

Nordiske Studier i Leksikografi

Nordiske Studier i Leksikografi

8

Rapport fra Konference om Leksikografi i Norden
Sønderborg 24.–28. maj 2005

Redigeret af

*Henrik Lorentzen og
Lars Trap-Jensen*

Skrifter udgivet af Nordisk Forening for Leksikografi
Skrift nr. 9

I samarbejde med Språkrådet i Norge og
Det Danske Sprog- og Litteraturselskab

København 2006

Nordiske Studier i Leksikografi, nr. 8
© Nordisk Forening for Leksikografi og forfatterne, 2006
Redaktion: Henrik Lorentzen og Lars Trap-Jensen
Bogtilrettelæggelse og sats: Ole Klitgaard, Forlagsbureauet
Sat med Stempel Garamond
Omslag: Pernille Sys Hansen
Tryk: Special-Trykkeriet Viborg
ISBN 87-7533-007-5
ISSN 0803-9313

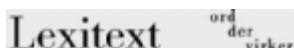


Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
www.dsl.dk

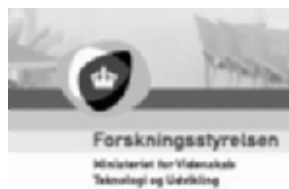
Udgivet med støtte fra:



Den Berlingske Fond



LETTERSTEDTSKA
FÖRENINGEN



Rapporten kan bestilles hos:

Språkrådet
Postboks 8107 Dep
NO-0032 Oslo
e-post: post@sprakradet.no

Indhold

Forord	9
Kristin Bakken Relevans, legitimitet og resultatkrav – eit komparativt blikk på utfordringane for dei store nasjonale ordboksverka i Norden i dag	11
Anna Braasch Den danske Sprogteknologiske Ordbase og dens anvendelse i værktøj til leksikografiske formål	25
Lars Brink Den fraseologiske terminologi	39
Ulrika Djärv Språkstadieordböcker nu och då	53
Øystein Eek Nyord og allmenne ordbøker	65
Ken Farø Leksikografisk radikalisme? 10 forslag til den bilingvale ordbog	85
Birthe Gawinski Spansk mad – en illustreret onlineordbog	97
Franziskus Geeb lookedup4you: studerende, kunder og online-leksikografi	107
Alexandra Granström (Hur) använder du ordböcker? En undersökning bland svenskläroarstudierande och svensklärare	121
Oddrun Grønvik Verknader av digitalisering på materialvurdering, redaksjonell metode og opplæring	129
Oddrun Grønvik og Lars Jørgen Tvedt Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy	143

Asgerd Gudiksen Ømålsordbogen og mejetærskergenerationen	151
Karin Hallén och Eva Thelin Ett svenskt dialektlexikon i startgröparna	163
Erla Hallsteinsdóttir Multilingval elektronisk idiomordbog	171
Bent Christian Jacobsen Muligheder for samarbejde mellem ordbøger der arbejder inden for det nordiske sprogområde	181
Knut E. Karlsen Geografiske hol i ei nasjonal ordbok – om geografisk representasjon av kjeldene i Norsk Ordbok	185
Guðrún Kvaran De første skridt til en islandsk fremmed- og/eller låneordbog	199
Lennart Larsson Varför förändras (inte) ordböcker? Exemplet Florinus	209
Rita Lenstrup Genveje i den elektroniske bilingvale ordbog illustreret ved hjælp af <i>man</i> → <i>you, one, we, they</i> og syntaktiske løsninger	223
Patrick Leroyer Ordbogseksterne henvisninger – fra belæg til funktioner	239
Henrik Lorentzen og Lars Trap-Jensen <i>ordnet.dk</i> – et nyt sprogligt opslagsværk på internettet	253
Erika Lyly Är ordboken <i>Svenskt språkbruk</i> en deskriptiv ordbok?	265
Sven-Göran Malmgren och Ruth Vatvedt Fjeld Om felaktiga särkrivningar i svenskan och norskan och om ordboksbaseade och statistiska program för att hitta dem	275
Sanni Nimb Adverbiers syntaks i ordbøger	283

Sanni Nimb, Ruth V. Fjeld, Maria Toporowska Gronostaj og Bolette Sandford Pedersen Leksikalsk beskrivelse af adverbiers semantik i norsk, svensk og dansk – LEXADV	301
Christian-Emil Ore og Lars Jørgen Tvedt Digital sats eller digital satsing?	315
Karen Margrethe Pedersen Sprogbrug og sprogsyn hos flertal og mindretal i den dansk-tyske grænseregion	323
Marie Bilde Rasmussen Nyt redigeringsystem – overvejelser og valg	347
Emma Sköldberg Idiom och semantisk variation – i autentiskt språkbruk och i ordböcker	359
Lars Törnqvist Uttalslexikon för talsyntes	373
Lars S. Vikør Skandinaviske lærebøger i leksikografi og leksikologi – tverrnasjonalt oversyn og vurdering	383
Dagfinn Worren Molbech som mønster for Aasen	391
Forfatterne	407
Stikord	411

Forord

Den ottende *Konference om leksikografi i Norden* blev afholdt på Idrætshøjskolen i Sønderborg i dagene 24.-28. maj 2005. Arrangør var Nordisk Forening for Leksikografi i samarbejde med Det Danske Sprog- og Litteraturselskab og Språkrådet i Norge. Den lokale organisationskomité har bestået af Carsten Hansen, Henrik Lorentzen og Lars Trap-Jensen, traditionen tro godt bistået af Rikke Hauge fra Språkrådets sekretariat.

Konferencen samlede i alt 104 deltagere, og der blev holdt 36 foredrag fordelt på to sessioner. Af disse er 32 foredrag omarbejdet til de artikler som udgives i denne konferencerapport. Som sædvanlig var der ikke noget bundet tema for konferencen, og programmet vidner derfor om spændvidden og mangfoldigheden inden for nordisk leksikografi. I tillæg til de tilmeldte foredrag var der to plenarforelæsninger ved inviterede gæster: Jørn Lund fortalte om nordisk sprogpolitik i europæisk perspektiv, og Karen Margrethe Pedersen orienterede om den sproglige mangfoldighed i det område hvor konferencen blev holdt, den dansk-tyske grænseregion.

Ud over det faglige program var der under konferencen indrettet et udstillingslokale hvor nordiske forlag præsenterede aktuelle ordbøger og anden litteratur, ligesom der var demonstration af et redigeringsystem til ordbøger.

Det sociale program indeholdt en udflugt til Dybbøl Banke, Rømø og Rudbøl samt konferencemiddag på Sønderborg Slot.

Redaktører af denne konferencerapport er Henrik Lorentzen og Lars Trap-Jensen. Artiklerne i rapporten er skrevet på enten norsk bokmål, nynorsk, svensk eller dansk efter forfatternes eget valg, og alle artikler er forsynet med et sammendrag på engelsk. Redaktørerne har gennemlæst og godkendt samtlige artikler. Bagest i rapporten er der et stikordsregister over centrale faglige ord og udtryk med sidehenvisninger til relevante artikler. Der er desuden en oversigt over samtlige forfattere med oplysning om arbejdssted, adresse m.m. Redaktørerne takker Carsten Hansen for hans indsats i forbindelse med rapportens tilblivelse.

Konferencen har modtaget økonomisk støtte fra en række fonde og forlag, og arrangørerne vil gerne udtrykke sin taknemmelighed over den velvilje de har mødt under forberedelserne. Uden den økonomiske støtte havde det ikke været muligt at gennemføre konferencen eller udgive konferencerapporten. Følgende har ydet økonomiske bidrag: Statens Humanistiske Forskningsråd, Nordisk Kulturfond, Nordplus Sprog, Den Berlingske Fond, Knud Højgaards Fond, Forlaget Gyldendal, Oticon Fonden, Letterstedtska Föreningen, Clara Lachmanns Fond og Politikens Forlag.

FORORD

Redaktørerne vil endelig takke alle som har bidraget med at skrive artikler eller på anden måde har hjulpet til med at færdiggøre og udsende denne rapport.

København, oktober 2006
Henrik Lorentzen, Lars Trap-Jensen

Relevans, legitimitet og resultatkrav – eit komparativt blikk på utfordringane for dei store nasjonale ordboksverka i Norden i dag

Relevance, Legitimacy and Demands of Completion – a comparison of the ways the national dictionaries in Scandinavia are challenged today. Five of the national dictionaries in Denmark, Sweden and Norway are presented in order to compare them with regard to the challenges they meet in contemporary society. Several of them were conceptualised several decades or even more than a century ago, meaning that they were motivated both by different cultural contexts and by different scientific ideals. At the same time these projects are extremely costly, thus demanding public funding. The paper focuses on four challenges pertaining to this situation: the challenge of securing the financial foundations, the challenge of giving the projects legitimate motivation, the challenge of enhancing their social and scientific relevance, and the challenge of dealing with demands of completion.

1. Innleiing

Emnet for føredraget blei valt fordi eg har ansvar for eitt av dei nordiske nasjonale ordboksprosjekta, og strevar dagleg med å sameine dei faglege måla for arbeidet med dei rammevilkåra vi har å halde oss til. Her ligg det mange utfordringar, for når eg brukar ord som ”streve”, impliserer det at dei faglege måla ikkje utan vidare lèt seg harmonisere med dei rammene som er sette for arbeidet vårt. I eit prosjekt som Norsk Ordbok 2014, ville ein kanskje som fagentusiast kunne ønskje seg så store og så oppdaterte og lett tilgjengelege samlingar som mogleg, uendeleg mykje papirplass å boltre seg på, og dessutan romslege økonomiske rammer som gjev rom for ”tilstrekkeleg” mange redaktørstillingar og støttepersonell. Dessutan ville ein jo gjerne ha optimale IT-verktøy, høvelege lokale, og ein ville helst sleppe stressande tidspress og ytre krav om ferdigstilling. På den andre sida er det klart at vi ideelt sett gjerne vil ha eit stort publikum, for vi vil jo også at ordboksproduktet skal vere etterspurt.

For Norsk Ordbok 2014 like lite som for dei andre nordiske ordboksverka, er situasjonen slik. I staden står vi overfor utfordringar på dei fleste av dei felte eg alt no har nemnt. Det gjeld utfordringar som har med grunnlagsmaterialet og arkivsituasjonen å gjere, plassproblem ved at line- og sidetalet ein har til disposisjon i dei trykte utgåvene er avgrensa, det er tids- og kostnadskrevjande å utvikle høvelege IT-verktøy, å sikre finansiering av prosjekta er ei konstant utfordring, og når det gjeld tilhøva til samfunnet, har vi relevans- og legitimitetsproblem samstundes med at vi har utfordringar med å formidle verksemda og produkta våre til potensielle brukarar.

Mot denne bakgrunnen tykkjer eg at det vil vere både interessant og lærerikt å samanlikne vilkåra for nokre av dei store ordboksverka i Norden i dag. Målet med ei slik komparativ tilnærming er for min eigen og ideelt sett også for andre sin del, å få idear til korleis ein best kan møte dei utfordringane det her er snakk om.

2. Ordbøkene

Dei fem ordbøkene eg vil samanlikne er forutan *Norsk Ordbok* (NO): *Norsk riksmålsordbok* (NRO) i Noreg, *Svenska Akademiens ordbok* (SAOB) i Sverige, og *Ordbog over det danske Sprog* (ODS) og *Den Danske Ordbog* (DDO) i Danmark. Eg ville ideelt sett gjerne ha inkludert ordboksprosjekt frå Finland og Island i samanlikninga, men det hadde gjort oppgåva for stor gjeve dei tidsrammene eg har hatt til rådvelde. Andre ordboksverk kunne også ha vore trekte inn, t.d. *Nusvensk ordbok* (1917-72), men medan arbeidet med denne ordboka er avslutta, er dei fem ordbøkene eg kjem til å drøfte, alle i meir eller mindre grad å rekne som igangverande prosjekt. Før eg held fram med sjølve samanlikninga, vil eg takke redaksjonane i Lund og København og NRO-redaksjonen i Oslo for all velviljug hjelp i samband med førebuinga av føredraget.

Statusen til dei fem aktuelle ordbøkene ser i utgangspunktet litt forskjellig ut. ODS, NRO og DDO er ”fullførde” ordbøker, ODS og NRO endåtil med supplementsband, SAOB er eit gammalt, vakse prosjekt med 34 band utgjevne (september 2005) og med bokstaven *t* under redigering. NO er eit yngre og mindre omfattande prosjekt enn SAOB; første band av NO kom alt i 1966, men pr. i dag (2005) er vi ikkje komne lenger enn til at fem av tolv planlagde band er utgjevne, og bokstaven *k* er under redigering. Samanlikninga av rammevilkår vil difor til ein viss grad bli basert på historiske tilhøve. På den andre sida er no arbeidet i gang med å vidareføre både NRO, ODS og DDO, og vilkåra for denne vidareføringa vil bli trekte inn i drøftinga nedanfor. Også NO og SAOB bør alt no planleggje revisjonsfasen av arbeidet. Tabell 1 gjev ytre karakteristikkar av dei fem ordbøkene.

Presentasjonen viser for det første at ordboksprosjekta har ganske ulik status. SAOB og NO er framleis under redigering, og dei har 2017 og 2014 som fullføringsår for denne primærredigeringa. Men dei er svært ulike med omsyn til ambisjonsnivå og omfang. ODS er SAOB sin næraste ordboksslektning i Skandinavia. Med supplementsbanda fyller ODS 33 band. Primærredigeringa av ODS blei fullført i 1956, men då var revisjonsarbeidet allereie i gang, og no hausten 2005 blir siste hand lagd på det femte og siste supplementsbandet. Samstundes har ein starta arbeidet med å lage ein elektronisk versjon av ordboka der supplementa skal vere fletta inn i hovudteksten. Den primære redigeringa av NRO var fullført alt i 1957, medan revisjonen var ferdig i 1995. Ein har no sett i gang arbeidet med å digitalisere verket, for deretter å førebu ny revisjon i den elektroniske versjonen. DDO, som er den nyaste av desse ordbøkene, er fullført i seks band no i 2005, og her er ein i gang med

å lage ein elektronisk versjon under overbygningen *ordnet.dk*, som tek mål av seg å integrere DDO, *Korpus 2000* og ODS i ein felles nettressurs (jf. Trap-Jensen 2004). Sidan SAOB alt er tilgjengeleg over internett, og sidan NO frå og med 2004 blir redigert som ein søkbar database (jf. Bakken & Ore 2005), kan ein generalisere og seie at alle desse store dokumentasjonsordbøkene er på veg mot det elektroniske mediet, men står på litt ulike steg i denne prosessen.

	SAOB	NRO	NO	ODS	DDO
Status	under red.	digit./revisjon	under red.	revisjon/digit.	digit.
Produksjonsperiode 1	1893-2017	1930-1957	1950-2014	1918-1956	1991-2005
Produksjonsperiode 2	?	1957-1995	?	1950-2005	
Materialgrunnlag	8,5 mill. setlar	1 mill. + 350.000 setlar	3,2 mill. setlar, korpus 22 mill.	2,5 mill. + 1,5 mill. setlar	korpus 40 mill.
Lemma	540.000	200.000	300.000	250.000	100.000
Band	34 (39)	4 + 2	5 (12)	28 + 5	6
Opplag	3600 >1000	?	2000 >1000	5000 + opptrykk	6400

Tabell 1. *Fakta om ordbøkene* (Kursiv viser til prognosetal)

Mot denne bakgrunnen er det interessant å kommentere opplagstala til papirutgåvene av dei ulike ordbøkene. Bengt Sigurd (1986) dokumenterer korleis SAOB blei lansert under stor entusiasme i 1892, og talet på abonnentar steig raskt til det nådde eit høgdepunkt i 1894 med 3600 (Sigurd 1986:146f.). Men ettersom utgjevingstakta dabba av, fall abonnentane frå, og opplaget er i dag på 1000 eksemplar. NO kom i byrjinga ut i eit opplag på 2000, men er også no nede i 1000. Det samla opplaget av NRO har det ikkje vore mogleg å finne fram til, det opprinnelege opplaget er ikkje kommentert i førsteutgåva og forlaget har ikkje lenger dei relevante arkiva, dessutan er ordboka seinare utgjeven i fotografiske opptrykk både i 1983 og 1991, og i opptrykk av einskilband innimellom og etter det. Det må til saman vere snakk om eit relativt stort samla opplag (10.000-20.000?). ODS sitt førsteopplag var imponerende nok 5000, og ordboka har dessutan blitt gjeve ut tre gonger seinare i uendra opptrykk (1966-70, 1975-77, 1981-83). DDO blir gjeven ut i eit opplag på 6400.

Ein ser klart ein samanheng mellom fullføring og opplagsstorleik. Det er ei trusprøve å abonnere på evigheitsprosjekt som SAOB og NO, medan salet av NRO og ODS, viser at fullførte ordboksverk er salbare, jamvel om det er snakk om ei

kostbar investering i 28 ordboksband slik tilfellet var med ODS. (Salet av ODS blei truleg hjelpt av det populære danske spørjeprogrammet *Fup eller Fakta* som blei sendt på TV i fleire rundar på 1960- og 1970-talet. Programmet brukte ODS som ”fasit”.) Vilkåra for sal av store kostbare papirordbøker er likevel i endring, og det store spørsmålet er å vite korleis dei fullførde utgåvene av NO og SAOB vil bli brukte og selde når det finst elektroniske parallellutgåver. Når ein no har valt å gje ut DDO i eit opplag på 6400, blir det spanande å sjå om dette er eit rett estimat i høve til marknaden, og korleis den framtidige elektroniske utgåva eventuelt vil påverke salet.

3. Institusjonell tilknytning

Dei fem ordbøkene har ulik institusjonell tilknytning. NO er organisert som eit forskingsprosjekt under Universitetet i Oslo, og er slik det einaste ordboksprosjektet med universitetstilknytning. Dei fire andre ordbøkene har det til felles at dei er organiserte og styrde av vitskapsakademi, men desse academia er rett nok svært forskjellige. Sjå tabell 2.

SAOB	Svenska akademien	Alm. akad. 18 medl.
NRO	Det norske Akademi for Sprog og Litteratur	Språkpol. alm. akad. 10-21 medl.
NO	Universitetet i Oslo	Universitet
ODS	Det Danske Sprog- og Litteraturselskab	Humvit. akad. 75 medl.
DDO	Det Danske Sprog- og Litteraturselskab	do.

Tabell 2. *Institusjonell tilknytning*

Svenska akademien har som kjent 18 medlemmer, og desse medlemene er for tida anten forfattarar eller professorar frå humaniora eller samfunnsfag. To av dei atten er språkvitskapsfolk. Redaksjonen av SAOB lever eit ganske uavhengig liv frå akademiet, men rapporterer jamnleg om framdrift og økonomi. Det Danske Sprog- og Litteraturselskab står som utgjevar av dei to danske ordbøkene. Dette er på den eine sida eit smalare samansett akademi, jf. at medlemene berre er vitskapsfolk frå språkvitskap, litteraturvitskap eller historie, på den andre sida har det breiare representasjon med så mange som 75 medlemmer. Den einskilde redaksjonen har eit fagleg tilsyn oppnemnt av DSL, og i tillegg eit pengetilsyn som dei rapporterer til om pengebruk, tidsbruk og framdrift. Det Norske Akademi for Sprog og Litteratur er til samanlikning like breitt samansett som Svenska Akademien, dvs. at det i tillegg

til språk- og litteraturforskarar også har medlemmer frå ånds- og kulturlivet meir allment. Det er på den andre sida ganske fåtalig, jf. at det til eikvar tid har mellom 10 og 21 medlemmer. Men den viktigaste forskjellen mellom det norske akademiet og dei tilsvarande nordiske, er at dette akademiet har eit eksplisitt språkideologisk siktemål, dvs. at det har teke side i den norske språkideologiske debatten, og er slik eit interesseorgan for riksmålsrørsla. Arbeidet med supplementsbanda gjorde redaktøren, Harald Noreng, for ein stor del aleine, utan at Akademiet la seg mykje opp i det, korkje i innhald eller ytre rammer. Felles for dei tre akademia er at dei er lukka selskap som regenererer seg sjølve.

NO har ei litt anna historie å vise til. Sjølv om initiativet til ordboka kom frå det store nynorskforlaget, Det Norske Samlaget, fekk prosjektet tidleg statsstøtte, anten rett over statsbudsjettet eller via forskingsrådet. I 1972 blei prosjektet lagt inn under Universitetet i Oslo, og i perioden fram til 2002 låg prosjektet inne i den vanlege instituttstrukturen. I 2002 blei redaksjonen omorganisert, og redaksjonen har no status som eit sjølvstendig prosjekt på sida av instituttstrukturen, men stadig med Universitetet i Oslo som overbygnad.

4. Utfordringar

Hovudfokuset her skal vere nokre av dei utfordringane dei store ordboksverka står ovafor, og eg har her valt å løfte fram fire utfordringar som utgangspunkt for samanlikning:

1. Finansiering
2. Legitimering
3. Relevans
4. ResultatkraV

4.1. Finansiering

Å sikre finansieringa for desse svært store, svært langvarige prosjekta er verkeleg ei utfordring, ikkje berre for ordboksredaksjonane sjølve, men for samfunnet som heilskap. Tabell 3 gjev eit lite oversyn over finansieringa av dei fem ordbøkene i den fasen vi no er inne i.

SAOB har sidan 1833 blitt finansiert ved å få utbyttet av den statlege publikasjonen *Post- och Inrikes Tidningar* (tilsvarande det norske *Lysingsbladet*). Det er altså snakk om ei fullfinansiering frå staten si side. Denne ordninga er no under re- forhandling i samband med at *Post- og Inrikes Tidningar* skal bli ei elektronisk netteneste. Men det er grunn til å tru at Bolagsverket på ein eller annan måte vil kompensere for dei tapte inntektene og vidareføre SAOB som statsfinansiert prosjekt. Både ODS og DDO i Danmark er fullfinansierte, men ikkje einssidig frå staten. Carlsbergfondet har bidrege på like fot med Kulturministeriet i finansieringa

av dei to prosjekta. ”Privat” finansiering kan kanskje ikkje Carlsbergfondet sine bidrag kallast; poenget her er at midlane ikkje kjem over statsbudsjettet. NO er til samanlikning fullfinansiert for statlege midlar: Noregs forskingsråd kostar trykkinga, Universitetet i Oslo dekkjer direkte og indirekte driftsutgifter og delar av lønsbudsjettet, medan størstedelen av lønsbudsjettet blir løyvd direkte over statsbudsjettet. NRO har dårlegare vilkår. Pr. i dag får prosjektet noko støtte direkte over statsbudsjettet, men ikkje slik at det er fullfinansiert i den noverande fasen. I tillegg spør NRO seg på fondsmidlar frå mindre fond som Fritt Ord eller Nansenfondet, og ordboka har tradisjonelt fått bidrag frå privatpersonar, dvs. privat finansiering i eigenleg mening. Forskringsrådet har også vore bidragsytar til NRO, og indirekte har Universiteta i Oslo og Bergen vore med på finansieringa av ordboka, sidan setelarkiva som NRO byggjer på, for ein del har vore bygde opp av tilsette ved universiteta. Men det er viktig å understreke at NRO pr. i dag ikkje er sikra full finansiering for å gjennomføre digitalisering eller revisjon.

SAOB	Inntektene frå <i>Post- og Inrikes Tidningar</i> . Reforhandling: Kompensasjon frå Bolagsverket?	Statleg. Fullfinansiert.
NRO	Kulturdepartementet, private fond, forlag, donasjonar	Priv./statleg. Ikkje fullfinansiert.
NO	Kulturdepartementet, Universitetet i Oslo, Noregs forskingsråd	Statleg. Fullfinansiert.
ODS	Kulturministeriet, Carlsbergfondet	Priv./statleg. Fullfinansiert.
DDO	Kulturministeriet, Carlsbergfondet	Priv./statleg. Fullfinansiert.

Tabell 3. *Finansiering*

Ordboksverka det her er snakk om, er ekstremt kostbare prosjekt, som ofte går over mange år, og som heilt sprenger dei økonomiske rammene ein vanlegvis set for forskings- eller bokprosjekt. Difor er ordboksverka heilt avhengige av statleg finansiering for å bli gjennomførde. Eit fellestrekk ved ODS, DDO og NO i den noverande fasen er at dei kviler på ei avtale om kostnadsdeling. Lis Jacobsen (1956) har gjort greie for bakgrunnen for at ODS fekk fullfinansiering og eit lønsnivå for redaktørane på line med andre statstenestemenn. Ho seier her rett ut at ”et forhåndstilsagn om bevilling fra fondet [Carlsbergfondet] var den løftestang, der gjorde det mulig at opnå statsbevillingen” (Jacobsen 1956:407). Same modell med kostnadsdeling mellom Carlsbergfondet og Kulturministeriet blei lagd til grunn for DDO. Diverre har ikkje svenskane eller vi i Noreg noko som liknar Carlsbergfondet, men modellen om kostnadsdeling blei likevel overteken då ein etablerte NO

som nytt prosjekt med ny finansiering. Rektor ved Universitetet i Oslo møtte kulturstatsråden, og Universitetet i Oslo forplikta seg til å dekkje ein del av utgiftene. Vilkåret var at Kulturdepartementet tok seg av auken i lønsbudsjettet. Dette har vist seg som ei gunstig ordning, for den verkar gjensidig forpliktande; det er vanskeleg for den eine parten å bryte sin del av avtala. Prosjektet har då også ved fleire høve eksplisitt appellert til partane ved å vise til den gjensidige bindinga.

4.2. Legitimering

Når ein er avhengig av statsløyvingar, slik som i praksis alle dei fem ordboksprosjekta er, blir spørsmålet om politisk legitimitet heilt avgjerande. Denne legitimiteten er ikkje ein permanent storleik, for i så langvarige prosjekt det her er snakk om, vil det truleg bli naudsynt å legitimere prosjekta fleire gonger og på forskjellige måtar, ettersom den politiske og kulturelle konteksten endrar seg.

Den norske situasjonen gjev ein god illustrasjon av dette poenget. Både NO og NRO får løyvingane sine rett over statsbudsjettet, dvs. løyvingane må vedtakast av Stortinget kvar haust. Det må argumenterast for løyvingssummane overfor regjeringa og i kulturkomiteen, og sidan det alltid er fleire ønskemål enn ressursar, må politikarane prioritere mellom dei kulturtiltaka dei har framfor seg. Korleis skal ein så greie å argumentere for at det å bruke 100 millionar kroner eller meir på ei ordbok som kanskje færre enn 1000 nordmenn er interesserte i å kjøpe, er rett prioritering av offentlege midlar?

Den norske stoda er interessant, fordi den avvik litt frå den danske og svenske, og fordi ordbøkene meir eksplisitt er legitimerede som kulturpolitiske tiltak. NO flyt på sin status som nynorsktiltak. Sidan nynorsken er eit statleg ansvar, og sidan målforma er under konstant trugsmål om marginalisering og på sikt uttradering, er det politisk korrekt å stimulere og støtte tiltak som kan betre nynorsken sin status i samfunnet. Dette kjem NO til gode. I tillegg har NO eit demokratisk eller distriktspolitisk aspekt, som gjer det kulturpolitisk omsetjeleg. For NO dokumenterer ikkje berre det nynorske skriftmålet, men også det ”norske folkemålet”, dvs. dialektane. I Noreg, der dialektane og distrikta står så sterkt, blir difor NO politisk korrekt også på dette punktet. Her kan heile folket og alle samfunnslag finne sitt eige kvardagspråk dokumentert. Om ikkje dette er nok, kan vi appellere til museumsaspektet av NO, dei gamle dialektorda i NO er eit aspekt av den folkelege og nynorsklitterære kulturarven som for ein del er på veg til å forsvinne. Mange ser behovet for å dokumentere denne språklege kulturarven før han er tapt for alltid.

Prosjektet NO 2014 skal i tillegg legitimere seg innanfor Universitetet i Oslo, for universitetet må også ville ønskje å vidareføre tilskotet til ordboksarbeidet innanfor sine pressa budsjett. I denne samanhengen har nok NO 2014 som vitskapleg digitaliseringsprosjekt mest appell og legitimeringstyngd. I og med digitaliseringsarbeidet som har gått føre seg dei siste åra, blir nemleg språkarkiva våre no lagde til

rette for anna utnytting enn den leksikografiske bruken. Setelarkivet og korpuset er tilgjengelege for ålmenta via internett, og ordboka er no ein elektronisk database som opnar opp for ei mengd nye utnyttingsmåtar som tener universitetsverksemda.

I møte med dei løyvande myndigheitene kan ein sjølv sagt også vise til kor mykje arbeid som faktisk alt er lagt ned i ordboka, og at det er umogleg å stanse ei ordbok i bokstaven *k*. Men røymsler frå grannelanda viser at dette kontinuitetsargumentet ikkje har avgjerande vekt. (Jf. OSD som no i praksis er terminert, og ONP i Danmark som etter band 3 (*de-em*) ikkje vil kome ut som papirbok.) Det er verdt å merkje seg at funksjonelle argument, slik som samfunnsnytte eller praktiske konsekvensar for den skrivande og lesande ålmenta i Noreg, er underordna i argumentasjonen for NO.

NRO utgjer eit interessant samanlikningspunkt. Ordboka skal jo legitimerast i den same kulturpolitiske konteksten som NO, men det har vist seg å vere vanskelegare å få gehør for dette prosjektet (jf. Guttu 1996:57f.). Dette heng truleg for ein del saman med statusen til riksmålet. I motsetnad til nynorsken har riksmålet ikkje offisiell status, og er slik ikkje ein konkurrent til nynorsken, men til bokmålet. Rett nok er dei reelle skilnadene mellom det offisielle bokmålet og riksmålet i ferd med å viskast ut, men i denne samanhengen er det underordna. Medan nynorsksaka har gjeve NO legitimitet og eit preg av politisk korrektheit, har riksmålsassosiasjonen truleg vore ei belastning for NRO i møte med dei løyvande myndigheitene. I praksis er då også dette tona ned i den eksplisitte presentasjonen av NRO, jf. Lars Bucher Johannessen og Lars Roar Langslet i forordet til tilleggsbanda som kom ut i 1995 (NRO 5:v). Derimot har ein appellert til forskingsrelevansen, dvs. ved å vise til faglege omsyn slik som behovet for å dokumentere utviklingane i det skriftlege notidsspråket. Denne argumentasjonsvegen er heller ikkje problemfri, for det hadde truleg vore lettare å argumentere for NRO som forskingsprosjekt om ordboka hadde hatt offisiell universitetstilknytning og tettare samheng med den språklege dokumentasjonsverksemda ved universiteta. NRO hadde truleg vore tent med å vere betre organisatorisk forankra som forskingsprosjekt, for slik det er, manglar NRO dei formelle forskingsgevanta det kunne ha vore opportunt å ikle seg.

Det sterkaste kortet NRO har å bruke andsynes dei løyvande myndigheitene, er bruksaspektet. Dette er den einaste dokumentasjonsordboka på bokmålssida (orsak termen!), NRO er fullført, og det har vist seg å vere behov for verket. Bokmålet er jo stetta av ei mykje større brukargruppe enn nynorsken, og dermed er det ein marknad for NRO. Ironisk nok kan kanskje dette også vere problematisk frå eit statleg finansieringssynspunkt, for forlagsinteressa for NRO er større enn for NO, og med god grunn.

Situasjonen i Danmark og Sverige er eg naturleg nok dårlegare kjend med, men i Danmark ser i alle høve ordboksprosjekta ut til å vere meir fagleg og funksjonelt motiverte. ODS blei lansert som eit nasjonalt dokumentasjonsprosjekt, avløysaren

til Molbechs då utdaterte ordbok. I utgangspunktet skulle ordboka dokumentere samtidsspråket avgrensa frå 1700 til redigeringstidspunktet, eit tidspunkt som flytta seg ettersom verket skreid fram (jf. Vikør 1999). Men i og med at ein sette kronologisk strek for ekserperinga ved sluttåret for primærutgjevinga (1955), har ordboka gjennom heile supplementsperioden blitt redigert som ei *historisk* dokumentasjonsordbok.

Ein tilskuv til at DDO blei sett i gang, var at Danmark gjekk inn i den europeiske unionen. På 1980-talet, i kjølvatnet av EU-avstemmingane, var det lett å få gehør for tiltak som styrkte det sær danske, og det blei advart mot at det danske språket kanskje ville bli truga innanfor rammene av den nye europeiske fellesskapen. DDO er altså delvis legitimert av behovet for å styrke det danske i møte med det framande. Fagleg sett var det lett å argumentere for at ein trong å dokumentere samtidsspråket, ODS inkluderer jo berre språk og språkleg utvikling frå før 1955. DDO blei difor nasjonalt og funksjonelt legitimert i møte med politikarane. Det galdt å dokumentere, men også ideelt sett å styrke det moderne danske språket i møte med EU. DDO måtte difor greie å sameine omsynet til eit vidt spekter av brukarar, for den alminnelege dansken måtte ikkje få eit så komplekst ordboksformat at han ikkje greidde å bruke det, samstundes med at ordboka også måtte kunne vere eit forskingsverktøy for lingvistar.

Når det gjeld DDO, og seinare *Korpus 2000* og *ordnet.dk*, ser det ut som DSL på ein vellukka måte fungerer som fagleg garantist opp mot finansieringsinstansane. DSL-prosjekta framstår slik som i utgangspunktet fagleg vellegitimerte, slik at ein kan bruke kreftene sine på å argumentere kultur- eller forskingspolitisk for desse prosjekta i møte med finansieringskjeldene. I Noreg manglar vi ein slik samlande fagleg koordinator av ordboksprosjekta, noko eg trur er ei ulempe. I alle høve kan det oppstå situasjonar der til dels konkurrerande prosjekt kan kome til å bruke krefter på å argumentere fagleg opp mot finansieringskjeldene, i staden for kultur- eller forskingspolitisk.

SAOB si opprinnelege legitimering var som eit språkleg og kulturelt dannelsingsprosjekt, jf. Svenska Akademiens valspråk *Snille och smak*, og Gustaf 3. sitt ønske om å dyrke fram "Svenska Språkets renhet, styrka och höghet" (Loman 1986:5ff.). Den svært tunge tradisjonen som no står opp under SAOB, og det så langt sikre finansieringsgrunnlaget, har nok gjort det mindre presserande å legitimere prosjektet i samfunnet. Det er blitt eit nasjonalt klenodium, og slik er det heva over kritiske eller kjetterske spørsmål. Det er kanskje også ei baksida ved denne situasjonen, for utan legitimeringsbehov mistar ein noko av kontakten med det samfunnet ein skal kommunisere med. Det er slik interessant at mange av dei entusiastiske abonnentane til dei første hefta av SAOB var lærarar og skulefolk, og forlaget lanserte ordboka som ei bok "för hela folket" (Loman 1986:14). Denne situasjonen er endra, og Bengt Sigurd (1986:176) skriv: "För allmänheten – även den s.k. bildade – är nog SAOB

okänd, och i ökande grad.” I tillegg ser ein at den tunge tradisjonen også lyfter SAOB over fagleg kritikk. Lundbladh (2003) dokumenterer korleis utgjevingane av SAOB-band og -hefte i svært liten grad har vore emne for faglege meldingar.

Oppsummeringsvis kan ein konkludere med at dei store nordiske ordbøkene er ulikt legitimerede. Ein ser at legitimeringsbehovet er ulikt til ulike tider, og at nasjonsbyggings- eller dannelsingsprosjekt kanskje ikkje er like ”salbare” no som for 100 år sidan. Det kan slik sett vere ei utfordring å endre argumentasjonen undervegs.

4.3. Relevans

Så langt har eg drøfta korleis offentleg finansiering kviler på kva slags politisk legitimering ordboksprosjekta får. Men ein får tru at politisk aksept også heng saman med kva slags sosial eller kulturell relevans ordbøkene *faktisk* har. Ein kan jo påstå at dette er eit prosjekt som bringer fram uunnverleg forskning, eller at dette er eit prosjekt som styrkjer den vanlege mannen sitt språklege medvit, men spørsmålet er om ordbøkene faktisk er relevante for dei brukargruppene ein ønskjer å nå, og i høve til den argumentasjonen som blir brukt for å legitimere ordbøkene politisk.

Vi har vel alle saman eit ideal om at ordbøkene vi produserer skal kjennast relevante for brukarane våre. På den andre sida er det eit empirisk faktum at vi har få brukarar. Både den ålmenne kulturelle konteksten og dei forskingsparadigma som var gjeldande då ordbøkene opprinneleg blei initierte, er endra, og dermed må ein spørje seg om ordbøkene våre faktisk og framleis er relevante.

Eit hovudinstrykk er at SAOB, ODS, NO og for ein del NRO har låg samfunnsrelevans. DDO har enno ikkje fått tid på seg til å stadfeste relevansen sin. SAOB og NO er lite kjende i samfunnet, og dei sel i små opplag. ODS og NRO som er fullførde, er meir kjende. Presseopplaga om NO får gjerne ei museal eller kuriøs, eller endåtil humoristisk vinkling. Dette gjeld då særleg framdriftstakta. Og det er klart at det formatet vi arbeider innanfor ser avvikande ut i ei ”tabloid” verd. I forsøka på å svare på utfordrande spørsmål frå pressa, kan vi dessutan ikkje lenger bruke ord som *danning*, *nasjonal* eller *kulturarv*, for dei er ikkje utan vidare honnørord i dag. Utfordringa blir difor å gjere ordbøkene meir samfunnsrelevante, og eg vil nemne nokre tiltak på vegen mot eit slikt mål.

Det er viktig å bruke media meir offensivt. Ofte er det slik at vi blir kontakta av pressa for at dei skal få stadfest sine fordommar om ordboksverksemd av oss. Her er det truleg mykje å hente ved å ta intitativ på eigne premissar. Lis Jacobsen ser ut til slik å ha brukt media svært medvite til beste for ODS i dei åra ho styrde ordboka. Stoffet vi arbeider med eignar seg desutan til mange andre og meir tabloide format enn som ordboksartiklar. Eg tenkjer her på Dagens Ord-spalter, éin redaktør hos oss har fast program på NRK kvar veke der han tek imot spørsmål frå lyttarane, og andre TV- og radiokonsept kunne også ha ordfokus.

Det norske TV-programmet *Typisk norsk* viser dessutan at språk og ord er noko

folk er svært interesserte i. I Noreg er det stor aktivitet rundt om i bygdene når det gjeld å registrere gamle dialektord. Etymologiske ordbøker sel godt både i Noreg og Sverige. Skandinavar er opptekne av litteratur og forfattarskapar. Og svært mange løyser kryssord eller spelar scrabble. Utfordringa vår er å knyte an til denne interessa på ein kreativ måte. Målet må vere å greie å kanalisere språkinteressa hos folk over på ordbøkene via dei meir tabloide formidlingsformata.

I denne samanhengen er digitaliseringsprosessane som alle dei nordiske ordboksprosjekta på ein eller annan måte står midt oppe i, avgjerande viktige. Ordboksverka våre er store, og dyre og plasskrevjande. Elektroniske versjonar av ordbøkene er ingen av delane, og dei inviterer dessutan til meir mangesidig bruk av ordbøkene (jf. Bakken & Ore 2005). I freistinga på å nå fleire brukarar i framtida vil det elektroniske mediet truleg bli heilt avgjerande.

Så langt har eg litt upresist snakka om samfunnsrelevansen. Eg vil avslutningsvis avgrense omgrepet litt, og sjå nærare på forskingsrelevansen av ordbøkene våre. Dei er trass alt vitskaplege dokumentasjonsordbøker, og forskarsamfunnet er truleg den primære brukargruppa for fleire av dei. Utgangspunktet for meg er ein påstand om at koplinga mellom den generelle språkvitskapen og den vitskaplege leksikografien blei gradvis svekt frå 1950 og i tiåra framover. Grunnane kan vere mange, men éin grunn er at nye fokus kom til innanfor språkvitskapen, og desse fokusa prioriterte syntaks og teori heller enn empiri og leksikon. Ein annan grunn kan vere at det blei færre formelle koplingar mellom ordboksverka og forskarsamfunna, det var ikkje lenger så vanleg å dele tida si mellom universitet og ordboksredaksjon, jf. det tette personalhøpethavet mellom Københavns universitet og den opprinnelege redaksjonen i ODS. Dessutan konsoliderte leksikografien seg som eige fagfelt, slik at det blei skipa til eigne leksikografitidsskrift, eigne undervisningstilbod, eigne konferansar og avhandlingar med reint leksikografisk fokus. Under éin synsstad er dette sjølvsagt eit gode, men det er grunn til å tru at det ikkje har fremja kontaktane mellom den generelle språkvitskapen og leksikografien. Sjå for øvrig Christensen 2002 for ei interessant samanlikning av korleis disiplinen semantikk blir praktisert innanfor SAOB og innanfor universitetet.

Mitt poeng med å peike på desse faktorane som tidlegare har skapt avstand og slik sett undergrave relevansen til ordbøkene for universitetssektoren, er å vise til at desse vilkåra på viktige punkt er endra, og at ein difor no har eit godt utgangspunkt for å knyte nye band mellom språkvitskapen og leksikografien.

For det første rår no nye språkvitskaplege forskingsparadigme med ny interesse for faktisk språkbruk og breie empiriske undersøkingar. Her er korpuslingvistikken truleg det beste dømet. Dessutan er det større interesse for leksikonet og for morfologi enn 1970- og 80-talet var prega av, og det innanfor ulike teoretiske overbygningar. Grensene mellom datalingvistikken og den databaserte leksikografien er i ferd med å viskast ut, noko som har vist seg å opne for nye spanande perspektiv.

Innanfor leksikografien har dessutan dei nye databaserte metodane gjeve oss eit heilt anna utgangspunkt for å møte andre språkforskarar på halvvegen. Viktigast her er det at datateknologien har brote ned grensene mellom ordbøker, databasar og kjelder. Innanfor prosjektet NO er t.d. no setelarkiva våre elektronisk tilgjengelege og søkbare, vi har etablert eit tekstkorpus som på same måte er tilgjengeleg og søkbart, og sjølve ordboka er ein database med direkte koplingar til kjeldegrunnlaget. Innanfor overbygningen *ordnet.dk* vil ein etter kvart finne koordinerte foredla språkressursar som innbyd til å utnytte desse ressursane på nye måtar. Ikkje minst viktig vil det vere at dei nye elektroniske formata opnar opp for systematisk og ikkje berre partikulær bruk av ordbøkene, noko som absolutt bør auke relevansen av dei for andre språkforskarar.

Eg trur altså digitalisering vil vere heilt avgjerande for meir forskingsrelevant bruk av ordbøkene. Men for å nå dette målet, er det i alle fall fire forhold som det trengst å arbeidast vidare med. Vi må arbeide med å bryte ned fordommane som finst mot ordboksgenren i forskarsamfunna, vi må gjere dei nye ordboksprosjekta synlege i forskarsamfunna, og vi må demonstrere for andre forskarar korleis ein ny vitskapleg ordboksbruk kan sjå ut. Målet må vere at auka forskingsrelevans og utnytting i neste omgang kan brukast som legitimeringsargument opp mot forskingssektoren, ikkje berre mot kultursektoren.

4.4. Resultatkrav

Til slutt vil eg stutt kommentere den utfordringa som gjeld resultatkrav. Det er no slik at alle dei fire fullfinansierte ordbøkene ha, eller har hatt resultatkrav stilte til seg. NO fekk eit slikt krav i og med omorganiseringa i 2002, medan SAOB fekk eit sluttår for arbeidet på 1990-talet. Med unntak av NRO, som altså ikkje er fullfinansiert, er resultatkravet formulert av finansieringskjeldene, som også fungerer som rapporteringsinstans. Når resultatkrav no er formulerte, er oppdragsaspektet av verksemda tydeleggjort. Det ser elles ut til at det er litt ulikt kor tett oppdragsgjeverane følgjer oss opp.

