den måde at der kun skal skrives én definition, mens der ved synonyme betydninger skrives en reference til betydnings-ID'et i målartiklen. For ordbogsbrugeren gør det som nævnt ingen forskel: På skærmen ser man blot to artikler med identiske betydningsangivelser.

Endelig er der plushenvisningerne. De bør selvsagt stadig have en plads i ordbogen, for brugeren kan ikke på forhånd vide at der findes relevante oplysninger i andre artikler. Det kan være oplysninger af encyklopædisk art, eller det kan være oplysninger om grammatiske former eller beslægtede ord.

Svaret på spørgsmålet om alle henvisninger skal nedlægges, er selvfølgelig et nej, det skal de ikke. Men der er god grund til at nedlægge rigtig mange af dem, frem for alt dem der kun er begrundet i pladsøkonomi. Dog skal man være opmærksom på at det kræver en vis omorganisering af data.

Det omvendte spørgsmål, om alt skal være henvisninger, kan ligeledes besvares negativt, i hvert fald i den forstand der her er talt om henvisninger. Derimod kan man godt operere med to forskellige slags henvisninger, eksplicitte hyperlinks der er fremhævet som man er vant til at se dem, fx understreget eller i en bestemt farve, og derudover en generel mulighed for at dobbeltklikke på et ord, hvorefter det slås op i ordbogen. Det er en facilitet som man også ser i andre ordbøger (fx *Cambridge Advanced Learner's Dictionary Online* eller LDOCE), og en mulighed som vi også overvejer. For at fungere hensigtsmæssigt kræves det dog at alle forekomster i ordbogen entydigt kan tilordnes den rigtige opslagsform, dvs. at bøjnings- og homografproblemer er løst.

## 5. Flerordsforbindelser

Det er ikke fordi flerordsforbindelser adskiller sig principielt fra resten af ar-

tikelementerne, men traditionelt er de et problembarn for papirordbøger på i hvert fald to områder. Det første er problemet med at finde ud af under hvilket af de indgående ord forbindelsen skal anbringes, det andet er udtrykkenes formmæssige variation. Et par eksempler fra den trykte DDO: Hvis man vil vide hvad *anabolske steroider* er, står forklaringen i artiklen *steroid*, mens der i artiklen *anabolsk* er en henvisning til udtrykket i artiklen *steroid*. Og hvis man vil vide hvad *som bare pokker* betyder, bliver man i artiklen *pokker* henvist til *som bare fanden*, som står i artiklen *fanden*, hvor man kan finde forklaringen. Det er fremgået at hvis pladsøkonomi er det eneste argument for at sende brugeren rundt ved hjælp af henvisninger, er der ingen grund til at bevare dem, og det gælder også i disse tilfælde. Det betyder at langt de fleste af henvisningerne kan nedlægges, og brugeren kan få forklaringen på stedet.

Med hensyn til datastrukturen kan man spørge sig om flerordsforbindelser egentlig behøver være knyttet til et bestemt af de indgående ord. Det kan forekomme mere hensigtsmæssigt med et databasedesign hvor flerordsforbindelserne er registreret i en selvstændig base, og hvor hvert enkelt af forbindelsernes indgående (betydende) ord peger på det tilsvarende opslagsord via ID-nummeret. Ved opslag på et bestemt ord er det derefter disse pegepinde der bestemmer hvilke ordforbindelser der vises under det pågældende ord. På den måde kan samme flerordsforbindelse vises i sin helhed under hvert enkelt af de indgående ord uden at man tager stilling til dens placering. Og der opstår ikke overflødig redundans fordi forbindelserne kun findes ét centralt sted. Ud over det redaktionelle arbejde med at omstrukturere data efterlader denne løsning blot et enkelt problem at tage stilling til: Hvordan håndterer man en søgning direkte på en flerordsforbindelse? Skal man bevare illusionen om artikeltilknytning og udnævne én artikel til "vært" for den pågældende forbindelse (som i fx NEO), eller skal man tillade et søgeresultat der består af andet end artikler (som i fx MEDO)? Det er indtil videre et uafklaret spørgsmål.

Endelig forudsætter løsningen ligeledes (jf. afsnit 4 nederst) at bøjnings- og homografiproblemer er løst.

## 6. Konklusion

Jeg har gennemgået nogle af de muligheder man har for at hjælpe brugeren med at finde frem til den ønskede information. Jeg har især interesseret mig for om de kræver ændringer i den eksisterende datastruktur. De muligheder og valg jeg har diskuteret, kræver forskellige indsatser fra redaktionens side.

En første redaktionel indsats vedrører konvertering af eksisterende data til en passende struktur. Den vil i en lang række tilfælde kunne foretages automatisk eller semi-automatisk for restgruppen af problematiske tilfælde, og gennemgangen har peget på tre typer af data. Den første er nedlæggelse af et antal artikler hvor der henvises fra én formvariant til en hovedartikel – dobbeltform, uofficiel

variant og sproglig eller ortografisk fejlform er de typer der blev opereret med i DDO. Nedlæggelse af henvisningsartiklerne vil kunne ske automatisk i ca. 80 % af tilfældene, mens de resterende 20 % kræver en manuel redaktionel indsats. Det skyldes at lavfrekvente fejlformer ikke har fået en beskrivelse i hovedartiklen i DDO, og derfor må oplysning herom manuelt overføres til hovedartiklen når henvisningen nedlægges.

Endvidere gælder det nedlæggelse af mange henvisninger på betydningsniveau, nemlig dem der ikke er hverken pædagogiske henvisninger eller plushenvisninger i DDO, men udelukkende har sin begrundelse i pladsøkonomi. Her er udfordringen at udskille de pædagogiske fra de ikke-pædagogiske. Det er ikke så svært i de tilfælde hvor der er tale om kontroversiel brug (eksempelvis kontamination), for dér er betydningen forsynet med en kommentar der gør opmærksom på det. Dog kan vi ikke isolere de tilfælde hvor en henvisning kunne være ønskelig pga. en ekstrem frekvensfordeling, til forskel fra de pladsøkonomisk begrundede henvisninger, og derfor bliver vi nødt til at træffe et valg. Alternativet er at gennemgå ca. 10.000 henvisninger manuelt for at tage stilling til om der skal bevares en henvisning af pædagogiske grunde, og det er usandsynligt at resurserne rækker til at prioritere denne opgave.

Endelig er der flerordsforbindelserne, som bør konverteres til en separat base og forsynes med pegepinde der henviser til de indgående opslagord via ID-numre. Også dette kan gøres automatisk for manges vedkommende, men der vil være en restgruppe af problematiske tilfælde (typisk homografiproblemer) som kræver manuel efterbehandling.

Den anden type af redaktionel indsats er redigering. Jeg har allerede nævnt at der vil være en række problematiske tilfælde som kræver en manuel bearbejdning. Så vidt jeg kan bedømme, er opgaven dog overkommelig så længe det er muligt at isolere disse tilfælde.

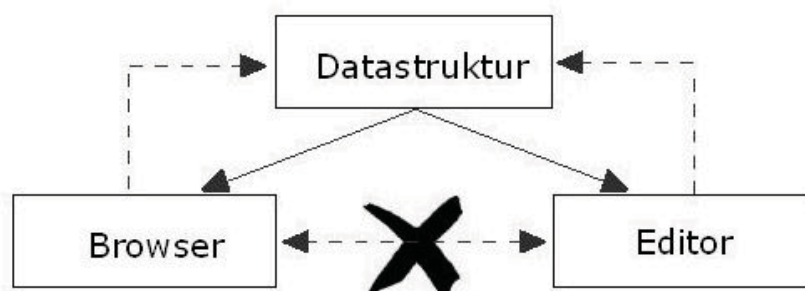
Den sidste type redaktionel indsats er programmeringsopgaver, og her drejer det sig især om visning af data i brugergrænsefladen. Blandt de opgaver jeg har nævnt, er hjælp til udvidet søgning, søgehjælp i form af trunkeeret søgning og en algoritme der giver meningsfulde bud på mulige alternativer. Derudover er det artikelresumé, teleskopvisninger og endelig en generel mulighed for at kunne slå op vha. dobbeltklik.

Hvis vi til sidst vender tilbage til diagrammet fra artiklens indledning, er vi måske blevet lidt klogere på sammenhængen mellem data, browser og editor. Jeg er hverken programmør, it-ekspert eller datalingvist, men vil gerne være med til at bygge bro mellem det der undertiden opfattes som to adskilte verdener, leksikografens og teknikerens.

I arbejdet med netordbogen går det i stigende grad op for mig hvor vigtig organiseringen af data er. Datastrukturen er det der bestemmer hvad det er muligt at vise både for redaktøren i redigeringsystemet og for brugeren i en browser. Derfor går påvirkningsretningen fra strukturen til grænsefladerne. Til gengæld



kan vi som brugere kun forholde os til data via det interface der afspejler data, altså gennem browserens grænseflade hvis vi er brugere af ordbogen, eller gennem editorens grænseflade hvis vi redigerer ordbogen. Editoren er også det sted hvor redaktørerne kan kode de data som lagres i basen. Så der sker en vis form for feedback fra det program vi ser data i. Derimod, og det er måske det vigtigste budskab i denne sammenhæng, er der ingen direkte forbindelse mellem det redaktøren ser når han/hun redigerer en artikel, og det slutbrugeren ser i sin browser. Redaktøren kan redigere data via et databasesystem, en simpel XML-editor, en webbaseret grænseflade eller et særligt redigeringsystem udviklet til formålet. Det kan alt sammen være mere eller mindre hensigtsmæssigt set fra redaktørens side, men det har i bund og grund ikke noget at gøre med det skærbillede som slutbrugeren kommer til at se på. Et bud på sammenhængen kunne derfor være som i figur 11.



Figur 11.

Det er i stigende grad vigtigt at vi som leksikografer også interesserer os for emner som databasedesign, XML-struktur og elementær programmering, ikke fordi vi skal overtage teknikernes arbejde, men fordi det er nødvendigt for at forstå hvilke muligheder de digitale medier stiller til rådighed, og ikke mindst er det nødvendigt i projektstyringen for at kunne vurdere hvor meget tid og hvor mange resurser en bestemt beslutning kræver for at blive ført ud i livet.

## Litteratur

### Ordbøger:

*Cambridge Advanced Learner's Dictionary Online*: <http://dictionary.cambridge.org>

*Chambers 21<sup>st</sup> Century Dictionary Online*: <http://www.chambersharrap.co.uk>

DDO = Ebba Hjorth, Kjeld Kristensen m.fl. (red.). *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab og Gyldendal, København 2003–2005.

*Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*: <http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/dwb/wbgui?lemid=GA00001>

*ellexiko*: [http://hypermedia.ids-mannheim.de/pls/ellexiko/p4\\_start.portal](http://hypermedia.ids-mannheim.de/pls/ellexiko/p4_start.portal)

LDOCE = *Longman Dictionary of Contemporary English Online*: <http://pewebdic2.cw.idm.fr>

*Lexin*: <http://lexin.nada.kth.se/lexin.html>

MEDO = *Macmillan English Dictionary Online*: <http://online.macmillandictionary.com>

*Musikordbogen*: <http://musikordbogen.dk>

NEO = *Nationalencyklopedins Ordbok*: <http://www.ne.se>

*Nordisk leksikografisk ordbok*, Henning Bergenholtz, Ilse Cantell, Ruth V. Fjeld, Dag Gundersen, Jón Hilmar Jónsson, Bo Svensén. Oslo: Universitetsforlaget 1997.

ODS = *Ordbog over det danske Sprog*: <http://ordnet.dk/ods>

OED = *Oxford English Dictionary*: <http://dictionary.oed.com/>

*Oxford Advanced Learner's Dictionary Online*: <http://www.oup.com/elt/catalogue/teachersites/oald7/?cc=dk>

*Retskrivningsordbogen*: <http://www.retskrivningsordbogen.dk>

SAOB = *Svenska Akademiens Ordbok*: <http://g3.spraakdata.gu.se/saob/>

*Wikipedia*: <http://da.wikipedia.org>

#### *Anden litteratur*

Burke, Sean Michael 2003: The design of online dictionaries. I: Piet van Sterkenburg (ed.): *A Practical Guide to Lexicography*. Amsterdam: John Benjamins Publishing Company, 240-249.

Lorentzen, Henrik og Lars Trap-Jensen 2006: *ordnet.dk* – et nyt sprogligt opslagsværk på internettet. I: Henrik Lorentzen og Lars Trap-Jensen (red.). *Nordiske Studier i Leksikografi, NFL-skrift nr. 9*. København, 253–264.

Svensén, Bo 2004: *Handbok i lexicografi*. Stockholm: Norstedts Akademiska Förlag.

Trap-Jensen, Lars 2005: Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: *LexicoNordica 12*, 109–122.

Lars Trap-Jensen  
cand.mag, MPhil, f. 1960  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
ltj@dsl.dk

LARS S. VIKØR

## Behandling av fraseologi i Norsk Ordbok

Treatment of phraseology is an important part of lexicography, which has been given extensive attention during the most recent decades. In *Norsk Ordbok* ('Norwegian Dictionary'), as in many sister projects, this has been done less systematically than desirable in the past. At present, when the project is being thoroughly revised, a systematization has been undertaken. This is not yet finished, but here I introduce a typology of Norwegian phrases related to three possible ways of presenting them in the dictionary: as separate lemmas (only recommended for foreign phrases imported integrally into Norwegian), as sublemmas (roughly for the most idiomatic phrases) and as editorial examples (mostly collocations or looser word combinations, and expressions with a normal sentence structure). The typology is based on grammatical criteria, and for each category examples are given, followed by an indication of the most suitable lexicographic treatment according to the view of the author.