Det har ofte blitt fokusert på to negative sider ved det å redigere ordbok med stramme tidsfristar. Det kan for det første vere stressande å arbeide med strenge tidsfristar, og for det andre oppstår det ofte ein konflikt mellom progresjon og kvalitet. Begge problema vil kjennast mest akutte når tidsfristane blir urimeleg strenge. NO 2014 har likevel insistert på å få rapportere om framdrifta direkte til Kulturdepartementet og ikkje berre til vårt eige styre. Vi ønskjer å fokusere på det faktumet at vi har fått eit oppdrag frå samfunnet, og slik brukar vi sluttåret som utgangspunkt for å argumentere eksternt for finansiering. Internt er sluttåret eit absolutt orienteringspunkt når vi må gjere val som gjeld framdrift, dimensjonering og avgrensing.

Innanfor rimelege grenser trur eg faktisk det er bra å arbeide med absolutte

tidsfristar. For det første må all vitenskap avgrensast; det gjeld å svare på oppgåva ein har sett seg på adekvat måte. Dessutan er det ofte slik at når eit prosjekt manglar tidsavgrensing, vil gjerne dei kvalitative og kvantitative normene endre seg undervegs, jf. den naturlege utviklinga av SAOB, ODS og NO opp gjennom åra. Dette viser også at kvalitet er ein variabel storleik som ikkje kan vurderast uavhengig av rammene for oppgåva. Eg trur dessutan at det å greie å innfri eit samfunnsoppdrag er heilt avgjerande for å gjere arbeidet vårt relevant, noko ikkje minst salet av dei fullførde ordbøkene NRO og ODS demonstrerer. Vi må demonstrere ein vilje til å møte samfunnet på dei premissane andre legg for oss. Eg må likevel presisere at det gjeld å finne fram til ei tidsavgrensing som står i rett forhold til omfanget av arbeidet.

5. Konklusjon

Denne samanliknande drøftinga av utfordringar som er sams for dei fem nordiske ordboksverka eg her har sett nærare på, hadde som primært mål å gje meg innsikter som kan kome NO til gode. Eg vil difor konkludere med dei personlege læringsresultata eg sjølv sit igjen med. Mi von er at dette også kan gagne andre.

På vegner av alle ordboksprosjekta trur eg for det første at vi må legitimere prosjekta på ein meir medviten måte enn tradisjonelt har vore gjort. For det andre er det viktig å betre ordbøkene sin samfunnsrelevans og forskingsrelevans. For det tredje må vi vere meir utoverretta om vi ønskjer å synleggjere og halde på det samfunnsoppdraget vi faktisk har fått, og for det fjerde må vi ta resultatkrava på alvor. For NO sin del vil eg leggje til at vi har mykje å gå på når det gjeld fagleg legitimering, medan vi ligg betre an når det gjeld kulturpolitisk legitimering. Fagleg legitimering må i siste instans forankrast i fagleg kvalitet, og vi bør difor halde fram med å styrkje den faglege kvaliteten internt i redaksjonen. Og til sist vil eg halde fram at digitaliseringsarbeidet er avgjerande i høve til mange av dei utfordringane vi no står overfor. Digitaliseringsarbeid synleggjer, skaffar nye brukarar, lyfter det interne medvitsnivået, inviterer til ny og meir systematisk bruk av ordbøkene, gjer ordbøkene meir vitskaplege og fremjar framdrift. Det er difor gledeleg at alle dei fem ordboksverka eg her har sett nærare på, no vidareutviklar ordbøkene sine langs dette sporet.

Litteratur

Ordbøker:

DDO 2003-2005 = E. Hjorth/K. Kristensen (red.): *Den Danske Ordbog*. 6 band. Udgivet af

Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

NO 1966- = *Norsk Ordbok*. 5 av 12 band. Oslo: Det Norske Samlaget.

NRO 1937-1957, 1995 = T. Knudsen/A. Sommerfeldt/H. Noreng (red.): *Norsk Riksmålsord-*

bok. 6 band. Utgitt av Det norske Akademi for Sprog og Litteratur. Oslo: Aschehoug/Gyldendal/Kunnskapsforlaget.

Nusvensk ordbok 1917-1972. 10 band. Stockholm: Wahlström & Widstrand.

ODS 1918-1956 = *Ordbog over det danske Sprog*. 28 band. Udgivet af Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

ODS-S 1992-2005 = *Ordbog over det danske Sprog. Supplement*. 5 band. Udgivet af Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

ONP 1989-2004 = *Ordbog over det norrøne prosasprog. Register*, 3 bind. Udgivet af Den arnamagnæanske kommission. København.

OSD 1991 = *Ordbok över Sveriges Dialekter*. 1 band. Uppsala: Arkivet för Ordbok över Sveriges dialekter.

SAOB 1898- = *Ordbok över Svenska Språket*. 34 av 39 band. Utgiven av Svenska Akademien. Lund: Gleerup.

Annan litteratur:

Allén, Sture, Loman, Bengt & Sigurd, Bengt 1986: *Svenska Akademien och svenska språket. Tre studier. Svenska Akademien 200 år*. (Svenska Akademiens handlingar från år 1986. Fjärde delen 1986.) Stockholm: Norstedts.

Bakken, Kristin & Ore, Christian-Emil 2005: Norsk Ordbok – også ei elektronisk ordbok? I: *LexicoNordica* 12, 7-18.

Christensen, Lisa 2002: Universitetssemantik och ordbokssemantik. I: Mattisson, Anki mfl. (red.): *Alla ord är lika roliga. Festskrift til Lars Svensson 28 februari 2002*. Svenska Akademien. Distr. Nordstedts: Stockholm, 24-39.

Guttu, Tor 1996: Norsk Riksmålsordbok V-VI. I: *Mål og Makt* 4, 53-58.

Jacobsen, Lis 1956: Da den danske ordbog blev en institution. I: Aakjær, Svend, Hald, Kr. & Hansen, Aage (red.): *Festskrift til Peter Skautrup 21. januar 1956*. Århus: Universitetsforlaget, 403-410.

Loman, Bengt 1986: En Inrättning, ägnad endast til Språkets förbättring. I: Allén, Loman & Sigurd 1986, 1-142.

Lundbladh, Carl-Erik 2003: Kritiken av SAOB. I: *LexicoNordica* 10, 99-118.

Sigurd, Bengt 1986: Ordboken, ordlistan och några andra av Svenska Akademiens språkliga insatser under 1900-talet. I: Allén, Loman & Sigurd 1986, 145-232.

Trap-Jensen, Lars 2004: Et net af ord – ordnet.dk. I: *Mål & Mæle* 4, 24-30.

Vikør, Lars Sigurdsson 1999: Fleirgenerasjonsordbøker og tida. I: *Nordiska studier i lexikografi 4. Rapport från Konferensen om lexikografi i Norden, Esbo 21-24 maj 1997*. Helsingfors: Nordiska föreningen för lexikografi, Helsingfors, 395-405.

Den danske Sprogteknologiske Ordbase og dens anvendelse i værktøj til leksikografiske formål

The Danish Lexicon for Language Technology Applications (STO) and its use in a tool for lexicographic purposes. This article deals with the largest and most comprehensive computational lexicon for Danish. Firstly, the development principles, the lexical coverage and the linguistic content of this lexicon are presented. This part focuses on the treatment of inflectional morphology by means of the Remove/Add computing method. Secondly, the development and functionalities of a flexible and effective lemmatiser program for Danish are discussed; the rules of the lemmatiser have been derived from the STO morphology data. A few examples illustrate the use of the lemmatiser in solving lexicographic tasks. Finally, the user-interface for online look-ups in the STO database is described: it transforms the computational lexicon into an electronic dictionary making it a useful source of lexical knowledge for lexicographers and other interested users. Also a number of useful web addresses, viz. to the STO database, the lemmatiser and relevant documentation, also in English, are provided.

1. Indledning

Den Sprogteknologiske Ordbase (STO) for dansk er udviklet til anvendelse som ordbogskomponent i programmer til datamatisk sprogbehandling, fx morfologisk eller syntaktisk analyse af tekster og applikationer hvori der indgår sådanne analyser. Dermed kan en sådan ordbase også udgøre kernen i nogle programmer som med fordel kan bruges i leksikografers og lingvisters arbejde, eksempelvis i et værktøj der automatisk identificerer lemmaer i en tekst, en såkaldt lemmatiser. For at opnå de bedst mulige resultater er det vigtigt at værktøjernes ordbogsmodul har en vis størrelse, og at det er leksikografisk og lingvistisk velfunderet.

STO er allerede blevet anvendt som leksikonmodul i flere værktøjer, foruden at den også er blevet brugt i en række lingvistiske forskningsprojekter, endda inden arbejdet med basen var afsluttet (fx Ørsnes 2004). For interesserede er der via internettet adgang til at søge i store dele af ordbasen, og der registreres allerede nu mange opslag – op til flere tusinde om ugen, hvilket tyder på stor almen interesse.

I det følgende beskrives først STOs indhold med hensyn til dens eksterne dækning (antallet og arten af lemmaer) og interne dækning (oplysningstyper). Derefter præsenteres lemmatiseringsværktøjet for dansk der er udviklet til det leksikografiske arbejde i STO-projektet, og som nu er tilgængelig for andre projekter. Til sidst beskrives kort hvordan denne ordbase der, selv om den er udarbejdet til datamatiske formål, også kan anvendes som elektronisk ordbog over internettet ved hjælp af en brugergrænseflade.

2. STO som dataleksikografisk produkt

STO er, som nævnt i indledningen, en ordbase der er udviklet til anvendelse i sprogteknologisk, datalingvistisk og dataleksikografisk forskning. Ordbasens materiale er korpusbaseret både med hensyn til lemmasektionen og til beskrivelsen af ordenes egenskaber. De metodiske overvejelser der ligger til grund for basen med hensyn til dens indhold, struktur og beskrivelse, er styret af de planlagte anvendelser.

Det er fælles for elektroniske ordbøger og datamatiske ordbaser (også kaldet *leksikon*) at de beskriver et nærmere afgrænset ordforråd i overensstemmelse med faste leksikografiske retningslinjer. En computer kan ikke udtrække oplysninger fra eksempler ved hjælp af analogier eller på anden måde udnytte tekstlig information (henvisninger, citater, forklaringer osv.), og derfor stilles der andre krav til en ordbase til datamatisk anvendelse end til en ordbog for mennesker. Forskellen mellem en elektronisk ordbog og en ordbase består primært i at den leksikografiske beskrivelse i et leksikon skal være meget mere detaljeret og opdelt i dens mindste bestanddele. Desuden skal den være formaliseret i et fast beskrivelsessprog og struktureret i klare oplysningstyper. Den skarpe grænse mellem de to typer leksikalske datasamlinger udviskes dog noget ved at der benyttes flere og flere datamatiske metoder og værktøjer i den traditionelle leksikografi som påvirker både arbejdet og produktet. Omvendt genbruges data fra ordbøger ved udarbejdelsen af ordbaser, foruden at leksikografens håndværk vinder indpas i leksikografi for datamater.

På adressen <http://cst.dk/sto/referencer/index.html> er der en række artikler om STO som beskriver forskellige aspekter i projektet. Yderligere information fås ved direkte henvendelse til Center for Sprogteknologi, Københavns Universitet.

2.1. Ordforrådet i STO

Ved udarbejdelsen af STO er der blevet lagt vægt på at medtage et bredt sammensat udvalg af ord. Ordbasen indeholder i alt mere end 81.500 lemmaer, fortrinsvis fra almensprog. En stor hjælp i selektionen af de 68.000 almensproglige lemmaer var Den Danske Ordbogs (DDO) foreløbige, frekvensbaserede lemmaliste som blev stillet til rådighed for arbejdet i 2001. De resterende ca. 13.500 lemmaer stammer fra fagsproglige tekster, men de er ikke egentlige eksperttermer. De seks udvalgte fagområder var edb/it, miljø, sundhed/helse, finans/økonomi, forvaltning samt handel/erhverv. Formålet med at inddrage fagrelaterede tekster fra internettet var at udbrede ordbasens ekstensionelle dækning sådan at ordforrådet også omfatter almene fagord, også kaldet gråzone-ord, der ligger tæt op ad det almensproglige ordforråd. En detaljeret oversigt over sammensætningen af ordforrådet og statistik over de enkelte ordklasser mv. kan ses på hjemmesiden www.cst.dk/sto.

2.2. *Lingvistiske oplysninger i STO*

2.2.1. *Principper*

Udgangspunktet for valget af oplysninger i den lingvistiske beskrivelse er styret af den datalingvistiske tilgang til behandling af sprog; denne opererer traditionelt med to typer grundmoduler, nemlig et ordbogsmodul (det såkaldte leksikon) og et grammatikmodul. Den traditionelle, skarpe grænse mellem grammatikken (der beskriver *generelle* – grammatiske – regler) og ordbogen (der indeholder ordene og beskrivelsen af deres *individuelle* egenskaber) udviskes dog mere og mere. Siden 1990'erne er oplysninger der beskriver ordenes syntaktiske konstruktionspotentiale, medtaget i den leksikalske beskrivelse; dette gælder i stigende grad også deres semantiske kompatibilitet. Det grundlæggende princip er at reglerne i et sprogteknologisk systems grammatik og oplysningerne i dets ordbog tilsammen skal udgøre en samlet formaliseret beskrivelse og dermed dække det ønskede segment af sproget. Dette princip stammer fra de leksikalistiske syntaksteorier, herunder den såkaldte Head-Driven Phrase Structure Grammar (HPSG, se Pollard & Sag 1994). Det udmønter sig i dag i den såkaldte leksikalisme (“the lexicalist approach”), der indebærer at en ordbase bør indeholde mange og detaljerede oplysninger om syntaktiske generaliseringer (Ørsnes 2004:213). Derved bliver grammatikken i stor udstrækning integreret i den leksikalske beskrivelse. Sådanne generaliseringer er fx den regelbundne dativalternation (1a), eller aktiv/passiv-alternationen (1b), som i traditionel lingvistik beskrives af grammatikken.

(1a) Marie gav Peter et kys/Marie gav et kys til Peter

(1b) Peter betaler udgifterne/Udgifterne betales af Peter

Mange moderne ordbøger for mennesker, som for eksempel Den Danske Ordbog (DDO, 2003-05), følger også denne tendens og opererer med såkaldte skabeloner eller konstruktionsmønstre, som er en slags formaliseret beskrivelse af ordets typiske nærkontekst, samtidig med at traditionelle brugseksempler illustrerer ordets konstruktionsmønstre.

2.2.2. *Oplysningstyperne*

STO indeholder en lang række strukturerede og formaliserede oplysninger fordelt på tre beskrivelseslag: morfologi (for hele ordforrådet), syntaks (for mere end 45.000 lemmaer, udvalgt efter frekvens) og semantik (for en mindre del af ordforrådet, ca. 8.000 lemmaer med i alt 10.000 læsninger) til eksperimentelle formål.

I det følgende beskrives først en del af det morfologiske lag i detaljer, om end ikke udtømmende. Derefter gives et overblik over oplysningerne der hører til det syntaktiske lag.

Administrative og andre ikke-lingvistiske oplysningstyper bliver ikke omtalt her.

Grunden til den detaljerede præsentation af de bøjningsmorfologiske oplysninger er at disse danner basis for det lemmatiseringsværktøj der skildres i afsnit 3. Præsentationen i nedenstående afsnit 2.3 og dets underafsnit er i høj grad baseret på dokumentet STOs Lingvistiske Specifikationer (Braasch et al. 2004-2005).

2.3. Morfologi

Den samlede morfologiske beskrivelse af et lemma er fordelt på flere blokke der indeholder hver sin type oplysninger som vedrører lemmaets ordklasse, stavning, bøjning, sammensætning (kun for substantiviske komposita) og “autonomi” (med værdien “NO” angives at ordet kun forekommer i faste udtryk som fx [*gå i skudder-mudder*]).

2.3.1. Ordklasseangivelserne

STO følger på dette punkt i alt væsentligt Retskrivningsordbogens (RO 2001) ordklasseinddeling med ganske få undtagelser. En sådan undtagelse er at STO behandler talord som adjektiver, med subkategorierne “cardinal”, fx *fem* og “ordinal”, fx *femte*. Desuden er der indført en kategori “unique” der dækker over subkategorierne formelt subjekt (*der*), infinitivmarkør (*at*) og lemmaet *som* i ikke-konjunktionsfunktionen.

2.3.2. Stavning

Hvis et ord har flere stavemåder, er disse anført i ordbasen. I visse tilfælde er også enkelte alternative stavemåder, der ikke er godkendt i RO 2001, medtaget. Begrundelsen herfor er følgende (jf. Braasch & Olsen 2005):

Det drejer sig først og fremmest om stavemåder der har ændret status fra godkendt til ikke-godkendt eller omvendt i de seneste udgaver af RO. Da STO skal kunne bruges til automatisk genkendelse [af ord i tekster], og da dette nødvendigvis også må omfatte ældre tekster end den sidste udgave af RO, er der i STO medtaget former i overensstemmelse med RO 86 og frem. Disse former mærkes som ikke-godkendte. Andre ikke-godkendte stavemåder i STO er fx ‘canarisk’ og ‘sclerose’ som begge er uhyre hyppigt forekommende. Også bøjningsmønstre kan være ikke-godkendte i tilknytning til visse ord, som er ligeledes meget hyppige, som fx ‘test’ som har aldrig måttet bøjes med ‘-s’ i pluralis, formen forekommer ikke desto mindre i mange tekster, og derfor er den medtaget i STO.

2.3.3. Bøjningsoplysninger

I den morfologiske beskrivelse er det væsentligste krav at beskrivelsesapparatet skal kunne rumme og håndtere alle danske bøjningsformer samt andre relevante, morfologirelaterede oplysninger. Dette krav er opfyldt ved at fastlægge det tilstrækkelige sæt af træk med tilhørende værdimængder, uanset om der er tale om regelmæssig eller uregelmæssig bøjning. Der opereres med morfemer (typerne *rod*

og *endelse*) og individuelle produktionsregler (af typen “fjern/tilskriv”) til beregningen af de enkelte ordformer.

Set fra en datalingvistisk synsvinkel er denne metode meget effektiv da den sikrer en ensartet og økonomisk håndtering af den samlede bøjningsmorfologi. Et væsentligt punkt er at ord der traditionelt anses for at have uregelmæssig bøjning, i STO håndteres på en meget enkel måde helt på linje med regelmæssig bøjning.

Bøjningsoplysningerne er udtrykt i bøjningsmønstre. Hver ordklasse har sin specielle kombination af formbestemmende træk, og for hvert træk er der defineret en liste af relevante værdier. Hvert mønster er unikt og omfatter ordets sammenhørende bøjningsformer. De enkelte bøjningsformer produceres ved hjælp af de såkaldte *beregningsregler*. Et ord kan naturligvis have mere end et bøjningsmønster, som fx ordet *tallerken*, med former uden (2a) eller med synkope (2b og 2c):

(2a) med beregningsreglerne (+en, +er, +erne) => tallerkenen/tallerkener osv.

(2b) med beregningsreglerne (+en, [en]ner, [en]nerne) => tallerkenen/tallerkner osv.

(2c) med beregningsreglerne ([en]nen, [en]ner, [en]nerne) => tallerknen/tallerkner.

En beregningsregel tager udgangspunkt i grundformen og udpeger *roden* ved at angive hvad der skal fjernes fra grundformen (notation i []), og hvad der derefter skal tilføjes for den pågældende form. Det specielle ved systemet er altså at begrebet *rod* her er forstået som den absolut længste del af et ord som er uforandret i den beskrevne bøjningsform. Således får man et ords operationelle rod ved at fjerne, begyndende bagfra, den del af ordet der ændres. I dette beskrivelsessystem er der ingen lingvistisk baserede regler for hvad *rod* er (det samme gælder begrebet *bøjningsendelse*); man går rent formalistisk til værks. Det er – set fra datamatisk synspunkt – en udmærket måde at håndtere ord med uregelmæssig bøjning på, nemlig på lige fod med de regelmæssigt bøjede, men det afviger fra den traditionelle lingvistiske anskuelsesmåde.

Nedenfor gives der et eksempel fra STOs Lingvistiske Specifikationer (op.cit.) på konstruktion af substantivers bøjningsmønstre, som omfatter de ordklassespecifikke egenskaber. I et mønster håndteres foruden bøjningsendelserne alle formrelaterede egenskaber, såsom synkope, fordobling af stamkonsonant og ændring af stammevo-kal, idet de inkluderes i beregningsreglerne.

Eksempelvis beskrives ordet ‘tid’ med mønster *MFG0016* som kombinerer oplysningerne om ordklasse (substantiv), køn (fælleskøn), bøjningsendelser for tal og bestemthed i umarkeret kasus (+0, +en, +er, +erne) som lægges til ordets rod der i dette tilfælde er identisk med opslagsordets grundform. Kasusendelsen (+s) for genitiv tilskrives hver af de fire nævnte former hvilket giver i alt 8 former. På denne måde laves et nyt mønster for hver unik kombination af træk/værdi-par; og der laves kun én enkelt udtømmende beskrivelse (ét mønster) af hver kombination. Der refereres under opslagsordet ved et

nummer til det passende mønster [...]. Mange substantiver bøjes på samme måde som 'tid', [fx 'stol', 'citron'] dvs. mønstret har et stort antal forekomster, såkaldte instantieringer.

Et mønster beskriver i de fleste tilfælde mange ords bøjning, mens en række mønstre kun har en enkelt eller nogle få instantieringer. Det er dem den traditionelle morfologi beskriver som undtagelser, fx *barnebarn* med flertalsformen *børnebørn*. Her fjernes intet fra grundformens rod (*barn*) for at danne formen ental/bestemt/genitiv, blot tilføjes *-ets*, hvorimod der for at generere formen flertal/ubestemt/umarkeret kasus skal fjernes *-arnebarn* og tilføjes *-ørnebørn*. I dette mønster er roden for pluralis ligeledes reduceret til et enkelt bogstav, *b*. Mønstret har kun denne ene instantiering, og i traditionel leksikografi håndteres ordet som særtilfælde eller undtagelse. En væsentlig fordel ved den her anvendte metode er, som allerede nævnt, at alle ord håndteres vha. samme mekanisme, hvilket sikrer en enkel og ensartet processering i forbindelse med praktiske anvendelser, eksempelvis i et lemmatiseringsværktøj. Der er naturligvis også en vis ulempe for leksikografen, nemlig det store antal bøjningsmønstre det er nødvendigt at etablere før systemet er fuldt udviklet.

2.3.4. *Sammensætningsoplysninger*

STO-basen indeholder mange afledte og sammensatte opslagsord. Afledte ord håndteres med samme mekanisme som simple, usammensatte ord, uden oplysning om orddannelse. På den anden side er håndteringen af sammensætningsmorfologi vigtig for dansk da den mest produktive metode for dannelse af nye ord netop er sammensætning af substantiver. Derfor er STO-materialet også forberedt til dynamiske anvendelser hvor eksempelvis nye substantiviske sammensætninger kan genkendes hvis de består af ord der er kodet i basen. Ordbasen indeholder nedenstående to oplysningstyper vedrørende substantiviske sammensætnings morfologi.

Fugelement i sammensætninger

Når et ord indgår som førsteled i sammensætninger, er der tre muligheder mht. hvordan dets form er i sammensætningen, jf. eksemplerne nedenfor: (3a) ordet forbliver uændret i sammensætninger; (3b) ordet afkortes i sammensætninger; (3c) ordet får tilføjet et fugelement¹

(3a) lampe => lampe[0]fod, lampe[0]skærm

(3b) maskine => maskin[e]mester, maskin[e]oversættelse

(3c) afdeling => afdeling+s+leder, afdeling+s+sygeplejerske

¹ Aage Hansen (1967) kalder dette *sammenbindingselement*, i RO 2001 hedder det *bindebogstav*.

Et simpleksord kan have et eller flere forskellige fugeelementer som alle registreres og udtrykkes i overensstemmelse med fjern/tilskriv-metoden (jf. beregningsreglerne ovenfor). Oplysningerne om fugeelementer ved fx ordet *mand* formuleres således på følgende måde:

- (4a) mand + 0 => mandtal
- (4b) mand + e => mandeår
- (4c) mand + s => mandsperson.

Denne oplysningstype er i høj grad korpusbaseret, og den er registreret mere udførligt og systematisk i STO end i Retskrivningsordbogen (jf. RO 2001, Indledningens afsnit 7.)

Dekomponering

Dekomponering er markering af et kompositums (sammensætning) primære bestanddele. Det foretages kun på det øverste niveau, nemlig i to dele: førsteled og sidsteled, også i de tilfælde hvor førsteledet i sig selv er et kompositum (5a). Et kompositums led markeres ved at sætte ‘+’ mellem leddene og mellem led og fugeelement (5b). I de tilfælde hvor noget fjernes, markeres det på samme måde som i bøjningsmønstrenes beregningsregler, som fx ved lemmaet *arbejdsfordeling* (5c).

- (5a) urtepotte + skjuler
- (5b) stat + s + sikkerhed
- (5c) arbejde + [e]s + fordeling

Der gælder to principielle betingelser mht. om en sammensætning dekomponeres eller ikke. For det første dekomponeres kun sådanne sammensatte ord som består af to dele der hver især er et selvstændigt ord. For det andet skal begge led beholde deres oprindelige betydning. Derfor dekomponeres ord som *makroøkonomi* og *urmager* ikke.

2.4. Syntaks

Kernen i den syntaktiske beskrivelse er valensmønstret. Det indeholder oplysningerne om hvor mange led der knytter sig til ordet (aritet), hvorvidt leddene er obligatoriske eller ikke, hvilken syntaktisk funktion (fx subjekt, objekt) og hvilken syntaktisk kategori det enkelte led har (fx nominal- eller præpositionssyntagme inkl. den styrede præposition, eller en ledsætning). Derudover er der en række oplysninger som fx vedrører verbers refleksivitet, partikel og brug af hjælpeverbum. Desuden er der korpuseksempler som belyser hver syntaktisk konstruktion. Eksemplerne kan ikke bruges af maskiner, men er medtaget for at lette leksikografens arbejde. Der er to typer eksempler, den ene type er standardeksemplet der knytter sig til en given

konstruktionstype. Det indeholder ikke selve lemmaet, det eksemplificerer blot den type nærkontekst som lemmaet kan indgå i. Den anden type er det individuelle eksempel, med lemmaet i den pågældende konstruktion. Standardeksempler er fortrinsvis anvendt i tilfælde af simple konstruktionstyper, eksempelvis ved monovalente substantiver. Hvis lemmaet indgår i en kompleks konstruktion med flere valensbundne led fra forskellige syntaktiske kategorier, er der i de fleste tilfælde indsat et individuelt eksempel. Dette er en fordel for brugeren når han/hun slår op i databasen. En mere detaljeret redegørelse over de syntaktiske træk kan findes bl.a. i Braasch & Pedersen (2002) og i STOs Lingvistiske Specifikationer (op.cit.)

De syntaktiske oplysninger anvendes eksempelvis i automatisk sætningsanalyse, den såkaldte parsning. Parsning benyttes bl.a. som delproces i leksikografiske værktøjer, eksempelvis til at genkende, opmærke og registrere et ords grammatiske strukturer i et tekstkorpus forud for ordets leksikografiske beskrivelse. Et sådant værktøj for engelsk der kan udtrække et ords såkaldte leksikalske profiler fra et korpus, beskrives i Kilgarriff og Rundell (2002).

3. Et sprogteknologisk værktøj for leksikografer: lemmatiser

I moderne datamatstøttet leksikografi bruges sprogteknologien på mange forskellige områder, eksempelvis i arbejdet med et tekstkorpus. En af de grundlæggende arbejdsprocesser er at gennemlæse relevante tekster og finde nye lemmaer til den ordbog der er under udarbejdelse eller opdatering. I denne proces er der god hjælp at hente fra forskellige sprogteknologiske værktøjer. På Center for Sprogteknologis hjemmeside, <http://cst.dk/online/index.html>, kan der afprøves en række af dem i kombination med hinanden (jf. "Seks værktøjer i tandem"). Kombinationen omfatter bl.a. en såkaldt POS-tagger (som beriger teksten med ordklasseopmærkninger), en navnegenkender og en lemmatiser. Eksempelvis kan der vha. lemmatiseren automatisk produceres en liste af lemmaer der forekommer i en given tekst. I det følgende fokuseres på lemmatiseren fordi den på flere måder er et godt eksempel på forholdet mellem datalingvistisk forskning, sprogteknologisk implementering og dataleksikografisk anvendelse. Nedenstående beskrivelse er mht. de tekniske detaljer baseret på dokumentationen af værktøjet.

Lemmatiseren er udviklet af Bart Jongejan og Dorte Haltrup Hansen i STO-projektet med det formål at dække behovet for et leksikografisk hjælpeværktøj til udtrækning af de ord fra fagrelaterede tekster der endnu ikke indgik i STOs ordforråd (som tidligere beskrevet, se afsnit 2.1). Målet var at lemmatiseren skulle være mere præcis og fleksibel end de traditionelle programmer til lemmatisering der normalt arbejder med trunkering. CST's lemmatiser er regelbaseret, og dens regler kan både håndtere regelmæssige og ikke-regelmæssige bøjninger. Lemmatiseringen omfatter tre opgaver:

- at føre hvert ord i en tekst tilbage til dets grundform (som er et kendt ord – et lemma i ordbogen)
- at vælge lemma hvis mere end et kendt ord er muligt lemma (homografer)
- at gætte lemmaer hvis grundformen ikke er kendt i ordbogen.

I lemmatiserens ordbogsmodul udnyttes STOs ordforråd og de bøjningsmorfologiske oplysninger der er anført for hvert lemma. Udvikling og træning af programmet blev gennemført i flere trin parallelt med udvidelsen af STO på følgende måde. Udgangspunktet i 2002 var ordforrådet på 50.000 lemmaer og deres bøjningsmønstre. Dette materiale blev udfoldet til en fuldformsordbog med i alt ca. 594.000 ordformer. Ud fra dette materiale udledes bøjningsreglerne (“flex rules”), disse bruges af lemmatiseren til at genkende ordformer og føre dem tilbage til det pågældende lemma. For en detaljeret beskrivelse af produktion og applikation af regler mv. henvises til den fulde dokumentation (Jongejan & Haltrup 2005).

Da det nye tekstkorpus fra det første fagområde, edb/it, først var blevet lemmatiseret, kunne lemmatiseren således sammenholde den producerede lemmaliste med STOs ordliste og identificere hvilke lemmaer der var nye i forhold til STOs almensproglige ordforråd. I processen anvendtes forskellige outputformater, med fokus på lemmaet eller listen over ordformer og deres morfologiske etikette til hvert lemma som output. Listerne blev sorteret efter forskellige kriterier, disse er beskrevet i detaljer i Jongejan & Haltrup (op.cit.).

De relevante nye ord blev derefter integreret i STO-basen og forsynet med en lingvistisk beskrivelse, hvorefter lemmatiseren kunne trænes med det udvidede materiale. Denne proces blev så benyttet i flere gennemløb til at udvide STOs ordforråd med lemmaer fra yderligere fem fagområder. Lemmatiseren er således dels baseret på STO-materialet, dels blevet benyttet i det leksikografiske arbejde i STO til at lemmatisere nye tekster og udpege lemmakandidater til udbygning af ordforrådet.

På grund af STO-materialets størrelse og kvalitet kan lemmatiseren nu beregne lemmaet med 94-98 procents nøjagtighed. Det bedste resultat opnås hvis inputteksten er ordklasseopmærket. Tabel 1 viser en sammenligning af testresultater for lemmatiseren brugt med forskellige optioner udført på et korpus bestående af 250.000 løbende ord (det såkaldte PAROLE-korpus).

	Correct lemmas	Time
Input <i>with</i> POS-tags Lemmatisation <i>with</i> dictionary = real lemmatiser	97,8 %	App. 1 min.
Input <i>without</i> POS-tags Lemmatisation <i>with</i> dictionary = discount lemmatiser	94,5 %	App. 25 sec
Input <i>with</i> POS-tags Lemmatisation <i>without</i> dictionary = good stemmer	97,4 %	App. 48 sec
Input <i>without</i> POS-tags Lemmatisation <i>without</i> dictionary = stemmer	88,4 %	App. 30 sec

Tabel 1. Sammenligning af testresultater for lemmatiseren. Kilde: the CST Lemmatiser (Jongejan & Haltrup 2005)

Det bør bemærkes at ordbogskomponenten kan udskiftes med brugerens egen ordbog. Den skal blot indeholde de nødvendige oplysninger om lemma og ordformer således at lemmatiseren kan generere bøjningsreglerne ud fra denne ordbog. Desuden kan lemmatiseren også arbejde uden et ordbogsmodul, blot med de regler der er genereret fra ordbogen. Lemmatiseren kan for øvrigt også trænes til at håndtere andre sprog med suffiksbaseret bøjningsmorfologi, fx engelsk og svensk.

Det er indlysende at en sådan fleksibel lemmatiser, der også har flere faciliteter og forskellige input- og outputformater, kan bruges på mange måder i leksikografisk arbejde; nedenfor nævnes blot nogle få konkrete eksempler foruden de ovennævnte generelle funktioner.

Lemmatiseren kan beregne frekvensen af både de enkelte ordformer og samle alle forekomsttal for et lemmas ordformer. Med udgangspunkt i denne facilitet kan man lave automatiske undersøgelser og få svar på spørgsmål som fx

- Med hvilken hyppighed forekommer et givent lemma i det valgte korpus? – vigtigt for at kunne afgøre om lemmaet skal medtages i en ordbog der er under udarbejdelse eller udvidelse.
- Hvilken bøjningsform forekommer hyppigst i teksten/korpusset, fx *tallerkenen* (uden synkope) eller *tallerknen* (med synkope)? – og hvad skal stå først i artiklen? Bruges den græsk/latinske flertalsform så hyppigt, fx *korpora* i stedet for *korpusser*, at den bør det medtages i ordbogen? (Formen er jo ikke RO-godkendt, men er alligevel hyppig, især i lingvistiske fagtekster).
- I hvilke bøjede former forekommer lemmaet *test* – er den danske eller den engelskinspirerede (ikke RO-godkendte) flertalsform *tests* hyppigst?

- Hvad er lemmaet til bøjningsformerne af et nyt fremmedord der forekommer i teksten, fx *beepere*, eller *wannabees/wannabeer*.
- Med hvilket køn bruges fremmedordet hyppigst i korpuset? – fx en/et *website*.

Svaret på spørgsmålene, dvs. resultatet af denne type undersøgelser, kan med fordel inddrages i udformningen af ordbogsartikler fordi oplysningerne baseres på empiri i stedet for på introspektion. Det er indlysende, og et velkendt faktum, at brugen af en lemmatiser og andre sprogteknologiske hjælpeværktøjer øger både effektiviteten og pålideligheden af det leksikografiske arbejde.

Den detaljerede dokumentation (Jongejan & Haltrup, op.cit.) findes på adressen <http://cst.dk/online/index.html>, hvor også lemmatiseren kan afprøves.

4. Online opslag i STO

Den Sprogteknologiske Ordbase er frit tilgængelig for online opslag over internettet. Leksikografer, lingvister og alle interesserede sprogbrugere kan derved få et indblik i den største del af STO-databasens indhold, idet semantikdelen dog ikke er tilgængelig. Desuden er der en række oplysningstyper som ikke vises på skærmen fordi de kun er relevante for datamatiske applikationer. Brugergrænsefladen er tilgængelig fra <http://cst.dk/sto/webinterface/index.html>. På hjemmesiden findes der også en brugervejledning og eksempler på interessante søgeord.

Grænsefladen har fire forskellige søgemuligheder (som det fremgår af figur 1, der viser et udsnit af sammenklippede skærbilleder); de skitseres kort her:

- *Ordsøgning* med et lemma eller en ordform. Søgeresultatet er opslagsordets bøjningsformer (i alle dets bøjningsmønstre) eller den søgte ordforms lemmaform og dens fulde bøjning. Desuden vises lemmaets – eller lemmaernes – syntaktiske konstruktionsmuligheder med eksempler.
- *Substantiviske sammensætninger*. Søgeresultatet er en liste med alle substantiver som søgestrengen indgår i, dvs. både sådanne ord som er registrerede og mulige sammensætninger (sidstnævnte er markeret med kursiv), fx ‘mand’ i ‘mandår’ og ‘havemand’ vs. ‘*konfirmand*’. Desuden kan man se sammensætningens struktur.
- *Korpussøgning* kan ske med et lemma, en enkelt ordform eller et ords udvalgte konstruktion. Søgeresultatet er forekomsterne i kwic-format fra et korpus bestående af artikler fra Berlingske Tidende, årgang 90-92.
- *Parameterbaseret søgning* med kombination af forskellige lingvistiske kriterier (fx ordklasse + aritet + styret præposition). Søgeresultatet er en liste af tilfældigt udvalgte ord (op til 30), hvis egenskaber svarer til den valgte kombination.

læse – verbum

Morfologi

<u>læs</u>	(imperativ - bydeform)
<u>læser</u>	(præsens, passiv)
<u>læser</u>	(præsens, aktiv)
<u>læste</u>	(præteritum, aktiv)
<u>læstes</u>	(præteritum, passiv)
<u>→læse</u>	→(infinitiv, aktiv)
<u>læset</u>	(infinitiv, passiv)
<u>læsende</u>	(imperfektum participium - isag tilæggsform)
<u>læst</u>	(perfektum participium - isag tilæggsform)
<u>læser</u>	(nominalisering, gerundium)
<u>læsende</u>	(adjektivering, imperfektum participium)
<u>læst</u>	(adjektivering, perfektum participium (ental, ubestemt, intetkøn))
<u>læst</u>	(adjektivering, perfektum participium (ental, ubestemt, fælleskøn))
<u>læste</u>	(adjektivering, perfektum participium (ental, bestemt))
<u>læste</u>	(adjektivering, perfektum participium (flertal))

(Korpusføring på alle måneder)

Syntaks

KONSTRUKTION:

SUBJEKT læse (OBJEKT) PARTIKEL

SUBJEKT

Optionelt = Nej

OBJEKT

Materiale = Nominativsyntagma

Optionelt = Ja

PARTIKEL

Partikel = op

(Korpusføring: læse_op)

Optionelt = Nej

(STO-kode: Dv2xW0-op)

EKSEMPEL:

Mehine læste (indtjeldet) op

sprogtycker og betoning, mens de læser hendes tekster op. Jeg aner intet om Ulla Haak, men har fornøjelse af, måske når de læser dem op for barnene. Et nyt skud på stamm efter Poulsen. Nogle af forfatterne læser deres egne tekster op for publikum. Andre leder sig ind.

Inger Christensen og Anker Schack læser digte op af 35 digtere i antologien. Læs test på . Jeg rytter på hovedet og læser grisebøger for mine kolleger, når jeg i dag om Peter Poulsen. Fredag formiddag læser han dagens digt op, som er skrevet specielt til leg!

smukke Bodios Pl. Om eftermiddagen læser han digte op, mens to jazz-musikere spiller sin Faurchou fortæller om Andersen og læser hans eventyr op, og indtæller dækker digteren se, han selv har skrevet, og som han læser op, inden den politiske ping-pong se

Figur 1. Mulighederne for søgning i STO-basen

Den sidstnævnte søgemulighed er nok den mest interessante for leksikografer, og den er beskrevet i Braasch & Olsen 2005 på følgende vis:

[...] man kan få vist grupper af ord der har samme lingvistiske egenskaber ud fra den valgte parameterkombination, eksempelvis en gruppe af divalente verber med partikel og/eller præpositionen *for*. Tabel 2 viser et udsnit af resultatet for denne søgning. Den forkortede beskrivelse af konstruktionen kan være lidt svær at gennemskue, derfor er det muligt at klikke på lemmat og få vist alle ordets konstruktioner i detaljer og med eksempler. I tilfældet *ængste* er nedenstående konstruktion en ud af de 5 registrerede for dette verbum. Ved samme opslag vises også ordets morfologi for at skabe sammenhæng mellem det morfologiske og det syntaktiske beskrivelseslag.

STO er leksikografisk og lingvistisk velfunderet, og grænsefladen på internettet gør den til en online sproglig oplysningskilde, også for leksikografer. For at man kan sammenligne søgeresultaterne fra STO med andre danske sproressourcer, er der etableret links på nettet til

- Retskrivningsordbogen (3. udgave 1996-2002, Dansk Sprognævn)
- Korpus 2000 (Det Danske Sprog- og Litteraturselskab)
- Google (søgning på danske internetsider).

Nr.	Lemma	Ordklasse	Forkortet beskrivelse af konstruktion (Klik på lemmaet for uddybning)
1	<i>stå, 1</i>	<i>verbum</i>	<i>divalent: NP, PP(NP/infinite subject equi control) prep=for</i>
2	<i>angste</i>	<i>verbum</i>	<i>divalent reflexive: NP, PP(NP/that-c/inf w. subject control) prep=for</i>
3	<i>udgive</i>	<i>verbum</i>	<i>divalent reflexive: NP, obligatory PP(NP/inf) prep = for</i>
4	<i>grue</i>	<i>verbum</i>	<i>divalent: NP, PP(NP/that-clause) prep=for</i>

Tabel 2. Søgeresultat for en parameterbaseret søgning (udsnit af skærmbillede)

5. Opsummering

Formålet med denne artikel har været at præsentere den nu færdige Sprogteknologiske Ordbase for dansk, STO, med fokus på de oplysningstyper der er relevante for ordbasens anvendelse i et værktøj for leksikografer. Den datamatiske leksikografi har to aspekter. Det ene er udarbejdelsen af leksikalske datasamlinger for sprogteknologiske anvendelser – som sådan er STO den mest omfattende for dansk. Det andet aspekt er anvendelsen af sprogteknologiske (datamatisk baserede) værktøjer i leksikografi. I dette indlæg har jeg forsøgt at vise hvordan disse to aspekter i dag er tæt forbundne med hinanden, og hvilket gavn leksikografien kan have af den sprogteknologiske udvikling der er baseret på datalingvistisk forskning.

Slutnote

Projektet har modtaget bevilling fra Ministeriet for Videnskab, Teknologi og Udvikling for perioden 1. marts 2001-29. februar 2004. Ordbasen er produceret i et samarbejde mellem Center for Sprogteknologi, KUA; Institut for Datalingvistik, CBS; Institut for Almen og Anvendt Lingvistik, KUA; Institut for Fagsprog, Kommunikation og Informationsvidenskab, SDU. Projektledelse og koordinering blev varetaget af CST. Brugergrænsefladens udvikling blev finansieret af Danmarks Elektroniske Forskningsbibliotek (DEF) i 2002.

Litteratur

Internetadresser der er nævnt i artiklen (pr. 15. september 2005):

<http://cst.dk/sto/referencer/index.html>

<http://cst.dk/online/index.html>

www.cst.dk/sto

Ordbøger:

DDO = *Den Danske Ordbog*. (Hovedred. Ebba Hjorth & Kjeld Kristensen). København: Det Danske Sprog- og Litteraturselskab og Gyldendal, 2003-2005.

RO = *Retskrivningsordbogen* (2001), elektronisk version.

STO = Braasch, Anna m.fl. (red.): *Sprogteknologisk Ordbase*. København: Center for Sprogteknologi, 2001-2004. www.cst.dk/sto

Anden litteratur:

Braasch, Anna, Costanza Navarretta, Sanni Nimb, Sussi Olsen, Bolette S. Pedersen & Claus Povlsen 2004-2005: *STOs Lingvistiske Specifikationer*. CST: København (Upubliceret dokumentation, kan udleveres efter aftale).

Braasch, Anna & Sussi Olsen 2005: Den SprogTeknologiske Ordbase (STO) for dansk – Leksikografiske aspekter. I: *LEDA-Nyt nr. 39 - Marts 2005*, 13-20. Leksikografer i Danmark, København: Center for Sprogteknologi.

Braasch, Anna & Bolette S. Pedersen 2002: Recent Work in the Danish Computational Lexicon Project "STO". I: Braasch, A. & C. Povlsen (red.), *The Tenth EURALEX International Congress, Proceedings, Vol. I*. Copenhagen: CST, 301-314.

Hansen, Aage 1967: *Moderne dansk*. København: Grafisk Forlag.

Jongejan, Bart & Dorte Haltrup 2005: *the CST Lemmatiser*, version 2.7 (23. august 2005). Center for Sprogteknologi. (Publiceret på <http://cst.dk/online/index.html>).