### 1 Bakgrunn

Fraseologi er ikkje berre eit spesialområde for ein bestemt ordbokssjanger, men eit kjerneområde for alle større ordbøker, og aller mest for store nasjonale definisjonsordbøker. I tradisjonell leksikografi har dette språknivået ofte vorte behandla ganske usystematisk, og mitt prosjekt, *Norsk Ordbok* er ikkje noko unntak i så måte. Dei siste åra har vi arbeidd med ei systematisering som blant anna har bestått i ei klassifisering av frasetypar. Det kan gjerast på ulike måtar og ut frå ulike kriterium, men eg legg her fram eit forsøk eg har gjort i *Norsk Ordbok*-samanheng. Eg har bygd på tidlegare forsøk og grunnlagsmateriale frå kollega Oddrun Grønvik i det same prosjektet, og på fraseologi-kapitlet i Valerij Berkovs *Norsk ordlære* (Berkov 1997: 93–107). Men slik denne typologien blir lagd fram her, er eg eineansvarleg for han. Han er ikkje "vedteken" i *Norsk Ordbok*-samanheng, men skal munne ut i eit omforeint sett retningslinjer for den leksikografiske behandlinga av frasar, som vi i *Norsk Ordbok* kallar *fleirordige uttrykk*. Og sidan dei nordgermanske språka vel er nokså like når det gjeld grunnmønstra i fraseologien, håpar eg at dette kan ha ein viss overføringsverdi til andre, i alle fall som eit diskusjonsgrunnlag.

### 2 Tre behandlingsmåtar

Fleirordige uttrykk kan i *Norsk Ordbok* i prinsippet behandlast på tre måtar: Dei kan vere *eigne oppslag* (fleirordslemma), dei kan vere *underoppslag* (sublemma),

dvs. halvveges sjølvstendige ordartiklar plasserte under eit vanleg eittordslemma, eller dei kan ha status som *redaksjonelt døme* i ein vanleg artikkel. Systemet kan setjast opp slik:

1. *Eige oppslag* (lemma) (sjeldan)
2. *Underoppslag* (UO, sublemma):  
Eit underoppslag er ein artikkelbolk som behandlar visse typar fleirordige uttrykk, og som inneheld dei funksjonelle delane av ein vanleg innhaldsartikkel, nemleg tydingsbolk(ar) med definisjon(ar) og eventuelle døme med kjeldefesting. Underoppslaget er i halvfeit. Eit underoppslag eller ei rekkje av underoppslag avsluttar alltid ein tydingsbolk i hovudartikkelen (dvs. at det eller dei også kan utgjere ein heil tydingsbolk i seg sjølv).
3. *Redaksjonelt døme* (RD; redaksjonelt eksempel):  
Eit redaksjonelt døme er ei gjengiving av ein frase (oftast kollokasjon) som er forholdsviss godt innarbeidd i språket og godt representert i materialet, eventuelt med varierende form. Redaksjonelle døme kan ha forklaring, men har ikkje kjeldemarkering, og står såleis på redaksjonens ansvar. Redaksjonelle døme i *Norsk Ordbok* står i kursiv.

Det nye i *Norsk Ordbok*-samanheng er omgrepet *underoppslag*. Tradisjonelt har behandlinga av frasar i ordboksartiklane skjedd ved bruk av *redaksjonelle døme*. Det vil, som sagt ovanfor, seie at frasen er gjengitt i kursiv, den vanlege måten å gjengi språkbrukseksempel og sitat på, og (om nødvendig) forsynt med ei forklaring og eventuelt eit kjeldefesta sitat eller fleire som viser bruken i kontekst. Redaksjonelle døme og sitat har altså same typografiske status i artikkelen, forskjellen er berre at sitat har ei kjeldemarkering etter seg og følgjer rettskrivinga i originalen, mens dei redaksjonelle døma står utan kjelde og har den rettskrivinga som ordboka bruker redaksjonelt. Vilkåret for bruk av redaksjonelle døme er at frasen finst i såpass mange eller ulike kjelder at han kan reknast som (meir eller mindre) vanleg brukt i språket.

I den gamle redigeringsmåten, som gjekk ut på rein manuell skriving av artiklane, var det ingenting i vegen for å vere fleksibel her. I den nye redigeringsapplikasjonen som er teken i bruk frå det sist utkomne bandet, band 6, har vi måtta lage strengare kriterium og meir firkanta reglar for artikkeloppsettet, og i den samanhengen oppretta vi ein ny kategori som vi kalla underoppslag til skilnad frå redaksjonelle døme. Underoppslaget står i halvfeit, men innbakt i ein artikkel med eit enkeltord som hovudoppslag. Her følgjer to eksempel på artiklar (utdrag), der den første har eit redaksjonelt døme og den andre eit underoppslag (med to sitat som eksemplifiserer underoppslaget). Parentes med prikkar "(...)" står for kjeldereferansar som eg her har utelate:

**katte-pine** f lei knipe, vanskeleg situasjon (...): *koma i kattepine*.

**klut** m ... **1** lite stykke vovi el sydd ty ... // i vend **vera ein raud klut** vera omstridd og utfordrande; vekkja harme (...): *for Elisabet var Wibe-namnet ein raud klut (...)* / iron.: *NRK (har) i mange år vore den raude kluten for gode nordmenn (...)* ...

Å finne kriterium for når ein frase skal plasserast i den eine eller den andre av desse kategoriane, har vore ei av dei store utfordringane i redigeringsarbeidet. Eg har derfor altså prøvd å lage ein typologi over frasar og fordele dei ulike typane på desse tre behandlingsmåtene etter bestemte kriterium. Det er den eg legg fram her. Anbefalingane av oppføringsmåtar for kvar kategori er så langt berre forslag, vi driv framleis og diskuterer dette i *Norsk Ordbok* og har ikkje vedteke noko. Så det er berre mitt private utkast eg legg fram. Men eg trur ein slik typologi kan ha fagleg leksikografisk interesse som eit diskusjonsgrunnlag og kanskje ein analyseriskeiskap også ut over *Norsk Ordbok*.

### 3 To synspunkt

Som nemnt kallast *frase* i *Norsk Ordbok* ofte *fleirordig uttrykk*, og det er ei samle-nemning på konstruksjonar, kollokasjonar og idiom. Utgangspunktet mitt er at slike fleirordige uttrykk kan sjåast under to hovudsynsvinklar. Det er:

- *uttrykkets synspunkt*: uttrykket skal med, men kor skal det plasserast?
- *ordets synspunkt*: ordet skal belysast, kva uttrykk trengst?

Uttrykkets synspunkt disponerer for underoppslag (eller eigne oppslag i særlege tilfelle), ordets synspunkt disponerer for redaksjonelt døme.

*Flyge i flint* 'eksplodere av raseri' må stå forklart i ordboka, men det er eit metaforisk uttrykk som verken involverer reell flyging eller flintestein. I *Norsk Ordbok* er det plassert under **flyga**:

*flyga (sund) i flint(estein)* springa sund i småstykke; overf sprekka av sinne ... / ofte i samb med adv som *mest, nesten* o l: *rysen vart so vond at han mest flaug i flint (...)*

Men det står også under **flint**:

vanl i vend *flyga i flint* sprekka av sinne, bli fykande sint: *om du vart så sint at du flaug i flint (...)* / \*eg ... gjordeest brått så vill og sint, at eg nesten flaug i flint (...)

Dette var den tradisjonelle behandlingsmåten, med uttrykket oppført som redaksjonelt døme med eksemplifiseringar i form av sitat med kjelder. I dag ville dette ha blitt eit underoppslag, og det ville berre ha stått under eitt av oppslagsorda **flyga** eller **flint**. Truleg helst under **flint**, som er det mest særmerkte ordet i frasen, og i artikkelen **flyga** ville det ha stått på ein semantisk passende plass: "*flyga i flint* sjå *flint*".

Når verbet *flyga* skal beskrivast – i ein lang artikkel – skal det også vere med

ei tyding "(stendig) farta, reika ikring, serl føremålslaust, resultatlaust el med lite sømeleg ærend", og da trengst det illustrerande uttrykk og setningar: "*han er ute og flyg / han flyg ikring og gjer ugagn / flyga med slarv*". Desse er ikkje kjeldefesta, og dermed er dei redaksjonelle døme; dei utdjupar og illustrerer tydinga. Nokon eigen funksjon i språket utover det har dei ikkje.

#### 4 Tre typar inndelingskriterium

Når ein skal klassifisere fleirordige uttrykk, kan ein byggje på fleire typar kriterium. Det kan vere:

1. *Oppbygginga av uttrykket*
2. *Funksjonen til uttrykket i setninga*
3. *Leksikaliseringsgrad*

Det første kriteriet er grammatisk. Eg skil mellom tre hovudtypar: *hypotaktiske uttrykk* eller *hypotagme*, *parataktiske uttrykk* eller *paratagme*, og *uttrykk i setningsform*. Hypotagme har ein kjerne eller basis, i norsk grammatikk også kalla overledd, og eitt eller fleire underordna ledd som vi kallar underledd eller adledd. Paratagme har to eller fleire (oftast berre to) sidestilte ledd med ein forbindar, vanlegvis *og*. Setningsform inneber det vanlege kriteriet: subjekt og verbal med eventuelle utfyllingar. Eg har ein fjerde kategori som er meir pragmatisk, nemleg *innlånte uttrykk* som fungerer som heilskapar. Eigentleg kunne dei inngå i ei overordna inndeling med alle dei opphavleg norske uttrykka som motsetning, men kompleksiteten i typologien blir stor nok utan enda eit nivå i strukturen.

Den andre kriteriet er også grammatisk, men perifert. Det dreiar seg om den funksjonen uttrykket samla får i setninga. Normalt vil ein nominalfrase fungere som eit nomen (subjekt, objekt osv.), ein verbalfrase som verbal osv. Kriteriet gir altså same inndeling som det førre, men vi har nokre få avvik i tilfelle av grammatikalisering, til dømes at eit adverbialt uttrykk som *så sant* blir brukt som subjunksjon og altså går over frå å vere innhaldsord (*sant* av *sann*) til å bli eit grammatisk funksjonsord (som prinsipielt kunne vore samanskrive: *så sant*, men ikkje blir det fordi rettskrivinga tilseier særskriving): *eg tek turen så sant det ikkje blir uvêr* (der *så sant* er ekvivalent med subjunksjonen *dersom* eller *om*). Dette kriteriet er så perifert at eg ser bort frå det her (men sjå 5.1.2.b nedanfor).

Det tredje kriteriet er derimot svært viktig, og det går på tvers av det første. Det er *leksikaliseringsgrad*, altså i kor stor grad uttrykket er ein fast og sjølvstendig frase og ikkje berre ein tilfeldig og utbytbar ordkombinasjon. Dette er fraseologiens område, og på dette feltet er det gjort mykje godt teoretisk arbeid (eg viser her berre til Svensén 2004). Vi har den kjende skalaen frå konstruksjon (*lese i ei bok, ete ei brødskiye*), over kollokasjon (*lese til eksamen, ete frukost*) til idiom (*lese nokon teksta, ete kirsebær med dei store*), med ulike gradar av fastheit og idiomatisk sjølvstende inn- anfor kvar kategori (fri vs. fast ordkombinasjon, open vs. lukka kollokasjon osv.).

Problemet med denne skalaen er at han er ganske diffus, med glidande overgangar mellom kategoriane, så ein må bruke ein stor grad av skjønn når ein plasserer enkeltuttrykk inn. Klare og prototypiske idiom, kollokasjonar osv, finst det sjølvsagt mange av, men for ei ordbok som prinsipielt skal dekkje eit heilt ordforråd, er det dei vanskelege overgangstilfella som er viktige. Det må finnast ein måte å behandle dei på.

Typologien min tek utgangspunkt i den *grammatiske oppbygginga* av uttrykka, og byggjer på den overordna firedelinga under kriterium 1 ovanfor. Det har den fordel at ein stort sett kan setje opp klare kategoriar bygde på faste kjennemerke. På grunnlag av det set eg opp ein typologi. For kvar enkelt kategori vurderer eg så kva oppføringsmåte kategorien naturleg disponerer for. Da må eg også trekkje inn kriteriet leksikaliseringsgrad. Somme kategoriar består av sterkt leksikalisererte uttrykk, som derfor blir tilrådde å førast opp som UO. Andre tenderer mot å bestå av lausare og meir eller mindre uetablerte ordkombinasjonar som følgjer bestemte grammatiske konstruksjonsreglar og opptrer saman av semantiske grunnar, dei blir RD. Setningsform disponerer alltid for RD.

Leksikaliseringsgrad blir såleis eit viktig, men ikkje heilt avgjerande, kriterium for fordelinga på dei to oppføringsmåtane. Men, som eg har sagt ovanfor, er leksikaliseringsskalaen diffus, og dermed vanskeleg å handtere. Mange av dei kategoriane eg fører opp, vil innehalde uttrykk av ulike leksikaliseringsgradar, slik at enkeltuttrykka må førast som RD eller UO alt etter kor idiomatiserte dei er. Da må redaktøren vurdere dette på bakgrunn av eigen språksans og framfor alt av korleis uttrykket er brukt i datamaterialet, og kanskje ha ei vanskeleg avveging å gjere. Full konsekvens i behandlingsmåten er neppe mogleg å oppnå når gråsonene er så store som her.

Men vi har eit par meir handfaste leksikografiske prinsipp for fordelinga mellom RD og UO, og dei er overordna også om dei går på tvers av den fordelinga eg foreslår her. Vi bruker UO der *ein frase har fleire tydingar* eller *vi vil eksemplifisere frasen med sitat*. Ein frase kan for eksempel vere eit idiom med overført metaforisk tyding, og samtidig førekomme som kollokasjon med ei meir konkret tyding. *Å vere på jordet* er eit typisk norsk døme. Konkret betyr det å vere på åkeren og arbeide der. Metaforisk betyr det, etter definisjonen i *Norsk Ordbok*, å "vera desorientert el uvitande". Da kan det jamvel graderast: *å vere heilt på jordet* – i den konkrete tydinga blir det ei umogleg samanstilling. Da lagar vi underoppslag, som det her er naturleg å plassere under **jorde** 'åker'. I det aktuelle tilfellet er den konkrete tydinga ikkje ført opp i ordboka i det heile, den er blitt oppfatta som sjølvsagd. Samfunnsutviklinga har dessutan ført til at den konkrete tydinga er blitt sjeldan, mens den metaforiske er ganske vanleg. Andre gonger fører vi opp begge tydingane, og gir gjerne eit brukseksempel i sitatform på den metaforiske eller på begge. Når eit uttrykk har to eller fleire tydingar og det er tilgang på fleire sitat med uttrykket, har vi i seg sjølv sterke indikasjonar på leksikalisering.

## 5 Typologi

### 5.0 Innleiing

Så går vi over til sjølve typologien. Under kvar kategori gir eg ein del konkrete eksempel på uttrykk, og drøftar eventuelle problem ved kategorien. Så gir eg ei anbefaling om oppføringsmåte, RD eller UO, og kva slags hovudoppslag uttrykket bør behandlast under, t.d. første ordet i frasen (*av og til* under **av**), kjerneleddet i frasen (*kvit dverg* under **dverg**), det mest innhaldstunge ordet i frasen (*leggja saka daud* under **sak**) osv. Ord i halvfeit refererer til hovudoppslag i ordboka, fleirordsuttrykk i halvfeit refererer til underoppslag.

Ein kan merke seg at mange av dei uttrykka eg siterer, allereie er behandla i dei utkomne banda i ordboka. Men det ser eg her bort frå: Dette er ikkje ei beskriving av tidlegare ordbokspraksis, men eit forslag til framtidig behandling. Tanken er jo at når papirordboka er fullført i 2014, skal datamaterialet forvalta vidare i form av ei elektronisk og stadig oppdatert ordbok, og da er det meininga at også dei tidlegare delane av ordboka skal tilpassast dei nye strukturane vi er i gang med å leggje fast no.

Eg siterer eksempeluttrykk i *Norsk Ordboks* rettskriving, som stammar frå 1938 og i ein del ord nyttar noko eldre former enn dei som er vanlege i nynorsk i dag. Men skilnaden er ikkje særleg stor. Det viktigaste er at infinitiv endar på *-a*, mens endinga *-e* er vanlegast i dag og også blir brukt i mi eiga tekst i denne artikkelen.

Det kan vere eit problem for ikkje-norske lesarar at ein del av dei idiomatiserte uttrykka kan vere vanskelege å forstå, sjølv om andre er meir fellesnordiske. Det har ikkje vore mogleg for meg å sortere vekk eller leggje inn forklaringar på det mest eksklusivt norske eller nynorske, også fordi eg ikkje alltid har full innsikt i kva som er særnorsk vs. fellesnordisk. Men hovudpoenget, sjølve struktureringa, håpar eg likevel går tilstrekkeleg klart fram.

### 5.1 Hypotaktiske uttrykk

#### 5.1.1 Nominalfrasar

- a. bestemmarledd + substantiv: *den gong, den tid, sine dagars ende, beste evne, inga årsak, dei rikes bord*
- b. beskrivarledd + substantiv: *fritt leide, god dag, grøn mann, kvit dverg, høg sigarføring, berr himmel, levande leksikon, blått blod, vårt daglege brød, forboden frukt*
- c. substantiv (nominal) + komplement: *litt av kvart, lus på tjørekost, smular frå dei rikes bord, spøk til side, ingen regel utan unnatak, ingen roser utan tornar, smør på flesk*



Kan ha alle leksikaliseringsgradar. Men bruk av naken ubestemt form (av teljelege substantiv) er ofte kriterium på at det er eit idiom. Døme: *grøn mann* 'trafikklys på grønt med bilete av ein gåande mann' motsett *ein grøn mann*, som t.d. kan vere ein mann i grøn uniform. Det same gjeld bruk av sjeldsynte ord (*leide*) eller (meir i nynorsk enn i bokmål<sup>1</sup>) genitivuttrykk (*dei rikes bord*). Helsingfrasar er vel alltid idiom.

*Oppføring*: normalt som RD under kjerneleddet. Ved reine idiom truleg UO under første innhaldstunge ord ( gjerne med tilvising frå det andre ordet).

### 5.1.2 Adjektivfrasar

- a. (forsterkande el. modifierande) adverb + adjektiv: *knekkjande likegyldig, ovende ven, grønaktig gul*: stort sett brukt predikativt.

*Oppføring*: helst som RD under kjerneordet (under begge dersom sambandet mellom orda er eksklusivt, som i *knekkjande likegyldig*, elles er det nok å nemne uttrykket som døme på bruken under eitt av (el. sjeldan begge) orda.

- b. avbleikt gradsadverb pluss nøytrumsadjektiv: *så sant, så snart, så vidt, så visst*. Desse kan opptre som funksjonsord (subjunksjonar).

*Oppføring*: truleg helst som UO under kjerneordet med tilvising frå *så*. Eit spørsmål for seg er om vi skal ha dei under **sann, snar, vid, viss** eller ha eige oppslag på nøytrumsformene. Det er gode argument for det siste, for dei fungerer nokså uavhengig av grunnformene både semantisk og funksjonelt. For dei mest grammatikaliserte kan vi vurdere eigne oppslag.

### 5.1.3 Preposisjonsfrasar

Desse består av preposisjon + nominal, og er her grupperte etter kva type nominal preposisjonen styrer):

- a. substantiv i bestemt form eller teljelege substantiv i ubestemt form med artikkel: *av verste skuffe, borti veggene, i grunnen, i lengda, i mellomtida, i siste liten, i vissa, på flygande flekken, på hangande håret (eller på eit hangande hår)*

*Oppføring*: normalt som RD under nominalet (somme av dei kan i tillegg vere med som døme på bruken under preposisjonane), men når det er eit klart idiom (som *i grunnen, i lengda, i mellomtida, i vissa*) kan det vere UO (framleis under nominalet).

- b. substantiv i naken ubøygde form (ubestemt form utan artikkel): *av ukjent opphav, i breidd med, i dag, i sumar, i stand til, på lag, på ståande*

1 Genitiv med endinga *-s* er umarkert i bokmål, men ikkje i nynorsk, som føretrakkjer preposisjonsuttrykk med *til* eller andre omskrivingsmåtar (heller *huset til presten* enn *prestens hus*). Men i høgtidleg stil eller i faste og idiomatiserte uttrykk kan s-genitiv brukast utan å bli stempla som dårleg språk også i nynorsk.

*for, til nytte, til side, under berr himmel.* Uklar overgang til dei to neste gruppene, c og d (som i *i blinde*). Denne typen tenderer til å vere leksikalisert som idiom.

*Oppføring:* gjerne som UO, under nominalet (men som RD når det er gjennom-siktig, som *av ukjent opphav*).

- c. substantiv (el. adjektiv) i stivna kasusform (dativ eller genitiv): *av garde, i hende, på fote, på langs, til dels, til gode, til lags, til rors, til sals, til stades*

*Oppføring:* UO under nominalet (grunnform) er vel mest naturleg her, altså **av garde** under **gard** osv. Men det bør vere med døme på typen med tilvising under preposisjonen, altså under **av** med tilvising til **gard**. Dette gjeld særleg ved *til*, sidan det er veldig mange uttrykk med *til* + substantiv med *-s*.

- d. substantiv som berre finst i slik frase: *i går, i hel, i mak, i miss (mist), i overmorgon, i senn, i svime, om kapp, til overs, til pers*

*Oppføring:* UO under nominalet vel mest naturleg her òg.

- e. adjektiv eller funksjonsord, t.d. pronomen: *for det, i alt, i det minste, i seg sjølv, på ny, på nytt, til sist*

*Oppføring:* også her helst som UO under nominalet (i grunnform: **i alt** under **all**, **på nytt** under **ny**).

#### 5.1.4 Verbfrasar

- a. verb + partikkel (adverb, preposisjon, reflektivpronomen ev. kombinert med annan partikkel): *gie seg, gripa inn, gå av, gå for seg, gå forbi, gå fri, gå seg vill, koma i hug, lyfta opp, ta av*

*Oppføring:* UO under verbet (samla og alfabetisert i eigen tydingsbolk til slutt i verbartikkelen).

- b. same typen med lengre utfylling: *byrja med to tome hender, koma på si rette hylle, setja seg på sin høge hest, ikkje vera fødd i går, betala i dyre dommar, slå til lyd for noko, få på pelsen, leita høgt og lågt, leva i sus og dus, slå seg for brystet, halda saman i tjukt og tunt, gå som varmt kveitebrød*

*Oppføring:* RD under verbet, i alle fall normalt. Men utfyllingane er ofte så innhaldstunge at det kan høve å behandle uttrykka der, og da kjem dei i andre kategoriar i denne oversikta (som nominal- eller preposisjonsuttrykk, eller som ordpar (*sus og dus*, til dømes)).

- c. verb + objekt + eventuell utfylling: *berga livet, leggja saka daud, gjeva nokon det glatte lag, snu kappa etter vinden, stikka noko under stol, gjera*

*ein mygg til ein elefant, føra nokon bak lyset, selja skinnet før bjørnen er skoten, mjøla si eiga kake, ha sine svin på skogen, kasta perler for svin, ikkje sjå skogen for berre tre*

*Oppføring:* helst som RD, men kan stå som UO når det er eit klart idiom. Det er vanskeleg å gi generelle retningslinjer for plasseringa, fordi uttrykka er så ueinsarta, men ein bør avvege mellom verbet og det mest innhaldstunge ordet i uttrykket når dette ikkje er verbet. Fleire av dei ordtakslignande sekvensane kan ein forøvrig sjå på som setningar utan subjekt og med infinitivform, og presentere dei som RD i setningsform under det semantisk tyngste ordet, t.d. *han gav gutane det glatte lag* under **glatt**, *ein skal ikkje selja skinnet før bjørnen er skoten* under **bjørn**, *disponenten hadde sine svin på skogen* under **svin**.

#### 5.1.5 Adverbfrasar

*fritt fram, nære på*

*Oppføring:* Dette er ein vanskeleg kategori, men slike leksikaliserte samband høver vel best som UO under det første ordet.

#### 5.1.6 Interjeksjonsfrasar

*ja visst, ja så menn*

Kanskje termen ikkje er så god, men det går på interjeksjonar som *ja, nei, jau* + utfyllande tillegg.

*Oppføring:* eigne oppslag (men *så menn* høver kanskje best som UO under *så*).

### 5.2 Parataktiske uttrykk

#### 5.2.1 Ordpar

- a. med innhaldsord: *fatta og begripa, gud og kvar mann, likt og ulikt, dag og natt, kort og godt, ros og ris, skikk og bruk, fyr og flamme, velja og vraka, himmel og jord, vidt og breitt, fattig og rik*

*Oppføring:* RD (men UO når dei er faste og har idiomkarakter, altså ikkje er gjennomslittige ut frå enkeltorda) under første ordet med tilvising frå siste.

- b. med innhaldsord innbakt i preposisjonsfrase: *frå spøk til alvor, frå topp til tå, gjennom marg og bein, i flokk og fylgje, i ny og ne, i tur og orden, med hud og hår, med nebb og klør, med råd og dåd, mellom borken og veden, på godt og vondt, på nåde og unåde, på rad og rekkje, på sett og vis, til odel og eige*; med konjunksjonar: *korkje fugl eller fisk*

*Oppføring:* RD (same atterhald som ovanfor), under første innhaldsordet, med tilvising frå siste.

- c. med ord som berre (eller nesten berre) blir nytta i dette uttrykket:  
*med brask og bram, rubb og rake, skalta og valta, kreti og pleti, klabb og babb*

*Oppføring:* UO under første ordet (tilvising frå det andre trengst kanskje ikkje – truleg vil brukarar flest sjå uttrykket som ein samla heilskap og berre slå opp på det første ordet).

- d. med eit slikt ord pluss eit vanleg innhaldsord (innhaldsordet normalt, men ikkje alltid, først): *med naud og neppe, på lukke og fromme, fri og frank, kvar smitt og smule, rubb og stubb*

*Oppføring:* UO under det vanlege innhaldsordet, med tilvising frå det andre ordet dersom det står først.