Kilgarriff, Adam & Michael Rundell 2002: Lexical Profiling Software and its Lexicographic Applications – a Case Study. I: Braasch, A. & Povlsen, C. (red.) *The Tenth EURALEX International Congress, Proceedings, Vol. II*. Copenhagen: CST, 807-818.

Olsen, Sussi 2005: STO – En sprogteknologisk database. I: Peter Widell & Mette Kunøe (red.): *10. Møde om Udforskningen af Dansk Sprog*. Århus, 289-297.

Pedersen, Bolette S. 2005: Datamatisk leksikografi i Norden – status og visioner. I: Ruth V. Fjeld & Dagfinn Worren (red.): *Nordiske Studier i leksikografi. Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24. mai 2003*. Oslo: NFL, 302-314.

Pollard, Carl & Ivan Sag 1994: *Head-Driven Phrase-Structure Grammar*. Chicago & London: The University of Chicago Press.

Ørnsnes, Bjarne 2004: Automatisk opbygning af et LFG-baseret datamatisk leksikon for dansk. I: Henrik Holmboe (red.): *Nordisk Sprogteknologi. Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. København: Museum Tusulanum, 211-237.

Den fraseologiske terminologi

The phraseological terminology. The purpose of this paper is to bring law and order to the phraseological terminology. Our current “terminology” is not only systematically undignified, we simply do not understand each other at conferences and the like because it is often impossible to decode the speaker’s terminology and the contents of what he is saying simultaneously. But the worst thing is that thinking is so much influenced by language that as long as our terminology is not well-established, we can hardly do creative empirical research in our field. I go through all the key concepts: *phrases* which are divided into *collocations* (divided into *open* and *closed*) and *idioms* (divided into *metaphorical* and *non-metaphorical* (further divided into *transparent* and *non-transparent*)). All the definitions are listed at the end of the paper. I have tried to avoid idiosyncrasy, by clinging to the best of our traditions.

1. Symptom og signal

Lad os for en gangs skyld begynde helt fra bunden – med grundforskellen mellem udtryk og betydning. Fænomener uden betydning i nogen som helst forstand kan vi kalde *asemantiske fænomener*. Det er fx en torden, en solnedgang, en hosten, en jagt. Oehlschläger skriver ganske vist

En sagte Torden dunder.
Hele Norden undrer.

som Naturens signal om, at den har skænket Danmark et helligt oldtidsminde. Det er også almindeligt at opfatte sorte katte og hundeglam som Skæbnens varsel om død og ulykke. I videnskaben vil vi dog fortsat betragte torden som torden og hundeglam som hundeglam. Dette fører straks over i et andet fundamentalt skel, som alle amatører ignorerer, nemlig skellet mellem *symptom* og *signal*. Hvis jeg ryster på hænderne, udsender jeg et symptom på nervøsitet, jeg signalerer den ikke. Tilsvarende kan Ofelia stille sig i en taleafstand af kun 10 cm fra Hamlet som et symptom på, at hun er forelsket i manden. – Alle fænomener er som regel symptomer på et eller andet. Den omtalte torden er symptom på elektriske spændinger i de højere luftlag, den hvidklædte morgenhave er symptom på, at det har sneet samme nat osv. Disse symptomer kaldes i dagligsproget *tegn på* noget, modsat *tegn før noget*, som er signaler. Dagligsproget er meget klogere end de halvstuderede røvere. Symptomerne er sprogvidenskaben helt uvedkommende. Hvis Ofelia derimod prøver at *meddele* Hamlet sin forelskelse ved at stå tæt ved ham, nærmer vi os det sproglige, men da dette tegn ikke er *etableret*, er det ikke rigtig kommunikation, kun *forsøg på kommunikation*. De asemantiske fænomener kan imidlertid undertiden som *afledt* funktion blive til signaler, endda etablerede signaler. Et *av!*

er primært et symptom, ikke et signal, men det er også blevet etableret med signalbetydningen 'Lad være med det dér, det gør faktisk ondt!' Og en tøvlyd, et *øh*, der primært er et symptom på, at afsender ikke kan finde ordene, kan også fungere som et etableret signal: 'Jeg har stadig taletur, lad være at afbryde mig.'

Foruden denne dobbeltfunktion er der også den *fingerede symptomfunktion*. At tage begge hænder op for sit bøjede hoved har grundfunktion som et symptom på fortvivelse. Men mennesker er ikke sene til at lyve med hænderne. Man kan således udmærket tænke sig folk, som ved en stor leders begravelse bruger denne bevægelse, skønt de i deres stille sind er godt tilfredse med dødsfaldet.

Mere raffineret end denne bastante vildledning, man kunne næsten sige løgn, er Ofelias nævnte variant: Symptomet er ikke rent fup, men samtidig udsendes det som forsøg på et signal, som man dog ikke tager ansvar for. I min skoletid var det almindeligt at gå med spejderuniform, hvis der efter skoletid var spejdermøde. Det var uimodsigeligt et symptom, men det fungerede udmærket som signal: 'Ja, jeg er spejder, I må godt beundre mine duelighedstegn!' Der var også engang en ung kvinde, der klædte sig af i sin lejlighed, skønt jeg sad i et andet vindue få meter derfra. Var det blot et symptom på lyst til at skifte tøj, eller signalerede hun: 'Du må gerne beundre min skønne krop.'? Kun hun selv vidste det. Spejderuniformen og afklædningen var under alle omstændigheder i det højeste *uetablerede signaler*; havde de været etablerede, kunne der ikke fingeres ren symptom-funktion. Lad os kalde denne ikke-svigefulde overgangsform mellem symptom og signal for *frivilligt offentliggjort sandt symptom*. Når man udsender et frivilligt symptom, som fx at tørre sveden af panden i andres nærværelse, så har man samtidig godkendt de andres normaltolkning af symptomet – og det véd de andre. Der er derfor ikke langt til signalet "Jeg har det varmt!". Men så længe jeg ikke har signaleret det tydeligt, tager jeg heller ikke ansvar for det. Man tager kun indholds-ansvar for signaler. Og derfor kunne den strippende dame med god ret spille himmelfalden, hvis jeg havde nærmet mig: "Har man nu ikke lov til at klæde sig af i sin egen lejlighed??" Så jeg foretog mig intet yderligere.

2. Udtryk

Semantiske udtryk – af enhver art – kaldes *udtryk*. De kan være *non-verbale* som fx en knytnæve eller det modsatte: tryglende foldede hænder. De *verbale udtryk* frembringes med taleorganerne. Louis Hjelmslev, der altid skulle være så original og dyb, kaldte dem *tegn*, men i dagligsproget er tegn jo forbeholdt symptomer (med præpositionen *på*) og non-verbale udtryk (med *for*), og man skal aldrig – aldrig – gå terminologisk imod dagligsproget, hvis man ikke har en eksplicit pointe med det. Det havde Hjelmslev ikke. Hvad han kaldte *tegn*, bør fortsat hedde *verbale udtryk*. Hvis udtrykket ikke består af flere udtryk, kaldes det et *morfem*. Morfemerne danner på alle sprog *ord*. Dem har jeg forsøgsvis og ufuldstændigt defineret i Brink 2000. Min definition fremgår af oversigten i afsnit 11. Men den behøver vi ikke at

tage stilling til her. Eller, rettere, det kræver en selvstændig konference. Vi må blot gøre os klart, at eftersom fraseologien per definition er et studium af *flerordsudtryk* (ordforbindelser), sejler vi med et lille lig, et ord-lig, i lasten, så længe vi ikke til bunds har defineret dét.

2.1. Etablerede udtryk

Vi har nu nærmet os fraseologien. De 3 sidste ord i sang-linjen *Se den lille kate-killing, nej, hvor er den sød!*, altså *er den sød*, er ubestrideligt et flerordsudtryk, men det er ikke *etableret*, dvs. *husket, lagret*, af nogen dansker. Fraseologien beskæftiger sig kun med *etablerede flerordsudtryk*. Jeg foretrækker klart termen *etableret* frem for *leksikaliseret*, både fordi den er mere mundret, og fordi *leksikaliseret* er stærkt misbrugt, næsten et fagligt modeord. Det ses i flere konkurrerende betydninger. Nogle bruger det kun om fx *sommerfugl, tørklæde, spille op, helt i skoven* modsat *svømmefugl, silkeklæde, spille klaver, gå i 'skoven*, altså om udtryk med en *pålydende betydning (face value/meaning)*, der afviger fra den *faktiske betydning*; andre om begge dele. Dét er højst uheldigt. Det fælles er, at udtrykkene er huskede, lagrede, de dannes ikke af den enkelte afsender. Og dét kalder jeg *etableret*.

2.2. Uforudsigelige etablerede udtryk

Det særlige *sommerfugle*-fænomen er jo helt centralt i fraseologien, og vi kommer ingen vegne uden en fast term for dét. Jeg foreslår derfor – lige ud ad landevejen – *uforudsigelig betydning*, som naturligvis ikke er forbeholdt flerordsudtryk. Både *flodhest* og *hest* i betydningen ‘bestemt gymnastikredskab’ og ‘bestemt skakbrik’ har uforudsigelig betydning. Samt naturligvis *hest* – og alle andre enkeltmorfemer – i grundbetydning.

Der er også den store svaghed ved *leksikaliseret* i den sidstnævnte brug, at det lægger op til, at det pågældende udtryk er *ligeså* relevant i et leksikon, en ordbog, som enkeltordene, men det er de allermest banale etablerede flerordsudtryk jo ikke. Udtrykket *at have for mange penge* er, skønt etableret, bestemt ikke ligeså relevant i en ordbog som *penge*.

Dagligsprogets betegnelse for (overraskende) uforudsigelig betydning er *overført betydning* (i ordbøger også: *figurligt*), men dét passer ikke godt, når der ingen billedtale er indblandet. Vi kan ikke godt sige, at *sommerfugl* har overført betydning, dels fordi udtrykket aldrig har haft nogen “uoverført” betydning, dels fordi der ikke er tale om metaforik. *Uforudsigelig betydning* er altså en betydning, der afviger fra den pålydende betydning, og *pålydende betydning* er den betydning, et udtryk har for dén, der kender alle sprogets regler undtagen netop den aktuelle faktiske betydningsvariant af det pågældende udtryk. Den pålydende betydning af *sommerfugl* er altså ‘fugl, der er karakteriseret ved noget sommerligt’. Og den pålydende betydning af *slå op* er ‘slå et slag opad’, mens den faktiske betydning jo er ‘bryde et

kæresteforhold'. Man må her gøre sig klart, at de enkelte morfemer i deres grundbetydning ikke har nogen pålydende betydning. For *hest* i sin grundbetydning betyder ikke noget for den, der ikke kender den faktiske betydning, hvorimod *hest* om gymnastikredskab har den pålydende betydning 'hest'. Gymnastikbetydningen er uforudsigelig. Det mærker vi tydeligt, når vi korrigerer en misforståelse og siger "Nej, jeg taler om *rigtige* heste." Vi kender alle den *rigtige* betydning af *heste*, *stjerner*, *torsk* og *negerkys*. Uforudsigelig betydning er uforudsigelig i forhold til komponenternes betydning på det øverste analyseniveau, ikke på et lavere: *At have sommerfugle i maven* har uforudsigelig betydning, men ikke i kraft af, at *sommerfugl* selv er uforudsigeligt.

Dagligsproget har også en anden måde at betegne den forudsigelige, dvs. den pålydende betydning end ved hjælp af *rigtig*. Nemlig: *den bogstavelige mening*. Dette opfindsomme udtryk, hvis egen betydning ikke skal tages bogstaveligt, er ikke så godt i videnskaben, fordi det aldrig bruges om enkelt-morfemer. Man taler ikke om den bogstavelige mening af *stjerner* (man siger *rigtige stjerner*). Men i sprogvidenskaben er det mht. forudsigelighed ligegyldigt, om der er tale om ét el. flere morfemer, om vi taler om den pålydende betydning af *hest* (i skak) eller *spille på flere heste*. Begge dele er metaforisk. Desuden: Et oversættelseslån er som bekendt et ord, hvis pålydende betydning – og kun den – er oversat til modersmålet, fx *frispark* < *free kick*. Her virker det akavet at sige, at ordet oversætter den engelske bogstavelige mening, for udtrykket har i engelsk aldrig sin pålydende betydning, dén er nemlig uetableret, og det er oven i købet svært at se, hvad den skulle være – 'et spark, man ikke bliver straffet for, et gratis spark, et spark efter forgodtbefindende'? Efter min opfattelse er *free kick*'s pålydende betydning = 'spark, der i en eller anden forstand er frit': Den, der ikke kender den faktiske betydning, véd, at det er et eller andet spark, der på en eller anden måde er 'frit'; han kan regne ud, at udtrykket er uforudsigeligt, men kan ikke sige, hvad det ud fra pålydende præcist skulle betyde. Den situation er ganske almindelig. *Et fransk frimærke* er et frimærke udstedt af den franske stat, men *et fransk kys* (som i min mors ungdom betød det samme som det, de unge i dag kalder en *tungeslasker*) kan umuligt være et kys udstedt af den franske stat, udvekslet i Frankrig, af franskmænd e.l. Den, der ikke kender den faktiske betydning, men samtlige sprogets andre regler, kan gå ud fra, at der her gemmer sig et teknisk underbegreb under 'kys', som på en eller anden ret irrelevant måde har forbindelse til de i sin tid så frivole franskmænd. Heller ikke her kan man heldigt tale om den *bogstavelige* mening – for der er ingen. Man kan derimod udmærket tale om den pålydende betydning – som er overbegrebslig – og den uforudsigelige.

3. Fraser

Vi kommer nu til spørgsmålet, om fraseologiens emne, fraserne, omfatter alle etablerede flerordsudtryk. Dvs. ikke blot de stensikre, uforudsigelige, som *købe katten i*

sækken og *spille fandango*, men også de mindre farverige, forudsigelige, som *Adam og Eva*; *gå en tur*; *Se nu dér!*; *lige akkurat*; *Jeg ved det ikke*; *Så er dét en aftale*. De er i hvert fald som oftest ude af fokus i fraseologien. Og grunden er indlysende. De har netop ikke uforudsigelig betydning. Hvis betydnings-uforudsigelighed skal være et krav til frasen, må vi altså gøre os klart, at sådanne meget faste udtryk falder udenfor.

Men vi skål selvfølgelig have en term for hele gruppen af etablerede flerordsudtryk, og dén betegnelse, jeg lige nu brugte, ligner ingen term, den er for tung og har karakter af dét, den er, en definition. Jeg vil derfor med Ken Farø kalde hele feltet for *fraser*. Faktisk kalder han dem *frasemer*. Men det vil jeg fraråde. Hvis udtryk på *-'em* ikke står i modsætning til ord uden *-'em*, så snylter de meningsløst på *fonfonem*, *graf-grafem*, ja, jeg vil endda sige, at de direkte vildleder til at se en sådan modsætning. Og så længe vore termer i almindelighed ikke ender på *-'em*, altså så længe vi ikke siger *subjektem*, *sætningem* og *udråbstegnem*, er der ingen god grund til at optage dem i sprogvidenskaben. Jamen, *fraser*, er dét ikke en dårlig term pga. dagligsprogets *tomme fraser*? Nej, ikke af dén grund. For det er indlysende, at vi ikke kan bruge *frase* fagligt i dén betydning. Dagligsprogets tomme fraser fordeler sig på mindst tre sprogvidenskabelige kategorier: 1) *Sproglige klichéer*, som jo er præcist, 2) *Høflighedsfraser*, som er en særlig kategori, vi kommer til om lidt, og 3) Etablerede ytringer af særlig høj frekvens som fx *Jeg elsker dig*, der jo ingenlunde er en kliché. – Der er i vor videnskab aldrig brug for dagligsprogets højst tågede begreb *fraser*. Og hvor man er bange for, at ens modtager skal forveksle det med dagligsprogets fraser, kan man udvide det til *sproglig/lingvistisk frase*. Så har man klart signaleret det som term. Langt værre er det, at *frase* fagligt i forvejen bruges om Chomskys *led*. Men dét mener jeg var et tilbageskridt. Vi har i europæisk tradition termerne *led*, *nominal*, *hypotagme*, *subjekt*, *prædikat* og *præpositionsled*. Og de er fuldt tilstrækkelige. Der er ingen brug for Chomskys *fraser*, som Chomsky skabte, fordi amerikansk manglede begrebet *led*, tysk *Satzglied*. Synd for amerikansk, men vi skal selvfølgelig ikke overtage deres erstatningsvarer. Der er imidlertid en tredje brug af *frase*, dog ikke særlig udbredt. I intonationslæren er vi nogle, der (ligesom i musikken) taler om *fraser* om *de sætnings-intonationsmæssige enheder*. Denne brug er imidlertid så sjælden og så klart forskellig fra fraseologiens brug, at det ikke gør ret meget. Jeg synes ikke, at denne brug kan berettige til oprettelsen af det vildledende og pointeløse *frasem*. Vi bør ikke lave en kunstig term med suffikset *-'em* uden at mene noget alvorligt med *-'em*.

Til sidst må det nævnes, at en frase i ét tilfælde kan udgøres af et enkelt ord – nemlig når dette udgør en hel ytring, som fx *Tak!* eller *Ud!* Disse skal selvfølgelig ikke blot være etablerede som ord, men også som ytringer. Det fælles for fraser bliver da strengt taget ikke, at materialet består af flere ord, men at det består af materiale af mere kompleks art end enkeltord.

4. Idiomer

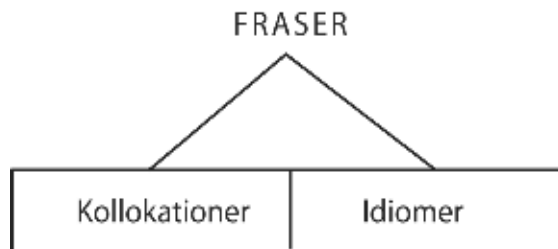
Vores næste term bliver da naturligt en term for de fraser, der har uforudsigelig betydning. Og det kan kun blive *idiom*. Under dette får vi så – lige akkurat – *høflighedsfraserne* med: *Det glæder mig at træffe Dem; Det er længe siden; Hvad går du så og laver?* De behandles i sprogsamfundet, som om de har fuld pålydende betydning, men dét, der gør dem til høflighedsudtryk, er, at vi alle ved, at vi skal trække en del fra. Hvis den pige, jeg længe har sværmet for, siger “Godt at se dig, Lars!”, så skulle jeg jo juble indvendig, men det gør jeg ikke. Vi får også de fleste *sproglige klichéer* med, dvs. de etablerede udtryk, der indebærer et fravalg af det normale, ligefremme udtryk i situationen til fordel for en tilsigtet stilistisk effekt. Fx *den gamle have* om Tivoli: Hvis man ikke kender den faktiske betydning, må man tro, at det er en gammel have, der lige er blevet omtalt. Eller vi har villaejeren, der vil diskutere et par ting med sin nabo over hækken, og på et tidspunkt siger: “Så går vi over til næste punkt på dagsordenen.” Når dette udtryk bruges i Folketinget, har det derimod sin fulde pålydende betydning og er således intet idiom, om end højst konventionaliseret. Det sidste kalder jeg en *rutine-ytring* ligesom auktionarius’ *Ingen højere?* eller danselærerens *Og holdt!* eller skiltens *Tobaksrygning forbudt*. De nævnte er dannet efter reglerne og er dermed ikke idiomer. Dét gælder til gengæld ikke rutine-ytringen *Passér gaden!* Dén er idiomatisk. Idiomerne lapper ganske tilsvarende ind over de egentlige *rituelle ytringer*, som er ytringer, der skål siges i bestemte situationer, fx den aronitiske velsignelse. De kan være idiomatiske eller rene kollokationer, men er naturligvis altid fraser.

I den gamle, slappe leksikografiske terminologi dækkede *idiom* normalt hverken “partikel-verber” eller ordsprog. Og visse leksikografer vægrer sig ved at lade *idiom* dække disse. Men jeg tror, at grunden var, at man ikke brød sig om overlappende termer. *Partikel-verber* og *ordsprog* er ret præcise termer, *idiom* derimod var uklart, og man syntes så, at diverse fraser måtte være enten-eller. Men hvis menings-uforudsigelighed er et relevant kriterium – og det er det naturligvis – så skal vi selvfølgelig have en term for alle etablerede uforudsigelige ordforbindelser, og det kan kun være *idiomer*. Vi må så se i øjnene, at visse partikel-verber som *kaste op* også er idiomer, mens andre som *kaste ned* ikke er det, ligesom visse ordsprog er idiomer, fx *Man skal ikke skue hunden på hårene*, mens andre ikke er det, fx *Lykken er ikke gods eller guld*. Først da får vi orden i sagerne. For skellet mellem meningsmæssig regelret forudsigelighed og uforudsigelighed er jo lige relevant, hvad enten udtrykket består af en sætning, en verbalforbindelse eller noget andet. Og hvis *idiomer* godt må omfatte hele sætninger – hvad alle tillader – er det usagligt at ekskludere netop ordsprogene.

5. Kollokationer

Dét var de uforudsigelige fraser, idiomerne. Hvad så med de forudsigelige? Dem er der en vis tradition for at kalde *kollokationer*. Og det er da en udmærket tradition, hvis den ellers overholdes. Kollokation betyder jo efter sit latinske pålydende ‘sammenstilling’, underforstået ‘rén sammenstilling’, men da vi i dag ikke er så gode til latin, er der intet i vejen for kun at bruge det om fraser. Som Farø så glimrende har gjort opmærksom på, er der to afgørende forskellige kollokationer, som jeg (efter Clausén & Lyly) vil kalde *åbne* og *lukkede kollokationer*. De åbne er fx *at gå på vandet*. Dét udelukker hverken *at spadserere på vandet* eller *at gå på vandoverfladen* eller noget andet med samme sandhedsbetydning (stilnuancer er der naturligvis). Her er der ingen problemer for hverken afsender el. modtager: Når man kender enkeltordene og sprogets generelle syntaks, giver konstruktionen sig selv uden yderligere regler. – Derimod kan *Adam og Eva* ikke hedde noget som helst andet, hverken *Adam plus Eva* eller *Adam samt Eva*, og da Adam nu engang er den ældste og et helt hoved højere end Eva, er det også unaturligt med *Eva og Adam*. Og tilsvarende med fx *Han måtte trække sin cykel hele vejen*. Betydningen er helt forudsigelig, modtager har ingen problemer, men afsender har. Han skål nemlig bruge dette udtryk, selvom *skubbe sin cykel* ville være ligeså naturligt. I Tyskland *skubber* man sin cykel, dér nede *gør* man en rejse – helt forudsigeligt, men i Danmark er dét blevet gammeldags, hér skal man helst *foretage* en rejse. Vi *laver* mad, vi *binder* en knude og *begår* et mord. Betydningen er helt forudsigelig, men konstruktionen som sådan er det ikke. Den er *lukket*, ja så lukket, at danske børn har haft problemer med dem de sidste generationer, hvor *lave* har ekspanderet voldsomt. Børn og unge *laver* undersøgelser, *laver* indbrud og *laver* et overfald, alt for mange overfald. De forvandler dermed lukkede kollokationer til åbne, for dé er nemmere at have med at gøre.

Lad os lige sammenfatte det vigtigste af det foregående i et skema, som jeg inderligt ville ønske, at vi lagde os fast på – for vi mærker jo på hver eneste konference vort terminologiske babylon. Meget ville være vundet, hvis vi blot kunne enes om den simple sammenhæng, der fremgår af figur 1.



Figur 1.

6. Metaforisk og umetaforisk idiom

Et vigtigt skel blandt idiomerne er dét mellem de *non-metaforiske* og de *metaforiske*. De sidste rummer en implicit sammenligning. De kaldes efter min mening bedst slet og ret *metaforiske idiom*er. De kan hverken kaldes *ikonografiske* eller *billed-idiom*er, eftersom der skål være tale om en *implicit* sammenligning. *Sort som kul* og *som et skib uden ror* er så ikonografiske, som det kan ønskes, men de er ikke metaforiske. De metaforiske idiomer har altid – det følger af deres definition – gennemsigthed (se nedenfor) mellem pålydende og faktisk betydning, men det er de ikke ene om. Det har de non-metaforiske idiomer *holde mund* og *spytte blod* også.

Da mange idiomer er metaforiske, kan det være gavnligt at pege på nogle, der ikke er det. Jeg har allerede nævnt *holde mund*. Et andet er *117 gange* = ‘overordentlig mange gange’. Svenskerne siger sjovt nok *femti-elva*. Eller i samme boldgade *et par stykker*; her siger spanierne *cuatro*! I spansk en ordbetydning, i dansk et idiom. Man kunne også nævne alle de (mere eller mindre) uigennemsigtige idiomer, der qua uigennemsigtige ikke kan være metaforiske. Fx [Hun] *har det med at* [spise sæbe]. Eller [Noget] *er* [nogen] *meget om at gøre*.

7. Gennemsigthed

Som antydet viser idiomerne forskellig *gennemsigthed*: Kollokationerne er per definition alle gennemsigtige, men idiomerne er det kun, når det er let at årsagsforbinde forholdet mellem deres komponenter og den faktiske betydning (uagtet dette forhold jo ikke er regelret). *At falde på halen for* ‘blive slået af beundring over’ er klart nok menings-uforudsigeligt, men vi kan sætte alle komponenter i naturlig forbindelse med den pålydende betydning og se denne som en *sammenligning* med den faktiske betydning. I idiomerne *holde mund* eller *give sig i kast med* oplever vi ikke nogen sammenligning; man oplever, hvis man ellers tænker lidt over sproget, at *mund* kan forstås som et naturligt udtryk for ‘tale’ og *kast* som en ‘pludselig begyndelse’. Helt úigennemsigtige fraser er der få af (men gennemsigthed går naturligvis i grader, jf. Farø 2003). Jeg har fundet *skyde papegøjen* og *ikke stikke op for bollemælk* – begge på vej ud af sproget, netop af den grund.

8. Integrerede og uintegrerede idiomer

Visse idiomer kan forveksles med enkeltord i sær-betydning. Nemlig de *uintegrerede* i modsætning til de *integrerede*. I et integreret idiom bærer hele udtrykket betydningen; ingen ord har forudsigelig faktisk betydning. I det uintegrerede idiom har visse ord forudsigelig faktisk betydning. Metaforiske idiomer er integrerede: *glide*, *ud*, *ad*, *tangenten* har ingen selvstændig forudsigelig betydning i det bekendte idiom; de danner den faktiske betydning i fællesskab. I det umetaforiske idiom *slå 'ud* ‘slå bevidstløs’ har *slå* sin normale betydning; kun *ud* er uforudsigelig. Men det

ér en frase, fordi forbindelsen er etableret, har mindst to ord og rummer en uforudsigelig betydning/syntaks. – Tilsvarende er *slå'ud* 'få udslæt' uintegreret, men her er det, lige omvendt, *slå*, der er uforudsigeligt, mens *ud* har sin normale retningsangivelse. – Til gengæld er *slå'op* (om mennesker, ikke om hopper!) integreret: Hverken *slå* eller *op* bidrager med noget selvstændigt og forudsigeligt til betydningen 'bryde et kæresteforhold'.

9. Frase og konstruktion

Det er vigtigt at indskærpe forskellen mellem frase og *konstruktion*. En frase er netop et helt bestemt udtryk, et eksplicit udtryk, om jeg så må sige. En konstruktion er et *uudfyldt skema*. Selve forbindelsen *at give ngn. ngt.* er en etableret *konstruktion* for alle, der kan dansk, men det er ingen frase. Dét er derimod *at give julegaver*: et etableret flerords-udtryk med et forudsigeligt indhold efter komponent-reglerne, altså en kollokation, og *at give bagerbørn hvedebrød*: et etableret flerords-udtryk med uforudsigeligt indhold, altså et idiom. Fraser kan *indgå* i en konstruktion: [Nogen/Subjekt] *giver bagerbørn hvedebrød*, men den udfyldte plads i konstruktionen hører ikke med til frasen.

Alle mine definitionsforsøg skal forstås som stipulative hensigtsmæssige præciseringer af dagligsprog eller fagtradition. Dermed hverken kan eller skal de være knivskarpe. Sproget har i modsætning til atompartikler alle mulige drilske overgangstilfælde. Det kan derfor godt være en dyd at lade en vis grænseuklarhed blive tilbage.

Lad mig give et eksempel på et overgangstilfælde. Jeg skrev engang: "Min lille datter kan Møllehave udenad." Det var en bevidst tvetydighed. Den spiller på to fraser: 1) *At kunne* [noget] *udenad*, 2) *At kunne* [nogen] *udenad*. Den første må kaldes en kollokation, komponenterne har normal betydning. Det er også regelret at indsætte et proprium i stedet for [noget], idet proprium da betyder 'NN's tekster', jf. "Jeg læser Bodelsen". Her er der intet uforudsigeligt (selvom ordet *udenad* i sig selv er uforudsigeligt). Den anden frase, *at kunne* [nogen] *udenad*, er en metaforisk brug af den første: Man sammenligner et menneske, fx Møllehave, med en tekst indeholdende, hvad dette menneske går rundt og siger, og hævder så, at nogen kan denne tekst udenad, dvs. kán den til kvalmepunktet. Metaforik, javel, men lige på grænsen, fordi det at nævne manden i stedet for hans mundtlige kæpheste mere minder om pars pro toto end om en normal sammenligning.

Et andet vanskeligt tilfælde er frasen *ikke sandt?* Ingen tvivl om, at den er en frase. Men er den et idiom? Dertil kræves jo betydnings-uforudsigelighed. Næppe nogen har nogensinde været i tvivl om betydningen, første gang de hørte den. Det sørger ordene *sand* og *ikke* for. Jeg tror, det er for pedantisk at sige, at betydningen ikke stemmer 100 % med komponenternes regelrette delbetydninger. Den faktiske betydning er snarest 'Jeg går ud fra, at du er med/enig så langt?', og dét ville et regelret,

uetableret udtryk som *Er dét ikke sandt?* også kunne betyde. Men syntaksen. Til-syneladende er det en simpel ellipse, men i virkeligheden en *unikal* ellipse. Vi kan jo ikke sige *ikke 'morsomt?*, *ikke 'løgn?*, *ikke 'trist?* osv. Vi hår altså med et idiom at gøre – alene pga. syntaksen. Og det afgørende er dermed ikke, at mennesker er så sprogbegavede, at de i praksis kan forudsige betydningen. Det afgørende er, at frasens betydning ikke er dannet efter reglerne.

10. Idiom-tryk

Til sidst vil jeg – måske overflødig – gøre opmærksom på, at dansk fonetisk markerer en hel del idiom. Tager vi det metaforiske (og dermed gennemsigtige) idiom *at tage tyren ved hornene*, så har det tryktab i verbet, men den eneste grund til dét er, at udtrykket er et idiom! Hvis det ikke var dét, ville det hedde **'tage tyren ved 'hornene*. Dét kan det ikke hedde, heller ikke, hvis det drejer sig om en rigtig tyr, for som en vis mand skrev i 1976, så gælder det som hovedregel, at idiom, der er ópstået som idiom, ikke kan bruges i forudsigelig betydning. Man vil da associere til idiomet og afvise et tilsvarende uetableret udtryk. Min søn hedder David, men jeg vil aldrig nogensinde seriøst kunne sige, hvor David købte øllet. Hvis idiomet derimod er overført fra et i forvejen eksisterende etableret udtryk, fx *løbe af stablen*, så kan vi, naturligvis, stadig bruge udtrykket i den gamle betydning. – Idiomatisk trykmarkering har vi også i *'ikke ligge på den 'lade 'side*. Det skulle regelret hedde *'ikke 'ligge på den 'lade 'side*, men det kan man ikke sige. Endvidere: *stå med 'hatten i 'hånden; føle ['nogen] på 'tænderne*. Jeg har desuden fundet idiomerne *stå på 'hovedet* og *ligge på 'ryggen*. Er dé virkelig uforudsigelige idiom? Ja. For dét hoved og dén ryg kan kun være subjektets. Hjemme hos ós leger vi undertiden “Døde Heste”. Jeg siger da fx til min datter: “Læg dig på 'sofaen”, hvortil hun kan svare: “Jamen, dér ligger David!” Jeg: “Så 'læg dig på 'ryggen [altså af sofaen]!”. Jeg siger naturligvis ikke: “Læg dig på 'ryggen!”. Die Sprache ist ein Wunder, das sich nie auswundern läßt.

11. Oversigt over fraseologiske definitioner

Udtryk = Betydningsbærende tegn for noget, som afsender tager ansvar for. (Betydningen kan være et betydningsbidrag til en tilkendegivelse eller en hel tilkendegivelse, og udtrykket kan være verbalt, non-verbalt, etableret eller uetableret). “*Fuckfingeren*”; *Du er jo; Grønne breve; Skrid!*

Etableret = I forvejen husket/lagret af afsender. *Børnehave*.

Uetableret = Ikke i forvejen husket/lagret af afsender, men skabt på stedet. *Børnetanker*.

Morfem = Verbalt udtryk, der ikke består af flere sådanne. *Klarinet*.

Ord = En gruppe af morfemer, der syfter på hinanden og kun kan have deres pågældende semantiske relation ved at stå side om side, dog evt. adskilt af andre sådanne morfemer. Eller et enkelt morfem, der ikke indgår i en sådan gruppe. – Det fælles er da kontinuerthed. *Risretterne; Nu.*

Flerordsudtryk (ordforbindelse) = Udtryk bestående af to eller flere ord.

Frase = Etableret flerordsudtryk / etableret ytring. *Hele verden; Adam og Eva; Gå hjem og vug!; Tak!*

Pålydende betydning = Den betydning, som et udtryk har for den, der kender alle sprogets regler undtagen netop den faktiske aktuelle betydningsvariant af dette udtryk. *'Foretag et løb i min bagdel!'*

Faktisk betydning = Betydning (modsat pålydende betydning). *'Det, du siger, er helt irrelevant!'*

Forudsigelig betydning = Den betydning, der regelret følger af et udtryks komponenters betydning og syntaks. *Havebænk; En stor cigar; En lille pige.*

Idiom = Frase med uforudsigelig betydning. *Lade 5 og 7 være lige; Hold din mund!*

Metaforisk idiom = Idiom med implicit sammenligning. *En god cigar.*

Kollokation = Frase med forudsigelig betydning. *Hele verden; Adam og Eva.*

Lukket kollokation = Kollokation, der udelukker et synonymt udtryk dannet efter de øvrige sprogeregler. *At trække en cykel.*

Åben kollokation = Kollokation, der ikke er lukket. *Stærkt beruset; At gribe en bold.*

Gennemsigtig = Det, at et udtryks faktiske betydning – uanset om den er uforudsigelig, altså ikke-regelret – let kan årsagsforbindes med den pålydende. *Stå til 'søs.*

Uigennemsigtig = Ikke gennemsigtig. *Ikke stikke op for bollemælk.*

Rutine-ytring = Ytrings-frase, hvis forekomst er ret forudsigelig i en bestemt situation. *Tak!; ↑ Og 'holdt!'*

Rituel ytring = Ytrings-frase, der skål siges i en bestemt situation. *Hermed erklærer jeg udstillingen for åbnet; Hvad er barnets navn?*

Høflighedsfrase = Rutine-ytring, der bruges per høflighedskonvention, og som afsender derfor ikke tager fuldt ansvar for. *Det er alt for meget!*

Citat = Anførelse af en ytring, der af afsender signaleres som skabt af en anden, og som han derfor ikke tager direkte ansvar for. (For afsender er det en frase). [Som min far altid sagde:] “*Kvinder og killinger skal forkæles.*”

Bevinget ord = Berømt ikke-anonymt citat. (Ergo frase). [Som det hed i Oraklet:] “*Kend dig selv!*”

Ordsprog = Citat, der af afsender fremføres som en gammel anonym almenyldig ytrings-frase, som afsender ikke hævder (thi det er et citat), men som han normalt bruger som argument og dermed præsupponerer sandheden af. [Det kræver naturligvis mod, men] *hvo intet vover, intet vinder.*

Sentens = Anonym citat-frase, der ikke er et ordsprog. (Herunder de *travesterede* ordsprog). *Tid er penge; Vanskeligheder er (som bekendt) til for at overvindes; Blondes have mere fun; Blåt er altid pænt; Smil til verden, og den smiler til dig; Brændt barn lugter ilde; Hvad jernet er fuldt af, løber munden over med.*

Slogan = Sentens med slagkraftig opfordring. *Danmark for folket; Sig Jolly til din cola.*

Allusion = Til et citat hentydende udtryk, som er integreret i afsenders tale, og som han derfor tager afsender-ansvar for. [Jeg tror, vi her er bedst tjent med et par] *smagsdommere*; [Filmen er] *knald'hamrende* [god].

Sproglig kliché = Etableret sprogræk (oftest en frase), hvormed afsender erstatter det i situationen normale, ligefremme udtryk for at opnå en stil-effekt. (Termen er naturligvis ikke værdiladet i faglig brug. Bemærk, at afsenders intention er afgørende; hvad der er effekt for én, kan være ligefremt for en anden). *Støvlelandet; Das Vaterland; I en lind strøm; Ikke en hujende fis.*

Konstruktion = Udtryk med uspecificerede variabler. *Ngn. giver ngn. ngt.*

Unikal = kun forekommende i ét bestemt poly-morfematisk udtryk. *Skorsten; Revl og krat; Til orde; En på kassen = 'hovedet'. Syntaksen i koste 'dyrt.*

Ovenstående udtømmer ikke den fraseologiske terminologi. Der mangler fx *talemåde*, (*fast*) *vending*, *fyndord*, *løsen*, *slagord*, *motto*, *diktum*, *devise*. De er vage i dagligsproget, og jeg tror, de kan undværes fagligt. Mere veldefinerede er *aforisme*, *epigram*, *anekdote*, *maksime*, *vittighed*, *wellerisme*. De er (ligesom ordsprog) alle *tekster* med æstetisk ambition og kan som sådanne (med utryghed) overlades til litteraterne.

Litteratur

- Brink, Lars 1979: Ordforrådets omfang og vækst. I: *Børnenes sprog – sprogene omkring børn*. København: Hans Reitzel, 109-175.
- Brink, Lars 2000: Kongen af Danmark s bolsjer. I: *Nydanske Studier 26-27*. København: Dansk lærerforening, 35-42.
- Burger, Harald 1998: *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Neuburg.
- Clausén, Ulla 2003: Jämförelsekonstruktioner i Svensk konstruktionsordbok – struktur, variation och funktion. I: *Nordiske Studier i Leksikografi 6*. Tórshavn: Nordisk Forening for Leksikografi, 75-85.
- Farø, Ken 2003: Ordsprog i nutidsdansk: funktioner og problemer. I: *Danske Studier*. København: C.A. Reitzel, 38-64.
- Farø, Ken 2004: Omkring det grønne bord: Dansk-tysk idiomatik og sprogforandring. I: *Studier i nordisk 2002-2003*. København: Selskab for nordisk filologi, 51-75.
- Hansen, Erik 1979: Den sproglige kliché. I: *Danske Studier*. København: Akademisk Forlag, 5-23.
- Hansen, Erik og Jørn Lund 1983: *Sæt tryk på. Syntaktisk tryk i dansk*. København: DLH forskningsserien, bind 6.

Språkstadieordböcker nu och då

Period dictionaries past and present. This paper deals with the typological distinction of synchrony and diachrony in period dictionaries, particularly K.F. Söderwall's *Ordbok öfver svenska medeltidsspråket* (*Sdw*) (1884-1918). First of all the established definitions of the lexicographic terms synchronic and diachronic are given. Then follows an argument for a different set of concepts regarding the descriptive methods in period dictionaries: either describing the language as a system or focusing on the changes it undergoes (essentially regardless of the length of the period). These concepts are also – with some hesitation – labelled as synchrony and diachrony respectively. In order to concretise these concepts, different synchronic and diachronic descriptive methods applicable to different information categories in period dictionaries are presented. Finally, an article in *Sdw* is briefly discussed. This article does not strongly demonstrate the use of either a synchronic or a diachronic descriptive method.

1. Inledning

I mitt pågående avhandlingsarbete om K.F. Söderwall som historisk lexikograf undersöker jag hans språkstadieordbok *Ordbok öfver svenska medeltids-språket* (1884-1918) (*Sdw*) ur ett lexikografihistoriskt perspektiv. Det övergripande syftet är att undersöka lexikografens och ordbokens förhållande till samtidens lexikografiska praxis. Grundantagandet är att lexikografisk praxis, och därmed lexikografiska produkter, har sin grund i samtidens språkvetenskapliga teoretiska och metodiska tankesystem och hypotesen är att dessa förhållanden låter sig utläsas genom analys av ordböckernas utformning och innehåll.

I denna artikel presenteras ett problemområde som har aktualiserats i avhandlingsarbetets inledningsskede, nämligen den grundläggande typologiseringen av språkstadieordböcker som Söderwalls medeltidsordbok, närmare bestämt distinktionen mellan synkronisk respektive diakronisk beskrivningsmetod. Först redovisas och diskuteras vedertagna definitioner av dessa termer samt den därmed sammanhängande distinktionen mellan samtidsordböcker respektive historiska ordböcker. Därefter definieras den beskrivningsmetodiska begreppsdistinktion som jag vill göra och som jag i brist på bättre termer benämner synkronisk respektive diakronisk. Slutligen presenteras ett uppslag till en undersökning av hur dessa synkroniska och diakroniska beskrivningsmetoder kan realiseras i språkstadieordböcker och hur de realiseras i *Sdw*.

2. Vedertagna definitioner

Söderwalls medeltidsordbok är tvivelsutan en historisk ordbok då den beskriver ett språk som är äldre än lexikografens samtidsspråk. Söderwall var även i allra högsta

grad inblandad i arbetet med *Svenska Akademiens ordbok (SAOB)*, som också ofta karaktäriseras som historisk, vilket känns intuitivt tilltalande. Den till synes okomplicerade termen historisk tycks emellertid inte användas enhetligt (jämför t.ex. Reichmann 1984:460 och Wiegand 1984:557f med Jonsson 1992:232 och Bandle 2002, se även Nerlich 1992:51). Nordisk lexikografisk terminologi har dock fördelen av att vara stadfast i *NLO* (1997) som tillsammans med Svensén (2004) utgör källan för följande definitioner av termerna samtida respektive historisk samt de därmed nära förknippade termerna synkronisk respektive diakronisk.

Man kan också klassificera ordböcker efter deras förhållande till tiden. Man skiljer mellan synkroniska ordböcker, som beskriver språkets tillstånd under ett begränsat skede, och diakroniska ordböcker, som beskriver språkets utveckling under en längre period. Man skiljer också mellan historiska ordböcker, som beskriver en äldre tids språk, och samtidsspråkliga ordböcker, som beskriver samtida språk. (Svensén 2004:28)

Tillsammans kan de fyra termerna kombineras för en grundläggande typologisk distinktion som i figur 1. Enstaka exempel i figur 1 kan möjligen förefalla kontrainuitiva, framförallt det faktum att *SAOB* definieras som en samtidsordbok, vilket dock ska förstås som historisk-samtidsordbok.¹ Då min avsikt inte är att ifrågasätta vedertagna definitioner – utan snarare att definiera andra begrepp – räcker det för den fortsatta framställningen att konstatera att man räknar med historiska ordböcker av två slag: å ena sidan synkroniska historiska ordböcker som både i *NLO* och av Svensén (2004:28) exemplifieras med författar- eller verkordböcker och å andra sidan diakroniska historiska ordböcker som exemplifieras med språkstadieordböcker. Det är ordböcker av det senare slaget som står i fokus i denna framställning.

	samtidsordbok	historisk ordbok
synkronisk	t.ex. <i>Svensk ordbok</i>	t.ex. <i>Ordbok till Fredmans epistlar</i>
diakronisk	t.ex. <i>SAOB</i>	t.ex. <i>Ordbok öfver svenska medeltids-språket</i>

Figur 1. Exempel på ordbokstyper (*NLO 1997 och Svensén 2004:28*)

3. Beskrivningsmetoder, språkstadieordböcker och språkstadier

Mer besvärande för mina syften är att de båda distinktionerna samtida–historisk och synkronisk–diakronisk uppges avse tidsdimensionen. De begrepp avseende språkbeskrivningsmetoden som jag vill diskutera kan frikopplas från tidslängden. Bo

¹ D.v.s. en ordbok som behandlar både det äldre och det samtida språket (Bo Svensén, personlig kommunikation 2005-05-27).