- e. med funksjonsord: *av og til, i og med, i og for seg, til og frå, eine og åleine, i eitt og alt, her og der, hundre og ein*

*Oppføring:* UO under første ord eller første innhaldstunge ord.

- f. særlege uttrykk, som banningar av typen *for svarte svingande*

*Oppføring:* helst som RD under første innhaldsordet.

### 5.2.2 Frasepar

*aust, vest – heime best; ute av auga, ute av sinn; dess fleire kokkar, dess meir søl; betre føre var enn etter snar.* Desse har ofte ordtak-karakter.

*Oppføring:* RD under første ordet, ev. tilvising frå seinare nøkkelord i frasen.

### 5.2.3 Samanlikningsuttrykk

*som lus på tjørekost, som sild i (ei) tunne, som eit olja lyn, som kua i ei grøn eng, som sokken i jorda, som eld i turt gras, som fot i hose, full som ei alke, fara eller koma som eit uvêr*

*Oppføring:* RD under første innhaldsord, med ev. tilvising frå seinare nøkkelord.

## 5.3 Setningar

### 5.3.1 Setningar med innhaldstungt subjekt

*lukka er betre enn forstanden; ein fugl i handa er betre enn ti på taket; det hjarta er fullt av, renn munnen over med; nye kostar feiar best; gamal vane er vond å venda; lita tuve kan velta stort lass; naud lærer naken kvinne å spinna; tjuv trur kvar mann stel; utakk er verdas løn; ein god ting kan ikkje seiast for ofte; tida lækjer alle sår; brent barn skyr elden; mange bekker små gjer ei stor å; skipet gjekk under med mann og mus*

*Oppføring:* RD, normalt under første innhaldstunge ord. Dei fleste av desse setningane er ordtak, og dei kan ofte stå som sitat med kjelde og tilvising til andre kjelder med andre variantar av ordtaket.

### 5.3.2 Setningar med upersonleg subjekt

*det ber reint av; det er ugler i mosen; det er hipp som happ; det er eitt feitt; det spelar inga rolle*

*Oppføring:* RD under første innhaldstunge ord (i ”det er”-setningar bør ein òg kunne sløyfe ”det er” og behandle kjernen i uttrykket som eige idiom i form av eit UO, som **hipp som happ** og **eitt feitt**).

### 5.3.3 Setningar med ubestemt pronomen som subjekt (og imperativsetningar)

*har ein sagt A, så får ein seia B; pris ikkje dagen før sola går ned; ingen blir profet i sitt eige land; som ein ropar i skogen, får ein svar; ein lærer så lenge ein lever*

*Oppføring:* RD under første innhaldstunge ord.

### 5.3.4 Ordtak

av typen: *ein må hjelpa seg med det ein har, sa kjerringa, ho turka bordet med katta*

*Oppføring:* RD eller sitat med kjelde under første innhaldstunge ord, her under **hjelpa**, gjerne plassert inn under eit UO: **hjelpa seg med**.

## 5.4 Lånte uttrykk

Med det meiner vi uttrykk som er importerte som heilskapar frå eit framandt språk:

*columbi egg, coming man, in extenso, prima vista, pro et contra*

*Oppføring:* eigne oppslag. I uttrykk som består av både lånte og norske ord, vil hovudord vanlegvis vere lånte og funksjonsord norske (*i mente*); da kan ein ha eige oppslag på hovudordet med eventuell tilvising frå det (første eller viktigaste) norske ordet i uttrykket.

## 6 Oppsummering

Denne typologien er altså eit preliminært forslag frå mi side, tilpassa *Norsk Ordboks* formål og struktur. Det tek utgangspunkt i den potensielle eller reelle konflikten mellom to ulike kriterium for gruppering av fleirordsuttrykk, nemleg *grammatisk oppbygging* og *leksikaliseringsgrad*. Det grammatiske kriteriet gir klart åtskilte kategoriar, mens leksikaliseringsgrad er ein skala med diffuse overgangar. Dermed er det grammatiske kriteriet best eigna som grunnlag for ein rimeleg fast og konsistent typologi, men det er leksikaliseringsgraden som bestemmer kva status eit uttrykk har i språket, og dermed i folks språkkjensle, og denne statusen vil i stor grad spele inn når ordboksbrukarane slår opp etter eit uttrykk for å finne rett form, tyding og bruksmåte. Derfor må leksikaliseringsgraden alltid spele med også i ein grammatisk basert typologi – så vi får ein matrise der uttrykket må plasserast i forhold til to dimensjonar. Eg meiner at ein typologi som den eg har framstilt her, kan fremme ei meir presis og konsekvent behandling av

fleirordige uttrykk i ordbøker. Men 100 % konsekvent blir det aldri; den vagheita som er innebygd i sjølve språket, kan inga ordbok trylle heilt bort.

### Litteraturliste

Berkov, Valerij 1997: *Norsk ordlære*. Oslo: Universitetsforlaget.

Svensén, Bo 2004: *Handbok i lexikografi*. 2. utgåve. Stockholm: Norstedts Akademiska Förlag.

Lars S. Vikør  
professor, hovudredaktør, f. 1946  
Norsk Ordbok 2014  
Universitetet i Oslo  
Postboks 1021 Blindern  
N-0315 Oslo  
l.s.vikor@iln.uio.no

## “Bjørn Haldorsens islandske Lexicon” og Ivar Aasens ordbøker

Ivar Aasen (1813–96) was Norway’s first modern lexicographer. *Ordbog over det norske Folkesprog* (Dictionary of the Norwegian dialects) (1850) and *Norsk Ordbog* (Norwegian dictionary) (1873) were his main publications, which showed that Norwegian existed as a Scandinavian language alongside Danish, Swedish and Icelandic. Ideologically, Aasen’s project was made possible because of the language view of romanticism, which gave prominence to the dialects. Linguistically, it was based on the historical-comparative method.

This article deals with the way in which the Icelandic dictionary *Lexicon Islandico-Latino-Danicum* (1814), edited by Björn Halldórsson (1724–94), revised, and published by Rasmus Rask (1787–1832), functioned as reference work for language history in Aasen’s project, since no such source was as yet available for Norwegian.

### 1 Innleiing

Emnet for denne artikkelen er samanhengen mellom “Bjørn Haldorsens islandske Lexicon” (1814) og Ivar Aasens ordbøker (1850 og 1873).

#### 1.1 Bakgrunn

Ivar Aasen (1813–96) innleidde den moderne leksikografien i Noreg med *Ordbog over det norske Folkesprog* (heretter: ONF) frå 1850 (Worren 2006:391). Av alle føresetnader som låg til grunn for at ONF vart den første moderne ordboka for norsk språk, var oppdraget som Ivar Aasen fekk frå Det kongelige Norske Videnskabers Selskab i Trondheim i 1842, det viktigaste. Oppdraget gjekk ut på å samle inn

- ordtilfang frå dei norske dialektane med tanke på ein sams grammatikk og ei tilsvarande ordbok frå dei norske målføra og
- dessutan ei ordbok “hvori vort nærværende Skriftsprogs Ord” skulle vere oppslagsord med ekvivalentar frå dialektane (etter Walton 1996: 77).

Direksjonen i Vitskapsselskapet rådde til tre hjelpemiddel for Aasen. Det var

- “Molbechs danske Ordbog,
- Bjørn Haldorsens islandske Lexicon og
- Rasks oldnordiske Grammatik af 1832 med tilhørende Læsebog.” (Op.cit. s. 78; punktinndelinga gjord av meg. DW)

### 1.2 Spørsmål med leksikografisk tilsnitt

I høve til emnet for denne artikkelen og oppgåva til Ivar Aasen melder det seg i alle fall tre spørsmål med leksikografisk tilsnitt:

1. Kvifor rådde Vitskapsselskapet nett til “Bjørn *Haldorsens* islandske *Lexicon*”?
2. Korleis brukte Aasen denne ordboka i prosjektet sitt?
3. Kor tenleg var dette “islandske *Lexicon*” til føremålet?

## 2 Den faghistoriske bakgrunnen

### 2.1 Folkeånd og språk

Både oppdraget og hjelpemidla er prega av samtida: Oppdraget til Aasen var eit prosjekt som voks fram av tankane i romantikken om folkeånda som hadde si sterkaste uttrykksform i det folkelege språket med sikre historiske røter og med nasjonalstaten som ramme (Torp og Vikør 1993: 139 ff). Attom ønsket om å finne fram til dei historiske røtene låg ideen om ein tidlegare gullalder for språket ikkje minst i formrikdom. Forståinga av at det gammalnorske språket var grunnlaget for dei nynorske dialektane, kom opp mot slutten av 1700-talet.

Prosjektet til Vitskapsselskapet hadde opplegg i seg til å bli eit samlande nasjonalt prosjekt. Det vart det også i og med at det resulterte i eit eige skriftmål for Noreg, som vi i dag kjenner som nynorsk.

### 2.2 Om å dokumentere samanhengen mellom gullalderspråk og folkemål

Eit av hovudspørsmåla til eit slikt dokumentasjonsprosjekt må då vere å få klarlagt kva slag språkhistorisk forståing som låg bak gullaldertenkinga, og i vår samanheng kva som skulle vere det leksikografiske referanseverket for den påreknna gullalderen.

### 2.3 Den historisk samanliknande språkvitskapen og Rasmus Rask

I tiåra før Aasen-prosjektet var Rasmus Rask (1787–1832) utan tvil den lysande lingvistiske stjerna i Norden. Saman med tyskarane Jacob Grimm (1785–1863) og Franz Bopp (1791–1867) var Rask ein av grunnleggjarane av den historisk samanliknande språkvitskapen. Det skal heller ikkje underslåast at svensken Johan Ihre (1707–1780) var ein særmerkt føregangsmann på same feltet. Tenkjemåten i denne språkteorien kravde både systematisk kunnskap om fleire samtidspråk og om tilsvarande kunnskap for eldre språksteg og evne til å sjå parallellutviklingar over tid (Hovdhaugen et al 2000: 158). Rask hadde særleg utvikla teorien på dei nordiske språka. Han kom jamvel fram til at det gamle nordiske språket var “mere grammatikalsk fuldkomment end Tysken” (Rask 1811: vi).



#### 2.4 Den nordiske utfordringa

Med Rask som autoritet på dei nordiske språka vart synet hans på dei språkhistoriske samanhengane retningsgjevande. Det galdt også nemningsbruken. Ottosson (2002: 787) har i den samanhengen ei oppklarande framstilling: I komparative studiar i germansk språkvitskap har ein sidan Grimm med tilvising til Rask nytta “Altnordisch” om språket i gullalderperioden, som gjerne er rekna frå tidbolken 1100 til ca. 1350. Det bygde på at Rask kalla “det gamle Nordiske Sprog” frå denne perioden for islandsk (jf. tittelen “Vejledning til det Islandske eller gamle Nordiske Sprog” (Rask 1811)). Frå sjølve gullaldertida er derimot den første kjende fellesnemninga *dansk tunge*, som seinare i perioden vart utfordra av “nórrønt mál”, som tyder ‘det norske språket’ for det som vi i dag kjenner som islandsk og norsk. Islandsk om målet på Island kom først i bruk kring reformasjonen (Ottosson 2002: 789 f). Men for Rask var det ikkje tale om ein uheldig nemningsbruk, men uttrykk for at han meinte at samtidsmålet på Island i hovudsaka svara til målet i gullaldertida og var fellessteget for det vi i dag skil mellom som gammalisländsk, gammalnorsk, gammaldansk, gammalsvensk (og gotisk). Holm-Olsen (1981: 49) seier jamvel at Rask med dette synspunktet “for en tid [førte] forskningen på villspor”. I *Kortfattet Vejledning til det oldnordiske eller gamle islandske Sprog* (1832) hadde Rask, som tittelen syner, korrigert denne oppfatninga. Med *Um frumparta íslenzkrar túngu í fornöld* frå 1846 av Konráð Gíslason var særleg den fonologiske utviklinga mellom gammalisländsk og moderne islandsk fullgodt klargjord (Ottosson 2002: 95).

### 3 Presentasjonar av “Bjørn Haldorsens islandske Lexicon” og ONF

På dette bakteppet skal det gjevast stutte presentasjonar av “Bjørn Haldorsens islandske Lexicon” og ONF.

#### 3.1 “Bjørn Haldorsens islandske Lexicon”

Første manuskriptet til “Bjørn Haldorsens islandske Lexicon” var ferdig i 1785. Oppslagsorda var på islandsk og forklaringane på latin (Kvaran 2000: 85). Etter opplegget høyrer ordboka til det som har vore kalla latintradisjonen, som var rådande i nordisk leksikografi til innpå 1700-talet (jf. Ralph 2000: 19). Men Kvaran peikar på at forfattaren, presten Björn Halldórsson (1724–94) frå Vestfjordane i Island, var typisk for opplysningstida. Det var interessa for det islandske språket som var grunnlaget for arbeidet. Halldórsson reiste til København med manuskriptet, og i 1786 bestemte Den arnamagnæanske kommisjonen seg for å gje ut manuskriptet. Men førebuinga til utgjevinga avslørte at lemmautvalet var ufullstendig. I tillegg kunne ekvivalentar mangle, eller dei var ikkje råkande. Dermed vart manuskriptet liggjande i 25 år.