Ralph (2000) diskuterar hur vedertagna språkstadieindelningar inverkar på språk-historikers behandling av begreppen synkroni respektive diakroni:

periodindelningen för svenska språkets historia [...] har tydligen en psykologisk effekt på oss, som får följer då vi formulerar våra problemställningar. Periodindelningen samverkar exempelvis i vissa resonemang med distinktionen synkroni–diakroni på ett sätt som kan vara ganska förvirrande. (Ralph 2000:371)

Ralph påminner också om ”att synkroni respektive diakroni är olika perspektiv på språket, dvs. egentligen utan någon tidsdimension inblandad, åtminstone då det gäller synkroni” (Ralph 2000:371) och refererar sin tidigare framförda något utvidgade användning av begreppen: ”Synkronisk språkhistoria’ skulle då innebära sådan språkhistorisk verksamhet som koncentrerar sig på synkroniska system i äldre tid, medan ’diakronisk språkhistorieforskning’ verkligen fokuserar tidsdimensionen, dvs. själva förändringen.” (Ralph 2000:372), vilket så vitt jag kan avgöra inte motsägs av de klassiska saussureska definitionerna av begreppen synkroni och diakroni och ligger mycket nära den användning jag vill förespråka.

Tidsdimensionen i språkstadieordböcker kompliceras således ytterligare av själva språkstadiebegreppet. Den vedertagna distinktionen mellan synkroniska respektive diakroniska historiska ordböcker såsom den återgivits ovan antyder att tidsdimensionen, d.v.s. längden på den tid från vilken ordbokens skriftliga källmaterial hänför sig, är avgörande för hur materialet beskrivs. Även Ralph vill begränsa längden på det språkstadium som ska kunna beskrivas synkroniskt:

Med ’synkront stadium’ kan man enbart avse ett momentant perspektiv på språket, dvs. en idealisering, och i praktiken kan i någon mån detta språkbruk utsträckas till principiellt synkrona beskrivningar av givna språkssystem betraktade under en mycket begränsad tidsrymd. Om man med ’synkront stadium’ mera regelmässigt menar ett språkssystem iakttaget under flera hundra år, så missbrukar man Saussures knäsatta termer. (Ralph 2000:372)

Saussures mest explicita definition av begreppen synkronisk respektive diakronisk lingvistik nämner visserligen inte tidens längd men förklarar synkronisk lingvistik som studiet av systemet i ett språkstadium vars utsträckning i tid inte är den huvudsakliga avgörande faktorn, det är istället språkets begränsade förändringar under denna tid som är avgörande (Saussure 1962:140ff), vilket förvisso snarare låter som en definition av språkstadium än av synkronisk lingvistik.

Det är också definitionerna hos Saussure samt idealiserade språkstadiers avsaknad av (system)förändringar som utgör utgångspunkt för den karaktäristik av språkstadieordböcker som flera tyskspråkiga lexikografer ger uttryck för. T.ex. förespråkar Reichmann (1990a:1417) en avgränsning av språkstadiet utifrån ett antal kriterier som avser att ringa in en tidsrymd när språket är så stabilt att det kan beskrivas som

ett system. I sin strängaste form är denna typologisering av språkstadieordböcker alltså beroende av en starkt idealiserad språkstadiedefinition. I realiteten tycks dock språkstadieordböcker typologiseras som synkroniska även om de språkstadier ordböckerna behandlar inte möter dessa strikta språkstadiedefinitioner (Reichmann 1984:463–465) och i detta sammanhang har även Reichmann definierat synkroniska historiska ordböcker som beskrivande lexikon vid en viss tid (Reichmann 1984:461). Det råder ingen enighet om typologiseringen av historiska ordböcker heller bland tyskspråkiga lexikografer, men huvudsakligen är synkroniska historiska ordböcker i denna tradition liktydigt med språkstadieordböcker, medan diakroniska historiska ordböcker representeras av ordböcker av *SAOB*-typ (Reichmann 1984:460ff, 1990b:1589, Wiegand 1984:557f, van Sterkenburg 1990:1430f). Samtidigt medger dock dessa definitioner den teoretiska möjligheten att språkstadieordböcker kan vara synkroniska och/eller diakroniska.

Inte heller *NLO*:s definition av språkstadieordbok tar ställning till distinktionen synkronisk–diakronisk (men omtalar systemet, vilket skulle kunna antyda en synkronisk beskrivning) (*NLO* 1997:214f) och avsaknaden av ställningstagande indikerar att distinktionen synkronisk–diakronisk inte är konstituerande för ordbokstypen och håller följaktligen öppet för att språkstadier kan beskrivas antingen med fokus på systemet eller på förändringen.

Trots att jag vill frikoppla synkroni–diakroni-distinktionen från tidsdimensionen står det givetvis klart att olika språkstadier kan vara mer eller mindre lämpade för endera av de båda beskrivningsmetoderna och att språkstadiet i *Sdw* (grovt räknat fr.o.m. slutet av 1200-talet t.o.m. början av 1500-talet enligt handskriftsdateringarna i källförteckningen) skulle kunna vara direkt olämpligt för en synkronisk beskrivning. De språkstadiindelningar som språkhistorikerna dras med är, som Ralph (2000:370) påpekar, dock ofta resultat av andra överväganden.

4. Begreppen synkronisk respektive diakronisk – definitionsförsök

Det råder således en diskrepans mellan å ena sidan de i Norden vedertagna termerna synkronisk respektive diakronisk och å andra sidan de båda begrepp gällande olika beskrivningsmetoder som jag åsyftar. Synkronisk respektive diakronisk handlar i min mening om hur språket under ett kortare eller längre historiskt (eller samtida) språkstadium beskrivs.

Språkstadiindelningen som sådan kan som framgått väcka två diametralt skilda föreställningar. Den ena föreställningen tar fasta på språkstadiet som en tid när språket är enhetligt och kan beskrivas som ett system. Beskrivningen i språkstadieordböcker som utgår från denna språkstadiesyn benämns i brist på bättre term *synkronisk* i denna artikel. Enligt mitt synsätt behandlar en synkronisk ordbok språket som ett system och ingen hänsyn tas till att språkbrukarna använder både arkaismer och neologismer. Diakronisk markering är bara en av många diasystemati-

ska markeringar av språkstadiets variation och behandlar endast sådana ord, former, betydelser m.m. som förekommer i systemet, om än perifert. Det språkstadium språkstadieordboken beskriver kan vara kort eller långt men i det senare fallet krävs naturligtvis en högre grad av abstraktion eller idealisering i förhållande till materialet.

Den andra föreställningen tar fasta på språkstadiet som en tid när språket genomgår sådana förändringar som med fördel beskrivs tillsammans. Beskrivningen i språkstadieordböcker som utgår från denna språkstadiesyn benämns här *diakronisk*. Enligt mitt synsätt behandlar en diakronisk ordbok språkets förändring och framställer denna som ett tidsförlopp oavsett om det behandlade språkstadiet är långt eller kort. T.ex. kan en semantisk förändring eller en lexikal novation ske snabbt men ska ändå beskrivas som en process om det är förändringen som är föremål för beskrivning.

5. Synkroni och diakroni – begreppens validitet och relevans

Det är givetvis grovt anakronistiskt att utifrån strukturalistiska begrepp och termer typologisera Söderwalls ordbok, som i allt väsentligt får betraktas som rentutav prejunggrammatisk, d.v.s. snarast som en produkt av det romantiska historisk-komparativa språkvetenskapliga paradigmet. Det kan vara befogat att fundera på huruvida äldre ordböcker överhuvudtaget kan typologiseras med hjälp av nutida begrepp och termer. Dagens metalexikografer karakteriserar således språkstadieordböcker som synkroniska eller diakroniska historiska ordböcker, men dessa båda beskrivningar behöver inte vara oförenliga. Reichmann (1984:463-465) kategoriserar några få språkstadieordböcker som både synkroniska och diakroniska och Svensén (2004:47) framställer synkroni-diakroni-distinktionen som en skala. En annan möjlighet är att *Sdw* och tidens övriga språkstadieordböcker varken är synkroniska eller diakroniska, eftersom distinktionen inte var tillgänglig för dåtidens lexikografer. Det är emellertid en alltför nihilistisk lösning. 1800-talets lexikografer saknade tillgång till flera av de typologiseringskriterier som används idag, ändå måste lexikografiska produkter från det förgångna kunna typologiseras och karakteriseras om lexikografihistorisk forskning alls ska kunna bedrivas. Trots det anakronistiska i terminologin och de bakomliggande begreppen måste begreppsapparaten med tillhörande terminologi kunna nyttjas för att beskriva även äldre ordböcker.

Även inom lexikografen får de rätt abstrakta och principiella språkstadie-indelningsavhängiga föreställningarna metodiska konsekvenser, som Ralph (2000) resonerade om mer generellt angående språkhistoriska beskrivningar. Anhängare av den synkroniska beskrivningen utgår från att språkstadieordböcker behandlar språkstadier, vilka (i idealfallet) utgör system, som per definition ska beskrivas synkroniskt. Förespråkare för den diakroniska beskrivningen menar att eftersom språkstadieordböckerna (oftast) behandlar tidsrymder på flera hundra år under vilka språket tvivelsutan genomgår förändringar måste detta beskrivas diakroniskt. Jag vill

konkretisera typologiseringen genom ett närgånget studium av innehållet i ordböckerna själva. Det vore t.ex. belysande för lexikografins förhållande till samtidens språkvetenskap att studera vilket genomslag distinktionen synkroni–diakroni får i historiska ordböcker efter att Saussures begrepp etablerats. Förutsättningen för en sådan undersökning av de strukturalistiska historiska ordböckerna är möjligheten till en jämförelse med beskrivningar av de verkliga förhållandena i de prestrukturalistiska.

6. Konkretion: synkronisk respektive diakronisk beskrivning

Den terminologiska exercisen utgör inget självändamål. Det är intressantare hur språket beskrivs i en ordbok än vilken beteckning en eller annan lexikograf sätter på den. Därmed är vi framme vid problemets kärna nämligen relationen mellan de abstrakta typologiseringskriterierna och den konkreta språkbeskrivningen i ordboken. Frågan är: Vari består den synkroniska respektive den diakroniska beskrivningen i språkstadiordböckerna?

I figur 2 sammanfattas hur en synkronisk respektive en diakronisk beskrivning skulle kunna skilja sig åt i några för språkstadiordboken framträdande informationskategorier. En utgångspunkt utgör Reichmanns (1984:470) beskrivning av den historiska lexikografins karaktäristika, där det hävdas att diakronin återfinns i ordboksartikelns redovisning av ”’Geschichte der Ausdrucksseite’ und ’Geschichte der Inhaltsseite des Wortes’, z.T. aber auch im Lemma, in den Positionen ’Morphologie/Grammatik’, ’Wortbildungsverhalten’, ’Bedeutungen’ und ’Belege’”. De synkroniska beskrivningsmetoderna ter sig sällan lika självklara som de diakroniska och Reichmanns (1984:470) presentation tar betecknande nog enbart upp den diakroniska beskrivningen, medan den synkroniska anges vara omvänd. En annan utgångspunkt är Hans Jonsson (1992) som använder termen historisk för att beskriva *SAOB*, men också sätter likhetstecken mellan kronologisk och historisk i sitt resonemang om *SAOB*:s mikrostruktur (Jonsson 1992:232). En tredje utgångspunkt är Robert E. Lewis (2004) som behandlar betydelsemomentordningen i historiska ordböcker och språkstadiordböcker (uppdelningen definieras ej). Lewis beskrivning gör de komplexa distinktionerna full rättvisa genom att först skilja på logisk respektive kronologisk (eller historisk) betydelseordning. Den logiska betydelseordningen indelas sedan i ”relational logic” (relationslogisk) respektive ”developmental logic” (utvecklingslogisk), vilket förefaller mig vara ett annat sätt att uttrycka distinktionen mellan synkroni och diakroni (utan att använda de alltför belastade termerna?). Den relationslogiska beskrivningen förevisar i Lewis återgivning klara prototypeteoretiska drag, som att betydelser kan uppstå ur flera befintliga betydelser samt att de viktigaste betydelserna anges först. (Det kan påpekas att t.ex. Geeraerts 1997 istället har förespråkats prototypeteoretisk metod för diakronisk beskrivning.) Den utvecklingslogiska beskrivningen exemplifierar Lewis med *Oxford English*

Dictionary, som i enlighet med det sena 1800-talets utvecklingsperspektiv tillämpar en (strikt) teori om regelbundenhet i semantisk förändring. Lewis poängterar dock att historiska ordböcker och språkstadieordböcker tenderar att blanda de båda beskrivningsmetoderna och att den äldre utvecklingslogiska beskrivningen alltjämt tillämpas, om än inte lika optimistiskt angående regelbundenheten.

	synkronisk beskrivning	diakronisk beskrivning
lemmat	lemmatisering på formell eller semantisk grund; explicit abstrakt (re)konstruktion som representant för idealiserat system	lemmatisering på etymologisk grund; (re)konstruerad äldsta form samt (re)konstruerad yngsta form
form- och variant- parentesen	exempel i frekvensordning	exempel på varje form anges i kronologisk ordning – rekonstruerat etymologiskt eller källans datering (text- eller handskriftsålder)
betydelsemoment- ordningen	frekvens; (psyko)logiska samband, t.ex. prototypeteoretiska betydelsekluster med centrum och periferi, d.v.s. relationslogisk ordning	kronologisk ordning – källans datering (text- eller handskriftsålder) eller rekonstruerat enligt någon teori om regelbundenheter i semantisk förändring, d.v.s. utvecklingslogisk ordning
språkprovs- ordningen	tydlighet, d.v.s. språkprovets förklaringskraft	kronologiskt inom varje betydelsemoment utifrån källans datering (text- eller handskriftsålder)

Figur 2. *Synkroni respektive diakroni i språkstadieordböcker*

I översikten i figur 2 framstår kronologi utifrån datering av former, betydelser och språkprov som den självklara diakroniska beskrivningsmetoden och det kan vara på sin plats att poängtera att sådana uppgifter i form av ren tidsfästning inte i sig utgör en diakronisk beskrivning. För detta krävs en tolkning av de daterade belägen och en framställning som fokuserar förändringsprocessen. De synkroniska beskrivningsmedlen är måhända mer svärgenomskådade än de diakroniska och frågan är hur ordboksforskaren, för att inte tala om ordboksanvändaren, avgör huruvida t.ex. former, varianter och betydelser återges i frekvensordning eller i relationslogisk eller utvecklingslogisk ordning. Det föreligger en uppenbar risk att ordboksanvändaren inte förmår tillgodogöra sig den information lexikografen med möda framställt eller

tvärtom utläser uppgifter som lexikografen aldrig avsett. Därtill kommer det faktum att lexikografen kan tillämpa olika principer beroende på ordens olika beskaffenhet och källsituation (se t.ex. Healey 2004:141).

7. Språkbeskrivningen i *Sd̄w*

För att avgöra vilken språkbeskrivningsmetod som tillämpats – synkronisk eller diakronisk – måste man således förstå lexikografens principer för t.ex. lemmats normalisering och syftet med återgivningen av former, varianter och språkprov samt principerna för betydelseindelning, -beskrivning och -ordning. Min okontroversiella hypotes är att en språkbeskrivning som syftar till att återge språket som ett system (d.v.s. synkroniskt) ser annorlunda ut än en beskrivning som syftar till att avspegla dess förändring (d.v.s. diakroniskt). Söderwall har inte meddelat sig i tryck om dylika spörsmål och ordboken är det enda till buds stående materialet för en sådan undersökning.

Här återges ett exempel på en förkortad version av en ordboksartikel i *Sd̄w*. Uteslutna är ett antal språkprov och nakna källangivelser vars citerade källor dock återges inom parentes. Flera angivelser till samma källa i samma betydelsemoment återges dock bara då dateringarna skiljer sig åt. Tillagda är dateringarna som är ungefärliga och återgivna enligt källförteckningen i *Sd̄w*. Den åtskillnad som Söderwall ibland gör mellan textens och handskriftens ålder återges så att textåldern ges inom hakparentes, även när handskriften är i original och text- och handskriftsålder sammanfaller. Tillagt är även numreringen av undermomentet.

plagha (plaagha: -ar *Lg* 809 1502; -as *VKR* 29 [1451-52]. **plägħa: -adhe** *Bil* 565 1420-50 [cirka 1270]; -ade *ib* 659 1420-50 [cirka 1270]. -ar, -adhe, -adher), *v.* [*Isl.* plága. *Mnt.* plagen. *Mlat.* plagare] 1) *eg. slå; gissla, tukta, späka.* huru vndirsamlica hon sin licamma plagadhe *KL* 334 cirka 1400 (Övriga källor: *Su* 350 [slutet av 1400-talet], *ib* 452 1450-1510 [cirka 1450], *Bo* 106 cirka 1420 [slutet av 1300-talet]) 2) *straffa.* formogha ey at rätta owir them ällir plagha (*corrigere ... aut punire*) *Bir* 3:343 1400-25 [slutet 1300-talet] (Övriga källor: *KS* (fem angivelser) 1634 [cirka 1330], *VKR* 29 [1451-52], *Ber* 62 [1480-1500], *Bir* (två angivelser) 1400-25 [slutet av 1300-talet]) 3) *straffa, hemsöka (om Gud).* plaghar waar herra . . . mannen *MB* 1:131 1430-50 [cirka 1330] (Övriga källor: *Bir* 3:451 1400-25 [slutet av 1300-talet], *Bil* 659 1420-50 [cirka 1270], *KL* 241 1385 [sannolikt något äldre], *Gr* 315 cirka 1420 [slutet av 1300-talet], *MB* 1:429 1430-50 [cirka 1330], *Bil* 565 1420-50 [cirka 1270]; *KS* 74 (181, 81) 1634 [cirka 1330]) 4) *pläga, pina.* *VKR* v. cirka 1450 (?) [1373-85]. pinlica plaghas *Bo* 194 cirka 1420 [slutet 1300-talet]. (Övriga källor: *Pa* 19 1420-50 [cirka 1270]; *Bir* 1:174 cirka 1450; *Bir* 3:312, 135 1400-25 [slutet 1300-talet]) 4:1) *besvärä.* at the ey warda plägħade medh fals wara ok flärd *KS* 74 (182, 81) 1634 [cirka 1330]

Av den förkortade ordboksartikeln kan utläsas att:

1. *lemmat* konstrueras till en tidig form (vilket också är den mest frekventa formen i språkproven). Mer sentida drag förekommer dock i språkproven: *-aa-*, *-å-*, *-g-*. (Homografsepareringen är inte aktuell i detta fall annat än för att skilja verbet från substantivet, men generellt i ordboken tycks en etymologisk princip ha varit styrande när formella kriterier inte räcker);
2. *formparentesen* återger (sannolikt) faktiska beläggformer, som oftast är daterbara men inte presenterade i kronologisk ordning;
3. *betydelsemomenten* är inte ordnade kronologiskt, åtminstone inte utifrån språkprovets datering. Men hur avgörs om frekvens, semantisk förändring eller (psyko)logiska samband är den ordnande principen? (Betydelsebeskrivningen är inte heller entydigt synkronisk då det knappast är den fornsvenska polysemi-strukturen som beskrivs.) Moment 1 tycks vara lånordets etymologiskt ursprungliga betydelse (p.g.a. det ovanliga "eg.") vilket antyder att artikeln är ordnad utifrån en utvecklingslogisk princip;
4. *språkproven* (och de nakna källangivelserna) anges inte heller kronologiskt, åtminstone inte utifrån källdateringen. Det är svårt att se att det skulle vara det mest belysande språkprovet som ges först (i exemplet är dessutom övriga språkprov uteslutna).

Det diakroniska inslaget är alltså inte särskilt framträdande i *Sdw*. Frågorna varför *Sdw* karaktäriseras som diakronisk och vad som gör språkbeskrivningen i en språkstadieordbok synkronisk respektive diakronisk kvarstår. Kräver en synkronisk framställning diasystematisk markering av systemets variation? Är det den etymologiserande (re)konstruktionen av lemmats form eller kanske upplysningarna i den s.k. etymologiparentesen som anses göra framställningen diakronisk? Det förefaller osannolikt. Redovisningen av variationen är inte konstituerande för en synkronisk beskrivning eftersom det är möjligt att ignorera variationen och bara presentera det därur abstraherade och idealiserade systemet, som självfallet inte ska uppfattas som en realitet utan just som en beskrivning. Däremot är redovisningen av förändringen konstituerande för en diakronisk beskrivning. Det är svårt att se att t.ex. ett (implicit) etymologiskt (re)konstruerat lemma tillsammans med upplysningar om motsvarigheter i besläktade språk skulle tillhandahålla tillräcklig information om förändringen i källspråket för att framställningen skulle betraktas som diakronisk.

Naturligtvis tillåter inte en enskild ordboksartikel några slutsatser om den eventuella synkroniska eller diakroniska beskrivningsmetoden i en hel ordbok. Inte heller i en mycket begränsad undersökning av åtta verb, av vilka *plagha* är ett, med fyra numrerade betydelsemoment och med en konkret betydelse, har jag kunnat utläsa någon återkommande framställningsprincip, t.ex. av det tidstypiska slaget: konkret före abstrakt, generell före specifik, bokstavlig före metaforisk o.s.v. (jfr Lewis 2004:153 och Nerlich 1992). För att kunna utläsa eventuella mönster i t.ex. bety-

delseordningen i *Sdw* skulle ett mycket större material behöva granskas. I en sådan undersökning vore det också önskvärt att låta ingå semantiska ordgrupper som t.ex. verb för verbal kommunikation eller egen förflyttning (se t.ex. Viberg 1992).

Ovan hävdades att ordbokstypologisering bör ske utifrån en ingående granskning av själva ordboken, men också ordboksexterna argument måste tillåtas. K.F. Söderwall har, av förståeliga skäl, ingen stor produktion utöver ordboken men den tidiga skriften *Hufvudepokerna af svenska språkets utbildning* (1870) äger särskilt intresse i en diskussion om lexikografens eventuella syn på synkroni–diakroni-distinktionen. I denna skrift indelar Söderwall den svenska språkhistorien i fyra epoker inspirerade av dansken N.M. Petersens indelning. Söderwall skriver: ”Genom att angifva hufvudepokerna af denna utbildning skola vi söka lemna en öfversigt af förändringarnas beskaffenhet och följd” (Söderwall 1870:[1]) vilket förefaller rätt diakroniskt inriktat och det är också den diakroniska beskrivningen av förändringarna under varje epok som präglar framställningen, samtidigt som olikheterna mellan språket i de olika epokerna också behandlas. Även lexikografens övriga språkvetenskapliga publikationer kan således belysa hans benägenhet att beskriva språket synkroniskt eller diakroniskt även om han inte har publicerat något metalexikografiskt arbete.

8. Konklusion

Den enda slutsats jag kan dra är att den här förespråkade distinktionen mellan en synkronisk och en diakronisk beskrivningsmetod i språkstadieordböcker svårligen låter sig benämnas med detta termpar, som blivit alltför belastat och därtill redan är upptaget för att göra en annan distinktion i den nordiska lexikografiska terminologin. De av Lewis (2004) lanserade termerna relationslogisk och utvecklingslogisk å andra sidan utgör inget fullgott alternativ eftersom de bara täcker en delaspekt av hela det beskrivningskomplex som det vore önskvärt att distinktionen omfattade.

Definitionen av de lexikografitypologiska begreppen synkroni respektive diakroni förtjänar en fördjupad behandling eftersom de för det första belyser förutsättningarna för att bedriva lexikografihistorisk forskning, d.v.s. huruvida metalexikografer kan använda dagens begreppsapparat för att typologisera äldre ordböcker. För det andra belyser ovanstående resonemang förhoppningsvis behovet av koppling mellan teori och praktik. Det kan ibland vara motiverat att fråga om typologiseringarna är gjorda utifrån metalexikografernas teoretiska hemvist eller om de är gjorda utifrån ordböckernas faktiska form och innehåll. Jag är inte i första hand intresserad av terminologin som sådan eller av att kritisera vedertagna definitioner. Jag efterlyser snarare klarhet i begreppen synkroni och diakroni i den historiska lexikografen då och nu eftersom det är viktigt att förstå enligt vilka principer en ordbok beskriver språket för att man som ordboksanvändare ska kunna veta vilken information som faktiskt kan utläsas ur ordboken.

Frågan varför Söderwalls medeltidsordbok karaktäriseras som diakronisk har varit utgångspunkten för denna artikel sprungen ur den lexikografihistoriska undersökning av *SdW* som utgör mitt avhandlingsarbete. Utifrån det perspektivet finns all anledning att föra samman definitionerna av synkroni respektive diakroni och en vetenskapshistorisk diskussion med en titt på ordboken. I ett vidare perspektiv kan däremot synkroni–diakroni-distinktionens relevans ifrågasättas. De båda beskrivningsmetoderna har så olika förutsättningar att det inte är självklart att se dem som varandras motsatser eller ens som grader på samma skala. Synkroni är en systembeskrivning utan tidsdimension medan en diakronisk beskrivning är en tolkning som framställer variationen i materialet som en förändringsprocess över tid. Som en konsekvens av ovan förda resonemang om svårigheterna med att utläsa huruvida en sådan tolkning har gjorts och i så fall enligt vilka principer uppstår en grundläggande fråga: Är det lexikografins uppgift att bidra till den diakroniska språkbeskrivningen? Eller annorlunda uttryckt: Vad är egentligen syftet med historiska ordböcker?

Litteratur

Ordböcker:

NLO = *Nordisk leksikografisk ordbok*. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 4). 1997. Oslo: Universitetsforlaget.
Söderwall, K.F. 1884-1918: *Ordbok öfver svenska medeltids-språket*. (*Supplement 1925-1973*). Stockholm: Svenska fornskriftsällskapet.

Annan litteratur:

Bandle, Oskar 2002: Diachrony and synchrony in Nordic language history. I: Oskar Bandle m.fl. (red.): *The Nordic Languages. An International Handbook of the History of the North Germanic Languages 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft 22). Berlin/New York: Walter de Gruyter, 25-30.
Geeraerts, Dirk 1997: *Diachronic Prototype Semantics. A Contribution to Historical Lexicology*. (Oxford Studies in Lexicography and Lexicology) Oxford: Clarendon.
Healey, Antonette diPaolo 2004: Polysemy and the Dictionary of Old English. I: J. Coleman & A. McDermott (red.): *Historical Dictionaries and Historical Dictionary Research. Papers from the International Conference on Historical Lexicography and Lexicology, at the University of Leicester, 2002*. (Lexicographica. Series Maior 123). Tübingen: Max Niemeyer Verlag, 137-147.
Jonsson, Hans 1992: De stora historiska ordböckernas särart. I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.-31. mai 1991*. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1). Oslo: Nordisk forening for leksikografi, 231-234.
Lewis, Robert E. 2004: Aspects of Polysemy in the Middle English Dictionary. I: J. Coleman & A. McDermott (red.): *Historical Dictionaries and Historical Dictionary Research. Papers*

- from the International Conference on Historical Lexicography and Lexicology, at the University of Leicester, 2002. (Lexicographica. Series Maior 123). Tübingen: Max Niemeyer Verlag, 149-156.
- Nerlich, Brigitte 1992: *Semantic Theories in Europe 1830-1930. From Etymology to Contextuality*. (Amsterdam studies in the theory and history of linguistic science. Series III, Studies in the history of the language sciences 59). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ralph, Bo 2000: Vanans makt i svensk språkhistoriebeskrivning. I: Lars-Erik Edlund (utg.): *Studier i svensk språkhistoria 5. Förhandlingar vid Femte sammankomsten för svenska språkets historia Umeå 20-22 november 1997*. Umeå: Institutionen för litteraturvetenskap och nordiska språk vid Umeå universitet, 360-379.
- Reichmann, Oskar 1984: Historische Lexikographie. I: Werner Besch, Oskar Reichmann, Stefan Sonderegger (utg.): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft 2). Berlin/New York: Walter de Gruyter, 460-492.
- Reichmann, Oskar 1990a: Das Sprachstadienwörterbuch I: Deutsch. I: Franz Josef Hausmann m.fl. (utg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie 2*. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.) Berlin/New York: Walter de Gruyter, 1416-1429.
- Reichmann, Oskar 1990b: Formen und Probleme der Datenerhebung I: Synchronische und diachronische historische Wörterbücher. I: Franz Josef Hausmann m.fl. (utg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie 2*. (Handbücher zur Sprach- und Kommunikationswissenschaft 5). Berlin/New York: Walter de Gruyter, 1588-1611.
- de Saussure, Ferdinand [1916] 1962: *Cours de linguistique générale*. Paris: Payot.
- van Sterkenburg, Piet 1990: Das Sprachstadienwörterbuch II: Niederländisch. I: Franz Josef Hausmann m.fl. (utg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie 2*. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.) Berlin/New York: Walter de Gruyter, 1430-1435.
- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. (Andra, omarbetade och utökade upplagan). Stockholm: Norstedts Akademiska Förlag.
- Söderwall, K.F. 1870: *Hufvudepokerna af svenska språkets utbildning*. Lund: C.W.K. Gleerups förlag.
- Viberg, Åke 1992: Universellt och språkspecifikt i det svenska ordförrådets organisation. I: *Tijdschrift voor Skandinavistiek* 13:2, 17-58.
- Wiegand, Herbert Ernst 1984: Prinzipien und Methoden historischer Lexikographie. I: Werner Besch, Oskar Reichmann, Stefan Sonderegger (utg.): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft 2). Berlin/New York: Walter de Gruyter, 557-620.

Nyord og allmenne ordbøker

Neologisms and general dictionaries. The article presents a survey that was carried out among editors at Kunnskapsforlaget concerning neologisms in the Norwegian language. The editors were asked to mark which of the neologisms they would insert in general monolingual dictionaries. At the NFL Conference in Sønderborg 2005, the same test was done with the audience. The results are compared with the editors' survey and some comments are made on how academic lexicographers and commercial lexicographers and encyclopaedia editors evaluate which neologisms are most useful.

1. Innledning

Kunnskapsforlaget har en redaksjon på noe over 20 faste redaksjonelle medarbeidere, som sammen med det eksterne nettverket av fagmedarbeidere og forfattere forvalter en stor portefølje av oppslagsbøker. Noen spesielle prosjekter har ekstern finansiering, men de fleste må være økonomisk selvberende. Av redaksjonsmedarbeiderne er det foruten lederne tre som arbeider hovedsakelig med språkordbøker, alle de andre driver først og fremst med fagleksikografi. For enkelhets skyld kaller vi dem nedenfor undertiden *encyklopedistene*, i motsetning til *leksikografene* som deltok på NFL-konferansen i Sønderborg.

Alle utgivelsene skal i prinsippet holdes à jour, og for både forlagsledelsen, redaksjonen og de eksterne medarbeiderne er det et permanent spørsmål hva ajourhold innebærer rent innholdsmessig, og hvordan vi skal arbeide med ajourholdet.

Den tradisjonelle utgivelsesformen er *trykte bøker*, men siden ca. 1990 har en god del titler også utkommet *elektronisk*. Siste generasjon elektroniske ordbokutgivelser hittil er nettstedet www.ordnett.no, en abonnementstjeneste med enkelte gratis-tjenester. Pr. august 2005 omfatter *Ordnett* 19 ordbøker. I tillegg utgir vi *Aschehoug og Gyldendals store norske leksikon* elektronisk, også dette som en abonnementstjeneste på nettet. Den neste store nettsatsingen er *Norsk biografisk leksikon*.

Etter hvert som mer av omsetningen kommer fra elektroniske produkter, der omfangsproblemet er mindre påtrengende, vil spørsmål om *selektering av oppslagsord* kanskje bli mindre relevant. Men så lenge alle utgivelsene skjer primært i bokform, er utvalg og utvalgs-kriterier ikke til å komme utenom.

Det er derfor fortsatt relevant å diskutere for eksempel *hvilke nye oppslagsord* som bør være med i en ny utgave av et oppslagsverk – enten det er et leksikon eller en ordbok.

I forbindelse med leksikografikonferansen i Sønderborg mai 2005 har vi foretatt en liten undersøkelse omkring innlemmelse av nyord i norske rettskrivningsordbøker. Resultatet ble lagt frem på konferansen og meddeles her i noe utvidet form.

En leksikografisk forsker ville sannsynligvis begynne med å problematisere nyordbegrepet. Vi tar utgangspunkt i et materiale hentet fra norske nyordbøker, og støtter oss derfor til de kriterier for utvalget som der er benyttet. Helt generelt kan disse kriteriene oppsummeres slik: Dels dreier det seg om (1) *oppslagsord*, dels om (2) *betydninger*, som ikke er belagt i eldre ordbøker, men som er belagt én eller flere ganger i skriftlig materiale fra den nære fortid. Vi har konsentrert oss om den første kategorien.

For et grundigere studium av nyordbegrepet henviser vi til Pia Jarvads *Nye ord – hvorfor og hvordan?* (Gyldendal 1995).

2. Vagleik Leiras nyordbok og standardordbøkene

Som ledd i et nordisk nyordprosjekt utkom *Nyord i norsk 1945-75* i 1982, med Vagleik Leira i Norsk språkråd som redaktør.

Vi bestemte oss for å sjekke litt rundt Leira-nyordenes *Nachleben* i to norske ordbøker: *Tanums store rettskrivningsordbok* og *Bokmålsordboka*. Vi hadde ikke tid til å jobbe med et stort materiale, men tok rett og slett de 50 første ordene på B (inkl. noen sammensetninger der sisteleddet begynner på B) som hadde førstebelegg fra og med 1967.

Vi sjekket om disse ordene fantes i den første utgaven av *Tanum* som med stor sikkerhet kan ha utnyttet hans materiale, nemlig 7. utgave fra 1989.

Det kom flere nye *Tanum*-utgaver mellom 1967 og 1989, så Leiras bok trenger ikke å ha vært kilde for nyordene, men det spiller for så vidt ingen rolle.

Av de 50 første ordene på B er 25 med i *Tanum 1989*, altså en opptaksprosent på 50. Dette er selvsagt et ørlite materiale, men viser en tendens.

I *Bokmålsordboka*, 1. og 2. utgave, finnes de samme 15 av de 50 ordene, og av disse tre som ikke finnes i *Tanum 1989*, nemlig *gjenbrukbar*, *espresso* og *bibelsmugler*. At det er de samme 15 ordene som er tatt med i begge utgavene, er kanskje ikke så rart; nyord i *Bokmålsordboka 1993* måtte hentes fra andre kilder enn den mer enn 10 år gamle *Nyord i norsk*.

<i>Oppslagsord</i>	<i>Forklaring</i>	<i>Med i Tanum 1989</i>	<i>Bokmålsordboka</i>
babysitter		x	
babystol			
backlash		x	x
badeheis		x	
baderomsmøbel		x	
bakkekontakt		x	x
baklengsmål		x	x
bakvendtstol			
ballongtest		x	
bananrepublikk		x	x

NYORD OG ALLMENNE ORDBØKER

idébank		x	
informasjonsbank			
kunnskapsbank		x	
bankomat			
bankøkonom		x	
gjenbrukbar			x
gjenvinnbar		x	x
innflyttbar			
nedsenkbar		x	x
påvisbar		x	
buksebar		x	
espressoer			x
hårbar	frisør		
platebar		x	x
slipsbar			
barbiturat		x	
barfotlege		x	x
barkemaskin		x	x
barmann	mann som betjener en bar		
barna-inne-system			
barnefyll			
barnehageonkel			
barselpermisjon		x	
battlejakke			
beboerhage			
bedriftsspionasje			
be-in	hippiesammenkomst		
beinmelstablett eller beinmjølstablett			
bensingjerrig			
bestillingstyv			
betjeningsknapp		x	
betjeningspanel			
bevisstgjøre		x	x
bevisstgjøring		x	x
bevissthetsutvidende		x	x
bibelsmugler			x
bibliotekbuss			
bifil		x	x
bikiniføre			
bilavgass			

Tabell 1.

Vi sjekket disse ordene på nettsider på norsk med søkemotoren *Google*. Dette ble gjort medio mai 2005. Søket på de 50 ordene fra *Leira 1982* viser at av de 26 ordene som har mer enn 100 treff, er det 7 som ikke er med i *Tanum 1989*. Den med høyest score som ikke er med, er *idébank*, som for noen tiår siden ble forbundet med Erik Byes TV-programserie, men som i senere tid har fått et mer allment innhold. Ordet gir hele 15 600 treff! Av de 24 ordene som har fått under 100 treff, er det 8 som er med i *Tanum*. *Buksebar* har 0 treff på nettet og *ballongtest* 1. Disse ordene bør kanskje kunne tas ut av *Tanum* på dette grunnlag. Interessant er *bibelsmugler*, som er med i *Bokmålsordboka* (ikke *Tanum*), og som var en god del brukt i den kalde krigens tid, men som nå bare gir 13 treff på nettet. Dette ordet er tydeligvis nesten gått av bruk.

Denne metoden er nok brukbar til å vurdere eldre nyord for nye utgaver av ordbøker: å identifisere nyord ved hjelp av nyordbøker, og så sjekke deres fortsatte bruk på nettet.

3. En liten redaksjonsenquete basert på Tor Guttus nyordbok-manus slik det var primo 2005

Neste bind i serien av nyordbøker, som skal dekke tiårene etter Leiras bok, foreligger i foreløpig manus. Denne gangen er det Tor Guttu som har ansvar for arbeidet, på oppdrag av Norsk språkråd, som i 2005 har skiftet navn til Språkrådet. Vi har fått låne hans manus, og har her plukket ut 100 ord på B, L og begynnelsen av M, inkludert noen sammensatte ord der sisteleddet begynner på en av disse bokstavene, med førstebelegg fra 1990 og senere.

<i>Oppslagsord</i>	<i>Kort forklaring/sitat</i>
babyboomer	person født under babyboom
backpacker	ryggsekk turist
bakbenkmann	backbencher
bakgårdsbedrift	uregistrert, mer eller mindre snusket fåmannsbedrift
bakgårdsland	Norge er "bakgårdslandet" i europeisk fotball
bakoversveis	
bakslask	backslash
ballerobring	breakdown (Drillo)
bambus-PC	PC fremstilt i lavkostland i Asia
bambuspære	?
bandog	farlig hund som må holdes bundet
barfoter	en som driver barfotkjøring (vannsport)

NYORD OG ALLMENNE ORDBØKER

barnekriminell	
barneran	
barneraner	
barrering	å slå hest med en stang eller stokk ved hinderpassering slik at den skal hoppe høyere neste gang
bekymringstelefon	krisetelefon
belteveske	
beng	gi beng; gi blaffen
benmargsregister	til bruk i behandling av benmargskreft
beretta	hodeplagg
bidus	samisk kjøttsuppe på reinsdyrkraft
bimedial	som angår eller omfatter to slags medier
biokrati	nazistenes forfølgelse av biologisk mindreverdige
biomagnetisk	i biomagnetisk forskning
biomagnetisme	magnetfelt i forbindelse med elektriske strømmer i kroppen
biopat	utøver av biopati, en naturmedisinsk terapi
bjose	alternativt navn på det kvinnelige kjønnsorgan
rollerblade(s)	
rollerblader	en som bruker rollerblades
blankskalle	snauklipt høyreekstremist
bleiegutt	sumobryter
blindernsk	som angår Universitetet på Blindern
blodfobi	skrekk for å se blod
blodtype	(verb)
blodtåke	nedsatt dømmekraft i pressen i forbindelse med volds- og drapssaker
blåruss	næringslivsmenneske
bobbittere	skjære penis av ektemannen
bodypacking	narkotikasmugling i kroppens hulrom
boke	booke, innføre i en fortegnelse, og dermed bestemme, fastlegge noe
bola	yppig, kraftig, oppblåst ved hjelp av dopingmidler
bomåpner	
bonsai	fornorsket skrivemåte av bonzai, japansk dvergbusk
borreliose	kronisk ledd- og muskelsykdom

brenne	ta opp på CD eller DVD
brevgiro	
browse	fare raskt igjennom
browser	søkeprogram i EDB
Brustad-bu	dagligvarebutikk som kan holde søndagsåpent
bry seg	gripe inn for å hindre overgrep og lignende
bulkvin	vin tatt hjem på fat og tappet på flasker i Norge
bulldosere	jevne med jorden, fjerne ved hjelp av bulldoser
bunntung	det motsatte av topptung
burka	muslimsk kvinnedrakt
bygdetango	
bysk	det motsatte av bondsk
bøyelaster	bøyelasteskip
laksekotelett	
landsmoder	
langventer	en som står lenge i behandlingsskø på sykehuset
lavprofil	lav profil mht. å markere seg
lederi	ordning eller praksis med stort antall (overflødige) ledere
lesotek	sted hvor man kan lese/bytte/leie bøker m.m.
lettsyltetøy	
likemannsarbeid	omsorgsarbeid utført av folk i samme situasjon som klienten
lilla	politisk: om allianse mellom blått og rødt
liposculptur	type fettsuging
lobbe	drive lobbyvirksomhet
lo/lo-trafikk	lift on/lift-off-trafikk (dvs. vanlig lasting og lossing)
lottomillionær	
lysark	transparent ark for overheadprojektor
lystvold	vold som begås av ren lyst
lytt	det å lytte en enkelt gang
lytterett	rett til avlytting
løsneområde	område hvor skred vanligvis løsner
løypelus	ikke-deltakere som beveger seg i løypene under et skirenn
macsk	som angår Macintosh datamaskiner

magekort	kort i livet (så det blir mellomrom til bukselinningen)
maglevtog	magnetsvevetog
magnetbandasje	bandasje med innlagte magneter hvis felt virker noen cm inn i kroppen
mail	e-post
mailbombing	nedsyting med uvedkommende e-post
mainstream	som appellerer til gjennomsnittet
maktspiller	person som driver maktspill
malfemisme	det motsatte av evfemisme
mandagsbarn	(barnehage)barn som er utslitt etter helgen
mannsglad	glad i menn
mannskvotering	
mannsvoldtekt	voldtekt på mann
mappeinnsyn	innsyn i egen sak (i hemmelig register)
matmølle	kjøkkenmaskin/food processor
matsminke	
medbygding	sambygding
mederkjenne	erkjenne noe som også andre erkjenner
mediesky	uvillig til å fremstå i massemedier
medietekke	
medievold	det å bli uthengt i pressen
medråderett	i forbindelse med norsk EU-medlemskap
megakul	svært kul
megamorder	person eller regime som står bak massemyrderi

Tabell 2.

Vi spurte våre kolleger (ved hjelp av et elektronisk spørreskjema) om de kjente disse ordene, om de brukte dem selv, om de mente de hører hjemme i neste utgave av *Tanums store rettskrivningsordbok*, i *Bokmålsordboka* og i en (uspesifisert) skoleordliste. Informantene kjenner *Tanum* ganske godt, og *Bokmålsordboka* litt mindre godt.

Våre informanter var vår egen leksikon- og ordbokredaksjon, inklusive oss selv. Dette er medarbeidere som hele tiden vurderer og skaper innhold i oppslagsverk, og som alle er vant med å fungere normerende: Det skal være et kvalitetsstempel på innhold som får plass i en oppslagsbok, enten det er en rettskrivningsordbok, en ordliste eller et leksikon. Vi legger altså ikke til grunn at standardutgivelsen er

elektronisk, der det bare kan pøses på med innhold fordi det ikke er noen omfangsbegrensninger. Da blir seleksjon mindre relevant.

De 100 ordene som vi har testet, har altså førstebelegg enten som ord eller i den aktuelle betydningen fra og med 1990, i det materialet som har vært benyttet til nyordboken. De fleste belegg er fra dagspressen.

Tanum er den etablerte oppslagsboken for ords-katten i moderne bokmål, med ca. 300 000 oppslagsord, herav selvsagt de aller fleste sammensetninger. Det er ikke en komplett definisjonsordbok. *Bokmålsordboka* er en standard definisjons- og rettskrivningsordbok med ca. 65 000 oppslagsord. Norske skoleordlister kan ha anslagsvis 40 000-50 000 oppslagsord.

Vi har ikke sjekket om noen av de 100 ordene faktisk er med i de nyeste utgavene av Tanum og Bokmålsordboka; det er ikke så interessant i denne sammenheng.

4. Resultater av redaksjonenqueten

Det viktigste resultatet av mini-undersøkelsen er at informantene ønsker mange av nyordene inn i alle ordbøkene. Mer enn halvparten av de spurte vil ha 28 av de 100 ordene inn i *Tanum*. Hvis dette resultatet skulle gjelde hele alfabetet, ville det bety at av et manus på 10 000 oppslagsord ville nærmere 3000 ønskes inn i *Tanum*.

De 28 ordene er vist i tabell 3.

backpacker
bakoversveis
barneraner
belteveske
beng
biomagnetisk
biomagnetisme
biopat
blankskalle
blåruss
bonsai
borreliose
brenne
brevgiro
browse
browser

bry seg
burka
laksekotelett
lettsyltetøy
liposculptur
lottomillionær
mail
mainstream
matsminke
mediesky
medietekke
rollerblade(s)

Tabell 3.

Hvis også vi tar med de ordene som nøyaktig halvparten av redaksjonsmedlemmene ville ha med, blir antallet 34. Tilleggsordene fremgår av tabell 4.

barnekriminell
lysark
medråderett
beretta
biokrati
lavprofil

Tabell 4.

I redaksjonen som helhet er det 8 ord som ikke har fått noen stemmer. Disse ordene fremgår av tabell 5.

bambus-PC
bambuspære
bleiegutt
bobbittere
langventer
lederi
macsk
magekort

Tabell 5.

De ordene som fikk flest stemmer (16 og flere), kan ses i tabell 6.

burka
brenne
mail
browser
mediesky
bonsai
medietekke
matsminke
brevgiro

Tabell 6.

Det er vel ingen stor overraskelse at det ord som har fått aller flest stemmer, er høyaktuelt i forbindelse med forholdet mellom islamsk og europeisk kultur, og har vært trukket mye frem også i forbindelse med den nye utgaven av *Store norske leksikon*, mens tre av de øvrige ordene stammer fra dataspråket.

Hvis vi ser på stemmer avgitt for oppføring i *Bokmålsordboka*, blir 25 ord foreslått av mer enn halvparten av deltakerne, mens de samme 8 ordene som for *Tanum* heller ikke får noen positive stemmer her.

Det er altså større forskjell mellom opptakspraksis i *Tanum* og *Bokmålsordboka* slik de foreligger, og forskjellen i vår redaksjons forslag: Det er større forskjell i ordforrådet i de to ordbøkene enn våre redaktører har oppfattet. Deltakerne i den lille undersøkelsen er seg enten ikke helt bevisst at utvalget av oppslagsord i *Bokmålsordboka* er langt snevrere enn i *Tanum*, eller så mener de at også *Bokmålsordboka* skal ta opp i seg mye nytt materiale, og relativ sett mer nytt enn *Tanum*.

Informantene har forstått at utvalget av oppslagsord i skoleordlisten skal være snevrere enn i de to andre, og det er bare 10 ord av de 100 som får mer enn halvparten positive stemmer i undersøkelsen. Flertallet av våre redaktører mener altså at 90 % av ordene i nyordbok-manuset ikke er relevante for en skoleordliste, iallfall ikke foreløpig. Størst forskjell i stemmegivningen mellom *Tanum* og skoleordlisten er det for de ordene som er oppført i tabell 7.

borreliose
liposculptur
bonsai
brevgiro
biomagnetisme

biomagnetisk
biokrati
laksekotelett
beretta
bidus
bøyelaster
barrering

Tabell 7.

Alle disse har 7 stemmer og mer i forskjell. For øvrig er det et trekk at det er stort sett de samme ordene som ønskes inn av særlig mange i alle tre utgivelsene.

Et *Google-søk* på alle de 100 ordene viser at de som et flertall i redaksjonen ønsker innlemmet i Tanum, så godt som alle er bredt belagt på dokumenter på norsk på nettet. *Biomagnetisme* er et unntak. *Blåruss* i den nye betydningen er naturlig nok ikke søkbart, men må eventuelt undersøkes i det allmenne funnmaterialet. Det har vi ikke tatt oss tid til. *Liposculptur* skrives ifølge Google oftere med k istedenfor med c.

<i>Oppslagsord</i>	<i>Antall treff</i>	<i>Kommentar</i>
burka	3420	
brenne	22200	
mail	796000	
browser	235000	
mediesky	503	
bonsai	1250	
medietekke	339	
matsminke	344	
brevgiro	3190	
bakoversveis	5750	
backpacker	24200	
rollerblade(s)	839	rollerblade 779
borreliose	424	
belteveske	770	
liposculptur	10	liposkulptur 134
biopat	1400	
lettsyltetøy	243	

browse	214000	
mainstream	11500	
blåruss	må analyseres	
biomagnetisme	2	
lottomillionær	5560	

Tabell 8.

Også mange av ordene som et flertall i redaksjonen ikke ønsker inn i Tanum, er godt belagt på nettet. I en særstilling står:

lysark (26600)
mappeinnsyn (3240)

Også de 11 ord i tabell 9 har over 100 treff.

likemannsarbeid
medievold
megakul
barneran
matmølle
blodtåke
bekymringstelefon
landsmoder
bøyelaster
mailbombing
ballerobring

Tabell 9.

Resten har mellom 10 og 100, som det ses av tabell 10.

Brustad-bu (med skrivemåten brustadbu i tillegg)
magnetbandasje
bandog
bidus
bimedial
bulkvin
bodypacking

mannskvotering
babyboomer
rollerblader
bunntung
bomåpner
lystvold
bakgårdsbedrift
lo/lo-trafikk
benmargsregister
bakslask

Tabell 10.

Etter å ha gjort disse søkene, vil vi måtte vurdere om ikke flere enn de 28 (eller 34) ordene som "vant" avstemningen, bør få plass i *Tanum* fordi de tydeligvis er i allmenn bruk. Hvor mange, blir det en redaksjonell oppgave å bestemme.

Ordbokforfattere og -redaktører kan selvsagt ikke vente på en nyordbok for å kunne samle inn nyord. Hvilke teknikker man skal benytte for å supplere ordforråd i ordbøker fra primærkilder, er et stort tema som jeg ikke skal gå inn på her, men metodikk i forbindelse med sekundærkilder som nyordbøker er en mer oversiktlig materie, og det er kanskje ikke vitenskap, men nyttig for oss som utgir ordbøker.

Med nyordbøker og Internett har vi fått nyttige arbeidsredskaper for å studere ordenes bruk over tid. Nyordbøkene kan gi oss et oversiktlig materiale som i en del år fremover kan sjekkes på nettet, og vi kan dermed få hjelp til å vurdere om ord som har vært betraktet som interessante nyord for en tid tilbake, har fått gjennomslag i bruk. Det er vanskelig å sjekke alle ord i en ordbok på denne måten, men via nyordbøkene kan vi begrense materialet til noe av det mest relevante.

Hvis vi regner Internetts gjennombrudd fra ca. 1995, er vel tekstmaterialet på Internett nå så stort at ord som ikke eller nesten ikke finnes belagt på nettet, har vært lite i bruk i skrift de siste 10 år.

Når vi er så heldige å ha en redaksjon som er vant til å foreta vurderinger av kvaliteten på oppslagsord, ville det være dumt ikke å bruke enquete-metoden som et korrektiv til andre metoder. Denne lille testen har i så måte gitt løfterike resultater, og metoden vil nok bli benyttet igjen.

5. En liten enquete blant nordiske leksikografer i Sønderborg

På NFL-konferansen i Sønderborg ble et sammendrag av denne undersøkelsen presentert muntlig. Etter en kort introduksjon delte jeg ut det samme skjemaet som medlemmene av redaksjonen i Kunnskapsforlaget hadde fått, og jeg bad tilhørerne

gjøre et forsøk på å vurdere nyordene på samme måte som våre ansatte hadde gjort. De hadde da ikke fått informasjon om svarene våre ansatte hadde levert. Utfyllingen måtte skje raskt for at programmet skulle kunne gjennomføres, og bare ca. 1/3 av tilhørerne leverte fullstendig utfylte skjemaer tilbake. Ca. 1/3 hadde rukket å vurdere en del av ordene. Dette kunne man se enten av at avkryssingene brått sluttet, eller av at informantene hadde markert på skjemaet hvor langt de var kommet. De som leverte, ble bedt om å skrive sin nasjonalitet på skjemaet.

Så vidt jeg kan erindre, var dette det eneste foredraget under konferansen der tilhørerne ble aktivisert. Jeg oppfattet aktivitetsmomentet som positivt. Det økte sannsynligvis tilhørernes interesse for foredraget som fulgte. Derfor oppfordrer jeg foredragsholdere på senere NFL-konferanser til å prøve noe lignende når det passer ut fra temaet. På den annen side er det selvsagt svært begrenset hva man kan utlede av undersøkelsen, når materialet er så lite og andelen besvarelser som er direkte sammenlignbare med dem våre ansatte leverte, er så lav.

Svarskjemaene fordeling fremgår av tabell 11.

	I alt	Fullstendig besvart	Ufullstendig besvart
Danmark	10	5	5
Norge	10	4	6
Sverige	5	4	1
Tyskland	1	1	
Ikke oppgitt	6	2	4
Til sammen	32	16	16

Tabell 11.

Ut fra sammensetningen av tilstedeværende under foredraget antar jeg at de fleste som leverte uten å oppgi nasjonalitet, var norske. I så fall var de fleste som ikke leverte, ikke norske. Det er også sannsynlig ut fra vanskelighetene vi vil ha for å vurdere ordforrådet i et nabospråk.

Ettersom det dreier seg om en uhøytidelig enquete og ikke en vitenskapelig undersøkelse, isolerer vi kun besvarelsene fra de norske informantene, ettersom de ikke bare kjenner ordforrådet bedre enn de øvrige, men også de ordbøkene som det er referert til. De øvrige informantene er slått sammen, også de som ikke har oppgitt nasjonalitet.

De utenlandske besvarelsene har størst interesse ved det samlede antall stemmer som er avgitt for hvert ord, ikke hvilke ordbøker de gjelder. Flest stemmer er avgitt for ordene i tabell 12.

	Sønderborg utlendinger	Tanum Kunnskapsforlaget
bonsai	19	17
brenne	18	18
mail	17	17
mainstream	16	12
burka	15	19
browser	14	17
rollerblade(s)	13	14
backpacker	10	15
borreliose	10	14
lysark	9	10
lottomillionær	8	11
belteveske	7	14
browse	7	13
bakoversveis	6	15
biopat	6	13
matsminke	6	16
mediesky	6	17
babyboomer	5	4
biomagnetisk	5	11
biomagnetisme	5	11
bry seg	5	11
laksekotelett	5	11
medietekke	5	16

Tabell 12.

Av alle ordene som den internasjonale gruppen har gitt flest stemmer, er det bare ett som har falt langt ned i den norske redaksjonens preferanser, nemlig *babyboomer*. Dette ordet hadde så mange treff på Google at det kan være aktuelt for iallfall *Tanum*. Men det er uansett meget godt samsvar mellom de norske og ikke-norske vurderingene.

Topp-rankingen fra redaksjonsinformantene er anført i tabell 13 (for *Tanum*).

burka
brenne
mail
browser
mediesky
bonsai
medietekke
matsminke
brevgiro

Tabell 13.

Det eneste ordet fra den ”norske” listen som ikke gjenfinnes på Sønderborg-listen, er *brevgiro*. Dette ordet fikk kun én stemme.

6. Norske leksikografer og nyordene

Det ørlille svarmaterialet fra norske leksikografer som hadde kommet igjennom hele skjemaet, er interessant fordi det kommer fra deltakere med en helt spesiell kompetanse og status. Disse leksikografene tilhører en deltakergruppe som hovedsakelig bestod av ansatte i Språkrådet eller ved Universitetet i Oslo. Den største andelen av de norske deltakerne var medlemmer av redaksjonen i *Norsk Ordbok 2014*.

Vi skal se på det ordforrådet som fikk halvparten av stemmene (2 av 4). Dette er 29 ord, som oppføres i tabell 14.

bakoversveis	4
brevgiro	4
lysark	4
belteveske	3
beng	3
biomagnetisme	3
blodtåke	3
boke	3
brenne	3
bry seg	3
laksekotelett	3
lottomillionær	3

mail	3
matsminke	3
medråderett	3
backpacker	2
bakgårdsbedrift	2
barneraner	2
benmargsregister	2
bidus	2
biomagnetisk	2
blåruss	2
landsmoder	2
lettsyltetøy	2
lo/lo-trafikk	2
matmølle	2
mediesky	2
medietekke	2
medievold	2

Tabell 14.

Blant Kunnskapsforlagets redaktører var det 34 av de 100 ordene som fikk halvparten av stemmene, altså marginalt flere enn i leksikografgruppen. Vi synes dette er et ganske godt samsvar i vurderingene, kvantitativt sett.

KF-redaksjonens vurdering og leksikografenes vurdering kan ses av tabell 15.

<i>KF-redaksjonen</i>	<i>Leksikografene</i>	<i>Med i begge grupper</i>
backpacker	backpacker	×
	bakgårdsbedrift	
bakoversveis	bakoversveis	×
barnekriminell		
barneraner	barneraner	×
belteveske	belteveske	×
beng	beng	×
	benmargsregister	
beretta		

	bidus	
biomagnetisk	biomagnetisk	×
biomagnetisme	biomagnetisme	×
biopat		
biokrati		
blankskalle		
	blodtåke	
blåruss	blåruss	×
	boke	
bonsai		
borreliose		
brenne	brenne	×
brevgiro	brevgiro	×
browse		
browser		
bry seg	bry seg	×
burka		
laksekotelett	laksekotelett	×
	landsmoder	
lavprofil		
lettsyltetøy	lettsyltetøy	×
liposculptur		
	lo/lo-trafikk	
lottomillionær	lottomillionær	×
lysark	lysark	×
mail	mail	×
mainstream		
	matmølle	
matsminke	matsminke	×
mediesky	mediesky	×
medietekke	medietekke	×
	medievold	
medråderett	medråderett	×
rollerblade(s)		

Tabell 15.

Av de 43 ordene i tabell 15 er det dermed 20 som er foreslått av begge grupper, og 23 av bare den ene gruppen.

Hvis vi sammenligner leksikografenes topp-ranking-liste med funnene ved Google-søk, ser vi at av de ord som gav over 100 000 treff, er *mail* med, mens *browser* og *browse* mangler på leksikografenes liste. Av ordene som gav mellom 1000 og 100 000 treff, mangler *burka*, *bonsai*, *biopat*, *mainstream* og *mappeinnsyn*. Det var bare én av de 4 som gav sin stemme til *biopat*, *bonsai*, *burka* og *mainstream*. Det er kanskje lov å antyde at utelatelsen av *browser* og *browse* kan henge sammen med en noe restriktiv holdning til å innlemme engelske ord i norsk? På den annen side blir altså *mail* og *backpacker* akseptert. Holdningen kan kanskje karakteriseres som tvetydig?

Et interessant tilfelle er også *bøke*, et forsøk på å fornorske *booke*. Bare 6 av 20 i Kunnskapsforlaget ønsker dette ordet inn i *Tanum*, men 3 av de 4 "sikre" norske leksikografene. (Ordet har heller ingen høy stjerne blant den internasjonale leksikografgruppen i Sønderborg, med i alt 3 stemmer, derav 1 for innlemmelse i *Tanum*.) Kanskje fornorskingsforsøket anses som relativt unødvendig hos andre enn norske leksikografer og språknormerere? Ordet er dessverre ikke så enkelt å underkaste et Google-søk.

14 av 20 informanter i KF-redaksjonen kjenner ordet *landsmoder*, men bare 5 av 20 vil ha det med i ordbøkene. Kanskje har de tenkt at ordet er for nært knyttet til Gro Harlem Brundtlands person og hennes aktive fase i norsk politikk, og at det ikke er av interesse for ettertiden. Leksikografene er derimot relativt sett mer positive, og tenker kanskje mer på semantikken og på analogien til *landsfader*. Ironisk nok ville vi i skrivende stund kunne hevde at dette ordet mer hører hjemme i et leksikon og i en historisk, registrerende ordbok enn i en normerende ordbok og en ordliste.

Bare 4 av 20 informanter i Kunnskapsforlaget oppgir at de kjenner ordet *matmølle*, og 5 vil ha det med i *Tanum*. 2 av 4 norske leksikografer vil ha det inn. Google-søket viser at ordet er i bruk (326 treff), så her er nok en litt tradisjonalistisk holdning i leksikonredaksjonen utslagsgivende. Om dette ordet vil få virkelig gjennomslag i forbrukersamfunnet, og bli tatt i bruk av hvitevareprodusentene i noen sælig utstrekning, er vel tvilsomt, men her synes vi Google-søket støtter opp under leksikografenes vurdering og bør være en tankevekker for encyklopedistene.

7. Konklusjon

Encyklopedister har et perspektiv på virkeligheten som i store trekk faller sammen med leksikografenes, men det er også interessante forskjeller, som denne meget uhøytidelige og snevre undersøkelsen bærer bud om. Den indikerer at et samarbeid mellom de to grupper om ordutvalget i *Tanum* kunne gi et bedre resultat enn om man bare overlater utvalgsarbeidet til den ene gruppen eller til representanter for

denne. Fagkompetansen og vurderingsevnen blant encyklopedister er uten tvil en viktig ressurs for norsk leksikografi. På den annen side er leksikografenes kompetanse og vurderinger også av stor verdi for leksikonarbeidet.

Professor Dag Gundersen har gjennom mange år har ytet en stor personlig innsats for å bygge bro mellom norsk leksikografisk og encyklopedisk arbeid, bl.a. som mangeårig medarbeider i *Store norske leksikon*. Det gir derfor interessante perspektiver at neste utgave av hans og Bjarne Berulfsens store *Fremmedordbok* vil bli revidert gjennom et samarbeid med leksikonredaksjonen. Hvis et tilsvarende samarbeid vil bli gjennomført for fremtidige utgaver av *Tanum* og *Bokmålsordboka*, vil det uten tvil være i tråd med resultatene av denne undersøkelsen.

Litteratur

Ordbøker:

Bokmålsordboka. Definisjons- og rettskrivningsordbok 1993. Utarbeidet av avdeling for leksikografi ved Institutt for nordistikk og litteraturvitenskap, Universitetet i Oslo, i samarbeid med Norsk språkråd. 2. utgave v/ Marit Ingebjørg Landro og Boye Wangenstein. Oslo: Universitetsforlaget.

Norsk språkråd 1982: *Nyord i norsk 1945-1975*. Oslo: Universitetsforlaget 1982.

Tanums store rettskrivningsordbok 1989, 7. utgave v/ Bernt Fossetøl og Marius Sandvei. Oslo: Kunnskapsforlaget.

Annen litteratur:

Pia Jarvad 1995: *Nye ord – hvorfor og hvordan?* København: Gyldendal.

Leksikografisk radikalisme? 10 forslag til den bilingvale ordbog

Lexicographic Radicalism? 10 Proposals for the Bilingual Dictionary. A tendency in bilingual dictionaries towards what could be called automatism rooted in traditionalism is investigated and a number of possible solutions are proposed under the heading “lexicographic radicalism”. This notion by no means indicates a new theoretical challenge to lexicography. Instead, it focuses on typical problems in bilingual dictionaries which – at least within a functional concept – may not be as self-evident as they may seem. The problems discussed could be divided into those of inclusion: *variance*, *paradigm copying*, and *transparency*; those of omission: *fictionyms*, *frequency information*, *error forms*, *pragmatics*, and *illustrations*; and those of treatment: *culture specific lexemes* and *idioms*. The list presented is neither new in every detail nor complete but should be revised as old problems are solved and new ones arise. “Lexicographic radicalism” could be seen as a complex of non-trivial critique towards wide-spread traditions in bilingual lexicography.

1. Indledning

“Radikalisme” har på dansk adskillige betydninger; men kun to af dem er relevante her. Den ene betydning opfattes i visse kredse overvejende positivt, mens den anden er klart negativ. Ordets positive betydning kan parafraseres med ‘fritænkning’ og manglende automatisk respekt for tradition og autoriteter, hvorimod den negative betydningsvariant snarere kan gengives med “ekstremisme”. Min intention med denne artikel er naturligvis ikke at plædere for ekstremisme, men derimod for større leksikografisk frisind, aktuelt i den bilingvale leksikografi, hvor omfanget af traditionalisme synes at være omvendt proportionalt med en mulig funktionalistisk begrundelse for den. Kendetegnende for tosprogsleksikografisk radikalisme er overordnet, at den har et selektivt forhold til de bilingvale ordbøgers gængse informationstyper. Den har det også i forhold til begreber som “norm” og “konsekvens”, der tilsyneladende spiller en stor rolle i den traditionelle tosprogsleksikografi. Men den har det på et fagligt og relevant grundlag.

Det radikalistiske koncept, jeg her skitserer, er nogle forholdsvis eklektiske forslag til, hvordan man kan gøre op med forskellig udbredt traditionstænkning i tosprogsleksikografien. Det er et forslagskatalog, man kan se som et koncept eller vælge enkeltelementer ud fra efter for godt befindende. Konceptets eklekticisme gør også, at de punkter, der gennemgås, på ingen måde skal opfattes som udtømmende. Andre kan komme med nye forslag, såfremt de lever op til det generelle krav for leksikografisk radikalisme, nemlig: ingen respekt for de overleverede principper, *bare* fordi de er overleverede. Det er altså snarere forslagenes ånd, end det er summen af de konkrete forslag, der gør dem radikalistiske.

De punkter, der bliver diskuteret her, passer tilsyneladende udmærket ind i et

overordnet funktionalistisk ordbogskoncept. Der er derfor ikke tale om nogen konkurrerende teori til funktionslæren (jf. fx Bergenholtz/Tarp 2002 og Tarp 2004), for leksikografisk radikalisme er ikke nogen teori, men et til enhver tid skiftende sæt af kritikpunkter af gældende genrekonventioner. Som sådan er den derfor både mere konkret, mere idiosynkratisk, og samtidig mindre grundlæggende end funktionslæren, som både kan ses som inspirationskilde og teoretisk grundlag for leksikografisk radikalisme. Gennemgående for et radikalistisk koncept er et krav om relevans, ja faktisk er det det eneste hensyn. Og det er det stadige fokus på relevans, der styrer medtagelse, udeladelse og behandling af informationstyper.

Jeg vil gennemgå i alt 10 radikalistiske forslag, som i sagens natur kun kan få en kursorisk behandling her. De kan opdeles i problemer vedrørende inklusion: *varians*, *paradigmetvang* og *transparent*; eksklusion: *fiktion*, *frekvensoplysninger*, *fejlformer*, *pragmatik* og *illustrationer*; og vedrørende udførelse: *idiomer* og *kulturbundne leksemer*.

2. Problemer vedrørende inklusion

2.1. Varians

Det første punkt, som er på dagsordenen i det radikalistiske koncept, er fjernelsen af varians i tosprogsordbøgernes L2-produktionsdel, hvor den er hyppig, i hvert fald i danske bilingvale ordbøger. Det er først og fremmest ukommenteret bøjnings- og genusmæssig varians, ortografi- og ækvivalentvarians, som kan ses som et problematisk element i denne del af tosprogsordbøgerne. En gennemsnitlig dansk bruger har fx næppe brug for at få at vide, at *Stau* i genitiv kan hedde både *Staus* og *Stanes*, og i pluralis både *Stane* og *Staus*. Heller ikke at *Blackout* kan være enten maskulinum eller neutrum (begge: Jensen et al. 2003). Og hvad stiller den samme bruger op med ukommenterede variant-ækvivalenter som disse:

bjælde: *Schelle*, *Glöckchen* (Jensen et al. 2003)

hardware: *material informático*, *soporte físico* (Gawinski/Winfeld-Hansen 1998)

Hansaforbundet: *la liga/confederación (h)anséatica* (ibid.)?

Hvornår vælger brugeren den ene, hvornår den anden ækvivalent? Uden betydningsdiskriminering er det tilsyneladende ikke bare unødvendig, men også forvirrende varians. Hvad man derfor strengt taget har brug for, er én korrekt og frekvent bøjningsform, genus eller ækvivalent, ikke en automatisk overførelse af informationer fra de monolingvale ordbøger. Medmindre der selvfølgelig er tale om *specificerede* forskelle.

Man kan selvfølgelig indvende, at man fx ikke kan anvende ordbogen til funktionen korrektur (jf. Tarp 2004b), hvis man fjerner korrekte variante former. Her mener jeg, at det nok er at stille for store krav til den almindelige bilingvale

ordbog, hvis man skal kunne udlede af dens oplysninger, hvad der *ikke* er muligt. Det vil medføre meget lidt fleksibilitet for udtænkningen af ordbogskoncepter, hvis mulige *ex negativo*-beslutninger skal tilgodeses (jf. Farø 2004b). Mht. funktionen korrektur ville jeg derfor nok henvise til en monolingval ordbog, selv om problemet for så vidt også eksisterer her.

Er der i øvrigt tale om eksplicit biskopale ordbøger som fx adskillige af det tyske forlag Langenscheidts, der både henvender sig til brugere med L1 og L2 som modersmål, så stiller det sig naturligvis anderledes med variansen, fordi en ikke-native speaker vil kunne få brug for at identificere et ord ud fra en bestemt bøjningsform, vedkommende har fundet i en tekst. Men den type ordbøger er ved at være sjældne på i hvert fald det danske marked, og man kan i mange tilfælde også kritisere både konceptet og udførelsen for at ville for meget uden at kunne indfri det i praksis.

Ideen bag at udrydde varians i produktionsdelen er så sige at gøre arbejdet færdigt og tilpasse den monolingvale ordbogs oplysninger – som de formentlig stammer fra – til den helt nye kontekst, en bilingval ordbog er. Det kan man bl.a. gøre ved at undgå at overlade valg til brugeren, som han i reglen intet grundlag har for at træffe. Det radikalistiske element består altså her i at fjerne information, brugeren alligevel intet kan stille op med. Ukommenteret varians er nemlig næppe *relevant* information for brugeren.

2.2. Paradigmatvang

I forlængelse af problemet med varians i produktionsdelen kommer et andet problem i det radikalistiske søgelys, nemlig den udbredte “paradigmatvang” i tosprogsordbøgerne. Ofte er leksikografer tilsyneladende kede af at fortie pluralisformer, som er beskrevet i monolingvale ordbøger. Man udfolder derfor hele bøjningsparadigmet, nøjagtig som det er beskrevet i etsprogsordbogen. Men her må man se på, om de pluralisformer, etsprogsordbogen beskriver, er relevante for ordbogens betydninger og grundlæggende *fokus*. Dækker ordbogen fx kun det centrale ordforråd, hvad der gælder langt de fleste ordbøger på markedet, så kan man diskutere, om ikke pluralisformer som disse:

birkes: *Mohn* (-s -e)
 kræft: *Krebs* (-es -e)
 beton: *Beton* (-s -s)
 plastic: *Plaste* (- -n)

(taget fra Jensen et al. 2003 og Bork et al. 1999) ikke bevæger sig langt ud over de fleste danske tysksprogede ordbøgers fokus. Pluralisformerne *Mohne*, *Krebse*, *Betons* og *Plasten* er selvfølgelig nok fagligt mulige, omend også dér temmelig sjældne. Hvis den konkrete ordbog ikke beskæftiger sig med udpræget fagsprog i øvrigt, så er det

i god overensstemmelse med leksikografisk radikalisme at undgå at beskrive pluralisformen, dvs. simpelthen at lade, som om den ikke er der. Det forekommer at være lige så lidt kritisabelt som at undgå at give en ækvivalent til en betydning af et ord beskrevet i en monolingval ordbog, hvis den pågældende betydning ikke er relevant for den aktuelle bilingvale ordbogs brugergruppe.

En bilingval ordbog for danskere skal næppe *beskrive* sproget i sin helhed, men snarere være et så godt værktøj som muligt til en dansksprogets problemer i forbindelse med et fremmedsprog, og det som regel inden for et begrænset fokus.

2.3. Udnyttelse af transparens

I L2-L1-delen af ordbogen kan fremmedsprogets transparente leksemer nok i højere grad undværes, end det er tilfældet nu. Det er et punkt, Kromann/Riiber/Rosbach allerede var inde på i deres grundlæggende artikel fra 1984 (jf. også Kromann 1991). Eksempler som:

Aluminium, Ananas, Babysitter, Freak, Kentaur, Kilo, Lettland, Palästinenser, Statue, Tampon (Jensen et al. 2003)

vil der i en tyskordbog næppe være et reelt ækvivalensbehov for hos en dansk bruger, fordi ordene er receptivt fuldstændigt gennemskuelige. I hvert fald ikke i en kombi-ordbog eller i ordbøger, hvor de to deltekster på anden måde indgår i et funktionelt samspil med hinanden.

Man kan, som det er sket i den leksikografiske litteratur (jf. Bergenholtz 1997), indvende, at man i tilfælde af inklusion jo fx kan få nyttige *bøjnings*oplysninger i L2-L1-delen. Men dertil kan siges, at man i det mindste i kombiordbøgerne kan få præcis de samme oplysninger i L1-L2-delen. Man kunne også kritisere, at argumentet kun er gyldigt for printordbøger², hvad der selvfølgelig har en vis berettigelse. Men udgangspunktet for leksikografisk radikalisme er snarere en mere teoretisk diskussion af principper for prioritering af ressourcer og relevans, end det er en diskussion om *konkret* plads eller ej. Man kan desuden argumentere for, at selv i elektroniske ordbøger med "ubegrænset" plads kan for meget redundans i informationsstrukturen være skadelig for informationssøgningen og dermed for ordbogen som værktøj.

3. Problemer vedrørende eksklusion

Et andet aspekt af den radikalistiske kritik drejer sig om elementer, som typisk *ikke* medtages i tosprogsordbøger, selv om der næsten altid kan findes forbilledlige undtagelser.

² Som Erla Hallsteinsdóttir gjorde efter fremlæggelsen af dette indlæg på konferencen i Sønderborg.

3.1. Fiktionymi

På Volda-konferencen plæderede jeg (Farø 2005b) for at optage *fiktionymer* i tosprogsordbøgerne, dvs. leksemer, der betegner bl.a. fiktive figurer og steder og også titler på film og bøger, som er oversat. Det er eksempler som *Tintin*, der på tysk hedder *Tim (und Struppi)* og på nederlandsk *Kuifje*. Og *Breaking the Waves*, der på italiensk hedder *Sulle onde del destino*. Det er også tv-serien *Dollars*, hvis engelske titel er *Dynasty*, mens den tyske er *Der Denver-Clan*. Og det er *Ondskabens Øjne* med originaltitlen *Silence of the Lambs* og den tyske titel *Schweigen der Lämmer*. Netop filmtitler finder man nærmest overhovedet ikke i tosprogsleksikografien.

Fiktionymer er et væsentligt indslag i den radikalistiske ordbog, som ikke skelner skarpt mellem sproglige og encyklopædiske informationer, men snarere fokuserer på, at den skal være et allround-hjælpemiddel for navigering mellem to sprog. Fiktionymer kan medtages i tosprogsordbøger i det omfang, de er kommunikativt relevante for den påtænkte brugergruppe. Kriterier kan være frekvens, klassikerstatus m.m. – eller overordnet set: *kommunikativ relevans*, hvormed de ikke adskiller sig fra andre typer leksemer. Fiktionymer er som kategori betragtet nemlig ikke mindre kommunikativt relevante end andre leksemer.

Som et positivt eksempel kan nævnes van Hees (2004), der giver oplysningen, at *Tintin* på nederlandsk hedder *Kuifje*. Ækvivalenter til *Tintin* får man tilsyneladende ingen andre steder i dansk tosprogsleksikografi. Helt til dørs følger “ordbogssættet” dog ikke figuren, for en dansker, der støder på *Kuifje* i en nederlandsk tekst, kan ikke slå ham op og finde ud af, hvem han er – ligesom den flamske dansk-lørner ikke kan få at vide, hvad han hedder på dansk (se Laureys et al. 2004). I øvrigt ser eksemplet *Kuifje* ud til at være en ret enlig svale i van Hees et al. (2004): De fleste af de højfrequente fiktionymer, der nævnes i Farø (2005), er heller ikke med her. Der er tale om væsentlige fiktionymer som *Joakim von And*, *Junglebogen* og *Stormfulde Højder*. *Wuthering Heights* finder man derimod i Kjærulff Nielsen (2000).

Skulle man i øvrigt medtage lemmaet *Tintin* i en dansk-tysk ordbog, måtte man tage hensyn til, at navnet på dansk er polysemt: Det betegner både figuren og tegneserien. Dét er måske ikke kontrastivt relevant i forhold til alle sprog, men til fx tysk er det; her kræves der en betydning 1 med ækvivalenten *Tim* for figuren selv og en betydning 2 med ækvivalenten *Tim und Struppi*, som selve tegneserien hedder på tysk.

3.2. Frekvensangivelser

At frekvensundersøgelser kan være et nyttigt redskab i den leksikografiske research-fase, er et velkendt fænomen (jf. Svensén 2004). Men brugsfrekvens kan også anvendes i den leksikografiske præsentation, som bl.a. Zettersten/Lauridsen (1999) og MAD (2002) har vist. På to punkter kan eksplicitte frekvensangivelser være til nytte i tosprogsordbøger: Dels kan de bruges under indlæringen af fremmedsproget i

forbindelse med “ordbogslæsning”. Dvs. frekvensangivelserne kan hjælpe til at økonomisere med kræfterne og tiden for brugeren: Hvis et ord er højfrekvent i sprogbogen, kan det formentlig bedre betale sig at prøve at memorere det, end hvis det er lavfrekvent. Dels kan frekvensangivelserne bruges, hvis der er signifikante forskelle på udbredelsen af et lemma og dets “ækvivalent” eller mellem to ækvivalenter i forhold til hinanden. Det sidste bruges selvfølgelig allerede i nogen grad i tosprogsordbøgerne – når opslagsordet eller ækvivalenten angives som *sj.* – men det kunne nok gøres mere fintfølede med de korpusværktøjer, der nu er til rådighed. Man kunne også udnytte frekvenstal i makrostrukturen, som MAD (2002) er et fint eksempel på.

3.3. Fejlformer *m.m.*

Den leksikografiske radikalist er også positivt indstillet over for ikke-normerede oplagsord og deciderede fejlformer, når de tjener det formål, som i forvejen er et af tosprogsordbogens primære, nemlig at finde et (forslag til) en ækvivalent, jf.:

**acidophilus, *diamentral, *o.lign, *pizzaria*

Mærkværdigvis er ideen er slået igennem flere steder i den monolingvale leksikografi, fx i DDO og i PRO, men ikke i den bilingvale leksikografi, om hvilken man ellers i endnu højere grad end om etsprogsordbøgerne kan påstå, at de kun har redskabskarakter og ikke er normerende. Det kan virke urimeligt, at den dårlige staver så at sige straffes dobbelt ved heller ikke at kunne finde den rette ækvivalent. Derfor er det i en radikalistisk optik snarere et spørgsmål om at finde den rette præsentationsform af fejlformerne, hvis funktion er at være rene henvisningsoperatorer, og ikke så meget et spørgsmål om, om man må eller ikke må medtage fejlformer i en tosprogsordbog. Det afgørende er, om det tjener et højere formål at gøre det. Og så må man naturligvis leve med risikoen for en avisoverskrift af typen “Fejl i ordbog”, som udnytter ordet *fejls* ambiguitet.

3.4. Pragmatik

Pragmatik er et overset formidlingsaspekt i tosprogsleksikografien. Selv om pragmatiske oplysninger synes at stride mod, hvad en ordbog “er” (jf. Tarp 2004), så er der mange gode grunde til at samle pragmatiske oplysninger i tematiske bokse, som fx Gawinski/Winfeld-Hansen (1998) og Jensen et al. (2003) i hvert fald ansatsvist gør det.

Tematisk samlede oplysninger om fx sproglig handlen kan fint kombineres med de alfabetiske artikler via henvisninger og på den måde være med til at aflaste dem. Der kan både være tale om specifikke situationer og mere almen sproglig adfærd. Fx kan man beskrive en typisk dialog i bestemte kommunikative sammenhænge eller

man kan give information om sproglig høflighed og brug af forskellige hilse- og takkeformler, der sjældent har så entydige ækvivalenter, som man kan få indtryk af i selve lemmalisten. For tysk ville det fx være relevant at forklare, hvordan *Sie/du*-differentieringen anvendes, og hvordan personbetegnelser eksplicit markerer bærenens køn. Man kunne også gengive et eksempel på, hvordan et typisk brev eller en e-mail skrives osv.

Den leksikografiske radikalisme skelner ikke så nøje mellem ord og pragmatiske oplysninger, så pragmatik af forskellig art, der også kan være med til at gøre ordbogen til et mere indbydende redskab at slå op og læse i, er i god overensstemmelse med et radikalistisk koncept.

3.5. Illustrationer

Illustrationer kendes især fra fremmedordbøgerne (fx Brüel et al. 1993; Hårbøl et al. 1999), hvor de anvendes, når det er svært at forklare et ord på anden måde end gennem et billede. Men det er fx de kulturbundne leksemer i tosprogssammenhæng jo også (jf. 4.1.), og for en leksikografisk radikalist ville det i en tysk-dansk ordbog for danskere være oplagt at have en illustration af fx den tyske *Bundesadler*, DDR's flag, logoet *der Grüne Punkt* fra tysk genbrugsemballage eller af en *Schultüte*, som tyske skolebørn får på første skoledag. Vi er ikke så vant til at se illustrationer i tosprogsleksikografien, men det er grundlæggende ikke noget argument for ikke at ændre praksis: Måske ville det endda gøre ordbøgerne "lækrere" at arbejde med for brugeren.

4. Problemer med *behandling*

I dette afsnit kaster jeg et blik på problemer, der ikke har med hverken inklusion eller eksklusion af lemmata at gøre, men derimod med behandlingen af de medtagne leksemer. Også her kan man anlægge et radikalistisk perspektiv.

4.1. Kulturbundne leksemer

Fiktionsymerne, som blev diskuteret i 3.1, knytter tilsyneladende an til problemet *kulturbundne leksemer* (Carstensen 1989, Mogensen 2004, Adersen 2005), men der er tale om en helt anden problemstilling, som jeg i Farø (2005b) har vist: Hvor fiktionens denotat i stor udstrækning er fælles i de to sprogsamfund, så er de kulturbundne leksemer netop kendetegnet ved *ikke* eller kun delvist at have et fælles denotat.

De kulturbundne leksemer ligger den leksikografiske radikalist meget på sinde. Det skyldes det paradoks, at de på én gang kan siges at tilhøre noget af det mest problematiske leksikalske stof i den interlingvale kommunikation, og samtidig får de i de gængse tosprogsordbøger tilsyneladende ikke den opmærksomhed, brugeren ikke mindst af den grund kunne have behov for. Paradokset skyldes formentlig to

ting, nemlig at de er både svære og pladskrævende at behandle adækvat. For den leksikografiske radikalist er devisen imidlertid: Enten inkluderes de kulturbundne leksemer og behandles med den respekt og forsøg på forståelse, det fremmede altid bør vises. Eller også tages de ikke med i erkendelse af, at der ikke er ressourcer til en reelt brugerorienteret behandling.

Nedenfor gengives eksempler på efter min opfattelse vellykkede og mindre vellykkede løsninger, markeret med hhv. + og -:

- (-) Fremskridtspartiet: (DK) *die Fortschrittspartei* (Bergstrøm-Nielsen 1996)
- (-) sankthansaften: *Midsummer Eve* (Vinterberg/Bodelsen et al. 1998)
- (-) kronik: (omtr.) *Feuilleton* (Bork et al. 1999)
- (+) Tivoli: *alter, gemütlicher Vergnügungspark im Zentrum Kopenhagens* (Farø 2005a)
- (+) Bundesrat: *Forbundsrådet* (det tyske parlaments andetkammer som skal godkende de lovforslag der også gælder for delstaterne; medlemmer er 3-7 ministre/senatorer fra hver af delstaterne, i alt 69) (Thomsen 2000)

Die Fortschrittspartei er som ækvivalent til *Fremskridtspartiet* problematisk, fordi der er tale om et "pseudotransparent" ord, som kun tilsyneladende forklarer sig selv. En dansk bruger, der intetanende bruger denne "ækvivalent" på tysk, vil ikke nødvendigvis vide, at de fleste tysktalende formentlig vil tage "Fortschritt-" for gode varer og lægge noget i partiet, det måske i virkeligheden ikke kan stå inde for. Det er omtrent lige så farligt som at "oversætte" Det Radikale Venstre til *die Radikale Linke*.

Sankthansaften er måske nok til en vis grad ækvivalent med *Midsummer Eve*, men på den anden side kunne det være nyttigt at få at vide, om det engelske ord har de samme de- og konnotationer som det danske. Formentlig er der forskel på indholdet, præcis som der fx er det mellem dansk og tysk *sankthans*.

Kronik hører til blandt klassikerne, der er blevet diskuteret igen og igen i dansk germanistik. Bork et al.'s forslag er efter min mening ikke overbevisende, for refererer man som dansker til en *kronik*, er det til en bestemt institutionaliseret genre, ikke bare til kultursektionen som helhed, som *Feuilleton* netop gør det. Brinks (2002) *Glosse* er her nok mere adækvat, men en forklaring af institutionen for den tysktalende (jf. Farø 2005a) vil mange gange være nyttig. Det udelukker naturligvis ikke, at den mere "syntaktiske" ækvivalent *også* gives.

Både *Bundesrat* og *Tivoli* ser ud til at kunne fungere godt: I begge tilfælde er den primære funktion formentlig reception (for hhv. danskere og tyskere), så her er forklaringer adækvate. Metateksten fylder ganske vist lidt mere, end hvis man bare havde givet en "syntaktisk" ækvivalent, men til gengæld kan den både bruges i praksis og gør samtidig brugeren *klogere*, en ordbogsfunktion, der ofte overses (videns-tilegnelse). Essensen i behandlingen af de kulturbundne leksemer er, at det ikke hjælper brugeren at ignorere reelle kulturforskelle. Så bør man nok hellere udelade

dem, som Brink (2002) i stor udstrækning har gjort det. I modsætning hertil har Farø (2005a) forsøgt at medtage og forklare kulturbundne leksemer, især i L1-L2-delen med henblik på reception.

4.2. *Idiomatik*

Idiomatikken er et andet område, hvor den leksikografiske radikalisme trænger sig på. En radikalistisk tosprogsordbog er "ikonoklastisk", når det drejer sig om idiommer og andre såkaldte billedlige udtryk. Ikonoklasme er som bekendt "billedomstyrning", og i en idiom-radikalistisk sammenhæng er fokus netop ikke "billeder", der som regel er et sekundært aspekt ved sproglige tegn, men i stedet deres synkrone kommunikative funktioner. Og et idioms ikonografi eller "billedlighed" er ikke lig med dets funktion, men snarere en slags semiotisk last, idiommet bærer rundt på. Man kan kun vide noget om idiomers funktioner ved at undersøge dem i den reelt producerede diskurs, ikke ved at studere deres billedkomposition. I den forstand er idiommer trods det umiddelbare indtryk også arbitrære tegn (jf. Farø 2006; Farø i trykken, b). Opgaven er altså ikke at finde et "vællignende" idiom i L2, men derimod at undersøge idiomets semantik, pragmatik, syntaks osv. og dernæst at lede efter L2-leksemer, der bedst muligt matcher de fundne træk. Og det er ikke nødvendigvis idiommer, der udfylder den rolle (jf. Dobrovolskij 1999, Farø, i trykken, a). Et andet punkt er, at idiommer i tosprogsordbøger indimellem får for meget vægt,³ forstået på den måde at idiommer dels ofte er ret lavfrekvente og på den anden side sjældent tilhører den basale del af ordforrådet (jf. Wandruszka 1979).