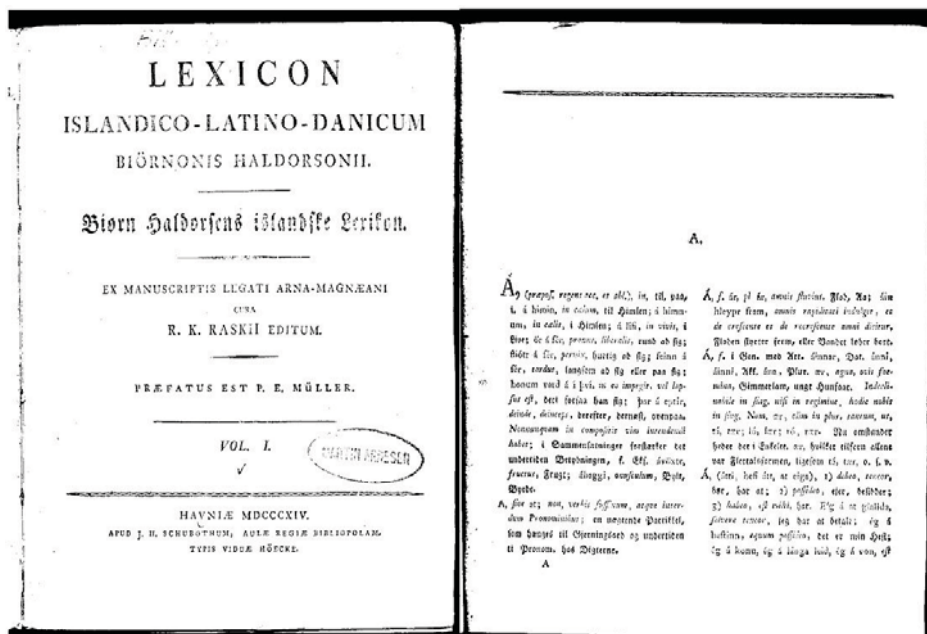
Som ein kuriositet, som syner at Vitskapselskapet i Trondheim hadde eit særleg tilhøve til “Bjørn Haldorsens islandske Lexicon”, kan det nemnast at sel-

skapet i biblioteket sitt har det første manuskriptutkastet til denne ordboka (Iversen 1928: 461), men når og korleis det kom dit, er ikkje kjent.

### 3.1.1 Utgjevinga i 1814

Manuskriptarbeidet vart teke opp att etter at P.E. Müller (1776–1834), professor i København, greidde å skaffe sponsorar til utgjevinga. Sponsorane var elles dei norske jarnverkseigarane Jacob Aall (1773–1844) og Carsten Anker (1747–1824) (jf. føreordet s. VIII).

Utgjevinga vart overlata til Rasmus Rask som fekk med seg fem islendingar til medhjelparar. I tillegg til å rette på veikskapane ved det opphavlege manuskriptet, vart det også sett til danske ekvivalentar (Kvaran 2000: 85), og ordboka fekk tittelen *Lexicon Islandico-Latino-Danicum* (heretter LILD).



Figur 1. Tittelsida og byrjinga av bokstaven A i LILD

### 3.1.2 Ordboksdelen

#### 3.1.2.1 Makrostrukturen

Dei tre språka i LILD er grafisk skilde. Islandsk står i halvfeit antikva, latin i kursiv antikva og dansk i gotisk. Ordtilfanget er i prinsippet ordna strengt alfabetisk etter oppslagsorda i det islandske utgangsspråket. Det særmerkte er samsorteringa av bokstavane U og V (U byrjar med *Vá f ... Vextr, Úfínn ... Uggvænlega, Via ...*).

### 3.1.2.2 Mikrostrukturen (lemmaordninga)

Som Kvaran (2000: 87) peikar på, er det ikkje gjeve opp ordklasser ved lemma, men substantiva er påførde genus. For verblemma er første person singularis i presens inngangsforma, her synt med verbet *aka*:

**Ek (ók, hefi ekid, at aka), ferre, veberre, currum agere, kjøre. At aka í móinn, difficultati non cedere, prop. per transversos grumos currum agere,** egentlig: at køre tværs over en opkastet Dyngje Jord, uegentlig: ikke give efter, vedblive sin Paasstand. **Aka segli, proferre, v. mutare pedem,** fire eller hale et Sejl.

### 3.1.2.3 Ordtilfanget (leksikon)

P.E. Müller gjev i føreordet (s. XI) uttrykk for ei tru på at LILD er nøkkelen for den gamle nordiske språkfelleskapen: “[Verket] aabner Adgangen til at forstaae de islandske Bøger; det oplyser tillige herligt de nordiske Sprogs ældgamle Skikkelse, og ligesom stiller os den for Øine.” Det var likevel ikkje denne tanken om islandsk som opphavsspråk (jf. pkt. 3.4) som vart drøfta i det som ser ut til å vere det einaste eigenlege faglege ordskiftet om LILD i samtida. Etter både Kvaran (2000: 85) og Magnússon (1992: xxv) skjedde det i ei bokmelding i *Dansk Litteratur-Tidende* for 1816. Meldaren var positiv til verket, men det vart klaga over at det ikkje var teke med nok ord “fra oldsproget og det gamle juridiske sprog.” Det vil seie at samtidsordtilfanget var for dominerande.

Rask svarte på kritikken, og sa at

[Bjørn Haldorsen] ... har ikke trukket nogen Grændse imellem det ældre og nyere Sprog; og vel ikke engang tænkt sig en saadan Adskillelse anderledes end i andre Sprog, hvor enkelte Ord forældes og enkelte nye opkomme, *uden at man derfor siger det er et nyt Sprog* (her sitert etter Kvaran 2000: 85 [mi kursivering, DW]).

Dette skal vel skjønast slik at ordtilfanget i LILD er bruksord i samtida utan omsyn til alder. Eit språk har alltid eit stabilt kjerneordtilfang, hevda dessutan Rask i andre samanhengar (her etter Hovdhaugen et al 2000: 161).

I nyutgåva av LILD er det elles opplyst at talet på oppslag er 28304 (Kvaran 2000:87). Dette ordtilfanget er jamført med lemmaregisteret til *Orðabók Háskólans* (heretter OH) på om lag 700 000 frå perioden 1540 til i dag. Jamføringa viser at LILD har 3664 lemma som ikkje finst i dette store lemmaregisteret. I alle fall somme av desse 3664 orda kan då vere eldre mål. Talet på belegg som lemmaregisteret berre har frå 1800- og 1900-talet, men som finst i LIDL, er på om lag 4300 lemma. Dette er venteleg det nyare ordtilfanget som kritikarane, irekna Aasen, gjerne skulle ha sett at det var skilt ut (jf. t.d. Blöndal 1924: [VII] og Walton 1996: 420). Müller hadde nok overdrivne tankar om LILD som språkhistorisk dokumentasjon, men for Aasen var LILD den tilrådde kjelda.

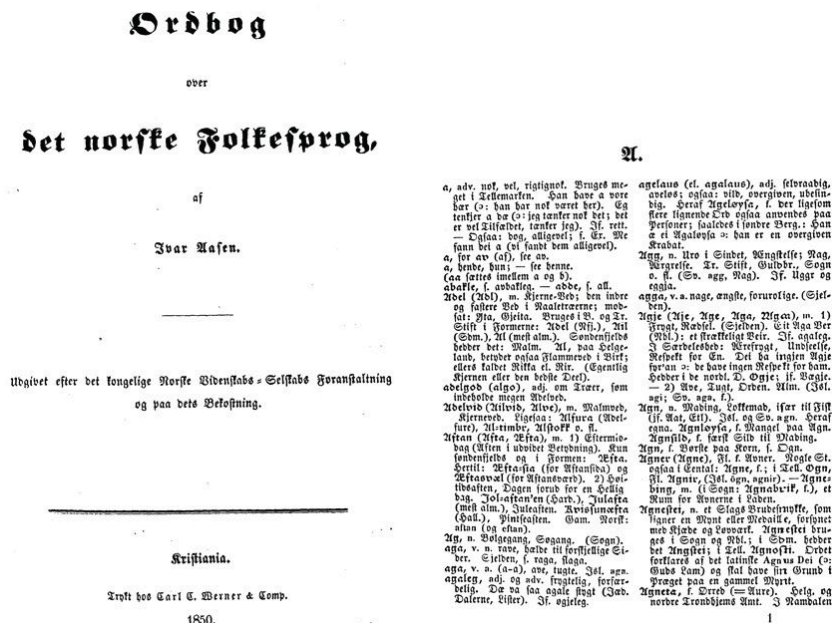
## 3.2 LILD og Aasen

For Aasen var det to hovudfasar i bruken av LILD. Det var

- i innsamlingsperioden 1842–1845 og
- i utgjevinga av *Det norske Folkesprogs Grammatik* (1848) og ONF (1850) på grunnlag av innsamlingsmaterialet

## 3.2.1 LILD i innsamlingsperioden

Eit vitnemål om at Aasen brukte LILD aktivt i innsamlingsperioden finst i dagboka der han for 28. mars 1843 har ført inn at han gjekk gjennom “Haldorsens isl. Ordbog og tilføiede isl. Ord til min søndmørske Ordbog” (BD III: 130). Den sunnmørske ordboka var grunnstommen i ordinnsamlinga hans (jf. Aarset 1994: XXXIII). Denne arbeidsmåten med å nytte andre kjelder til å identifisere nytt tilfang har Walton (2005: 351) karakterisert som Aasens effektive kumulative innsamlingsmetode.



Figur 2. Tittelsida og byrjinga av bokstaven A i ONF

## 3.2.2 Aasens vurdering av LILD som språkhistorisk kjelde for ONF

I tråd med Rask meinte Aasen i 1848-grammatikken at “den største Del af det norske Ordforraad findes igjen i de islandske Ordbøger” (Aarset 1996: 11).

Éi av desse ordbøkene Aasen siktar til, er altså LILD. Den nemner han eksplisitt i innleiinga i ONF i samband med den etymologiske informasjonskategorien. Her orsakar han seg for at han har brukt islandsk og gammalnorsk usystematisk, “uagtet det ikke kunde være min Hensigt at gjøre nogen Forskjel paa Norges

og Islands gamle Sprog, hvorimod jeg i enkelte Tilfælde havde tænkt at henvise til det Ny-Islandske” (siteret etter Kruken og Aarset 2000: 11). Men stoda var at “[d]en fuldstændigste Ordbog, som nu havdes ved Haanden, var Haldorsons islandske Lexikon, ved hvilket der imidlertid er den Mangel, at Kilderne ikke findes angivne, og at man altsaa ikke deraf kan see, om Ordet er nyt eller gammelt” (op.cit. s. 12).

### 3.2.3 Stutt presentasjon av ONF

ONF er ei eittbands tospalta ordbok på 634 sider. Talet på oppslagsord om lag 25 000, irekna ord på ikkjealfabetisk plass (Kruken og Aarset 2000: XXII). Oppslagsorda er norske målføreord som er lett standardiserte etter uttale, historisk opphav og skriftspråkskonvensjonar. Forklaringane er på dansk.

## 4 LILD i ONF

Det kan finnast spor i ONF frå LILD i alle informasjonskategoriane i ein ordartikkel, som *Nordisk leksikografisk ordbok* nemner som dei formelle, dei semantiske, dei kombinatoriske, dei ensyklopediske og dei historiske. Hovudoppgåva her er å sjå på i kva grad LILD har hatt innverknad på den historiske informasjonskategorien. I den samanhengen skal det gjerast greie for ei lita eiga gransking.

### 4.1 Dei ulike referansemåtane

Dei ulike måtane som LILD er brukt til språkhistorisk referanseverk på i ONF, må sjåast i samanheng med kjennskapen Aasen hadde til det mangslungne ordtilfanget i LILD. På den bakgrunnen er det i alle fall fire tenkjelege tilvisingsmåtar med utgangspunkt i LILD når Aasens hovudmål med ordskyldskapen var å føre flest mogleg ord attende til gammalnorsk:

- beinveges referanse til verket LILD
- referanse til islandsk
- referanse til nyislandsk
- referanse til gammalnorsk

Dei tre første referansemåtane skulle vere sjølvgejevne. Når referanse til gammalnorsk er med, er det særleg på grunnlag av Aasens utsegn om at nemningane islandsk og gammalnorsk er usystematisk brukte, slik at ein gammalnorsk referanse kan ha utgangspunkt i LILD.

Med hjelp av word-filer som var prentegrunnlag for nyutgjevinga av ONF i 2000 (Kruken og Aarset 2000), er det gjort søk på desse fire typane. Søkjestrengane har utgangspunkt i korleis Aasen nemnde desse fire kategoriane, særleg i lista over forkortingar.