Det viser sig som regel også, at der til et idiomatisk "billede" ikke svarer én bestemt funktion, som er sproguafhængig, men en række funktioner, som er helt konventionelle – igen et udslag af idiomets arbitraritet. Det gør, at det nederlandske *zuipen* kan være en udmærket ækvivalent til *bøje armen* i betydningen 'drikke alkohol' (som i van Hees et al. 2004), mens *holde ørerne stive* og *være langt ude i tovene's* ordbogsækvivalenter måske ikke, som det er nu (hhv. Bergstrøm-Nielsen et al. 1996 og Bork et al. 1999), bør være de oplagte *die Ohren steif halten* og *in den Seilen hängen*. Det tyske idiom er nemlig en pragmatisk formel, der fungerer som afskedshilsen, mens det danske bruges med betydningen 'skærpe sanserne'. Og *være ude i tovene* bruges om den psykiske konstitution, mens betydningen af det tyske idiom snarere handler om fysisk træthed.

5. *Sammenfatning*

Jeg har fremlagt en holdning til tosprogsordbøger, som man kan kalde "leksikografisk radikalisme". Kernen i leksikografisk radikalisme er, at der ikke tages hensyn til traditioner, som ikke kan begrundes med andet end netop traditioner eller

³ Et eksempel er muligvis Farø/Voller (2003).

leksikografiske autoriteter. I modsætning til i andre sammenhænge gælder næppe, at genretraditioner bør forhindre tosprogsordbøgers fortsatte udvikling hen imod et maksimalt gennemtænkt værktøj. Leksikografisk radikalisme er altså snarere en konsekvent rationalistisk tilgang til leksikografi, end det er udtryk for den ekstremisme, der betydningsmæssigt også ligger i ordet "radikalisme".

Som nævnt er forslagene ikke noget udtømmende katalog af radikalistiske elementer. Jeg har af pladshensyn holdt mig til 10 eksemplariske punkter, men man kunne nævne mange flere. Et er fx *ingen kontrastiv retouchering*, fx af leksemers markeringsgrad. Typiske er de berømte *F-words*, som jævnlige får en slags "ækvivalent-light" i L2-L1-delen (jf. også Adersen 2005), nok af hensyn til leksikografens egen blufærdighed. Men er man leksikografisk radikalist, gør man naturligvis ikke sproget pænere, end det er. Alternativet giver nemlig brugeren farlige redskaber i hånden.

Leksikografisk radikalisme er ikke nogen teori. Man kan formentlig heller ikke henvise til nogen konkret tosprogsordbog, der er virkelig radikalistisk. Men derfor kan man jo godt have det som ideelt mål. Leksikografisk radikalisme er som sagt heller ikke identisk med den leksikografiske funktionslære, selv om der er lighedspunkter. Funktionslæren er en teori om ordbøger, men den fokuserer mindre på den konkrete udfyldning af ordbogen, som har været stærkt i fokus her. Den leksikografiske radikalisme er snarere en kritik, som både kan indeholde positive og negative elementer. Det er en kritik, som ikke retter sig mod bestemte ordbøger, men derimod mod tilsyneladende unødvendige traditioner i leksikografien, både traditioner for at indeholde bestemte elementer, traditioner for at udelukke andre og traditioner for at behandle indholdet på en bestemt måde. Traditioner, som sagtens kan tåle at blive set radikalistisk efter i sømmene med jævne mellemrum.

Litteratur

Ordbøger:

- Bergstrøm-Nielsen, Henrik et al. 1996: *Dansk-Tysk Ordbog*. København: Munksgaard.
- Bork, Erik/Fleischer, Holm/Mogensen, Jens Erik/Molly, Helmut/Møller, Elisabeth 1999: *Dansk-Tysk Ordbog*. Gyldendal.
- Brink, Karl-Erich 2002: *Pons Praxiswörterbuch Dänisch mit Sprachführer*. Stuttgart: Klett.
- Brüel, Sven/Nielsen, Niels Åge/Plon, Lilian 1993: *Fremmedordbog*. DDO = Hjorth, Ebba/Kristensen, Kjeld (led. red.): *Den Danske Ordbog* (6 bd.) København: Gyldendal/DSL 2003-05.
- Farø, Ken 2005a: *Universalwörterbuch Dänisch*. København: Langenscheidt.
- Farø, Ken/Voller, Inge 2002: *Tysk-Dansk/Dansk-Tysk Ordbog*. København: Gyldendal.
- van Hees, Annelies et al. 2004: *Dansk-Hollandsk Ordbog*. København: Gyldendal.

- Hårbøl, Karl/Schack, Jørgen/Spang-Hanssen, Henning 1999: *Dansk Fremmedordbog*. Gyldendal.
- Jensen, Erik/Outzen, Svend Aage/Bienz, Karin et al. 2003: *Tysk-dansk/Dansk-tysk ordbog*. København: Politiken.
- Kjærulff Nielsen, Bernhard 1998: *Engelsk-Dansk Ordbog*. København: Gyldendal.
- Laureys, Godelieve (led. red.) et al. 2004: *Hollandsk-Dansk Ordbog*. København: Gyldendal.
- MAD = Rundell, Michael (led. red.) et al. 2002: *Macmillan English Dictionary*. Macmillan: Oxford.
- PRO = Appel, Vibeke et al. 2001: *Politikens Retskrivningsordbog*. København: Politiken.
- Thomsen, Knud Troels 2000: *Tysk kombi*. København: Gad.
- Vinterberg, Hermann/Bodelsen, C.A./Pedersen, Viggo Hjørnager 1998: *Dansk-Engelsk Ordbog*. København: Gyldendal.
- Gawinski, Birthe/Windfeld-Hansen, Johan 1998: *Dansk-spansk ordbog*. København: Munksgaard.
- Zettersten, Arne/Lauridsen, Hanne 1999: *Engelsk-dansk ordbog*. København: Politiken.

Anden litteratur:

- Adersen, Sofie Brix 2005: *Om kulturspecifika og deres oversættelse. En undersøgelse og diskussion af kulturspecifikke udtryk og deres oversættelse fra dansk til tysk*. København: Københavns Universitet. [Speciale].
- Bergenholtz, Henning 1997: Polyfunktionale ordbøger. I: *LexicoNordica* 4, 15-28.
- Bergenholtz, Henning/Tarp, Sven 2002: Die moderne lexicographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. I: *Lexicographica* 18, 253-263.
- Carstensen, Stine 1989: Zur Darstellung kulturgebundener Wörter in der zweisprachigen Lexikographie. I: *Kopenhagener Beiträge zur Germanistischen Linguistik* 25, 5-21.
- Dobrovolskij, Dmitrij 1999: On the Cross-Linguistic Equivalence of Idioms. I: Beedham, Christopher (red.): *Langue and Parole in Synchronic and Diachronic Perspective*. Amsterdam etc.: Elsevier.
- Farø, Ken 2004a: Hvornår går man over åen efter vand? Idiomaticke ækvivalensproblemer i leksikologi og leksikografi. I: *LexicoNordica* 11, 85-108.
- Farø, Ken 2004b: En ny tyskordbog i Danmark. Anmeldelse af Erik Jensen, Sven Aage Outzen, Karin Bienz: *Politikens Tysk-Dansk/Dansk-Tysk Ordbog*. Politikens Forlag 2003. I: *LexicoNordica* 11, 235-244.
- Farø, Ken 2005b: *Fiktionymmer i bilingval leksikografi*. I: Ruth Vatvedt Fjeld/Dagfinn Worren (red.): *Nordiska Studier i Leksikografi 7. Rapport frå Konferanse om leksikografi i Norden. Volda 20.-24. mai 2003*. Oslo: Nordisk forening for leksikografi, 120-132.
- Farø, Ken 2006: *Idiomatisitet – Ikonisitet – Arbitraritet. Beitrag zu einer funktionalistischen Theorie der Idiomäquivalenz*. København: Københavns Universitet [ph.d.-afhandling].
- Farø, Ken (i trykken, a): *Dogmatismus, Skeptizismus, Nihilismus und Pragmatismus bei der Idiommübersetzung*. I: Häcki Buhofer, Annelies (red.): *EUROPHRAS 04*. Basel.
- Farø, Ken (i trykken, b): Warum Idiome im mehrfachen Sinn arbiträr sind. I: *EUROPHRAS 05*. Slovenien.
- Kromann, Hans-Peder/Riiber, Theis/Rosbach, Poul 1984: Überlegungen zu Grundfragen der zweisprachigen Lexikographie. I: *Germanistische Linguistik* 3-6, 159-238.
- Kromann, Hans-Peder/Riiber, Theis/Rosbach, Poul 1991: Principles of Bilingual Lexico-

- graphy. I: Hausmann, Franz-Josef et al. (red.): *Dictionaries. Wörterbücher. Dictionnaires* (HSK 5.3). Berlin etc.: de Gruyter, 2711-2728.
- Mogensen, Jens Erik 2004: Den såkaldte uoversættelighed og tosprogsleksikografien. I: Bandle, Oskar et al. (red.): *Verschränkung der Kulturen. Der Sprach- und Literaturtausch zwischen Skandinavien und den deutschsprachigen Ländern*. Tübingen/Basel: Francke, 165-179.
- Svensén, Bo 2004: *Handbok i lexikografi: Ordböcker i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Tarp, Sven 2004a: Hvad er en bilingval ordbog? I: *LexicoNordica* 11, 5-22.
- Tarp, Sven 2004b: Korrektur og retning som leksikografiske funktioner. I: *Hermes* 33, 117-147.
- Wandruszka, Mario 1979: Kontrastive Idiomatik. I: M. Höfler/H. Vernay/L. Wolf (Hg.): *Festschrift Kurt Baldinger zum 60. Geburtstag*. Tübingen: Niemeyer, 951-963.

Spansk mad – en illustreret onlineordbog

Spanish food – an illustrated online dictionary. The question of equivalence has always created problems for lexicographers when we work with bilingual dictionaries. I have wondered whether information technology might offer a solution to this problem. The question that I have tried to answer is the following: “Can encyclopaedic information including illustrations rectify missing equivalence in bilingual, web-based dictionaries within cultural subject areas?” I have chosen to limit the subject to gastronomy, because this subject has a very culture-specific terminology. The target group of the dictionary is not limited to one particular group. Because of the hypertext structure that allows the user to work at various levels, the dictionary may meet the needs among Danish-speaking people with varying knowledge of Spanish. My conclusion is that the user of an online dictionary will have access to a lot more relevant information using an online encyclopaedic dictionary than it is possible to offer in a printed dictionary.

Spansk mad - en illustreret ordbog

Søg: DK A-Z Verber

Dette er en **spansk-dansk ordbog** for dig, der er interesseret i spansk mad og som gerne vil vide lidt mere end blot, hvad de enkelte madretter hedder på dansk.

Hvis det er første gang, du besøger denne ordbog, vil det være en god idé at læse brugervejledningen igennem.

Ordbogen indeholder:

A-Z	Alfabetisk ordliste	Måltider	med ordbog over de kendteste spanske retter
Madkort	tybiske madvarer og retter fra Spaniens regioner	Verber	Verber med relation til mad
Denominación de Origen	spanske produkter med officielt kvalitetsstempel	Faste vendinger om mad	

1. Indledning

Formålet med ordbogen er at hjælpe ordbogsbrugeren med at få en bedre fornemmelse af, hvad et givet opslagsord dækker over, især i de tilfælde, hvor der ikke findes en ækvivalent. Det er mit håb, at definitioner og/eller forklaringer på både kilde- og målsproget samt illustrationer vil yde denne hjælp. Denne type informationer vil man kun sjældent finde i papirudgaver af bilingvale ordbøger.

Ordbogen er blevet oprettet i forbindelse med et masterspeciale¹ på 'IT, Sprog og Læring'.

Der er to vigtige grunde til mit valg af emne for denne ordbog. Dels er den gastronomiske terminologi kendetegnet ved mange kulturbundne og dermed nul-ækvivalente ord. Dels har interessen for Spanien som rejsemål og for spansk mad været støt stigende gennem de seneste årtier, og adskillige spanske madvarer og retter er efterhånden kendt i Danmark. Der skyder spanske restauranter og tapas-barer op rundt omkring i landet, flere og flere spanske specialiteter kan købes i danske butikker.

Denne ordbog er tænkt som en hjælp til danskere med interesse for spansk mad og spanske produkter, og for nogle brugeres vedkommende vil der måske efterhånden opstå et behov for at få større viden om de enkelte retter og produkter end de informationer, man kan få i en almindelig ordbog eller i en kogebog. Det kunne være noget om, hvor i Spanien man dyrker hvilke produkter, dyrkningsmetoder (konventionelt/økologisk) etc.

Målgruppen er derfor dansksprogede med et vist kendskab til spansk, som har brug for at forstå (evt. også oversætte) en spansk tekst, som indeholder gastronomisk terminologi. Kendskab til spansk vil give et større udbytte, da der ud over ækvivalenter eller oversættelsesforslag vil være definitioner på spansk samt links til dansksprogede websteder, men ordbogen kan i princippet også bruges af ikke-spanskkyndige, som blot ønsker at kunne slå et ord op og få en dansk oversættelse eller definition/forklaring.

Jeg opfatter ikke min målgruppe som en entydig størrelse, men ser den mere som målgrupper på forskellige niveauer eller med forskellige behov. Hypertekststrukturen i en onlineordbog giver gode muligheder for at lade brugeren selv beslutte, hvilke informationer og hvor mange han/hun har brug for i den givne situation.

2. De vigtigste teoretiske aspekter

2.1. Ækivalens-problematikken

Jeg tager udgangspunkt i NLO's (Bergenholtz et al. 1997:15-18) liste over ækvi-valentyper:

- 'låneord', dvs. direkte brug af kildesprogsordet (eks. *guide* fra engelsk)
- 'oversættelseslån' (eks. *gæstearbejder* fra tysk)
- 'kulturel ækvivalent' (eks. *Folketinget* som ækvivalent til f.eks. det norske *Stortinget*)
- 'parafraseækvivalent' (eks. *Arztshelferin* på tysk kan gengives 'lægesekretær med laboratorieudannelse')

¹ Hele specialrapporten kan læses her: <http://www.gawinski.dk> – klik på 'master'.

- ‘encyklopædisk forklaring’ (eks. *ajo blanco* på spansk kan gengives med ‘kold hvidløgs- og mandelsuppe’)

Og her er det især sidstnævnte, der er relevant for min ordbog, selv om de andre også forekommer.

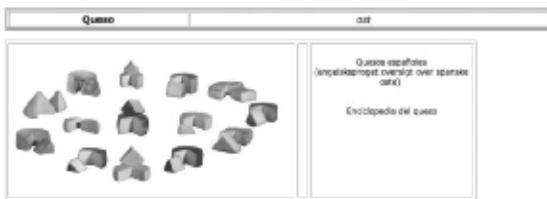
2.2. Encyklopædiske informationer inkl. illustrationer

Her kan der f.eks. være tale om definitioner og/eller illustrationer af forskellig art. Jeg har valgt at tage udgangspunkt i Svenséns (1987:165-170) illustrationskategorier:

‘en enkelt genstand eller en del af en genstand’



‘flere typer inden for samme klasse’



‘genstanden i sit miljø’



‘delaspekter af genstanden’



‘centrale genstande og begreber inden for et emneområde’



Svensén nævner yderligere to illustrationskategorier: ‘genstanden i funktion’ og ‘miljø med typiske genstande’, som jeg dog ikke har eksempler på i min ordbog endnu.

2.3. Ordbogen som læringsmiddel

Ud over at være et nyttigt redskab for folk med interesse for spansk mad og spanske produkter er denne ordbog også tænkt som et muligt læringsmiddel i spanskundervisningen.

Ute Rampillon (2001) nævner nogle eksempler på ordbogen i sprogundervisningen:

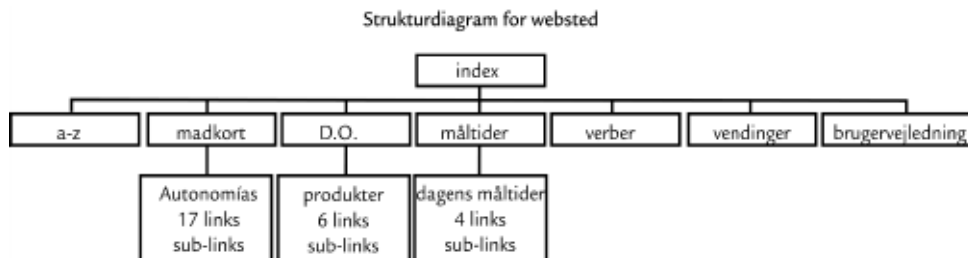
- udvikling af læringskompetence
- hjælp til indlæring af et fremmedsprog
- takling af fremmedsproglige situationer i dagligdagen
- ‘tidsfordriv’

I denne specifikke ordbog forestiller jeg mig i første omgang disse forløb (med mulighed for at udvide med flere hen ad vejen):

- i et undervisningsforløb om Spaniens geografi kunne man tage udgangspunkt i siden med Spanienskortet (<http://spanskmad.gawinski.dk/madkort.html>). Her får man informationer om de regionale forskelle med hensyn til, hvilke afgrøder man dyrker, hvilke produkter man fremstiller – og dermed hvilken mad man spiser i de pågældende områder.
- et andet undervisningsforløb kunne tage udgangspunkt i siden med Måltider (<http://spanskmad.gawinski.dk/maaltider.html>). Her får man informationer om dagens måltider, og hvad der er karakteristiske for dem (tidspunkter, retter etc.).
- et tredje undervisningsforløb kunne handle om fænomenet Denominación de Origen, dvs. det spanske landbrugs-, fiskeri- og fødevarerministeriums kvalitetskontrol af forskellige landbrugsprodukter (<http://spanskmad.gawinski.dk/do.html>)

3. Ordbogens opbygning

Ordbogens opbygning ses i følgende strukturdiagram:
men fremgår i øvrigt også af forsidens menulinje:



Menulinjen indeholder forskellige makrostrukturelle elementer:



Index-side.

Søg:

Søgefelt, hvor alle ordbogens ord kan slås op.

A-Z

Den alfabetiske ordliste, som er samlet på én side, inddelt i grupper på 3-4 bogstaver. Dette er muligt, så længe ordlisten ikke er meget omfattende.

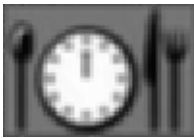
Den systematiske makrostruktur ses i de tre følgende menupunkter:



Det klikbare Spanienskort, som udgør den ene måde at systematisere emnet spansk mad på, nemlig ud fra de regionale forskelle, der findes inden for såvel produkter som madretter.



Vin-ikonet symboliserer den officielle spanske kvalitetsstempeling af visse landbrugsprodukter, Denominación de Origen (D.O.). Jeg har her valgt nogle af de mest kendte ud, men vil i forbindelse med en udbygning af ordbogen tage alle med.



Dagens måltider er endnu en måde at strukturere systematisk på. Hvad spiser man typisk til hvilke måltider og hvornår. Alt sammen noget, der adskiller sig fra danske forhold, og derfor nyttig information i forbindelse med emnet spansk mad.

4. Onlineordbøger – nu og i fremtiden

I min problemformulering til masterspecialet stiller jeg følgende spørgsmål:

Kan encyklopædiske informationer herunder illustrationer afhjælpe manglende ækvivalens i bilingvale, netbaserede ordbøger inden for kulturbundne fagområder?

og min konklusion herpå er, at det ikke umiddelbart kan lade sig gøre at ændre på den manglende ækvivalens imellem ord og begreber på to sprog – det er et faktum,

at visse kulturbundne ord og begreber på ét sprog simpelthen ikke findes på et andet sprog.

Hvis man er oversætter, er dette naturligvis et problem, for man kan ikke bare undlade at skrive noget i en oversættelse, hvis man støder på et nul-ækvivalent ord.

For alle andre ordbogsbrugere derimod kan problemet løses på forskellig vis. Når det drejer sig om at forstå en tekst, hvori der forekommer delvist ækvivalente eller nul-ækvivalente ord, vil det være tilstrækkeligt at få en definition eller forklaring på ordet eller begrebet. Desuden vil en eller anden form for illustration også være nyttig.

Inden for leksikografien har det hidtil ikke været særlig acceptabelt at inkludere informationer af encyklopædisk art i bilingvale, almensproglige ordbøger. Grundene hertil har været dels principielle forestillinger om, at sådanne informationer ikke hører hjemme her, dels økonomiske grunde fra forlagenes side.

I forbindelse med den teknologiske udvikling og den deraf følgende mulighed for at udvikle hypertextbaserede ordbøger, er især det økonomiske aspekt blevet mindre fremtrædende, men det har ikke nødvendigvis betydet, at vi har set en voldsom udvikling inden for leksikografien. I mange tilfælde indeholder de hypertextbaserede ordbøger stadig nøjagtig de samme elementer som de papirbaserede ordbøger. Det gælder de elektroniske (offline)ordbøger, og det gælder i de fleste tilfælde også de netbaserede udgaver af disse ordbøger.

Inden for de encyklopædiske produkter har vi derimod igennem det sidste årti set en stadig mere udbredt brug af hypermediets muligheder. Som eksempler kan nævnes Encyclopaedia Britannica og Microsofts Encarta, som begge udgives på dvd og indeholder alle tænkelige multimediale og interaktive ressourcer.

Min konklusion ud fra problemformuleringen vil være, at det for en ordbogsbruger, uanset dennes sproglige niveau på kildesproget, vil være en fordel med såvel encyklopædiske informationer som illustrationer for at kunne få den bedste forståelse for, hvad et givent ord eller udtryk dækker.

Man kunne godt forestille sig, at brugeren kunne få de samme informationer i en papirbaseret ordbog – men et meget væsentligt element ville mangle, nemlig hypertextualiteten, dvs. muligheden for at kunne modtage informationerne systematisk ordnet og ikke i en lineær form med eventuelle supplerende informationer fordelt over bogens forskellige dele. Man kan sige, at informationerne også er 'spredt' i onlineudgaven, brugerne mærker det bare ikke, for de er aldrig længere end et klik borte. Dette gælder naturligvis kun, hvis webstedet er logisk og brugervenligt opbygget (se f.eks. afsnit 2.3. om Ordbogen som læringsmiddel).

En papirordbog med lige så mange informationer ville blive alt for stor, og man ville ikke have adgang til hyperlinks ud over URL-lister.

Det er mit håb, at det vil være muligt at få overbevist forlagene (eller andre, der kan se idéen i at investere i tidssvarende leksikalske produkter) om det nyttige i at

give ordbogens bruger enhver supplerende information til et givent opslagsord – også selv om det bevæger sig ud over de hidtil benyttede medier (papir, cd-rom, dvd) og ‘går’ online.

Jeg har hørt det argument, at det er på grund af internettets foranderlige karakter, at man ikke tør satse mere på det. Jeg tror på, at der må kunne spares nogle ressourcer i forbindelse med den konventionelle ordbogsfremstilling, som kunne benyttes til at ansætte nogle personer til at sørge for opdatering af onlineordbøgerne. Det gælder både indsættelse af nye opslagsord og tjek af webstedseksterne links.

Jeg er ikke tilhænger af, at papirbaserede ordbøger skal afskaffes her og nu, men jeg tror på, at internettet kan tilføre leksikografiske produkter af enhver art nogle dimensioner, som de trykte medier ikke kan. Det ville være dejligt, hvis forlagene ville være med til at udvikle sådanne produkter i den nærmeste fremtid!

Litteratur

Ordbøger:

- Bergenholtz, Henning et al.: *Nordisk Leksikografisk Ordbok* (NLO). Oslo: Universitetsforlaget 1997.
- Bratli, C.: *Spansk-dansk ordbog*. København: Fischers forlag 1947.
- Brink Andersen, H. & Windfeld Hansen, J.: *Spansk-dansk ordbog*. København: Munksgaard (nu Gyldendal) 1999.
- CLAVE *Diccionario de Uso del Español Actual*, Madrid, 2002.
- Hornby, A.S.: *Oxford Advanced Learner's Dictionary* (OALD). Oxford: Oxford University Press 1990.
- Hårbøl, K. et al.: *Dansk Fremmedordbog* (DFO). København: Munksgaard 1997.
- Martínez de Sousa, José: *Diccionario de Lexicografía Práctica* (DLP), VOX, Bibliograf, Barcelona, 1995.
- Merriam-Webster *Spanish-English Dictionary* – iFinger-version.
- Norstedts *Spanska Spansk-svensk/svensk-spansk ordbok*. Stockholm: Norstedts 1999.
- Nudansk Ordbog med etymologi*. København: Politiken 2001.
- PONS *Lexiface Global Spanisch-Deutsch*, Ernst Klett Verlag GmbH – iFinger-version.
- Smith, C.: *Collins Diccionario Inglés Español>Inglés English>Spanish*. Glasgow: Harper Collins Publishers 1998.
- Taschenwörterbuch Spanisch*, Bertelsmann-Larousse, 2001.
- The Oxford Spanish Dictionary Spanish-English English-Spanish*. Oxford: The Oxford University Press 1994.
- The Oxford-Duden Pictorial Spanish and English Dictionary*. Oxford: Clarendon Press 1995.
- Vater, P.: *Spansk-dansk ordbog*. København: Gyldendal 2000.

Anden litteratur:

Burke, Sean Michael 1998: *The Design of Online Lexicons*.

<http://interglacial.com/~sburke/ma/>

Hupka, Werner 1989: *Wort und Bild – Die Illustrationen in Wörterbüchern und Enzyklopädien*, Lexicographica, Series Maior, 22. Tübingen: Niemeyer.

Lysemose, Marit 2001: *Fra papir til cyperspace: ordbogsstrukturer i online ordbøger*, Århus: Handelshøjskolen i Århus, Fransk Institut. <http://www.maritlysemose.person.dk/>

Martín Sánchez, Manuel 1997: *Diccionario del español coloquial – dichos, modismos y locuciones populares*, Tellus.

Refranero clásico español y otros dichos populares – antología, Madrid: Taurus, 1974.

Schnorr, Veronika 1986: Translational Equivalent and/or Explanation? The Perennial problem of Equivalence. I: *Lexicographica* 2. Tübingen: Niemeyer, 53-60.

Svensén, Bo 1987: *Handbok i lexikografi*. Stockholm: Esselte Studium & Tekniska Nomenklaturcentralen.

Wonsild Nielsen, Birthe 1992: *Echar la siesta og 2002 andre spanske udtryk og talemåder*. København: Høst.

Kogebøger og andre kilder brugt til lemmasektion:

Aris, Pepita: *Mad og drikke Spanien*, København: Gad Guide, 1992.

Collins: *Language Survival Guide*. Glasgow: HarperCollins Publishers, 2001.

de Juana: *Guía de pescados y mariscos de consumo usual en España*. Barcelona: Ediciones Omega 1987.

Knorr, Petra & Paz, Victoria: *Den spanske kogebog*. Köln: Komet (u.å.).

La cocina española, Bonechi, Italia, 2003.

Productos con Denominación de Origen, Susaeta (s.a.): *Quesos españoles, Aceites de España, Productos del campo, Legumbres, arroces y hortalizas, Vinos de Galicia, Carnes y embutidos*.

Trutter, Marion: *Spanske specialiteter – en kulinarisk opdagelsesrejse*, Könemann, 2000.

Vazquez, A.: *Pocket Menu Reader Spain*. N.Y.: Langenscheidt, 2000.

lookedup4you: studerende, kunder og online-leksikografi

lookedup4you: Students, clients and online lexicography. Seventeen BA students at the University of Applied Sciences, Hamburg, Germany, have developed a lexicographic online reference tool with young users as the target group for a client, Verbraucherzentrale Nordrhein-Westfalen, Germany. The online reference tool is based on a specialist system, an alphabetic approach and search options. For this project, lexicography was defined as an opportunity to structure information and as a tool that would meet the client's needs. The project included all the processes involved in the production of lexicographic online reference works; the definition of a concept designed to persuade the client; the development of a prototype; testing this prototype in consultation with the future users of the reference work; lemmatisation; writing the articles; the production of images and films; the conceptual design and production of quizzes for the reference work; and all the computer programming and design work necessary for a project of this nature.

1. Rammen omkring projektet

University of Applied Sciences i Hamborg, Tyskland (HAW 2005) har ca. 14.000 studerende og uddanner bl.a. bachelor-studerende i Medien und Information samt i Information and Library Services (DMI 2005). De færdige BA'ere arbejder inden for mange forskellige områder som bl.a. biblioteker, forlag, videnskabelige institutioner, informationsafdelinger i større virksomheder og ikke mindst inden for new media-området. Uddannelsen formidler frem for alt viden om forskellige metoder til strukturering af informationer og viden, såsom klassifikationer, tesauri, ontologier og ikke mindst lingvistiske metoder. Halvdelen af det tredje studieår bruges af de studerende udelukkende på et projekt (DMI 2005a), hvor de udfører en konkret opgave for en kunde. Projektgruppen omfatter typisk 12-15 studerende og 2-3 undervisere fra forskellige fagområder – nogle gange også eksterne lektorer med specialviden til netop den i projektet påkrævede opgaveløsning. De studerende eller universitetet får ingen betaling for opgaveløsningen, men de får til gengæld på denne måde en enestående chance for at lære at planlægge, budgettere og gennemføre en afgrænset arbejdsopgave i løbet af afgrænset tid med afgrænsede ressourcer. Ud over den store faglige andel er dette en vigtig erfaring, som svarer til nutidens arbejdsform i mange virksomheder.

Kundens fordel ved denne ordning er, at der bliver fremstillet et produkt, som, i det mindste hvad angår personaleomkostningerne, er gratis. På den anden side er afviklingen af projektet set fra kundens side tit kreativ, men ikke altid strømlinet. – Set fra de studerendes side er et sådant projekt en god mulighed for at bruge de på dette tidspunkt lærte færdigheder samt lære nyt gennem opgaven på den ene side og

samarbejdet med de andre studerende på den anden side. Projekterne har meget forskellige emner, men er altid relateret til strukturering, bearbejdelse og præsentation af informationer – dette er jo kerneområdet i uddannelsen.

Projektstyringen ligger hos de studerende. Undervisernes opgave i projektet er hovedsagelig at være faglig rådgiver. Kun hvis projektet udvikler sig i en fuldstændig forkert retning, eller hvis der opstår alt for store forsinkelser i forhold til tidsplanen, er det underviserens opgave at gribe ind. Der bliver ikke udpeget en projektleder, men projektet styres af demokratiske afgørelser på de ugentlige projektmøder. Hver studerende har mindst en gang i projektet opgaven som mødeleder og på et andet tidspunkt som referent fra et projektmøde. Mødereferaterne har en afgørende betydning for, at der ikke gentagne gange sættes spørgsmålstejn ved afgørelser, som man allerede har truffet.

En vigtig del af uddannelsen inden projektfasen er it-fagene. Efter at have lært forskellige grundlæggende ting omkring hard- og software samt produktion af websider (HTML etc.) uddannes de studerende – delvis i valgfag – i emner som XML/XSLT, programmering af dynamiske websider, SQL- og XML-databaser samt i screendesign og produktion af multimedia (film, lyd etc.). Undersøgelser i (website-)usability er ligeledes en del af disse it-relaterede fag. Dermed har i det mindste en del af de studerende, som deltager i projekterne, et godt grundlag for at producere rammerne samt dele af indholdet (især det multimediale) af indholdet i det endelige produkt.

Lingvistikken er også en del af uddannelsen inden projektet. Nogle af hovedområderne er her automatisk indeksering baseret på bl.a. morfologiske og statistiske metoder, text-mining, eller brug af andre til f.eks. billeder. De nævnte (it-)fag er ligeledes tit relateret til lingvistiske emner eller nærmere betegnet til praktisk eller teoretisk leksikografi eller terminologi. Dermed er et typisk kursus, som tager hensyn til såvel leksikografi/terminologi som til it-området, programmering af en internetordbog ved hjælp af et nærmere bestemt programmeringssprog eller i XML.

2. Projektet

Verbraucherzentrale Nordrhein-Westfalen (VZ 2005), som varetager forbrugernes interesser i den tyske delstat Nordrhein-Westfalen (18 millioner indbyggere), har en forbrugerportal på internettet til unge forbrugere (Checked 2005). Emner fra mobiltelefoni til købeloven og fra sport til sundhed behandles her målrettet mod unge forbrugere fra 14 til 18 år eller lidt ældre. Ud over emnerne er også stilen i teksterne samt layoutet af hele portalen skræddersyet til denne målgruppe.

De studerendes projekt gik ud på at udvikle samt implementere en online-ordbog (lookedup 2005). Da terminologien omkring ordet ordbog fra teoretisk side er meget omdiskuteret, og frem for alt i den almindelige brugers forståelse kun omfatter sprogordbøger, oftest i trykt form endda, besluttede projektets team at bruge

betegnelsen opslagsværk. Denne betegnelse er mere neutral, men har samtidig stadigvæk en tæt tilknytning til leksikografisk teori og praksis. Kravene til “online-opslagsværk til unge forbrugere”, som den foreløbige titel blev, var:

- Artikelteksterne skulle være målrettet mod unge forbrugere i stil og sprogbrug, men samtidig bruge og formidle veldokumenterede fakta – ikke mindst da de nævnte områder fra opslagsværket delvis indeholder juridiske aspekter
- Artiklerne skulle ikke kun bestå af tekster (og billeder), men derudover gøre brug af andre multimediale og interaktive elementer, som formidler den ønskede information og fanger de unges interesse i at læse denne information og dermed blive til mere velinformerede forbrugere
- Teknikken bag opslagsværket skulle være flytbar, dvs. det skulle være muligt at flytte ordbogen fra et teknisk område til et andet uden større vanskeligheder
- Opslagsværkets tekniske system skulle indeholde mulighed for at redigere og videreudvikle indholdet.

De studerende valgte følgende trin som arbejdsplan:

- Definition af brugergruppen
- Lemmatisering
- Definition af makro-, mikro- samt tilgangsstruktur
- Definition af multimediale midler samt placering af disse ved forskellige lemmata
- Indsamling af fakta til tekstproduktion samt verificering af kildematerialet
- Definition og implementering af en relational database
- Definition og implementering af brugergrænsefladen samt grænsefladen til redaktionssystemet
- Design og layout
- Evaluering af brugernes arbejde med en testudgave af opslagsværket, som indeholder udvalgte testartikler

Til grund for disse strukturerede arbejdsopgaver ligger en del afgørelser for eller imod arbejdsprincipper eller redskaber. Således bestemte man hurtigt, at en fordeling i grupper var nødvendig, hvis tidsplanen (3 måneder) skulle overholdes. Det blev til en teknisk arbejdsgruppe, en design- og layout-gruppe samt en film- og multimedia-gruppe. De udvalgte lemmata blev fordelt på alle projektdeltagere. Dermed havde alle gennem deres erfaringer med at skrive leksikografiske tekster en tæt tilknytning til og viden om de problemer, som opstår ved leksikografisk tekstproduktion. Den i tilsvarende projekter tit oplevede skillelinje mellem programudvikling og tekstarbejde opstod derfor ikke i projektgruppen.

Da det drejede sig om et projekt til en kunde, var der yderligere opgaver, som

skulle løses:

- Præsentation af et startkoncept til kunden, som skulle overbevise denne om de i projektet valgte midler og metoder
- Slutpræsentation af produktet i en større kreds med kundens deltagelse
- Overdragelse af projektet til kunden
- Dokumentation af projektet (historisk) samt produktet (teknisk, brugergrænseflader).

3. Leksikografien i projektet

3.1. *Det leksikografiske koncept*

Ud fra et metaleksikografisk synspunkt kan valget af metoder og midler i projektet bedst beskrives som en intuitiv vej til målet (Geeb/Spree 2005). Ved projektets start fik de studerende en grundig introduktion til den leksikografiske teori og praksis. Dette indebar også en gennemgang af den leksikografiske terminologi samt de forskellige leksikografiske produkter. Også områder som fagleksikografi samt terminologividenskab var en del af denne introduktion. Selve introduktionen byggede på de studerendes viden om forskellige slags informationsstrukturering. De havde for eksempel allerede på dette tidspunkt en omfattende viden og praktisk erfaring med fagsystematikker og klassifikationer, som ikke mindst i fagleksikografien har stor betydning. Hovedvægten i introduktionen til den leksikografiske arbejdsproces lå på en gennemgang af de leksikografiske strukturelle komponenter samt definitionen af og arbejdet med et tekstkorpus. Det sidstnævnte emne var dog ikke helt ukendt for de studerende på grund af deres viden fra det informationslingvistiske arbejdsområde automatisk indeksering. Arbejdet med den leksikografiske teori blev underbygget af et bredt udvalg af eksempler – mest fra online-opslagsværker.

I den efterfølgende diskussion af mulige opgaveløsninger blev de ovenfor nævnte delmål formuleret. På trods af introduktionen til den leksikografiske teori og på trods af den dermed lærte fagterminologi blev diskussionen i projektet hovedsagelig ført af to forskellige grupper: en gruppe, som gjorde brug af sin leksikografiske viden, og en gruppe, som valgte at bruge terminologi og redskaber til strukturering af information fra andre discipliner. Det lykkedes dog at skabe en fælles platform i diskussionen ved hele tiden at fokusere på produktets rammer, som ligger i det for alle velkendte world wide web. Desuden viste det sig, at deltagerne, som havde en leksikografisk viden fra andre kurser, blev inspireret af andre deltagere, som var specialister i programmering, grænsefladedesign eller usability engineering.

Da leksikografiske produkter – online eller på tryk – er en velkendt ting fra hverdagen, kunne alle projektdeltagere føre en saglig og målrettet diskussion, selv om deres viden inden for leksikografien var meget forskellig. Diskussionen viste

dermed, at leksikografien er en meget almindelig måde at strukturere information på, selv om man som bruger uden leksikografisk viden ikke direkte er klar over det. Og det viste ligeledes, at et leksikografisk projekt – frem for alt på internettet og med en forholdsvis ung målgruppe – langt fra kan klare sig med lingvistisk viden alene.

3.2. Lemmasektion

Ved lemmatiseringen fravalgte man på grund af det dertil estimerede tidsforbrug muligheden for at skabe et eget tekstkorpus til formålet. Ud fra materialet i selve forbrugerportalen mente man at kunne aflede de relevante lemmata. Dette mindre tekstkorpus blev suppleret af forbrugercentralens skriftlige materiale såsom brochurer og undervisningsmateriale. – Derudover så de studerende sig i stand til selv at vurdere lemmalistens omfang og evt. manglende dele, idet de alle befandt sig i starten af tyverne – altså ikke mange år fra brugergruppens alder. Men de unge forbrugere skulle også direkte spørges om deres mening vedrørende lemmalisten som en direkte validering af dette udvalg. Til dette formål udvalgte man nogle klasser i forskellige skoler samt foreninger og andre ungdomsgrupper. Som metode valgte projektgruppen spørgeskemaer samt interviews i små grupper.

Denne metode viste sig at være ret tidskrævende, men til gengæld blev det klart, at der er en del emner, som hverken forbrugercentralens materiale, portalen eller projektdeltagernes lemmaliste indeholder. Lemmalisten og mediostrukturen, som allerede på dette tidspunkt viste sig at være en vigtig del af overvejelserne ved lemmatiseringen, blev en del udvidet på grund af denne kontakt med målgruppen.

3.3. Teknikken

Blandt de mange muligheder med hensyn til den tekniske produktion af opslagsværket, som blev diskuteret, viste der sig at være to hovedalternativer: enten at bruge XML/XSLT (Geeb 2001, Geeb 2003) eller at bruge en (relational) database (Geeb 2005). Årsagen til, at netop disse to muligheder forblev som de to sidste i diskussionen, er på den ene side kundens behov vedrørende systemets flytbarhed, altså muligheden for uden problemer at flytte produktet fra ét sted til et andet eller fra ét operativsystem til et andet. På den anden side er disse to teknologier også meget anvendte og derfor veludviklede – ikke mindst i lingvistisk henseende og med henblik på leksikografiske behov. Dette er også grunden til, at disse to elementer er en væsentlig del af den it-undervisning, som de studerende har fået – en yderligere grund til at diskussionen omfattede de to nævnte alternativer.

XML som værktøj har store fordele på redaktionssiden. Det er nemt og overskueligt at bruge XML til tekstproduktion efter at være blevet introduceret til selve dokumentets grammatik, altså hierarkiet af elementer og attributter, som kan, må eller skal bruges. Strukturen og hermed også den leksikografiske struktur ville i XML være særdeles synlig for tekstproducenterne, og man ville kunne starte tekst-

produktionen uden yderligere programmering. Ved brug af tilsvarende retrieval-sprog (XSL, XPath) er XML desuden meget egnet til delvis og bearbejdet visning af indhold, altså f.eks. søgning.

Selve XML har dog intet layout og intet brugerinterface, men er kun en standard til produktion af serialiserede hierarkier. Der ville altså være behov for en transformation til et andet medium – f.eks. HTML. Også det er en overskuelig opgave ved hjælp af den XML-beslægtede standard XSL. Dog er der en del ulemper forbundet med denne løsning:

- Transformation af dokumenterne vil enten foregå på klient-computeren (browser), hvilket gør, at brugeren har adgang til alle data, som ligger i XML-dokumenter, også de data, som måske slet ikke er beregnet til brugeren.
- Transformationen på klientsiden vil ikke altid give det ønskede resultat, da det ikke er muligt at tage hensyn til alle former for software (browser) eller operativsystem, som brugeren kan have på sin maskine
- Hvis transformationen skal foregå på serversiden, og brugeren dermed kun skal kunne få HTML-sider, som laves på grundlag af XML-dokumenterne, kræver dette noget programmel, som på projektets tidspunkt ikke fandtes standardiseret. De programmer, som er i stand til at løse denne opgave (f.eks. AxKit eller x.pl), kræver alligevel en del it-viden ved installationen og gør det derfor sværere at opfylde kundens ønske om en let flytbarhed af produktet.
- Hvis de enkelte artikler til lemmalisten hver lå i en XML-instans (dokument), ville det betyde, at det ikke ville være muligt at generere alfabetiske eller systematiske lemmalister eller tilbyde en søgning uden yderligere programmel på serversiden.

For at kunne tilbyde denne tilgangsstruktur, eller en brugbar mediestruktur, ville det være nødvendigt enten selv at udvikle tilsvarende programmer eller at bruge en XML-database. Selvudviklede XML-programmer, som under alle omstændigheder skal bygge på en XML-parser på serversiden, kræver en del forudgående installation, som ikke er en del af standardinstallationerne på serversiden. – XML-databaser på den anden side er enten meget dyre (f.eks. Tamino) eller er stadigvæk på eksperimentel basis og kræver desuden specialviden inden for serveradministration (f.eks. eXist, Xindice, NatiX)

På baggrund af disse overvejelser med udgangspunkt i kundens ønske om, at der skulle bruges en meget udbredt teknologi, som gør, at produktet nemt kan flyttes og videreudvikles, faldt valget på en relational database, som følger standarden SQL. Derudover skulle der bruges et programmeringssprog, som ligeledes er meget kendt, udbredt og samtidig stabilt og velfunderet. Begge dele skulle være velimplementerede i verdenen af web-servere og ikke bundet til et bestemt kommercielt produkt. Dermed faldt de studerendes valg på den svenske database MySQL og program-

meringssproget PHP. Begge dele kan betegnes som særdeles integrerede med hinanden samt med den webserver, som har en markedsandel på over 70 %, og som desuden er et frit softwareprojekt: Apache.

Ud over at de på denne måde havde valgt en meget kendt platform, lykkedes det også de studerende at forenkle udviklingen af produktet, i og med at det nævnte sprog, databasen samt webserveren findes i en pakke, som er nem at installere til videre udvikling (WAMP/XXAMP).

3.4. Designet

Det er ikke kun i leksikografien, at ressourcerne til produktdesign typisk afmåles alt for knapt. Sådan skete det også i projektet lookedup4you. Designgruppen vakte i begyndelsen kun ringe opmærksomhed på projektmøderne og havde kun få medlemmer, som på grund af denne undervurdering heller ikke formåede at levere gode resultater. Først da konceptet og en prototype skulle præsenteres for kunden, blev det klart for projektgruppen, at produktdesignet er en væsentlig del af succesen, også den leksikografiske succes. Det blev hurtigt klart, at kunden ikke forstod strukturbegreberne eller den leksikografiske diskussion og heller ikke kunne bedømme et entity relationship-diagram vedrørende databasen. Derimod havde kunden en helt bestemt ide om produktets helhed, sammensat af struktur, indhold, teknik og også design.