## 4.1.1 Beinveges referanse til LILD

Av moglege søkjestrengar som kunne gje direkte tilslag til ordboka LILD (tittel, forfattar), var det berre søkjestringen “Haldorson” som faktisk gav tilslag, og det var berre fire slike tilslag i ordboksdelen i ONF: *gil* m, *ia* f, *kleiv* f og *munnga* n. Figur 3 jamnfører *ia* i ONF og *yða* i LILD:

ONF	LILD
<p><b>Ia</b> (for <i>Ida</i>), f. Strømning eller Hvirvel i Vand; især Bagstrøm, den tilbagegaaende Strømning i Bugterne af en Elv. Meget udbredt Ord, men i forskjellig Form, nemlig: <i>Ia</i> (Nhl. Hard.), <i>Ie</i> (Sfj. Tell. Rbg.), <i>Ea</i>, <i>Bak-ee</i> (Helgeland, Trondhjem), <i>Eaa</i> (Ndm.), <i>Odo</i>, for <i>Eda</i> (Namd.), <i>Udu</i> (Ørk.), <i>Bak-vudu</i>, og <i>Kringvudu</i> (Indr.); ogsaa <i>Bakeie</i> forekommer ved Trondhjem. Isl. <i>iða</i> (Gretters Saga, Kap. 69), <i>yða</i> (hos Haldorson). I svenske Dial. <i>ea</i>, <i>ia</i>, <i>ier</i>. Tingen hedder ellers <i>Evju</i> (Ag. Stift) og <i>Ave</i> (Helg.). I Sdm. hvor det skulde hedde <i>Ide</i> (aab. <i>i</i>), er dette Navn ubekjendt; jf. <i>Atterekje</i>, <i>Atterenme</i>, <i>Attebera</i>. I Nhl. er <i>Ia</i> en mindre Strøm end <i>Attebera</i>. /Ogsaa i svensk: <i>ida</i>./</p>	<p><b>Yda</b>, f. <i>æstus aquarum v. fluminum</i>, Vandhvirvel, stærk Strøm i Floder og Elve; å <b>unn</b>.</p>

Figur 3. Beinveges referanse til “Haldorson” i ONF og motsvaret i LILD

Dei få tilvisingane til “Haldorsen” dokumenterer i alle fall at dette tilslaget er tett knytt til islandsk, som er nemnt ved alle fire tilslaga.

## 4.1.2 Referanse til islandsk

Den viktigaste strengen på søket for islandsk var forkortinga “Isl.”. Her var tilslaga så mange at søka vart avgrensa til bokstaven *A* i ONF, som har om lag 540 hovudartiklar. 73 av desse artiklane hadde ein referanse til islandsk, der somme er jamføringstilvisingar (med “Jf.”). Alle desse tilslaga har eit motsvar i LILD. Eit opplagt døme kan vere:

ONF	LILD
<p><b>avla</b>, v.a. (<i>a - a</i>), 1) avle, indhøste, faae af Jorden. (Lidet brug.) 2) formaae, s. <i>evla</i>. 3) forvolde, paaføre En noget, især noget ondt. Sogn, Sdm. og fl. <i>Han avla dei ei Skjæmsla</i>): han forvoldte dem en Irettesættelse, klagede paa dem saa at de bleve overskjendte. (Isl. <i>afla</i>, skaffe).</p>	<p><b>Afla</b> (<i>afladi</i>, at <i>afla</i>), <i>paro</i>, <i>comparo cum Gen.</i>, erhverver, skaffer sig. 2) <i>possum</i> formaar, kan. <b>Hann fær þess ecki aflad</b>, <i>non potest</i>, det er han ikke i Stand til.</p>

Figur 4. Referanse til islandsk i ONF og motsvaret i LILD

Når ein jamnfører både skrivemåte, ordklasse bøyning og tyding, er det ingen sterk

påstand å hevde at grunnlaget i ONF for den islandske referansen er oppføringa i LILD.

#### 4.1.3 Referanse til nyislandsk

I heile ONF er det berre ei tilvising til nyislandsk. I bolken *Rettelser og Tillæg* under oppslaget *Vare* med forma *Vara* (i 1850-utgåva) heiter det at det blir brukt på same måten i nyislandsk, “f. Ex. *vara-alþingismaðr* (Suppleant til Althinget), *vara-uppstúninga* (subsidiært Forslag)”. Til *vara*-‘nest-’, ‘i staden for’ ser det ikkje ut til å vere noko motsvar i LILD.

#### 4.2.4 Referanse til gammalnorsk

Avgrensa til bokstaven *A* er det 13 referansar til gammalnorsk (“G. N.”). Også ved desse orda finst motsvara i LILD. Samansetninga *andlát* kan stå som døme på det nære sambandet mellom oppslag i ONF med gammalnorsk referanse og eit klart motsvar i LILD:

ONF	LILD
<b>Andlaat</b> , n. Dødsstund, de sidste Aandedræt. <i>Han laag i Andlaataa</i> (dat. pl.) B. Stift. G. N. <i>andlát</i> , af <i>láta</i> , tabe.	<b>Andlát</b> , n. mors, <i>amissio spiritus</i> , Død, Aandens Opgivelse.

Figur 5. Referanse til gammalnorsk i ONF og motsvaret i LILD

Vel kan det peikast på tydingsnyanse, men i form er likskapen så godt som fullkomen.

## 5 Om kor tenleg LILD var som språkhistorisk referanseverk

Både momenta frå den faghistoriske skissa og Aasens bruk av LILD som kjelde gjev grunnlag for å drage nokre slutningar om den rolla LILD spelte som fagleg vegvisar.

### 5.1 Openberre bruksmåtar

Det pragmatiske svaret på kor tenleg LILD var til føremålet som språkhistorisk referanseverk, er at det eigenleg var den einaste brukeleg kjelda som kunne nyttast som praktisk hjelpemiddel særleg om vestnordisk.

Den faglege metoden var å nytte oppføringa i LILD som grunnlag for at norske dialektord hadde historiske røter til gammalnorsk. Om dei til saman 95 (73+12) referansane til islandsk og gammalnorsk av meir enn 540 oppslag under bokstaven *A* i ONF godtgjorde at dei norske dialektane er etterkomarar etter eit gullalderspråk, skal ikkje avgjerast her. Men fordelinga mellom dei to nemningane “Isl.” og “G. N.” syner i alle fall at ONF hadde vore svært mangelfull i dokumentasjonen av den viktige historiske kategorien om ikkje LILD-referansane til “Isl.” kunne takast som indisium på den historiske samanhengen til gammalnorsk.

Jamvel om Molbechs *Dansk Ordbog* var hovudmønsteret i utforminga av ordartiklane (Worren 2006), var også LILD ei kjelde som synta aktuelle utformingar både i den formelle og semantiske kategorien. For Aasen var det ei utfordring å velje ei tenleg form for dei norske målføreorda. I LILD kunne han finne utgangspunkt i samsvarande ord, jf. t.d. mangfaldet ved *Ia*.

### 5.2 Avløysarar

Med den openberre veikskapen at LILD ikkje skilde mellom nytt og gammalt ordtilfang, var det sjølvstøtt ikkje uventa at det kom nye ordbøker for det eldste skriftfeste steget både for islandsk og norsk. I 1864 kom *Oldnordisk Ordbog* av Eiríkur Jónsson, og i 1867 hadde Johan Fritzner fullført *Ordbog over Det gamle norske Sprog*. Dette verket fekk stor innverknad på nyutgåva av ONF som kom i 1873, også med ny tittel: *Norsk Ordbog* (NO).

For nemningsbruken “islandsk” i høve til gammalnorsk som opphavsspråk fekk tilgangen til Fritzners ordbok dramatiske følgjer. Ved 53 av dei 73 orda under bokstaven A som hadde “islandsk” som referanse i ONF er islandsk bytt ut med gammalnorsk, som t.d. ved *agn*:

ONF	LILD	NO
<b>Agn</b> , n. Mading, Lokkemad, især til Fisk (jf. <i>Aat, Etl</i> ). Isl. og Sv. <i>agn</i> . Heraf <i>egna</i> .	<b>Agn</b> , n. <i>esca, illecebræ, inescamentum</i> , Madding, Lokkemad 2) <i>piscatura, capatura ferarum</i> , Fiske- Fugle-Fangst, Jagt.	<b>Agn</b> , n. Mading, Lokkemad, især til Fisk. (Jf. <i>Aat, Aata, Etsl.</i> ). Ogsaa G.N. og Sv. Heraf <i>egna</i> , v.

Figur 6. Eit døme på føringa av *agn* i ONF, LILD og NO

Berre eitt av orda har fått behalde “Isl.”, og det er jamføringsforma *álfabruni* under *alveld* som er ein hud- og dyresjukdom. På den andre sida har NO tilvisingar til islandsk ved elleve andre ord under bokstaven A, der ONF ikkje hadde det.

## 6 Avslutning

Når Vitskapsselskapet i Trondheim tilrådde LILD for Aasen, var det eit val utifrå både ideologiske, faglege og pragmatiske vurderingar. Med det nære sambandet mellom islandsk og norsk var LILD den leksikografiske kjelda som kunne syne den språkslektskapen som ville gje prestisje til eit eige norsk mål. I vitskapleg samanheng hadde ordboka også prestisjen frå Rasmus Rask. Under innsamlingsferda til Aasen galdt også den praktiske grunnen at LILD hadde eit format og ein storleik som gjorde det mogleg å ha henne med på innsamlingsferda. Faglege forsvarleg leksikografi i høveleg format blir aldri avleggs.

Jamvel om det kan visast til veikskapar ved LILD, vil denne fleirspråklege islandske ordboka, som endra forklaringsperspektivet frå latin til eit nordisk mål, for alltid stå som ei av grunnkjeldene for norsk morsmålsleksikografi. I notidsperspektivet er LILD også ei påminning av verdien av mangfaldet i det nordiske



språkfellesskapet. Dette mangfaldet er det både faglege og kulturelle grunnar til føre vidare.

## 7 Litteraturliste

- BD III = Djupedal, Reidar 1960: *Ivar Aasen. Brev og Dagbøker*. Oslo: Det Norske Samlaget.
- Blöndal, Sigfús 1924: *Islandsk-dansk ordbog*. Reykjavík: Prentsmiðjan Gutenberg.
- Bondevik, Jarle, Oddvar Nes og Terje Aarset 1998: *Målsamlingar frå Trondhjems og Tromsø Stifter av Ivar Aasen*. Bergen: Norsk Bokreidingslag L/L.
- Holm-Olsen, Ludvig 1981: *Lys over norrøn kultur*. Oslo: J.W. Cappelens Forlag.
- Hovdhaugen, Even, Fred Karlsson, Carol Henriksen og Bengt Sigurd 2000: *The History of Linguistics in the Nordic Countries*. Helsinki: Societas Scientiarum Fennica.
- Indrebø, Gustav 1951: *Norsk målsoga*: John Griegs boktrykkeri: Bergen
- Iversen, Ragnvald 1928: Et islandsk ordboksverk fra slutten av 1700-tallet. I: Johs. Brøndum-Nielsen, Elof Hellquist, O. F. Hultmann, Sigurður Nordal og Magnus Olsen (red.): *Festskrift til Finnur Jónsson*. København: Levin & Munksgaards Forlag, 455–62.
- Jónsson, Jón Aðalsteinn 1992: Formáli. I: *LILD 1992*, ix–xxix.
- Kruken, Kristoffer og Terje Aarset 2000: *Ivar Aasen. Ordbog over det norske Folkesprog*. Ny utgåve. Oslo: Det Norske Samlaget.
- Kvaran, Guðrún 2000: Det attende århundredes islandske leksikografi. I: *LexicoNordica 7*: 75–90.
- LILD = [Halldórsson, Björn] 1814: *LEXICON ISLANDICO-LATINO-DANICUM*. Vol. I–II. Havniæ: J.H. Scubothum.
- LILD 1992 = Björn Halldórsson. *ORÐABÓK Íslensk – latnesk – dönsk*. [Reykjavík:] Orðabók Háskólans.
- Magnússon, Friðrik 1992: Athugun á orðaforða orðabókar Björns Halldórssonar. I: *LILD 1992*, xxxi–xlvi.
- Nordisk leksikografisk ordbok* 1997. Oslo: Universitetsforlaget.
- Ottosson, Kjartan 2002: Old Nordic: A definition and delimitation of the period. I: Oskar Bandle (red.): *The Nordic Languages*. Berlin: Walter de Gruyter, 787–793.
- Ralph, Bo 2000: Svensk lexikografitradition. I: *LexicoNordica 7*: 5–22.
- Rask, Rasmus Kristian 1811: *Vejledning til det Islandske eller gamle Nordiske Sprog*. København: Schubothes Forlag.
- Torp, Arne og Lars S. Vikør 1993: *Hovuddrag i norsk språkhistorie*. Oslo: Ad Notam Gyldendal.
- Venås, Kjell 1996: *Då tida var fullkomen. Ivar Aasen*. Oslo: Novus forlag.
- Walton, Stephen 1996: *Ivar Aasens kropp*. Oslo: Det Norske Samlaget.
- Walton, Stephen 2005: Ivar Aasens kumulative metode og leksikografien. I: Ruth Vatvedt Fjeld og Dagfinn Worren (red.): *Nordiske studiar i leksikografi 8. Rapport frå Konferanse om leksikografi i Norden. Volda 20.–24. mai 2003*. Oslo, 348–355.

Worren, Dagfinn 2006: Molbech som mønster for Aasen. I: Henrik Lorentzen og Lars Trap-Jensen (red.): *Nordiske studier i leksikografi 8. Rapport fra Konferanse om leksikografi i Norden. Sønderborg 24.–28. mai 2005*. København, 391–406.

Aarset, Terje 1994: Innleiing. I: Jarle Bondevik, Oddvar Nes og Terje Aarset (red.): *Målsamlingar frå Sunnmøre av Ivar Aasen*. Bergen: Norsk Bokreidingslag, XVII–XLIV.

Aarset, Terje 1996: *Ivar Aasen. Det norske Folkesprogs Grammatik*. Volda: Høgskulen i Volda.

Dagfinn Worren  
professor, hovudredaktør, f. 1944  
Norsk Ordbok 2014, Universitetet i Oslo,  
Postboks 1021 Blindern  
NO-0315 Oslo  
dagfinn.worren@iln.uio.no

## Stikkordregister

### A

alfabetiske ordbøger 409  
 algoritmer 453  
 allmennordbok 132, 246  
 antonym 275, 279, 363  
 anvendarkontexter 147  
 anvandarstudier 143, 145  
 anvandarundervisning 149  
 aoristisk betydning 222  
 arbejdsgrænseflade 204  
 argumentstruktur 45, 140  
 artikelinterne tæk 73  
 artikelnavigering 456  
 artik(k)elstruktur 36, 70, 155, 410  
 aspekt 161  
 autentisk citat 176, 177, 241  
 autentisk eksempel 125, 316, 347, 441

### B

baggrundssøgning 455  
 baglænsordbog 450  
 barneordbog 372  
 begrebsordbøger / begrebsordbøger 271, 363  
 begrebssystemer 416  
 begrepsfelt 271  
 begrepsoverskrifter 276  
 belæg 330  
 beskrivende ordbøger 232  
 bestemthet 225  
 bestemthetsmarkør 225  
 betydelse / betydning  
   aoristisk betydning 222  
   delbetydninger 136, 270  
   perfektivisk fortidsbetydning 222  
   perfektivisk nåtidsbetydning 222  
 betydelsevarianter 332  
 betydningsbeskrivelse / betydelsebeskrivning 232, 245  
 betydningsdefinition 232  
 betydningsdifferensiering 270

betydningsforklaring 280  
 betydningskonstruktion 115  
 bevingede ord 9, 119, 373  
 bilingvale ordbøger / bilingvala ordböcker 183, 301, 324  
 bitransitiv konstruktion 318  
 bokmål 131  
 Bokmålsordboka 132  
 browser 213, 448  
 brugergrænseflade 206, 363, 448  
 bruk i referenssituationer 146  
 brukseksempel 92, 136, 220, 231, 469  
 bøjningsinformasjon / böjningsinformati-  
 on 91, 430  
 bøjningskoder 220  
 båndoptagelser 175, 178

### C

Center for Sprogteknologi 354  
 collocations 77  
 context-induced reinterpretation 77

### D

DanNet 45, 56, 353  
 dansk-latinske ordbøger 409  
 dansk tunge 479  
 dataintegritet 32, 39, 40  
 datamaskinell leksikografi 133  
 deduktiv tilnærming 36  
 defekte verb 226  
 definitions-/definisjonsordbok 244, 270, 310, 372, 465  
 delbetydninger 136, 270  
 Den Danske Ordbog 46, 135, 231, 353, 421, 448  
 deskriptiv leksikografi 29  
 deskriptiv ordbok 100  
 Det Danske Sprog- og Litteraturselskab 354  
 dialekt 4, 36, 101, 173, 210, 240, 376, 399, 478

dialektordbog/-ordbok 173, 209, 376  
 digital ordbog/ordbok 105, 209, 353, 447  
 digital sorterar 32, 35  
 ditransitiva konstruksjoner 332  
 divalente verber 56  
 dokumentasjonsordbok 40, 156  
 dokumentasjonsprosjekt 478  
 DOXA 138

## E

e-ordböcker 246  
 editor 140, 448  
 eksportere data 75  
 eksempel / eksempel 139, 176, 201, 245, 343, 426  
   autentisk eksempel 125, 316, 347, 441  
   brukseksempel 92, 136, 220, 231, 469  
   exempelsatser 330  
   konstruert eksempel 23  
   konstruksjonseksempel 454  
   redaksjonelt /redaksjonelt eksempel 176, 466  
   språkbruks-/sprogbrugseksempel 177, 197, 466  
   språkekspepel 244, 316, 440  
 eksterne henvisninger 215  
 ekvivalent / ækvivalent 61, 479  
 elektronisk korpus 91, 131  
 elektronisk ordbok 34, 108, 137, 199, 243, 272, 305, 438, 470  
 elektroniska ordbokskällor 148  
 elektroniskt lexikon 443  
 Emergent Grammar 77  
 emnesøgning 362  
 empirisk etterprøvbart 40  
 encyklopedisk informasjon 346, 435  
 encyklopedisk søkning 249  
 enkeltord 273  
 enspråk(l)ig ordbok 141, 220, 333  
 entydiggjøring 275  
 ettordslemma 140, 276, 466  
 etymologisk informasjon 140  
 etymologisk ordbok 102  
 EuroWordNet 55, 354  
 eksempelsatser 330  
 explicit informasjon 441

## F

fagleksikografi 385  
 fagordbøger 458  
 fag- og stilmarkører 456  
 fasta konstruksjoner 249  
 faste ordforbindelser 272  
 faste vendinger 114  
 fleirordsuttrykk 474  
 fleirspråkleg ordbok 486  
 flerordsenhet/-hed 3, 115, 273, 435  
 flerordsforbindelse 114, 203, 277, 448, 461  
 fler-/fleirordslemma 137, 465  
 flerordssøkning 250, 275  
 flerordstermer 436  
 flersprogede idiomordbøger 113  
 flertydig lemma 140, 275  
 forkortelser 72  
 frasem-lemma 10  
 fraseologi 114, 207, 271, 465  
 fraseologisk enhed 1  
 fraseologisk ordbog/ordbok 207, 273  
 fraseologiske relasjoner 272  
 fraseologisme 2, 120  
 fraseverb 157  
 fremmedord 19, 94, 181, 183, 287, 373  
 fremmedordbok 138  
 fritextsøgning / fritextsøkning 255, 456  
 function relation 50  
 funksjonalitet 283  
 funksjonel leksikografi 113  
 funksjonel valens 47  
 funksjons-/funksjonsord 93, 157, 176, 311, 468  
 funksjonsunderstøttende data 386  
 funksjonsunderstøttende dokument 396  
 funksjonsunderstøttende omtekst 383  
 funksjonsverbsfraser 318  
 fuzzy search 453  
 fyllnadsled 328, 330  
 færdigredigering 70

## G

gammaldansk 479  
 gammalislandsk 479  
 gammalnorsk 479

gammalsvensk 479  
 genuine troponymy 50  
 genus proximum 357  
 grammaticalization / grammatikalisering  
 77, 155, 468  
 grammatikalske konstruksjonar 166, 185  
 grammatikk 156  
 grammatiske konstruksjonsmønster 162  
 grensesnitt 141  
 grundstruktur 327  
 grunnord / grundord 72, 93, 134, 160  
 grunnskolen 372  
 grænseflade 204, 212, 363, 448  
 arbeidsgrænseflade 204  
 brugergrænseflade 206, 363, 448  
 redaktionel grænseflade 204

## H

hapax legomenon 139  
 henvisning / hänvisning 28, 97, 137, 201,  
 211, 220, 238, 250, 270, 318, 415, 427  
 eksterne henvisninger 215  
 interne henvisninger 215  
 kildehenvisninger 139, 216  
 krysshenvisninger 98  
 henvisningsartikler 453  
 henvisningsstruktur 447  
 historisk informasjon/information 102,  
 125, 137  
 historisk ordbog 376, 429  
 historisk samanliknande språkvitskap 478  
 hjelpeverb 157  
 homonym 205, 211, 277, 312  
 homonymi 275, 313  
 homonymindeks 277  
 homonymseparering 272  
 hyperlinks 121, 447  
 hyperonym 114, 140, 363, 438  
 hyperonymi 57  
 hyponym 363, 438  
 kohyponymer 140, 365  
 hyponymi 44  
 hyponymirelasjon 273  
 hypotaktiske uttrykk 468  
 håndordbog 132

## I

idiom 4, 37, 75, 81, 113, 137, 185, 201,  
 241, 250, 274, 315, 422, 438, 468  
 idiomatiseret forbindelse 230  
 idiomatisk betydning 114, 233  
 idiomatisk uttryk 4, 180, 221  
 idiomatiske vendinger 3  
 idiomleksikografi 116  
 idiomordbog 113  
 implicit information 435  
 inanimat referent 225  
 individuelt bruk 146  
 induktiv tilnærming 36  
 infinitiv 163  
 infinitivsobjekt 164  
 informasjonskategori 483  
 informasjonstype 37, 138  
 information / informasjon 143, 159, 185,  
 210, 219, 269, 376, 388, 439, 456  
 bøyninginformasjon / böjningsinfor-  
 mation 91, 430  
 encyklopedisk informasjon 346, 435  
 etymologisk informasjon 140  
 explicit information 441  
 historisk informasjon / information  
 102, 125, 137  
 implicit information 435  
 kombinatorisk informasjon 221  
 leksikografisk informasjon 119  
 leksikalisk informasjon 147  
 metaspråklig informasjon 443  
 orddelingsinformasjon 260  
 språk-/språklig informasjon 147, 244  
 valensinformasjon 327  
 informations-användning 144  
 informationsbehov 144, 291  
 informationsbeteende 143  
 informationsreserver 150  
 informasjonssystem 150  
 informasjonssökning/-søgning / infor-  
 masjonssøking 58, 133, 144, 202  
 informasjonssökningsprocess 144  
 informasjonssökningsstigar 149  
 inlärningsordböcker 151, 246  
 innhaldsverb 157

interkulturel viden 303  
 interne henvisninger 215  
 Internet-ordbok 243  
 islandsk-nordiske ordbøger 197  
 ISLEX-databasen 202  
 ISLEX-ordbogen 197

## J

jokertegn 448  
 Jysk Ordbog 1, 210

## K

kildehenvisninger 139, 216  
 kildeprog/-språk 183, 197, 220, 284, 409  
 kjerneordtilfang 481  
 kohyponymer 140, 365  
 kollokasjon / kollokation 3, 37, 137, 177, 205, 224, 241, 250, 272, 434, 468  
 kollokasjonsordbøker 272  
 kollokat 41, 157  
 kollokator 434  
 kombinatorisk informasjon 221  
 kompetenskontroll 147  
 konkordans 139, 203  
 konsistens 156  
 konstruert eksempel 23  
 konstruksjoner/-ar / konstruksjoner 223, 272, 332, 467  
   ditransitiva konstruksjoner 332  
   fasta konstruksjoner 249  
   grammatikalske konstruksjoner 166, 185  
   passivkonstruksjoner 164  
   perifrastiske konstruksjoner 162  
   syntaktiske konstruksjoner 331  
   tillfälliga konstruksjoner 249  
   verbale konstruksjoner 276  
   verbkonstruksjoner 309  
 konstruksjonsmønster 46, 278  
 grammatiske konstruksjonsmønster 162  
 syntaktiske konstruksjonsmønstre 43  
 verbale konstruksjonsmønstre 278  
 konstruksjonsordbøker 272  
 konstruksjonseksempel 454  
 konstruksionsgrammatik 332

konstruksjonsordbøker 284  
 konstruksjonsprocess 151  
 konventionaliseret forbindelse 230  
 korpus 143, 170, 436  
   elektronisk korpus 91, 131  
   talespråskorpus 135  
   tekstkorpus 92, 198, 382  
 Korpus 2000 424  
 korpussøgning 390  
 kotekstpartner 279  
 krydsordshjælp 366  
 krysshensvisninger 98  
 kulturonymer 303  
 kvaliastruktur 133  
 kvalitetskontroll 34, 39

## L

L1-ordbog 301  
 L2-L1-ordbog 284  
 latinsk-danske ordbøger 409  
 latintradisjonen 479  
 learner-/lørnerordbøker 301, 354, 365  
 leksikalisering 469  
 leksikaliseringsgrad 468  
 leksikalsk beskrivelse 133  
 leksikalsk folkeopplysning 132  
 leksikalsk resurse 43  
 leksikalske huller 61  
 leksikografi / lexikografi 116, 156  
   datamaskinell leksikografi 133  
   deskriptiv leksikografi 29  
   fagleksikografi 385  
   funksjonel leksikografi 113  
   idiomleksikografi 116  
   metaleksikografi 283  
   preskriptiv leksikografi 29  
   teoretisk leksikografi 140  
   tospråklig leksikografi 219  
   valenslexikografi 327  
   virksomhedsleksikografi 385  
 leksikografisk beskrivelse 90, 270