Design og layout er dét, som opfattes før struktur og før indhold, og som dermed er en vigtig del af kundens og brugernes måde at danne sig en mening på (Spool 1999). Efter dette mindre chok fik projektgruppen mere forståelse for værdien af design og layout, og den tilsvarende arbejdsgruppe blev større og fik mere plads i de almene diskussioner. – Denne øgede bevågenhed viste sig dog at være problematisk. Da design og layout først var blevet et emne, man diskuterede, viste det sig, at næsten alle projektdeltagere havde en mening om design i det hele taget og frem for alt om de små detaljer i projektet.

Det krævede en yderligere justering i projektets forståelse af designprocesserne at forstå, at også dette arbejde skal fordeles til enkelte eksperter i gruppen, hvis arbejde man så har tiltro til og ikke behøver at diskutere hver uge på ny. Set fra projektplanlægningen var der på dette tidspunkt allerede gået mere end halvdelen af den samlede tid. – Til den sidste projektfase var der planlagt betatests samt en omfattende revurdering af produktet. Disse 2-3 uger skulle ligeledes trækkes fra den tid, som var tilovers til designarbejdet, som så kun bestod af ca. en tredjedel af den samlede tid.

Designet omfattede opgaver såsom screendesign, valg af former og farver samt konceptet til forsiden, som brugeren ser lige før makrostrukturen. En væsentlig del var også produktionen af et logo, som skulle markere opslagsværket som et selvstændigt produkt, men som – i stil med designet i det hele taget – skulle vise en

tydelig tilknytning til forbrugerportalen. Desuden var det designgruppens opgave at definere rammerne for udseendet af multimediale samt interaktive elementer i opslagsværket.

Design består dog ikke kun af et koncept, men i lige så høj grad af fremstillingen og brugen af designet i f.eks. opslagsværkets tilgangsstruktur (navigation etc.). – For at lette denne vej fra koncept til produkt blev der fra enkelte specialister på området givet korte introduktioner til særlige emner som f.eks. farvesystemer.

3.5. Leksikografisk struktur

Rammestrukturen som del af den leksikografiske informationsarkitektur (Rosenfeld/Morville 2002; Wodtke 2002) defineres til dels gennem selve mediet world wide web. – I selve navigationen, som er kernen af rammestrukturen, findes en systematisk makrostruktur, en søgemulighed, tilgang til hjælpetekster samt oplysninger om forfatterne. Desuden er der tilgang til en samlet oversigt over filmsekvenser samt quizzes. Disse elementer fra rammestrukturen samt de nævnte makrostrukturelle elementer er synlige fra alle sider i opslagsværket eller med andre ord fra hver artikel.

Undersøgelsen af fordelingsstrukturen viser at informationerne er fordelt på tre områder:

- Selve artiklen til det enkelte lemma
- Filmsekvenserne, som viser lemmainformationerne som sammenhæng
- Quizzer, som giver lemmainformationerne som en quiz.

Leksikografiske oplysninger fra film eller quiz er dog altid også tilgængelige som almindelig tekst. Man behøver dermed ikke se filmen eller færdiggøre en quiz for at få den leksikografiske information. Begge dele er derimod tænkt som supplement netop for målgruppen unge forbrugere. Fordelingsstrukturen er dermed om end ikke nyskabende så dog på et indtil videre kun lidt brugt niveau.

Makrostrukturen (Geeb/Spree 2004) foreligger i tre forskellige og ligeberettigede tilgangsmuligheder. Allerede på indgangssiden (home page) er der en systematisk tilgang til lemmalisten med 16 forskellige kategorier. Denne makrostruktur er ligeledes tilgængelig fra alle artikler i opslagsværket. – Ligeledes direkte på hjemmesiden – og på alle andre sider (artikler) – er der en søgemulighed, som kan bruges på to måder. Der er enten en fuldtekstsøgning i stil med de af brugerne kendte søgefaciliteter. Derudover er der mulighed for at indskrænke søgningen gennem valg af et fagområde og derefter bruge fuldtekstsøgningen. Selve fuldtekstsøgningen er dog ikke bare en almindelig fuldtekstsøgning, men de viste links rankes i et system i tre trin. Første prioritet og dermed den højeste placering har søgeresultater, hvis den søgte tekststreng er del af lemmaet, dets pluralformer eller synonymmer. Andet trin

på ranking-listen fås ved, at den søgte tekststreng er del af eller identisk med søgeord, som artiklerne har fået tildelt. Disse søgeord er ellers ikke synlige for brugere og har udelukkende deres formål i denne ranking. De er en yderligere del af artikelindholdets systematisering og blev udvalgt efter aftale i hele projektgruppen og fordelt med megen omhu. Tredje trin i ranking-systemet opnås ved en almindelig fuldtekstsøgning i artikelteksten. Da det tit giver mange resultater, skal dette søgeresultat vises som det sidste og mindst specifikke. – Søgningen resulterer altid i en linkliste, hvor de fundne lemmata er vist i den nævnte rækkefølge. Alle yderligere leksikografiske informationer findes ikke i denne liste, men i mikrostrukturen.

Den tredje del af makrostrukturen er tæt knyttet til de to foregående: den alfabetiske liste, som bliver synlig når brugeren vælger et fagområde. Alle alfabetets bogstaver er synlige, men kun de brugte bogstaver er et link til det tilsvarende sted i lemmalisten. Den makrostrukturelle vej gennem fagområderne ender dermed altid i en alfabetisk liste over fagområdets lemmata. Hvis et lemma har en film eller en quiz tilknyttet, vises det allerede i denne alfabetiske liste.

Mikrostrukturen var nok den mest omdiskuterede strukturelle enhed i opslagsværket. Her har brugerundersøgelsen haft en betydelig indflydelse. Lemmaet vises i fed og stor skrift. Derefter følger i en ny linje evt. sproglige oplysninger såsom forkortelser eller synonymmer samt oversættelser, hvis det drejer sig om f.eks. engelske ord. Formelle elementer såsom udtale, genus, ordklasse etc. blev ikke brugt, da brugere ikke var interesserede i denne oplysning og til dels ikke engang kunne tyde den, men blev forvirrede eller afskrækket fra at læse videre.

Det væsentligste element i mikrostrukturen er en kort definition på nogle få sætninger lige efter lemmaet og dets sproglige oplysninger. Denne korte definition, som delvis indeholder oplysninger, som også er indeholdt i den efterfølgende lange semantiske kommentar, blev efterlyst af brugere. Deres fremgangsmåde kan beskrives som en hurtig søgning med ønske om et hurtigt resultat uden nærmere interesse for, hvorfor og hvordan resultatet blev fundet frem. Derfor vil flertallet af brugere åbenbart kun læse den lange semantiske kommentar, hvis de inden da har fået en kort oplysning, som viser dem, at selve den lange kommentar er relevant.

Sproget og stilen i artiklerne blev tilpasset til brugergruppen. Hvor man ellers på tysk ville bruge Sie eller slet ikke bruge en personlig tiltale af brugeren, bliver her lagt vægt på at bruge du-formen. Selv den lange kommentar skulle opdeles i mindre enheder med korte overskrifter for at lette læsningen. Ungdomssprog skulle dog ikke bruges for at synliggøre den rådgivende og kompetente side af opslagsværket samt den bagved liggende forbrugerorganisation.

Da den semantiske kommentar tit indeholder juridiske eller i det mindste juridisk vejledende oplysninger, skulle alle tekster godkendes af en ekspert på området. Efter aftale med kunden blev derfor en jurist fra forbrugercentralen knyttet til projektgruppen. Hans opgave var at læse samt redigere teksterne ud fra juridisk fagtermino-

logi, men med hensyntagen til den aftalte sprogbrug og den aftalte sproglige stil.

Mediostrukturen dannes – ud over de i makro- og mikrostrukturen nævnte elementer – af to typer links: interne links samt eksterne links, som fører hen til eksterne sider. Disse eksterne links er altid forsynet med korte kommentarer. Hvis url'en ikke giver en tydelig oplysning om kildens navn og betydning, bliver den også forklaret. Interne samt eksterne links står altid sidst i artiklen for at opnå, at brugerne først læser hele artiklen og ikke bliver forstyrret af links midt i teksten. De to former for links er lavet på forskellig vis for dermed at vise deres forskellige mål. Interne links har i modsætning til eksterne links ingen videre forklaring ud over lemmaets tekststreng.

Links vil altid åbne et nyt, mindre vindue, hvis ikke lookedup4you allerede har to vinduer åbent. Dermed vil der aldrig være mere end to vinduer – i forskellig størrelse – åbne på samme tid.

Tilgangsstrukturen er i princippet kundens opgave. Hans opgave er at tilpasse opslagsværket til portalen efter overdragelsen og dermed sørge for en direkte tilgang. Derudover blev selve opslagsværket, som har fået et eget navn og et selvstændigt domæne (www.lookedup4you.de), registreret i forskellige søgemaskiner. Det viser sig i logfilerne samt gennem kommentarer, som findes i forskellige blogs eller fora på nettet, at siden også bruges en del uden om kundens portal.

3.6. *Multimedia og interaktive elementer*

Multimedia og interaktive elementer i opslagsværket er som beskrevet en vigtig del af formidlingen af indholdet. De er tilgængelige fra makro- og mikrostrukturen og har en vigtig placering i fordelingsstrukturen. Dog passer de ikke ind i beskrivelsen af leksikografisk tekst og må derfor beskrives særskilt.

Til enkelte lemmata som f.eks. *Praxisgebübr* har projektgruppen lavet korte film på et eller to minutter, der tilbydes som streaming i forskellig kvalitet. Streaming er en fordel for brugeren, men en ulempe for kunden, som dermed ved siden af webserveren også skal kunne tilbyde en streamingserver. Dog syntes det usandsynligt, at en bruger vil downloade og ikke streame en film og dermed blev det klart, at en særskilt, men alligevel gratis streamingsserver blev nødvendig.

Filmene er enkle og ikke professionelle, hverken i brugen af lys eller skuespillere. De er derimod direkte, forståelige og også sjove – målrettet mod målgruppen og alligevel inden for budgettet.

Quizzer blev valgt som et interaktivt element. Det viste sig gennem samtale med de fremtidige brugere, at det ikke er nok at tilbyde statisk tekst og måske billeder. Selv om der i denne sammenhæng nok ikke vil blive tale om ren underholdning, er enkelte elementer fra infotainment også interessante i leksikografiske produkter som dette. Der findes quizzer til udvalgte lemmata, og de kan dermed bruges til at afprøve, om man har forstået det læste. Men de kan også bruges som test inden

læsningen – hvor meget ved man om et emne? De er opbygget som multiple-choice-spørgeskemaer, hvor det rigtige svar altid vises efter besvarelse af et spørgsmål, og hvor man til sidst får en vurdering af det samlede resultat. Quizzerne er tilpasset designet og den sproglige stil. De skal ikke betragtes som en særskilt del, men som en del af det samlede leksikografiske koncept.

4. Usability-studier

Brugerne fik indflydelse på opslagsværkets koncept allerede i starten. Men de fik også lejlighed til at bedømme opslagsværket under udviklingen. Disse tests blev defineret ud fra grænsefladedesignet, men med hensyn til leksikografiske spørgsmål. Projektgruppen havde dog ikke tid til at bruge et usability-laboratorium med optagelse af brugeren på video og efterfølgende vurdering og diskussion af resultaterne. Derimod blev der brugt en heuristisk evaluering (Nielsen 1993; Nielsen et al. 1994) ud fra kendte usability-principper samt thinking aloud-tests: Udvalgte brugere får stillet en researchopgave, som de skal løse ved hjælp af opslagsværket og uden yderligere hjælp fra andre (Dumas et al. 1999). De fortæller dog hele tiden om deres valg af redskaber og veje gennem opslagsværket til testlederen, som skriver dette ned. Denne metode er nem at gennemføre og giver alligevel sammen med en heuristisk evaluering et godt billede af de usability-problemer, som opslagsværket har.

5. Lessons Learned

Meget synes at være åbenlyst, og alligevel fandtes der kun lidt konkret viden om emnet i opslagsværker på internettet. – På grundlag af de studerendes forskellige baggrund blev der skabt et produkt, som er et online-opslagsværk ud over det almindelige. Selv om leksikografien og dens teori ikke altid var hovedemnet i projektgruppens diskussioner, kunne næsten alle åbne spørgsmål besvares ud fra en leksikografisk synsvinkel, og løsningerne kunne udføres med leksikografiske midler. Dog ville projektet ikke have været muligt uden det brede faglige grundlag, som projektgruppen i sin helhed stod for.

Ud fra et leksikografisk perspektiv er de mest relevante resultater, som går ud over den leksikografiske praksis, flertallet af online-ordbøger står for:

- Kundens og brugernes indflydelse på projektets udformning giver opslagsværket et solidt fundament
- Brugerundersøgelser i starten af projektet og direkte knyttet til lemmaselektionen øger chancen for en leksikografisk succes betydeligt
- Brugerundersøgelser hen ad vejen med leksikografiske spørgsmål, men efterhånden også rettet mod opbygningen af grænsefladen (usability tests) understøtter denne proces
- Leksikografiske metoder og principper kan med fordel også bruges i andre

projekter med informationsformidling for kunder uden interesse i eller viden om leksikografi. I dette tilfælde må leksikografien ikke være et tema i præsentationerne, som afholdes for kunden.

- Leksikografiske principper kan hurtigt læres, mens et projekt kører, hvis bare projektdeltagerne kender til andre metoder til at strukturere data og information
- Leksikografisk succes er afhængig af produktets helhed. Dette gælder frem for alt for leksikografiske produkter, som skal konkurrere med mange andre informationsservices på internettet. Helheden dannes af indhold, struktur, design og teknik. Alle fire områder skal bearbejdes fra projektets start.
- Leksikografer behøver ikke at være designere, programmører eller journalister for at skabe et godt leksikografisk produkt. Men i en leksikografisk arbejdsgruppe med internettet som mål burde viden om design, programmering og datamodellering samt om at skrive tekster til internettet være en naturlig del. Online-opslagsværker, som bliver produceret uden denne viden, kan i længden nok kun tilfredsstille brugernes behov og ønsker, hvis de har et ualmindelig godt indhold.
- Der er stadigvæk – også efter mere end 10 år med grafisk internet – mulighed for at være nyskabende i delområder på det online-leksikografiske område
- Selv om afgørelser principielt med fordel kan træffes af hele projektgruppen, skal der også være tiltro til de enkelte eksperter vurdering og afgørelser inden for delgrupperne. Dette gælder ikke mindst tekniske detaljløsninger eller afgørelser vedrørende designet.

Og endelig har underviserne lært eller fået bekræftet, at leksikografi i en universitetsuddannelse ikke bare er en lingvistisk disciplin, men frem for alt en meget gammel og afprøvet metode til strukturering af data og informationer, som på trods af dens alder er særdeles velegnet til internetprojekter. Moderne opslagsværker og informationsprodukter, som svarer til kundens ønsker og brugernes behov, kan meget vel skabes med leksikografiske metoder og viden.

Litteratur

- Checked 2005: Das Online-Jugendmagazin der Verbraucherzentrale Nordrhein-Westfalen. <http://www.checked4you.de> (28.8.2005).
- DMI 2005: Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Design, Medien und Information, Department Information. <http://www.bui.haw-hamburg.de> (28.8.2005).
- DMI 2005a: Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Design, Medien und Information, Department Information: Projekte. <http://allekto.bui.haw-hamburg.de/ueberuns/studienprojekte.php> (28.8.2005).

- Dumas, Joseph S./Redish, Janice C. 1999: *A Practical Guide to Usability Testing*. – 2. Aufl. Exeter: Intellect.
- Geeb, Franziskus 2001: leXeML – Vorschlag und Diskussion einer metalexikographischen Auszeichnungssprache. I: *Sprache und Datenverarbeitung. International Journal for Language Data Processing*, 27-61.
- Geeb, Franziskus 2003: Lexikographische Informationsstrukturierung mit XML. I: *nfd 7/2003*, 415-421.
- Geeb, Franziskus 2005: Perspektiven der Rechneranwendung in der Lexikographie. I: Henrik Gottlieb, Jens Erik Mogensen and Arne Zettersten (udg.): *Symposium on Lexicography XI: Proceedings of the Eleventh International Symposium on Lexicography May 4-6, 2002 at the University of Copenhagen*. Tübingen, 227-239.
- Geeb, Franziskus/Spree, Ulrike 2004: Wörterbücher und Enzyklopädien. I: Kuhlen, Rainer/Seeger, Thomas/Strauch, Dietmar: *Grundlagen der praktischen Information und Dokumentation. Bd. 1: Handbuch zur Einführung in die Informationswissenschaft und Praxis*. München, 481-493.
- Geeb, Franziskus/Spree, Ulrike 2005: lookedup4you – Mikrostruktur und Makrostruktur und all das. Produktion eines Online-Nachschlagewerks als studentisches Projekt. Ein Erfahrungsbericht. I: *nfd 3/2005*. 133-142.
- HAW 2005: Hochschule für Angewandte Wissenschaften Hamburg. <http://www.haw-hamburg.de> (28.8.2005).
- Lookedup 2005: Lookedup4you. Lexikon für junge Verbraucher. <http://www.lookedup4you.de> (28.8.2005).
- Nielsen, Jakob 1993: *Usability Engineering*. – San Diego: Academic Press.
- Nielsen, Jakob/Mack, Robert L. 1994: *Usability Inspection Methods*. New York: Wiley & Sons, Inc.
- Rosenfeld, Louis/Morville, Peter 2002: *Information Architecture for the World Wide Web*. 2. oplag. Beijing: O'Reilly.
- Spool, Jarred M. 1999: *Web Site Usability: A Designer's Guide*. – San Francisco: Morgan Kaufman Publishers.
- VZ 2005: Verbraucherzentrale Nordrhein-Westfalen. <http://www.vz-nrw.de> (28.8.2005).
- Wodtke, Cristina 2002: *Information Architecture: Blueprints for the Web*. – 1. opl. – Boston: New Riders.

(Hur) använder du ordböcker? En undersökning bland svensklärarstuderande och svensklärare

(How) do you use dictionaries? A survey among teacher students and teachers of Swedish. This article is based on an essay written during the autumn semester in 2004 at Uppsala University. The essay has a lexicographic approach and the purpose is to examine the use of dictionaries among the category of student teachers and teachers of Swedish. Age and education were presumed to affect their use. A questionnaire was handed out to the informants. The most important results show that the majority of the informants, especially teachers, use dictionaries, but almost exclusively printed editions. Although the informants claim to look up word explanations for the most part, it is still dictionaries that focus on grammar that are actually used. More than half of the informants said they had not had any form of education in dictionary use, but experienced a positive attitude towards the subject. An interesting question is why they have not been taught to use dictionaries and what the attitude of other kinds of teachers, for example teachers of Swedish as a second language education, is.

1. Inledning

Använder du ordböcker? Och i så fall hur? Det är två frågor som öppnar en bred kunskapslucka inom svensk språkforskning. Antingen ordboken finns till hands i tryckt eller idag kanske oftare i elektroniskt format, så finns det oftast tillgång till ett flertal olika ordböcker både i hemmet och på arbetsplatsen. Däremot är frågan om dessa språkverktyg faktiskt används och inte bara står i hyllan som obligatoriska böcker utan att tjäna sitt syfte. Idén att undersöka hur saken kan förhålla sig inom en kategori ordboksanvändare fick jag genom Svenska språknämnden och den rubriklista som de utarbetat för ämnesval till uppsatser. Rubriken finns under ämnet lexikografi, ett område som jag insåg sällan bidrar med ny kunskap om ordboksanvändning och då minst av allt ur ett användarperspektiv. Inte heller har någon större undersökning på svenskt håll påbörjats, vilket i sin tur ökade mitt intresse och min motivation för det valda uppsatsämnet. Föreliggande artikel baserar sig på min uppsats och jag har här valt att presentera mitt tillvägagångssätt samt de viktigaste resultaten av undersökningen.

1.1. Syfte

Jag vill undersöka hur svensklärare och svensklärarstuderande förhåller sig till och använder sig av enspråkigt svenska ordböcker och liknande hjälpmedel i textbearbetning på svenska. De är viktiga språkförebilder för elever i olika åldrar och jag vill därför i första hand ta reda på och beskriva hur den förebilden kan se ut vad gäller lärarnas användning av ordböcker. I andra hand utgår jag från att ålder och utbildningsbakgrund utgör påverkande faktorer på bruket av ordböcker. Min hypotes är

att yngre användare främst skulle nyttja elektroniska alternativ och då slå upp grammatiskt inriktade problem. Äldre användare skulle använda tryckta ordböcker och i sin tur slå upp moderna språkaspekter, exempelvis nyord och slang.

1.2. Material och metod

Enkäterna delades i första hand ut bland svensklärare i en medelstor mellansvensk stad med en så stor spridning som möjligt enligt undervisningsnivå. Av dessa svarade endast 13 stycken gymnasielärare, varför även enkäter delades ut till 20 stycken universitetsanställda svensklärare. För att ytterligare komplettera materialet valde jag också ut lärarstudierande som informantgrupp. Samtliga av dessa, 45 stycken gick vid undersökningstillfället – hösten 2004 – lärarprogrammet. 18 stycken studenter läste första terminen vid en institution för nordiska språk och hade undervisning i den första kursen i svenska, Svenska 1. De övriga, 27 stycken, läste tredje terminen och hade undervisning i kursen Svenska 3. Informantgrupperna benämner jag i det följande och i uppsatsen i övrigt för "Svenska 1", "Svenska 3" samt "Lärare".

2. Resultat

Jag redogör i det följande för mina viktigaste resultat eftersom en genomgång av alla enkätfrågor inte ryms inom ramarna för denna artikel. Till redogörelsen fogar jag relevanta statistikuppgifter i tabellformat för att på så vis tydliggöra resultaten.

Grupp	<u>Använder</u>		<u>Använder inte</u>		<u>Totalt</u>	
	Antal	%	Antal	%	Antal	%
Svenska 1 (n= 18)	8	44	10	56	18	23
Svenska 3 (n= 27)	20	74	7	26	27	35
Lärare (n= 33)	31	94	2	6	33	42
Totalt (n= 78)	59	76	19	24	78	100

Tabell 1. *Användarfrekvens av ordböcker*

Som tabell 1 visar används ordböcker, men behöver inte vara en självklarhet för alla. Användningen ökar ju längre hunden man är i sin utbildning, därmed också ju äldre man blir. En förklaring till detta kunde vara att kravet på den egna textbearbetningen stiger i takt med utbildningsgraden och därmed ökar även medvetenheten om det egna språket eftersom man tvingas att fundera på ordval, formuleringar och struktur. Då blir ordboken en nyttig och ofta ofrånkomlig tillgång. Att knappt någon av lärarna inte använder ordböcker skulle i så fall bekräfta resonemanget.

Ett andra viktigt resultat som undersökningen gav kullkastade min givna hypotes. Samtliga informanter, oberoende ålder, tycks föredra tryckta ordböcker framom elektroniska varianter, d.v.s. de som finns inlagda i datorer eller publicerade på Internet. Ytterst få hade markerat för dessa motsvarigheter, varför jag valde att fortsättningsvis koncentrera min resultatgenomgång och diskussion kring tryckta ordböcker. Här bör tilläggas att jag i efterhand insåg att listan på valda ordböcker i enkäten var angivna på ett ostrukturerat sätt, eftersom de skiljer sig åt innehålls- och målgruppsmässigt. Av slutresultatet kan ändå en intressant tendens utläsas, varför jag anser att listan i detta sammanhang är godtagbar. Som en förklaring till tidigare nämnda resultat kunde vara att man inte i första hand tänker på datorns ordböcker utan den fysiska som står i bokhyllan och därför också hellre använder den. Tryckta ordböcker är dessutom lättillgängligare för gemene man.

Vad som slås upp i ordböcker skiljer sig inte åt mellan informantgrupperna. Problemsökningen är således genomgående densamma. Det intressanta i sammanhanget ger ett tredje viktigt resultat i undersökningen vilket tydliggörs av tabell 2.

Uppslag	Svenska 1 (n= 18)		Svenska 3 (n= 27)		Lärare (n= 33)		Totalt (n= 78)	
	Antal	%	Antal	%	Antal	%	Antal	%
Ordförklaringar	17	94	26	96	29	88	72	92
Stavning	15	83	19	70	28	85	62	79
Synonymer	12	66	19	70	27	82	58	74
Ordklass- beteckning	11	61	13	48	17	52	41	53
Bestämd/ obestämd form	11	61	13	48	13	39	37	47
Avstavning	10	56	13	48	13	39	36	46
Motsatser	9	50	14	52	11	33	34	44
Annat (etymologi, konstruktioner)	1	6	6	22	7	21	14	18

Tabell 2. Vad som slås upp i ordböcker

Vad som slås upp går inte i hand med vilka ordböcker som används. Informanterna sade sig nämligen främst slå upp ordförklaringar, men istället är det mer grammatiskt inriktade ordböcker som faktiskt används. En förklaring till detta kunde bero på antaget och faktiskt beteende. Man tror sig använda ordböcker i ordförklarings syfte ("det som ordböcker är till för", som en informant uttryckte det), men i stället är det annat man letar efter, t. ex. stavning eller ordklassbeteckning.

Det fjärde resultatet är ett i mitt tycke det mest intressanta. Det kretsar kring frågan om eventuell undervisning i ordbokskunskap. Många har en ordbok hemma eller på sin arbetsplats, men hur många har egentligen tillräcklig bakgrundskunskap för att bäst utnyttja det innehåll ordboken har? Informationsupplysningen skiljer sig avsevärt från andra slags uppslagsböcker, exempelvis encyklopedier eller den mer vardagliga telefonkatalogen, och jag tror att många ordboksägare går miste om mycket av den information som finns efter ett uppslagsord, d.v.s. i en ordboksartikel. En sådan kunskap anser jag bör inhämtas i skolan, därför är det märkligt att en majoritet av mina informanter, 60 % så som tabell 3 visar, säger sig aldrig ha fått tidigare undervisning i ordbokskunskap. Detta kan visserligen förklaras med att de tillfrågade inte har någon minnesuppfattning om sådana inslag, att ordbokskunskap kanske förekom i någon form i svenskundervisningen, men i så fall inte särskilt djupgående. Mina resultat kan dock bekräftas av en annan undersökning gjord 1998 av B. T. Sue Atkins och Krista Varantola (1998:21-81), där även 60 % av deras informanter uppgav samma svar om frånvaro av undervisning eller instruktioner i användning av ordböcker. Detta väcker för mig en rad nyfikna och intressanta frågor som en framtida forskning kanske kan ge svar på. Vad orsakar detta undervisningsbortfall och kanske ännu viktigare, hur råda bot på det på bästa sätt?

Grupp	Med delaktighet		Utan delaktighet		Totalt	
	Antal	%	Antal	%	Antal	%
Svenska 1 (n= 18)	7	39	11	61	18	23
Svenska 3 (n= 27)	7	26	20	74	27	35
Lärare (n= 33)	19	58	14	42	33	42
Totalt (n= 78)	33	42	45	58	78	100

Tabell 3. Antalet informanter med eller utan tidigare delaktighet i ordboksundervisning

Sammanfattningsvis verkar den allmänna bilden av läraren som en tillförlitlig språkauktoritet vara mycket positiv. Ordböcker används och relativt aktivt dessutom. Lärarna hyser stort självförtroende för den egna språkkänslan vilket är positivt inom undervisningssituationer. Eleverna kan gott vända sig till sin lärare för att få ett säkert och korrekt svar, i annat fall hänvisar läraren till (tryckta) ordböcker!

Litteratur

Atkins, B. T. Sue & Varantola, Krista 1998: Language Learners Using Dictionaries: The Report on the EURALEX/ AILA Research Project on Dictionary Use. I: *Lexicographica. Series Maior 88*. Tübingen: Max Niemeyer Verlag, 21-81.

Enkät om ordboksanvändning

1. Kön

- Man Kvinna

2. Ålder 20 – 35 36 – 50 51 – 65+

3. Vilken (vilka) nivå undervisar du på?

- Låg/mellanstadium
 Högstadium
 Gymnasium
 Universitet/högskola

4. Vilken utbildning har du? Kryssa för lämpligt alternativ.

- Folkskola
 Yrkesskola
 Grundskola
 2-årigt gymnasium
 3-4 årigt gymnasium
 Universitet/högskola (inriktning, examen): _____
 Annan utbildning, nämligen: _____
Denna utbildnings längd: _____

5. Hur länge har du varit yrkesverksam lärare? _____

Varav hur lång tid anställd på nuvarande arbetsplats? _____

6. a Hur många kortare/längre texter skriver du och/eller bearbetar i genomsnitt per dag?

Räkna in både elevtexter och eget skrivande.

- Inga, men ____ i veckan
 1
 2-3
 4-5
 Annat antal, nämligen: _____

6. b Hur många av dessa är elevtexter respektive egna texter?

___ elevtexter ___ egna texter

7. a Använder du svenska ordböcker under din textbearbetning?
 Ja Nej (gå vidare till fråga 7. d)
7. b Om ja, vilka? Sätt ett P för de ordböcker du använder i pappersformat och ett E för de du använder i elektroniskt format, samt P/E om du använder både och.

- | | |
|---|--|
| <input type="checkbox"/> SAOL | <input type="checkbox"/> Etymologisk ordbok |
| <input type="checkbox"/> NEO | <input type="checkbox"/> TNC |
| <input type="checkbox"/> Synonymordbok | <input type="checkbox"/> Ny svengelsk ordbok |
| <input type="checkbox"/> Svenskt avstavningslexikon | <input type="checkbox"/> Nyordsboken |
| <input type="checkbox"/> Antonymordbok | <input type="checkbox"/> Konstruktionsordbok |
| <input type="checkbox"/> Slangordbok | <input type="checkbox"/> Baklängesordbok |
| <input type="checkbox"/> Främmande ord | <input type="checkbox"/> Förkortningsordbok |
| <input type="checkbox"/> Svenskt språkbruk | <input type="checkbox"/> Stora fula ordboken |
| <input type="checkbox"/> Oordboken | <input type="checkbox"/> Svenska skrivregler |

7. c Andra, vilka? (T. ex. utländska) _____

7. d Om nej, varför inte? _____

7. e Använder du funktioner i ordbehandlingsprogram istället för ordböcker? I så fall, vilka?

8. Kände du till att alla de ovannämnda ordböckerna fanns? I så fall, vilka? (Oavsett om du använder dem eller inte.)

- | | |
|---|--|
| <input type="checkbox"/> SAOL | <input type="checkbox"/> Etymologisk ordbok |
| <input type="checkbox"/> NEO | <input type="checkbox"/> TNC |
| <input type="checkbox"/> Synonymordbok | <input type="checkbox"/> Ny svengelsk ordbok |
| <input type="checkbox"/> Svenskt avstavningslexikon | <input type="checkbox"/> Nyordsboken |
| <input type="checkbox"/> Antonymordbok | <input type="checkbox"/> Konstruktionsordbok |
| <input type="checkbox"/> Slangordbok | <input type="checkbox"/> Baklängesordbok |
| <input type="checkbox"/> Främmande ord | <input type="checkbox"/> Förkortningsordbok |
| <input type="checkbox"/> Svenskt språkbruk | <input type="checkbox"/> Stora fula ordboken |
| <input type="checkbox"/> Oordboken | <input type="checkbox"/> Svenska skrivregler |

9. Hur ofta använder du ordböcker? Sätt ett P för pappersformat, E för elektroniskt format samt P/E för både och.

- Flera ggr dagligen
- Någon gång per dag
- Flera ggr i veckan
- Någon gång i veckan
- Flera ggr i månaden
- Någon gång i månaden
- Mycket sällan/aldrig

10. Vad slår du upp i ordböcker? Rangordna från mest uppslag (1) till minst uppslag (8).

- Ordförklaringar

(HUR) ANVÄNDER DU ORDBÖCKER?

- Synonymer
- Stavning
- Avstavning
- Ordklassbeteckning
- Bestämd/obestämd form
- Motsatser
- Annat, vad? _____

11. a Hittar du alltid svaret vid sökningen?

- Ja Nej

11. b Om inte, vad gör du då?

- Rådfrågar kollega
- Rådfrågar intern språkexpert
- Rådfrågar extern språkexpert (T. ex. Svenska språknämnden)
- Använder Internet
- Litar på egen språkkänsla
- Annat, vad? _____

12. a Förstår du all information i ordboken?

- Ja Nej

12. b Om inte, vad är det som kan skapa problem? (T.ex. uttalsbeteckningar, otillräcklig information, okunskap) _____

13. a Om flera alternativ finns vid ordförklaringar, hur gör du då?

- Väljer det första möjliga, utan att se till resten
- Prövar alla alternativ, men väljer det första ändå
- Prövar alla alternativ och väljer det som känns bäst
- Annat, vad? _____

13. b Läser du hela uppslagsordet, d. v. s. all information som står?

- Ja, allt
- Nej, bara början
- Nej, bara början, men skummar resten till slutet
- Annat, vad? _____

13. c Vad avgör om du läser allt eller inte?

- Längden på stycket
- Tiden
- Intresset
- Annat, vad? _____

14. a Hur upplever du det egna ordboksanvändandet? Kryssa för flera alternativ vid behov.

- Tidsödande (stressfaktor)
- Lärorikt
- Nödvändigt ont
- Roligt

(Inlörd) vana

Annat, vad? _____

14. b Har du någon gång tagit del i någon undervisning i ordböksanvändning?

Ja Nej

14. c Om ja, hur och var skedde det? _____

14. d Om nej, skulle du vilja ta del i det och lära dig mer? Motivera gärna.

14. e Ger du själv undervisning i ordböksanvändning?

Ja Nej

14. f Om ja, hur stor vikt lägger du vid den? Motivera gärna. _____

15. a Håller du dig uppdaterad med de senaste utgåvorna av ordböcker?

Ja Nej

15. b Om nej, varför inte? _____

16. a Vad anser du om ordböcker som hjälpmedel i allmänhet? Motivera ditt svar.

16. b Tycker du något borde förbättras, läggas till eller omstruktureras vad gäller ordböcker?

16. c Vilka ordböcker är bra/dåliga? Saknas någon? Motivera ditt svar.

17. Övriga synpunkter.

Verknader av digitalisering på materialvurdering, redaksjonell metode og opplæring

Effects of digitisation on the use of source materials, editorial method and editorial training. Digitisation of Norway's largest academic dictionary project, Norsk Ordbok 2014, through an interlinked set of relational databases has led to new ways of using source materials in lemma selection, manuscript management and quality control. Editorial method is radically changed by increased transparency, both in looking at source materials and in handling dictionary and entry formats. These changes are prerequisites for training linguists to become productive editors within a minimal time span, with six months for basic training and full editorial responsibility after one year. The aim of the exercise is to complete seven volumes of Norsk Ordbok in nine years, with no loss of, and preferably an increase in, quality.

1. Innleiing

Norsk Ordbok (NO) er eit vitenskapleg ordboksverk over nynorsk skriftmål og norsk talemål, der forfattarskap til ordartiklar er akademisk meriterande for kvar einiskild redaktør. Redigeringsstad er Universitetet i Oslo, der redaksjonen er organisert som prosjekt under Institutt for lingvistik og nordistikk. Digitaliseringa av NO 2014 skjer i tett samarbeid med Eining for digital dokumentasjon ved Historisk-Filosofisk fakultet.

I perioden 2003-2014 er ordboksprosjektet refinansiert og omorganisert med det føremålet at alle tolv band skal vere utgjevne ved utlaupe av 2014, som del av 200-årsjubileet for de norske Grunnlova. Verket skal (1) vere konsistent i materialutnytting og redigeringspraksis, (2) halde høg vitenskapleg standard når det gjeld lingvistisk kategorisering og innhald, og (3) vere dokumentert ned til detaljnivå, slik at alle opplysningar lett kan etterprøvast.

Dette er bakteppet for det store hamskiftet i Norsk Ordbok, som er full digitalisering av heile prosjektet, kombinert med rekruttering og opplæring av eit tjuetals nye redaktørar. Den følgjande drøftinga tek for seg verknader av digitalisering på materialvurdering, redaksjonell metode og opplæring, sett frå redaktørsynsstad og frå leiinga si side.

2. Materialvurdering

Konsistent materialvurdering i høve til verket som heilskap er vesentleg mellom anna for lemmaseleksjon og omfangsdimensjonering. Første oppgåve for prosjektet var derfor å skaffe seg oversyn over underlagsmaterialet, både i høve til lemma (mengd, type; normaliseringsgrad) og kjelder.

2.1. Underlagsmateriale for Norsk Ordbok

NO er ei dokumentasjonsordbok som opphavleg vart bygd opp på

1. nedskrivne talemålsopplysningar (ikkje lyd)
2. litteraturekserpt

Samlingane, som Norsk Ordbok byggjer på (i hovudsak skapte før prosjektstart i 2001) er svært ueinsarta. Eit slående drag ved dei, som samlingar, er mangelen på retningslinjer for oppbygginga, og den svært ujamne utnyttinga av ulike kjelder. Dette heng saman med den ujamne sjangerdekninga i nynorsk skriftmål – noko som er typisk for minoritetsspråk – og den like ujamne dokumentasjonen av norske målføre. I dei eldre samlingane har ein måtta bruke alt tilgjengeleg tilfang som var av verdi (Grønvik 1997:33).

Det gjeld generelt for ekserperte ordsamlingar at dei har underdekning av frekvent ordtilfang, og av ordtilfang frå nyare kjelder. Det galdt også dei nynorske ordsamlingane. NO 2014 har derfor bygd opp eit supplerande korpus av nynorsk tekst som i 2005 er på om lag 23 mill. ord. Om lag 80 % av tekstmengda er frå etter 1975. Hovudvekta ligg på sakprosa, men der er også skjønnlitteratur og ein del tekstar frå klassisk nynorsk, mellom anna den første fullstendige nynorske bibelutgåva.¹

Korpuset er særskilt nyttig til å gje betre og meir nyansert dekning av det frekvente ordtilfanget. Lemma frå korpuset som ikkje finst i samlingane, er for det meste samansetningar – utan at dei derfor er uinteressante. Korpuset gjev dessutan høve til å utnytte ein del av dei eldre litterære kjeldane på ein mykje grundigare måte, i og med det gjev tilgjenge til heile tekstar, ikkje berre ekserpt.

2.2. Materialoversyn – Metaordboka

Det manglar førebels oversyn over kva for tilfang som er utnytta, og korleis, i den delen av Norsk Ordbok som var redigert på gamlemåten (band 1-4, utgjevne 1965-2002, og band 5 harm-håvøyll). Fullt oversyn over dette får vi ikkje før band 1-4 er ferdig digitaliserte.² Men den gamle redaksjonen var medviten om at samlingane var ujamne og til dels tynne. Parolen var derfor å utnytte tilfanget til siste detalj. Bruksprinsippet var (noko forenkla) at alt av verdi skulle inn i ordboka. Mal for omfangsrekning av manus mangla. Resultatet vart redaktøravhengig, og dermed inkonsistent, materialutnytting og redigering.³ I tillegg mista ein kontroll over totalvolumet av manus (Bø 1989:90).

For å skaffe oversyn over grunnlagstilfanget vart det i perioden 2000-2003 bygt

¹ For korpusstatistikk sjå <http://folk.uio.no/danielr/no2014/nn/dist25.html>.

² Planlagt ferdig i 2006.

³ Som døme kan det nemnast at materialutnyttinga for verba *gjera*, *gjeva*, *ha*, *halda*, *gå* varierer frå om lag 120 til om lag 280 belegg per prenta side.

opp eit sams indeks for dei nynorske språksamlingane, den såkalla Metaordboka (MO), der tilfanget vart gjennom-normalisert ein gong til (om dette sjå Svardal 2003). MO synte at språksamlingane har om lag 550 000 oppslagsord. Om lag halvparten av desse oppslagsorda har eitt einaste belegg, men det belegget kunne vere alt frå ein etter måten stor artikkel i Grunnmanuskriptet⁴ for NO til eit einstaka litterært ekserpt. Lemmatal kunne derfor ikkje vere einaste utvalskriterium.

2.3. Omfang og styringsbehov

NO 2014 band 1-4 har om lag 100 000 artiklar. Plassomsyn og omsyn til kvaliteten på redigeringa tilseier at eitt band (på 800 sider) ikkje bør ha meir enn 25 000 ordartiklar. Dette gjev eit totalanslag på 300 000 artiklar fordelt på tolv band.

NO 2014 trong derfor retningslinjer for utsjalting av uaktuelle oppslagsord ut frå mengd og kvalitet på det underliggjande tilfanget. Vidare trongst det ein mal for artikkelstorleik innanfor totalramma på tolv band. Både retningslinjer for ordutval og dimensjoneringsmal måtte kunne automatiserast gjennom redigeringsprogrammet, men det måtte også vere mogleg for redaktørane å overprøve resultatet dersom utslaget vart vurdert som fagleg urimeleg.

2.4. Materialvurdering som grunnlag for lemmaseleksjon

Lemmaseleksjon i eit verk som NO må dels byggje på generelle omsyn, dels på omsyn som er spesifikke for verket. Generelle omsyn tilseier at grunnord får grundigare behandling enn avleingar og samansetningar, og at det nordiske ordtilfanget får ei grundigare handsaming enn importord (Venås 1989:162 f). Dei verksspesifikke omsyna har å gjera med nynorskens historie, status og tilblivingsgrunnlag, der relasjonen til norske målføre er sentral. Nynorsk er eit ungt skriftmål som ikkje kan seiast å vere fullt ut standardisert.⁵ Gjennom arbeidet med NO kjem redaktørane stadig over ordformer i målføra som ikkje er bestemte. NO har derfor eit særleg ansvar for å behandle eldre ordtilfang og unormert målføretilfang, og gje relativt større plass til mindre frekvente norske grunnord som ein ikkje finn i vanlege bruksordbøker.

Lemmaseleksjonen vart frå 2003 knytt til ei validering i kjeldegrunnlaget for NO. Vi hadde alt ein bibliografidatabase over alle skriftlege kjelder. Her var det høve til å leggje inn tilleggsopplysningar som i sin tur kunne nyttast i lemmaseleksjonen. Alle skriftlege kjelder for NO 2014 vart derfor sjangermerkte i bibliografidatabasen. Sjangermerkinga, kombinert med tal belegg, blir så nytta til å skilja mellom tre lemmagrupper, (1) lemma som skal redigerast, (2) lemma som kan redigerast, (3) lemma som er frårådde.

⁴ Sjå omtale på <http://no2014.uio.no/tekster/sok/index.php>.

⁵ Jf. det omfattande systemet med parallellformer og tillatne sideformer (klammeformer) i nynorsk rettskriving.

Obligatoriske lemma er (1) Lemma frå Grunnmanuskriptet, Skards rettskrivingsordliste og *Nynorskordboka*, og (2) Lemma frå målføretilfanget som ikkje er med i andre normerte ordbøker.

Dei viktigaste gruppene av frårådde oppslagsord er (1) lemma med tospråklege ordbøker, skuleordlister eller spesialordbøker (frå slang til yrkesterminologi) som einaste kjelder ("ordboksord"), og (2) lemma som er hapax frå skjønnlitterære omsetjingar eller som framstår som tilfeldige (skjønnlitterære) spontanlagingar. I den siste gruppa er det svært mange samansette ord.

Redaktøren kan overprøve automatikken ved å leggje til nytt tilfang, til dømes frå korpuset.

2.5. Dimensjonering av band og artiklar

Etter at sjølve redigeringskjemaet med utskriftfunksjon var på plass i 2004, var det dei redigeringsadministrative funksjonane som stod høgst på prioriteringslista. Denne programmodulen er ikkje komplett enno, men eitt av dei viktigaste elementa, dimensjoneringsmål pr. artikkel, er på plass.

Det tilfanget som må redigerast, er ordna gjennom MO, og dette tilfanget er gjennomarbeidd og normalisert.⁶ Derfor er det MO som er lagt til grunn for dimensjoneringa av dei attståande banda av NO. Dimensjoneringa for einskildartiklar byggjer på ei manuell utrekning av setelprogresjon og artikkeltypar i dei ferdigredigerte banda, som så vart brukt til å lage ein utrekningmodell for programmet, med vektning for beleggmenge pr. artikkel. Modellen gjev noko større plass til lemma med få belegg attom seg, enn lemma som byggjer på eit større tilfang. Modellen vart testa og justert, og er no teken i bruk. Kvar gong ein artikkel blir generert opp, kjem det fram eit framlegg om manusslengd i tal linjer for denne artikkelen.