- leksikografisk dokumentskabelonmodel 384, 386
- leksikografisk information 119, 269
- leksikografisk konsept 126
- leksikografisk konvergensmodel 383, 386
- leksikografisk relevans 300
- leksikografisk repræsentation 232
- leksikografiske behov 298
- leksikografiske data 124, 269
- leksikografiske funktioner 117, 395
- leksikografiske oplysninger 199
- leksikografiske vidensbehov 301
- leksikologisk analyse 125
- leksikologisk beskrivelse 122
- lemma 134, 156, 270, 312
  - ettordslemma 140, 276, 466
  - fler-/fleirordslemma 137, 465
  - flertydig lemma 140, 275
  - frasem-lemma 10
  - metalemma 99
  - sublemma 155, 230, 317
  - verballemmaer 47
  - verblemma 161
- lemmabestand 270
- lemmakandidat 139
- lemma-lexemmodellen 429
- lemmaliste 135, 278, 294
- lemmaluckor 247
- lemmaregister 481
- lemmaseleksjon/-seleksjon 270, 283
- lemmastruktur 22
- lemmategnform 139
- lemmatisering 270, 303
- lemmatisering av flerordsforbindelser 277
- lemmatisk enhet 270
- lemmautval(g) 93, 138, 479
- Levins verbalklasser 55
- lexem 421
- lexicalization 78
- lexical verbs 77
- lexikaliserade fraser 422
- liknelser 422
- logfil 70, 243
- lommeordbøker 219
- læreplan 372
- læseordbøger 284
- lântagarbruk 146
- ## M
- makrostruktur 275, 286, 334, 411, 447
- markør 303, 456
- maskinlæsbar ordbok 137
- materialanalyse 32
- maximalistisk 331, 334
- mediostruktur 447
- meronymer 363
- metalemma 99
- metaordbok 91, 106
- Metaordboka 31, 378
- metaspråklig information 443
- metodisk mangfold 37
- mikrostruktur 271, 415, 441, 447
- minimalistisk 334
- modale hjelpeverb 161, 163
- modalitet 161
- monolinvale ordbøger 302
- morsmålsfaget 373
- morsmålsordbøker / modersmålsordbøger 158, 285
- multisproglig ordbog 198
- mönstersökning 442
- målføreord 486
- målsprog 197
- ## N
- netbaserede ordbøger 214
- netordbog / nätordbok 215, 248, 447
- Nordisk leksikografisk ordbok 447, 483
- nordiske middelalderordbøker 91
- normativ ordbok 100, 132
- normering 19, 132, 378
- nórrønt mál 479
- Norsk elektronisk bokmålsordbok (NEBO) 137
- Norsk ordbank 135
- Norsk Ordbok 31, 96, 105, 131, 155, 380, 465
- nynorsk 375, 478
- Nynorskordboka 102, 132, 163, 403
- nyord 134
- nyordsmateriale 134

nyordsordbøger 286  
 nærsynonym 275, 362  
 nøkkelordregister 274

## O

online-ordbog 201, 448  
 onomasiologisk 138, 276, 353, 436  
 ontologi 138, 354  
 ontologiske modeller 354  
 ontologisk type 357  
 opphavsspråk 486  
 ordbog /ordbøger (typer mm.)  
   alfabetiske ordbøger 409  
   allmennordbok 132, 246  
   baglænsordbog 450  
   barneordbog 372  
   begrebsordbøger / begrepsordbøker 271, 363  
   beskrivende ordbøger 232  
   bilingvale ordbøger / bilingvala ordböcker 183, 301, 324  
   dansk-latinske ordbøger 409  
   definitionsordbok / definisjonsordbok 244, 270, 310, 372, 465  
   deskriptiv ordbok 100  
   dialektordbog/-ordbok 173, 209, 376  
   digital ordbog/ordbok 105, 209, 353, 447  
   dokumentasjonsordbok 40, 156  
   e-ordböcker 246  
   elektronisk ordbok 34, 108, 137, 199, 243, 272, 305, 438, 470  
   enspråk(l)ig ordbok 141, 220, 333  
   etymologisk ordbok 102  
   fagordbøger 458  
   fleirspråkleg ordbok 486  
   fraseologisk ordbog 207, 273  
   fremmedordbok 138  
   historisk ordbog 376, 429  
   håndordbog 132  
   idiomordbog 113  
   inlæringsordböcker 151, 246  
   Internet-ordbok 243  
   islandsk-nordiske ordbøger 197  
   kollokasjonsordbøker 272  
   konstruksjonsordbøker 272

konsultationsordbøger 284  
 L1-ordbog 301  
 L2-L1-ordbog 284  
 latinsk-danske ordbøger 409  
 learner-/lørnerordbøger 301, 354, 365  
 lommeordbøker 219  
 læseordbøger 284  
 maskinlæsbar ordbok 137  
 metaordbok 91, 106  
 monolinvale ordbøger 302  
 morsmålsordbøker / modersmålsordbøker 158, 285  
 multisproglig ordbog 198  
 netbaserede ordbøger 214  
 netordbog / nätordbok 215, 248, 447  
 nordiske middelalderordbøker 91  
 normativ ordbok 100, 132  
 nyordsordbøger 286  
 online-ordbog 201, 448  
 pappersordböcker / papiroidbög/-bok 243, 283, 305, 448, 470  
 passivordbog 284  
 produktionsordböcker 337  
 receptionsordbog 284, 337  
 rejseordbøger 291  
 rettskrivningsordbok / retskrivningsordbøger 19, 259  
 rimordbog 456  
 samtidsordbok 422  
 skolordbok 244  
 språkteknologisk ordbok 136  
 synonymordbøker 222, 271  
 sætningsordbøger 292  
 tematiske ordbøger 409  
 tospråklig ordbok / tosproglig ordbog 207, 219, 372  
 tosprogsordbøger 285  
 tryckta ordböcker / trykte ordbøker/-bøger 148, 270, 447  
 turistordbøger 291  
 valensordbok 327  
 virksomhedsordbøger 385  
 webordbog 197  
 ordbogs-/ordboksbase 139, 163, 198, 416  
 ordbogsfunktion 297  
 ordbogslokalisering 299



ordbogsstandardisering 298  
 ordboksanvåndare 143, 316, 427  
 ordboksanvåndning 143, 243  
 ordboksartikkel / ordbogsartikel 177,  
 199, 221, 232, 250, 270, 283, 309, 333,  
 421  
 ordboksbruk 373  
 ordboksbrukar(e) / ordbogsbruger 168,  
 224, 261, 287, 292  
 ordboksgrammatikk 220  
 orddelingsinformasjon 260  
 ordfelt 279  
 ordforråd 202, 259, 272, 284, 366, 399, 416  
 ordpar 274, 422  
 ordsamband 155  
 ordsamling 37, 105, 176, 449  
 ordspråk/-sprog 1, 373, 422  
 organisationskultur 384  
 overbegreb 65, 74, 356

## P

pappersordböcker / papirordbog/-bok  
 243, 283, 305, 448, 470  
 parataktiske uttrykk 468  
 parlør 292  
 Parolekorpusen 316  
 partikelförbindelse 311  
 partikelverb 48  
 passiv 163  
 passivdannande hjelpeverb 161  
 passivkonstruksjonar 164  
 passivordbog 284  
 perfektivisk fortidsbetydning 222  
 perfektivisk nåtidsbetydning 222  
 perifrastiske konstruksjonar 162  
 personlig referent 225  
 polysemi 157, 275, 313, 332, 334, 353  
 preskriptiv leksikografi 29  
 problemorienterat informationsbehov  
 147  
 produktionsordböcker 337  
 progressivity marker 77

## R

receptionsformål 363

receptionsordbog 284, 337  
 redaktionel grænseflade 204  
 redaksjonelt døme 466  
 redaksjonelt/redaktionelt eksempel 176,  
 466  
 redigeringsapplikasjon 31, 155, 470  
 redigeringsmetoder 69  
 redigerings skjema 155  
 redigerings system 91, 448  
 redigeringsværktøj 217  
 redigeringsværktøjet iLEX 217  
 referanseverk 478  
 referent 65, 138, 225, 331, 341  
   inanimat referent 225  
 reiseleksikografi 305  
 reiseordbøger 291  
 renæssancen 409  
 retningsadverbier 229  
 rettskrivningsordbok / retskrivningsord-  
 bøger 19, 259  
 riksmål 132  
 rimordbog 456  
 rullemenu 448

## S

samtidsordbok 422  
 satsmodeller 333  
 seddelsamling 173, 176, 212, 236  
 selektion 333  
 semantic shift 77  
 Semantic Units (SemU) 140  
 semantisk argumentstruktur 47  
 semantisk etikette 52  
 semantisk felt/fält 241, 279, 438  
 semantisk klassifisering 353  
 semantisk nærhet 45  
 semantisk ordnett 273  
 semantisk valens 333  
 semantiska roller 331  
 semantiske områder 74  
 semantiske relasjonar 271  
 SIMPLE 133, 354  
 skolordbok 244  
 små ordklasser 72  
 social kontekst 145  
 sorterar 35

Sorteraren 31, 32  
 sorterarverktøy 32  
 sprogteknologi 353  
 SprogTeknologisk Ordbase 43  
 sprogteknologiske værktøjer 354  
 sprogbrugs-/sprogbrugseksempel 177, 197, 466  
 sprogfelleskap 487  
 sprogshistorie 375  
 sprog-/sproglig information 147, 244  
 sprogteknologi 133  
 Sprogteknologisk bokmålsordbok (STEBO) 137  
 sprogteknologisk ordbok 136  
 standard-oversættelse 62  
 statistik 71  
 statistisk søgning 232  
 stavekontrol 72  
 store verb 158  
 strukturering av ordboksteksten 271  
 ställighet 330  
 sublemma 155, 230, 317  
 synonym 140, 222, 279, 362, 440  
   nærsynonym 275, 362  
 synonymi 279, 353, 362  
 synonymordbøger 222, 271  
 synonymrekke 222  
 synset 354  
 syntagmatiske relasjoner 271  
 syntaksmønster 48  
 syntaktisk beskrivelse 43  
 syntaktisk enhed 49  
 syntaktisk lighed 45  
 syntaktisk struktur 12, 438  
 syntaktisk valens 47, 279  
 syntaktiske konstruktioner 43, 331  
 syntaktiske konstruktionsmønstre 43  
 syntaktiske konstruktionspotentiale 47  
 systematisk 70  
 sætningsordbøger 292  
 søgefelt 448  
 søgeresultat 452  
 søgning / søkning / søking 70, 206, 243, 278, 364, 440, 448  
   begrebsbaseret søgning 354  
   emnesøgning 362

encyklopedisk søkning 249  
 flerordssøgning 250, 275  
 fritekstsøgning / fritextsøgning 255, 456  
 informationssøgning/-søgning / informationssøking 58, 133, 144, 202  
 korpussøgning 390  
 mønstersøgning 442  
 statistisk søgning 232  
 tesauristisk søkning 440  
 trunkeret søgning 448  
 websøgning 389  
 sökprocess 147  
 sökstrategier 247

## T

talemåde 1  
 talesprog/-sprog 19, 185, 266, 376  
 talespråkskorpus 135  
 tecken 332  
 tekstbehandling 75  
 tekstgenre 384, 387  
 tekstgenredata 390  
 tekstgenreorienterede behov 384  
 tekstkonventioner 384  
 tekstkorpus 92, 198, 382  
 tekstproduktion 363  
 teleskopvisning 458  
 tematiske ordbøger 409  
 temporale hjelpeverb 161  
 teoretisk leksikografi 140  
 terminologi 168, 422  
 tesauristisk søkning 440  
 tidssparing 38  
 tillfälliga konstruktioner 249  
 tilgangsstruktur 278, 447  
 tilläggsled 330  
 tjeck 70  
 to do-liste 70  
 topikaliserings 171  
 top-ontologi 357  
 tosprogede idiomordbøger 118  
 tosprogsordbøger 285  
 tospråklig leksikografi 219  
 tospråklig ordbok / tospråklig ordbog 207, 219, 372

trivalente verber 56  
 trunkeret søgning 448  
 tryckta ordböcker / trykte ordbøker/-  
 bøger 148, 270, 447  
 turistordbøger 291  
 tyding 156  
 tydingsbolkar 158  
 tydningsspekter 165  
 typiske samforekomster 232

## U

udvidet søgning 455  
 uetableret definition 62  
 underbegreb 61, 74, 356  
 underoppslag 155, 465  
 undervisningsrettleiing 372  
 unika konstituentter 422

## V

valens 207, 278, 327  
     funktional valens 47  
     semantisk valens 278, 333  
     syntaktisk valens 47, 279, 330  
     verbvalens 327  
 valensangivelser 137, 315  
 valensbeskrivning 328  
 valensbestämningar 315  
 valensbundethed 55  
 valensbärare 330  
 valensinformation 327  
 valenskonstruktion 332  
 valenslexikografi 327  
 valenslexikon 328  
 valensmønster 235  
 valensoplysninger 237  
 valensordbok 327  
 valensstruktur 58, 332  
 valensteori 327, 330  
 valenstype 50  
 valensuppgifter 309  
 valensvariation 330  
 verb / verbum  
     divalente verber 56  
     fraseverb 157  
     partikelverb 48

trivalente verber 56  
 verbale konstruksjoner 276  
 verbale konstruksjonsmønstre 278  
 verballemaer 47  
 verbforbindelser 276  
 verbkonstruktion/-konstruksjon 161,  
 309, 311  
 verblemma 161  
 verbpartikel 316  
 verbvalens 327  
 vestnordisk 485  
 vidaregåande opplæringa 372  
 vidensformål 363  
 videnstyper 301  
 virksomhedsleksikografi 385  
 virksomhedsordbøger 385

## W

web-baseret database 197  
 webordbog 197  
 websøgning 389  
 wordnet 353  
 WordNet 354

## X

XML-standard 214

## Y

ytringssituationer 230

## Æ

ækvivalent / ekvivalent 61, 201, 479

## Ø

Ømålsordbogen 173, 231