Dei resterande 6 banda av NO er også rekna gjennom i bolkar på 20 sider, slik at ein har første og siste lemma pr. tjuesiders bolk frå og med band 6 og ut band 12.

Redaktørane får tildelt manus til redigering i alfabetbolkar på 3-4 bokstavar (t.d. *kly-*, *kok-*). Programmet kan enno ikkje gje eit linjetalet for ein alfabetbolk, så det må redaktørane rekne ut sjølve ved å summere linjetalet for artiklar. Innanfor linjetalet for ein alfabetbolk kan redaktøren omdisponere artikkellengder – leggje til litt her og knipe litt der – men det er ikkje høve til å gå utover manusslengd for den tildelte alfabetbolken.

Dimensjoneringmodulen kom inn i redigeringsprogrammet ved påsketiden 2005, og verknaden vart synleg med ein gong. Før dette hadde mange redaktørar store problem med å halde omfangsgrensene, med påfølgjande lange retterunder og omskrivingar. Etter at dimensjoneringsmodulen kom inn, held dei fleste linjetalet sitt i første omgang, og dei kjem dermed fortare fram til leveringsklart manus.

⁶ MO i-å vart normalisert i 2000-2003. Bokstavane a-d er no normaliserte, medan normalisering pågår for bokstavane e-g. I gjennomsnitt blir tal oppslagsord pr. bokstav redusert med 15 %.

3. Redaksjonell metode

Medan redaktørane før produserte ordboksmanuskript som seriell tekst, skal dei no tilordne tilfang og skrive inn i felt med førehandsdefinerte rammer. Kva følgjer har dette hatt for arbeidsmetodar, produksjon og kvalitetskontroll?

3.1. Redigering i Norsk Ordbok før prosjektstart

Produksjonen av ordboksmanus har tre komponentar, uavhengig av hjelperåder. Den første, og morosamaste, er den som er rein filologi og lingvistik, som omfattar vurdering og ordning av tilfang, utforming av definisjonar, utveljing og kommentering av døme osv. Den andre fasen er utforming og nedskriving av manus, med kjelder, kryssreferansar, spesialteikn og konvensjonar – og dei er det mange av i eit akademisk ordboksverk. Den tredje fasen er klargjering av manus for produksjon, med rett sats og typografi osv. Ein stad imellom fase 2 og 3 ligg ansvaret for at manuskriptet er feilfritt, både tekstleg og reelt (kontroll av tilvisingar osv.).

Den tradisjonelle arbeidsdelinga i mange akademiske ordboksredaksjonar har vore at filologar har ansvar for den første fasen, medan ymse hjelpepersonell har teke seg av mest mogleg innanfor fase 2 og 3. For tjue år sidan vart NO-manus dels skrive for hand, dels på pc, av redaktørane. Alt den gongen vart manuskriptet feltinndelt (etter ei blanding av typografiske og leksikografiske kriterium) med kodar før setjing (sjå t.d. Hjulstad 1986; Worren 1992). Men det vart gjort av ein sekretær, og det var berre ho som kunne kodesystemet. Redaktørane var mest opptekne av at typografien vart rett i korrekturen, og ville neppe oppdage innhaldsmessig feilkoding dersom det ikkje vart avslørt av feil i typografien.

Dei fleste ordboksverk har hjelpedokument til støtte for redigeringa. For NOs del er det vel rett å seie at hjelpedokumenta før 2003 var nokså einsidig innretta mot akribien i verket, dvs. tekstkonvensjonar, pluss kjeldekunnskap og -kontroll.

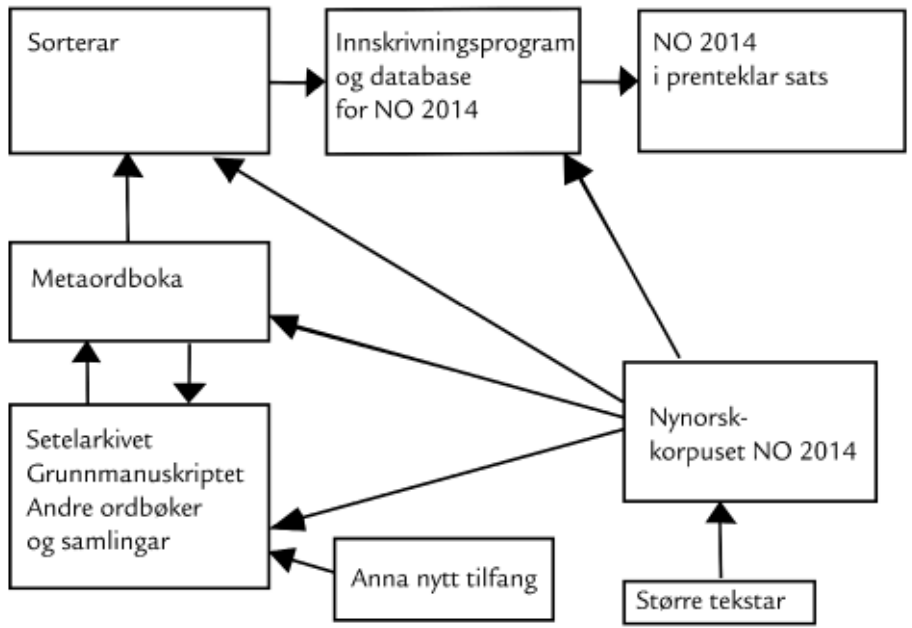
Dette heng dels saman med kva det er mogleg å kontrollere manuelt i ein flat tekst som har mange forfattarar. Det skal også seiast at NO har uvanleg krevjande konvensjonar, som skal ta høgde for både tale og skrift, og både synkron og diakron formvariasjon. Redaksjonelt fokus og redaksjonell metode kom derfor til å bli prega av behovet for korrekt handtering av ordboksakribien, medan dei lingvistiske kriteria nok var føresette, men kom meir i bakgrunnen.

3.2. Krav til redigeringsprogrammet for NO 2014

For NO 2014 var det eit overordna mål at innskrivingsprogrammet så langt mogleg skulle fristille redaktørane til å samle seg om det lingvistiske og leksikografiske i arbeidet, medan konvensjonane skulle handterast av programmet. I høve til kjedeføring og tilvisingar var det også ønskeleg å ikkje berre yte hjelp til effektiv redigering, men også førebyggje og hindre feil.

Både behovet for datatryggleik og behovet for kontroll tilsa at ordboka skulle

skrivast inn i ein relasjonell database, feltinndelt for ulike opplysningskategoriar, og at heile dokumentasjonssystemet rundt ordboka (som alt var lagra i andre databasar) skulle lenkast mest mogleg direkte til relevante felt i artiklane (meir om dette i Grønvik og Tvedt 2006). Figur 1 syner korleis dei viktigaste komponentane i redigeringsapplikasjonen er kopl saman.



Figur 1. NO 2014: Materialorganisering mellom databasar og andre komponentar

Eit anna krav var at det skulle vere minst mogleg nytt å lære for kvar ny programdel som kom på plass. Grensesnitt, ikon og basisfunksjonar er derfor mest mogleg like, og likt plassert, frå programdel til programdel og skjermbilete til skjermbilete. For eit meir IT-basert perspektiv på dette, sjå Grønvik og Tvedt 2006.

3.3. Innskrivningsprogrammet som modell av ordartikkelen

Dei fleste ordartiklane NO er stutte og etter måten enkle, men dei store artiklane kan vere kjempedigre og særns innfløkte. Det var opplagt at innskrivningsprogrammet måtte spegle ordboksartikkelen i maksimumsformat. Det har vore drøfta om det i ettertid skulle lagast eit redusert ”kvardags”-skjema, men sidan kategoribehovet kan variere sterkt frå små-artikkel til små-artikkel har det ikkje blitt følgt opp.

Det første som vart gjort, var å analysere tidlegare praksis i band 1-4, og skilje ut

hovuddrag i struktur. På grunnlag av dette vart det laga framlegg til artikkelskjema, som utgangspunkt for programmeringa (om dette sjå Grønvik 2005).

3.4. Endringar i arbeidsmetode

Akademisk leksikografi dreiar seg om å sortere, systematisere og kommentere språklege data. Når artikkelskrivinga går føre seg ut frå papirarkiv, og manus blir skrivne som laupande tekst, vil arbeidet arte seg slik:

- sortering av papirbelegg i haugar
- ordning (systematisering) av setelhaugane (utover skrivebord eller i sorteringskasser) (gjerne med påhefta lappar med ”overskrift” og nummerering)
- forming av tekst ut frå kvar liten bunke, i rett rekkjefølgd
- plassering av tilfang i arkiv igjen, helst ordna slik at det speglar artikkelstrukturen, med dei siterte setlane først og i rett rekkjefølgd.

Det vil seie at ein øvd redaktør skal ha i hovudet (1) heile kategorisystemet, (2) ordninga av kategoriane i hierarki, og (3) alle konvensjonane som gjeld for verket.

Med redigeringskjema som inngang til ein ordboksdatabase blir dei viktigaste endringane i arbeidsmetode desse:

- Kategorisystemet (artikkelstrukturen) er synleg på skjermen ned til detaljnivå før ein tek til å arbeide
- Ordninga (systematisering) av kategoriar i hierarki er innebygd i systemet
- Rekkjefølgja i innskriving er langt mindre relevant enn den var – det som tel, er å skrive inn teksten i rett felt. Rangering og rekkjefølgje kan takast i ein separat prosess til slutt.
- Tilfanget er digitalisert, og kan sorterast og tilordnast digitalt ned på sitatnivå, slik at kven som helst kan kontrollere materialordninga attom kvar artikkel. Ein slepp å lagre sorteringa i ein eigen prosess etter redigeringa.

3.5. Grensesnittet

Som nemnt kan dagens NO-redaktørar sitje og sjå på det som førre generasjon måtte ha i hovudet. Det blir her gjeve ein kort gjennomgang av korleis innskrivingsdelen av applikasjonen er ordna, visuelt og innhaldsmessig:

Grensesnittet i innskrivingsprogrammet er delt i tre, med (1) ein sams toppdel for heile Oracle-systemet (ikonrekkje for basisfunksjonar osv.), (2) artikkelen sett opp i tredigram til venstre, og (3) innskrivingsdelen til høgre. Innskrivingsdelen har form av ei rekkje skjema, eitt for kvar hovuddel ein artikkel kan ha. På skjerm er desse skjema ordna slik at det ser ut som arkivkort, jf. tabell 1.

[trediagram]	Artikkel	Oppslag	Hakeparentes	Tydingsbolk	Tilvising	Sorterar
--------------	----------	---------	--------------	-------------	-----------	----------

Tabell 1.

Under kvart av desse ”korta” finn ein så kortsystem som dekkjer dei aktuelle underkategoriane. Tabell 2 viser underkategoriar pr. kort.

1 [trediagram]	2 Artikkel	3 Oppslag	4 Hakeparentes	5 Tydingsbolk	6 Tilvising	7 Sorterar
[tre – viser artikkel-element ned til kjelde. Viser også kva for artiklar som har tilvising til den artikkelen ein står i – jf. kort for Tilvising]	Dimensjonering Statushistorie Redaktørhistorie Merknad	Homografnr Oppslagsord Vokalparentes Leddanalyse Ordklasse Ordklassetilstand Tabell for bøyingsssystem	Uttale Eldre kjelder Avvikande skriftformer Målføreformer Etymologi	Definisjon Sitat/Døme Samansetting s-rekkjer Undertydingar Underoppslag	[skjema for vising av artikkel som viser til den artikkelen ein står i]	[område for eigendefinerte tabellsett over tilfanget bak kvar ordartikkel, der redaktøren kan sortere og lagre sorteringar]

Tabell 2. *Underkategoriar pr. kort*

3.5.1. Trediagram

I utgangspunktet var det ein viss skepsis til tredigrammet som arbeidsreidskap, truleg av di det var uvant. Den store styrken ved tredigrammet er at det gjev godt oversyn, og er lett å omorganisere. Flyttar ein rundt på tydingar, så følgjer alt ”utstyr” med. Alle endringar synst med ein gong. Denne føremonen har stadig fleire kome til å verdsetje i det daglege arbeidet.

3.5.2. Artikkel

Her ligg dei redigeringsadministrative opplysningane pr. artikkel. Dimensjonering er kommentert ovanfor. Statushistorie syner kvar ein artikkel er i produksjonsgangen, redaktørhistorie syner om ein eller fleire har vore innom artikkelen osv. I merknadsfeltet kan redaktøren notere særskilte tilhøve ved ein artikkel, og eventuelt sende desse vidare til gruppeleiar eller hovudredaktør.

3.5.3. Oppslag

Her ligg skjema for oppslagsdelen av artikkelen. Dette skjemaet skil seg frå dei andre ved at det er menyar eller føringsrestriksjonar for samtlege felt utan feltet for oppslagsord og bøyingsformer, og oppslagsordet blir generert opp. Berre bøyingsformer blir skrivne inn av redaktøren.

3.5.4. Hakeparentesen

Fem opplysningskategoriar har plass her, og tre – eldre kjelder, avvikande skriftformer og målføreformer – krev kjeldeføring. Kjeldeføringa er menystyrt, og heimfestingane for målføreopplysningane har no også fått automatisert hierarki, slik at områdenemne overtek for punktheimfestingar når dekningsgraden er god nok. Det er ikkje mogleg å føre både områdenemne og punktheimfesting. Uttaledelen er ikkje menystyrt, men dei fleste aktuelle føringsmønster er opplista i redigeringshandboka, for å sikre konsistens.

Formatet for hakeparentesen vil bli justert ein gong til i samband med at band 1-5 skal leggjast inn i ordboksdatatabasen. Da blir det høve til å leggje ei siste hand på etymologiskjemaet, der formatet førebels gnissar litt mot materialet.

3.5.5. Tydingsbolken

Her blir hovudtyngda av det skapande redaktørarbeidet lagt inn. Kjeldefesting, krysstilvisingssystem og tilvisingar til lemma som ikkje blir redigerte i papirutgåva (semantisk gjennomsynlege samansetningar med oppslagsordet som førsteledd) er også her lagt inn som menyar og lenker, med innebygde restriksjonar mot t.d. sirkelkoplingar. Under redigeringa har ein fullt oversyn over desse artikkel-elementa så langt ein vil, dei kjem opp som peikarar i tredigrammet etter kvart som dei blir hekta på.

Ein kan sjå på papirordboka som ein simulert database, der tilvisingane skal uttrykkje lenker m.a. i det semantiske hierarkiet som finst i språket. Men papirordboka kan aldri bli verkeleg stringent, for feilkoplingar straffar seg ikkje direkte. Med NO som database, straffar det seg straks å kople feil. Lagar ein sirkelkoplingar, får ein ikkje lagre. Blir artikkelar feilplasserte i førsteleddsrekke, forsvinn dei frå oppgenerert sats. Ingen redaktør gjer den typen feil mange gonger.

Tydingsbolken var ferdig spesifisert alt i 2003, og er ikkje endra når det gjeld innhald, men det har vore somme justeringar av funksjonaliteten. Det nye (2005) er at materialsorterar og korpus er koplade inn i applikasjonen, slik at sorteringsresultat kan lagrast etter kvart, anten samla på artikkelnode, eller spesifisert på tyding (definisjonselement) og sitat eller døme. Det viser seg at tilfanget attom ulike lemma kan invitere til ulik kategorisering. Kvar redaktør rår over eigne sorteringar og utval, ut frå sitt faglege skjøn. For somme artikkelar blir det laga førebelse tydingsskjema i trestrukturen med direkte innsortering av tilfang, for andre blir det sortert etter meir empiriske kriterium (t.d. kollokat). Ein kan lage – og lagre – eit uavgrensa tal sorteringar for kvar artikkel.

3.5.6. Tilvising

På dette ”kortet” kan ein sjå den artikkelen som har ei tilvising til det lemmaet ein arbeider med. Tilvisingskoplingane ligg i tredigrammet som ei liste med peikarar

under lemmaet, men er også distribuert på tydingsbolkane i treet, slik at ein skal kunne gå rett på ved første forsøk.

3.5.7. Sorterar

Dette er som før nemnt det ferskaste modulen i redigeringsprogrammet. Sorteraren er tilgjengeleg via innskrivingsapplikasjonen, men er ikkje del av den. Den er meir å likne med ei handlekorg, der ein kan hente inn tilfang frå MO, frå korpuset, og etter kvart, om vi skulle ønskje det, frå andre kjelder. Sorteraren er forma som eit rekneark der tilfanget ligg i første kolonne, og så kan ein leggje til kolonner, subkategorisere og ordne av hjartans lyst. Her kan ein også skilje homografar og kopiere over sorteringar til andre artiklar.

Sorteraren var eit svært etterlengte hjelpemiddel i arbeidet med dei store artiklane, og har alt synt seg svært anvendeleg i fleire fasar av redigeringsprosessen. NO 2014 har ein satelittredaksjon ved NTNU i Trondheim, som bere har tilgjenge til samlingane over nett. Her var sorteraren særskilt etterlengta, ettersom den fristiller redaktørane frå omfattande kladding og sortering i vanlege tekstbehandlingsprogram.

3.6. Eit ord om søkjesystemet

I og med at heile NO-systemet er eit sett av databasar, er også alt søkbart. Dette gjer det lett å finne modellar og leite etter parallellar når ein står fast. Ein kan t.d. søkje opp alle definisjonar under alle lemma på ”-skap m” og sjå om det finst mønsterformuleringar (det bør det jo vere). Søkjesystemet blir nytta i redigering, til administrasjon og i konsistenskontroll – og med god effekt. Når manglande stringens kan finnast med eit par tastetrykk, er det bryet verdt å rydde godt opp etter seg.

3.7. Oppsummering

Dei store endringa i redigeringsmetode som applikasjonen med tilhøyrande databasesystem har ført med seg, er at redaksjonen no har eit samla overblikk over tilfang, artiklar og dokumentasjon, ikkje berre frå ein synsvinkel, men frå mange. Alle kan gå inn kvar dei vil i systemet og sjå på kva dei vil. Sett på spissen kan ein seie at medan kvar redaktør før sat aleine og neddykka i tilfanget ”sitt”, så sit redaktørkollegiet no høgt oppe og ser på korleis kvar og ein ordnar opp med sin teig. Medan spørsmålsstillinga før ville vera ”korleis skal eg handtere dette ordet?” er spørsmålet no (i større grad i alle fall) ”korleis handterer NO denne typen ord?”

Den endra spørsmålsstillinga aktualiserer dermed meir generelle faglege problemstillingar, som i sin tur gjer det lettare å sjå tilfanget frå ein språkvitskapeleg ståstad. Det er lett å opprette eit artikkelskjelett (trestruktur) på grunnlag av ein hypotese, og prøveplassere tilfang inn ut frå ei generell lingvistisk tilnærming til ei gjeven ordgruppe. Dermed er det også t.d. lettare å sjå møtepunkta mellom grammatikk og semantikk, og ikkje minst sjå det semantiske innhaldet i dei større gram-

matiske kategoriane. Det er sannsynleg at dette vil gje NO meir systematiske, og derfor betre, ordartiklar over dei store funksjonsorda (t.d. preposisjonane, ein del store verb og adverb) enn det som var mogleg å få til på gamlemåten. Også behandlinga av ordlagingsselement som prefiks og suffiks vil få eit naudsynt lyft på denne måten.

Produksjonen av ordartiklar blir langt meir målretta og effektiv når tvil om konvensjonar osb. er rydda vekk. Berre det at band 5 vart ferdig til rett tid (desember 2004), med ca. 270 sider redigerte av for det meste nyttilsette redaktører, syner det.

Når det gjeld kvalitetskontroll, må den vera tvifelt. At konsistensen i verket er blitt betre, veit vi, men konsistens er ikkje nok. Påstanden at ordboksmanuset held minst like høgt fagleg nivå som før byggjer på intern kontroll, ved at gruppeleiarar og hovudredaktør les og korrigerer fortløupande, og på ekstern kontroll, ved at redaktørar presenterer ord og ordgrupper som dei har arbeidd med i vitskaplege føredrag og artiklar – m.a.o. normale akademiske kvalitetsprøvingar.

4. Opplæring

Før har det vore vanleg å rekne med at det tek to til fem år å lære opp ein redaktør i eit vitskapleg ordboksverk. Norsk Ordbok 2014 set nye redaktørar i funksjon etter svært kort opplæringstid, ved hjelp av innskrivingsprogrammet og databasesystemet.

Kva følgjer har dette for fokus og innhald i redaktør opplæringa? Lat oss opne med konklusjonen: I denne perioden har NO 2014 tilsett opp mot tjue nye redaktørar og redaksjonsassistentar. Det representerer ein stor og variert lingvistisk kompetanse, men det dei har sams er at ingen av dei hadde arbeidd i eit ordboksverk tilsvarande NO 2014 før dei byrja i stillingane sine. Alle måtte derfor ha opplæring frå botnen – og alle er blitt produktive innanfor stipulert tid.

4.1. Tradisjonell redaktør opplæring

Den tradisjonelle opplæringsmetoden i dei akademiske ordboksverka har vore at ein har kome inn i redaksjonane på ein slags lærlingstatus, og så har trødd i fotefara til meistrane, utan å få noka organisert opplæring. Opplæringsinitiativet låg hos nybyrjaren, som helst skulle lære ved å lese det som var skrive før, og så absorbere detaljane under marsjen. Ein skulle redigere, gjere feil, få feilen påpeikt, rette den, få nye feil påpeikte, osb. Tida frå første manus vart skrive og til utgjeving kunne vera svært lang (åtte-ti år er ikkje uvanleg), noko som igjen fører med seg tidkrevjande revisjon og innarbeiding av nytt tilfang i samband med korrektur og utgjeving. Med ei slik tilnærming er det ikkje rart at folk treng minst fem år for å bli varme i trøya.

4.2. Kunnskapskrav før og no

Vi kan jamføre kunnskapskrav til ordboksredaktørar i akademiske ordboksverk før og no, for å sjå kvar skilnadene ligg. Før måtte ein ha

1. Solide lingvistiske kunnskapar, ensyklopediske interesser og god formuleringsevne
2. Svært godt minne for detaljar (tilfang, redaksjonelle konvensjonar, lister, namn, ortografi ...)
3. Svært god evne til grannsemd i detaljar, også utan streng ytre kontroll
4. Evne til å tenke strukturelt er ein føremon
5. Grenselaust tolmød!

Etter kvart sat ordboksredaktøren på ein enorm detaljkunnskap om eit grunnlagstilfang som få andre kjende.

Krava i dag er dels dei same, men ikkje heilt. No trengst det:

1. Solide lingvistiske kunnskapar, ensyklopediske interesser og god formuleringsevne
2. Kunnskap om korleis ein handterer og vurderer store datamengder
3. Eit skarpt blikk for det meiningsberande i detaljar, men mindre krav til minne (mykje av det som før måtte hugsast er no bygd inn i redigeringsprogrammet)
4. Evne til å tenkje strukturelt er heilt nødvendig
5. Godt syn og god finmotorikk – elles blir det mange uhell i applikasjonen!

Ein kunne leggje til at der ordboksredaktøren før måtte tole å sitje myke aleine, må han no tole kritisk innsyn i alt han eller ho gjer.

Punkt 1, nøkkelkravet, det som kan lyfte definisjonsskriving frå handverk til kunst, er ikkje endra med tida. På dei andre punkta er det endring, og dei endringane speglar at vi lever i eit informasjonssamfunn med overflod på data og godt med pengar, men veldig lite tid. I dagens universitetssamanheng verkar det mest komisk å snakke om prosjekt som varer i meir enn ti år, slik NO 2014 gjer.

4.3. Krav til redaktør opplæring

I NO 2014 diskuterer vi ikkje lenger om det er behov for organisert opplæring av nye redaktørar. Spørsmålet er kva slag opplæring dei treng, som gjev prosjektet dyktige, produktive redaktørar på stuttast mogleg tid. Maksimal opplæringstid og prøvetid har til no vore eitt år, etter det er det normale produktivitetskrav til alle.

I NO 2014 har dei nye redaktørane gått denne løypa, med mindre variasjonar:

1. Først ei økt med materialarbeid (normering i MO i 2-3 månader)
2. Deretter rett på redigering og definerings
3. Får tildelt alfabetbolk
4. Får opplæring i gruppe, individuelt og gjennom seminar (førelesingsrekke knytt til redigeringshandboka, ved kolleger eller innhenta ekspertise)

5. Får støtte i detaljert redigeringshandbok, redigeringsprogram med manual og i open arbeidskultur: Det er greitt å spørje – og alle treng korrigerings.
6. Får oppmuntring og oppfølging når det gjeld å fylle faglege lakuner, t.d. ved å følgje aktuelle kurs eller seminar på UiO, konferansedeltaking osv.

4.4. Progresjon i redaktøropplæringa

Dei fleste nye redaktørar byrjar med å redigere innhaldsord, og går laus på funksjonsord først når dei har noko røynsle. Men dette gjeld ikkje i same grad for alle. Mellom rekruttane er det fleire utprega grammatikarar som raskt tek tak i større funksjonsord (som dei fekk tildelt som særskilde oppdrag).

Generelt har ordninga likevel vore slik: Opplæringa har gjeve ein gjennomgang av leksikalske særdrag og tilsvarende leksikografisk behandling ordklasse for ordklasse. Dei nye redaktørane skriv dei første artiklane sine om konkrete substantiv med 5-20 belegg og minst ein godt avgrensa sakdefinisjon. Deretter følgjer adjektiv, så verb. Ved verb bør ein ta seg tid til å gjennomgå heile verbfrasen med valens og ulike typar kollokat, av di dette er tydingsskiljande kriterium som ber strukturen i verbartiklar. Gjennomgangen av ordgrupper tek til med grunnord, deretter følgjer avleiingar og samansetningar.

Dette høyrer kanskje ut som ein repetisjon av grunnkurs i leksikalsk semantikk, men i høve til ordboksarbeid blir perspektivet annleis. Det er ikkje spørsmål om å meistre omgrepa eller forstå tenkinga. Perspektivet er ”korleis operasjonaliserer vi denne kunnskapen i materialbehandlinga, og kva følgjer får den for sortering, systematisering og framstilling i ordartikkelen?” Dette er ikkje triviell kunnskap, og det er ikkje purt handverk. Ein skal vite kva ein gjer, og korleis, men ein må også vite kvifor – og kva for alternativ som kunne vore aktuelle. Dermed er ein langt inne i generell lingvistisk teori, som jo bør tole slik materialtesting.

4.5. Prioriterte mål for NO 2014

Redigeringa av NO i database tok til våren 2003. Dei måla prosjektet då hadde for staben, var (1) meir effektiv opplæring, (2) auka tempo i manusproduksjon, (3) auka kontroll med resultatet, (4) auka konsistens i redigeringa, (5) tettare samarbeid innanfor redaksjonen og med Eining for digital dokumentasjon.

Dette har gått om lag som planlagt. Ikkje utan motbakkar, men motbakkar gjev god treningseffekt. Utan redigeringsprogrammet med tilliggjande modular, endra redaksjonell metode og effektiv opplæring hadde NO 2014 aldri fått ut band 5 til rett tid. Redigeringsprogrammet med tilhøyr blir karakterisert som sjølvforklarande av dei nyaste redaktørane, som har kome til ferdige program og slik sett dekkja bord. NO 2014 meiner derfor å kunne slå fast at redigeringsprogrammet står for ”den gode systemtvangen”.

5. Konklusjon

Digitalisering av NO 2014 har gjeve (1) konsistent materialbehandling som grunnlag for lemmaseleksjon, dimensjonering og kjeldeføring, (2) meir systematisk og kollektiv leksikografisk metode på grunnlag av lingvistisk skoloring, med prosjekt-målet som ramme og redigeringsapplikasjonen som hjelpemiddel, og (3) grunnlag for effektiv redaktør opplæring – og dermed synt at ordboksredigering er eit fag som kan lærast.

Litteratur

- Bø, Reidar 1989: Arbeidet med Norsk Ordbok. I: Almenningen, Olaf og Grønvik, Oddrun (red.): *Ord og mål. Festskrift til Magne Rommetveit*. Oslo: Kringkastingsringen, 80-94.
- Grønvik, Oddrun 1997: Om kjeldegrunnlaget for Norsk Ordbok. I: Vikør, Lars S.: *Norske Ordbok – nynorskens leksikografiske kanon?* Universitetet i Oslo, Institutt for nordistikk og litteraturvitskap, 23-37.
- Grønvik, Oddrun 2005: Redigeringsprogrammet for Norsk Ordbok 2014. I skjæringspunktet mellom menneskeleg skjøn og automatisering. I: Fjeld, Ruth Vatvedt, og Worren, Dagfinn (red.): *Nordiske studier i leksikografi 7. Rapport fra Konferanse om leksikografi i Norden Volda 20.-24. mai 2003*. Oslo, 156-165.
- Grønvik, Oddrun og Tvedt, Lars Jørgen 2006: Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy. I denne rapporten.
- Hjulstad, Håvard 1986: Mikro FELTED. Feltinnskrivingsystem for mikromaskin. Universitetet i Oslo, Avdeling for namnegransking. 16 s.
- Svardal, Terje 2003: Norsk Ordbok – om normering av Metaordboka. I: Hansen, Zakaris Svabo og Johansen, Anfinnur (red.): *Nordiske studier i leksikografi 6. Rapport fra Konferanse om leksikografi i Norden Tórshavn 21.-25. august 2001*. Tórshavn, 351-356.
- Venås, Kjell 1989: Kva ventar vi oss av Norsk Ordbok? I: Almenningen, Olaf og Grønvik, Oddrun (red.): *Ord og mål. Festskrift til Magne Rommetveit*. Oslo: Kringkastingsringen, 153-166.
- Worren, Dagfinn 1992: Om feltinndeling av ordartiklane i Norsk Ordbok. I: Vikør, Lars S. og Worren, Dagfinn: *Feltskjema for innskriving av ordmateriale*. Universitetet i Oslo, Institutt for nordistikk og litteraturvitskap. 18 s.

Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy

Norsk Ordbok 2014 – presentation of a complex lexicographic tool. The dictionary writing (DWS) system of Norsk Ordbok has been developed by the project Norsk Ordbok 2014 and the Unit for digital documentation (EDD) at the University of Oslo. This paper briefly describes the main modules of this DWS.

1. Innleiing

Norsk Ordbok skal gjerast ferdig innan 2014. For å nå dette målet vart prosjektet Norsk Ordbok 2014 (NO 2014) oppretta, med ein utvida redaksjonsstab. Det vart naudsynt å innføre nye metodar og nye verktøy for å effektivisere redigeringsprosessen.

Målsettinga med arbeidet vart også endra. I tillegg til å vere ei papirordbok, er det no eit mål at Norsk Ordbok skal gjerast tilgjengeleg i digital form.

Med dette som utgangspunkt har prosjektet NO 2014 og Eining for digital dokumentasjon (EDD) ved Universitetet i Oslo utvikla eit spesiallaga redigeringsssystem for ordboka.

Redigeringsystemet har vakse fram i eit tett samarbeid mellom NO 2014 og EDD. Spesifiseringsarbeid og utviklingsarbeid har gått parallelt, og nye modular har blitt utvikla og innarbeida i systemet samstundes som redigeringsystemet har vore i bruk i redaksjonen.

Vi vil i denne artikkelen kort skissere hovudfunksjonane til dette nye redigeringsystemet. Dette er ikkje ein fullstendig dokumentasjon eller presentasjon av systemet, men ei kortfatta innføring i kva hovudmodular vi har vald å ta med i systemet.

2. Generelt system

EDD utviklar system for ei rad ulike fagområde, og i avtalen mellom NO 2014 og EDD vart det lagt som premiss at dei systema som vart utvikla i samarbeid mellom NO 2014 og EDD skulle kunne brukast også av andre. EDD valde difor å basere redigeringsystemet på det rammeverket som alt var utvikla. Dette avsnittet skildrar difor dette rammeverket i generelt.

2.1. *Metadatabase*

Etter som talet på ulike fagdatabasar vaks innanfor ramma av EDD, vart det stilt større og større krav til gjenbruk av komponentar for å halde kostnadane ved systemutviklinga lågast mogleg. Det vart også stilt krav om at ein burde kunne søkje

i ulike databasar utan å måtte bruke ulike databaseapplikasjonar. EDD fann det difor naudsynt å vurdere om ein kunne utvikle eit system som sentraliserte alle felles delar av eit større databasesystem, men samstundes gjorde det mogleg å utvikle det spesielle for kvar einskild fagdatabase som EDD fekk ansvaret for.

Løysinga på dette problemet var å identifisere alle objekt i databasen med ein objekttype og ein identifikator. Med objekt meiner vi til dømes ein ordboksartikkel, ein setel i ein seteldatabase, ein kataloginnføring frå eit museum, eller ei lydfil frå ein musikkdatabase.

EDD utvikla så ein felles metadatabase med informasjon om strukturen for dei ulike fagdatabasane, eit felles Windows-rammeverk med tilsvarende web-rammeverk, og eit system for å kople spesialutvikla framvisarar/redigeringsskjema til dei ulike objekttypane.

Metadatabasen som ligg til grunn for systemet, er basert på ein relasjonell databasestruktur. For kvar ny database ein utviklar, registrerer ein metadata om tabellar og relasjonar slik at ein har meir informasjon om dette enn det som ligg i den normale metadata-strukturen for eit databasesystem. Denne informasjonen brukar systema dynamisk til å generere søkjeskjema, og søka mot databasen vert genererte på basis av metadata og data frå brukaren.

Resultata av slike spørjingar vert presenterte som resultatlistar som får si form via data i metadatabasen. I tillegg vert resultata vist i spesialkonstruerte redigeringsskjema. Koplinga mellom resultatlistar og redigeringsskjema er også styrt via metadatabasen.

Spørjesystemet representerer i utgangspunktet relasjonsmodellen som ligg til grunn for dei ulike databasane. Alle datafelt i alle tabellar er i utgangspunktet tilgjengeleg som søkjefelt for sluttbrukaren. For dei fleste brukarar av systemet vil dette framstå som komplekst og utilgjengeleg. EDD har difor utvikla eit system i metadatabasen som gjer det mogleg å forenkle strukturen slik at den passar for kvar einskild. Dei fleste databasane er difor utstyrte med ein eller fleire spesiellaga søkjeskjema som brukaren kan velje mellom. I tillegg er det fritt for brukaren å lage sine egne søkjeskjema. Dette er mogleg fordi ein har basert seg på dynamisk bruk av data frå metadatabasen i staden for å generere statisk programkode eller skjemastruktur, slik mange tilsvarende system gjer.

Den einskilde brukaren kan også lagre søkjekriteria eller resultat av søk i det ein har kalla lister. Dette gjer det mogleg å kome lett tilbake til dei objekta i basen som ein arbeider med, eller som har spesielle karakteristikk. Dette har vore mykje brukt i statushandtering i samband med registrering, innskriving, osv.

Sidan databasane til EDD varierer i stor grad, fann ein ut at det ikkje ville vere naturleg å lage felles redigeringsskjema for alle basane, eller forsøke å lage dynamisk generering av redigeringssystema. I staden har EDD utvikla eit generelt grensesnitt mellom det dynamiske søkjesystemet og spesiellaga redigeringsskjema for dei ulike

databasane. Søkjesystemet bruker metadatabasen til å kontrollere kva redigerings-skjema som høyrer til kvar database.

Ved å kombinere dynamisk generering av søkjesystema med spesiellaga redigerings-skjema, har EDD utvikla ein struktur som gjer det mogleg å fokusere på det spesielle ved kva einskild base, samstundes som det meste er generelt. EDD kan difor utvikle avanserte system på kort tid.

Basert på same metadatabase-struktur har EDD også utvikla eit web-system med same funksjonalitet som redigeringsapplikasjonen har. Søkjeskjema og resultat er lett gjenkjenneleg for brukaren, men i staden for redigerings-skjema, får ein her fram ei HTML-vising av objekta ein leiter fram.

2.2. Teknologi

I botnen av fellessystemet ligg ein Oracle-database. Oracle er vald både ut frå funksjonsomsyn, men også fordi Universitetet i Oslo har ein avtale med denne databaseleverandøren.

På klientsida har EDD så langt berre utvikla løysingar for operativsystemet Microsoft Windows og for web. Grunnen til dette er at universiteta i Noreg stort sett brukar Windows. EDD har nytta Delphi frå Borland som utviklingsplattform.

Sidan EDD må handtere meir enn berre standardtekst, har ein tatt i bruk diverse ulike teknologiar. I databasen lagrar EDD både strukturert tekst, lyd, bilete og kartdata, og alle desse datatypene vert manipulert både i databasesystemet og på klienten.

3. Redigeringsystemet

Redigeringsystemet til Norsk Ordbok er utvikla innanfor det rammeverket som er omtalt i førre avsnitt. Strukturen for å bygge opp søkjesystem og kopling til andre databasar låg difor klar før modelleringa av ordboksstrukturen tok til.

3.1. Grunnlagsmateriale

Mykje av grunnlagsmaterialet for Norsk Ordbok er digitalisert og ligg som egne fagdatabasar i det felles rammeverket som er utvikla ved EDD. Norsk Ordbok 2014 har brukt ein god del ressursar på å normalisere dette materialet ved å kople det saman via Metaordboka som også er ein fagdatabase i rammeverket. Dette er mogleg fordi objekta er unikt identifisert via objekttypar og identifikatorar.

3.2. Struktur

Norsk Ordbok er modellert strengt hierarkisk. Dette vart gjort fordi ein i utgangspunktet rekna med at dette er den rette strukturen for ordboka. Etter kvart har ein vel kome fram til at det finst strukturar som bryt med denne føresetnaden, og dette har skapt visse problem.

Under føresetnad av ein hierarkisk struktur på artiklane i ordboka, er redigerings-systemet for artiklane basert på ein trestruktur som speglar databasestrukturen. Redigering av dei einskilde delane av artikkelen vert gjord i skjema knytt til den einskilde noden. Mykje av kontrollen med at redigeringsreglane vert følgde har vi lagt inn i programvaren i databasetenaren. Dette har vi gjort for å kunne sikre dataintegritet sjølv om ein nyttar ulike verktøy for å manipulere data i databasen.

Ved å dra nytte av at alle objekt er identifisert via objekttypen og ein identifikator, kan ein i teorien kople alle typar data til aktuelle nodar i artikkelreet. Dette gjer det mogleg å dokumentere alle tydingar, eller andre delar av artikkelen, ved å peike til andre data, som til dømes setelarkiv, andre ordbøker, osv.

3.3. Presentasjon

Det er lagt vinn på at så lite som mogleg av skiljeteikn, bindetekstar, og liknande, skal leggjast inn i databasen. Presentasjon av artikkelen i form av trykte boksider eller i form av webpublisering, skjer ved at data frå databasen vert eksportert til eit format som er tilpassa presentasjonsformatet. Vi har basert oss på at artiklane vert eksportert som XML. For generering av websider vert XML transformert ved hjelp av XSLT til HTML, og mest mogleg av formatering av HTML vert gjort ved hjelp av CSS.

For papirpresentasjon, har vi basert oss på å bruke PDF. Artiklane vert konvertert frå XML til XSL-FO ved hjelp av XSLT, og dette vert så formatert ved hjelp av XSL-FO-programvare til PDF. Vi har forsøkt å bruke gratisprogramvare til dette, mellom anna FOP, men på grunn av spesielle krav som NO 2014 stilte, har vi enda opp med å bruke XEP frå RenderX for transformasjon til PDF. Heile trykksatsen for band 5 vart generert på denne måten som ei PDF-fil.

Eit av problema i samband med papirversjonen av ordboka, var handtering av spesialteikn. Norsk Ordbok valde å nytta fonten Lexicon frå The Enschedé Font Foundry, og Enschedé konstruerte dei manglande teikna i denne fonten.

Eit anna stort problem var handtering av orddeling. I ei ordbok med mange avstyttingar er det ikkje mogleg å lage rutinar for korleis dette skal gjerast. I tillegg til dei regelbaserte systema kom vi difor ikkje unna store ordlister med unntaksreglar. Vi måtte også lage reglar for å hindre linjedeling mellom ord. Alt dette vart lagt inn i XSL-FO-systema våre.

3.4. Redigeringshjelp

3.4.1. Autogenerering av artikkelskjelett

Redigeringsystemet til Norsk Ordbok drar nytte av arbeidet som er gjort med Metaordboka ved at ein automatisk genererer opp eit rammeverk for ein artikkel basert på innhaldet i Metaordboka. Systemet tolkar kjeldematerialet som er kopl til metaordboksartiklane, og gjev redaktørane råd om dei bør lage artiklar i Norsk

Ordbok basert på dei einskilde metaordboksartiklane. Vel ein å lage ein ny artikkel i Norsk Ordbok, lagar systemet ein artikkel med oppslagsformar og ordklassemerking på plass. I tillegg vert peikarar til grunnlagsmaterialet kopla til artikkelen, slik at store delar av dokumentasjonsarbeidet er gjort alt før redaktøren har skrive ei einaste tyding.

3.4.2. Sortering og kopling av grunnlagsdata

I redigeringsystemet til Norsk Ordbok er det lagt inn eit sorteringsverktøy som gjer det mogleg å sortere grunnlagsmaterialet etter dei kriteria som til ei kvar tid er nyttig for den aktuelle artikkelen. Sorteringsinformasjonen vert teken vare på så lenge redaktøren sjølv ønskjer, og basert på sorteringsinformasjonen kan redaktøren bygge opp tydingsstruktur med kopling direkte til dei einskilde kjeldene. I tillegg kan ein brukar av ordboka sjå kva setlar frå setelarkivet, eller kva korpusbelegg som ligg til grunn for ein kvar definisjon i ordboka.

Redaktøren kan også hente inn anna grunnlagsmateriale undervegs i redigeringa av artikkelen. På dette viset kan redaktøren mellom anna kople korpusbelegg direkte til tydingar i artikkelen.

3.4.3. Krysstilvisingar

Ein sentral funksjonalitet i eit redigeringsystem er å halde kontroll på alle tilvisingar i systemet. Redaktørane kan berre vise til artiklar eller tydingar som reelt eksisterer i ordboka. Dette er sikra ved at alle tilvisingar er reelle koplingar til artiklar som er redigert i databasen. Det er ikkje mogleg å slette ein artikkel som ein annan refererer til. Den hierarkiske visinga av artikkelen syner tilvisingar i begge retningar, slik at det mellom anna er mogleg å sjå kva artiklar som viser til den artikkelen ein redigerer.

3.4.4. Handtering av kjelder og heimfestingar

For å sikre konsistens i handteringa av heimfesting av tydingar, døme og belegg, er det lagt inn eit stadhierarki i systemet som automatisk skjuler heimfestingar som er omfatta av andre heimfestingar for same tyding. Redaktøren kan også be systemet om å sjekke om det finst så mange ulike heimfestingar at det bør slåast saman til ei heimfesting på høgare nivå. Til dømes kan systemet automatisk slå saman ei rad kommunar til eit fylke.

Det er også meininga at systemet etter kvart skal kunne lese ut alle heimfestingar og kjelder frå grunnlagsmaterialet, og på dette viset spare redaktørane for arbeidet med å redigere dette sjølv.

3.4.5. Dimensjonering

Norsk Ordbok er etter kvart komen i grenselandet mellom å vere ei digital ordbok

og ei prenta ordbok. I den digitale verda er ikkje volum eit stort problem, men den trykte utgåva må halde seg innanfor visse rammor. I redigeringsystemet er det difor lagt inn eit system for automatisk å foreslå storleiken ein artikkel skal ha. Utrekninga av dette er basert på innhaldet i Metaordboka for den tilsvarende artikkelen. Talet på belegg for ein metaordboksartikkel vert lagt til grunn, og basert på diverse vektingsreglar vert storleiken for artikkelen i Norsk Ordbok rekna ut. Det talet som kjem fram er ei rettesnor for redaktørane, men dette talet kan sjølvstendig overprøvast. Undervegs vert den reelle storleiken på artikkelen registrert, og dette kan danne grunnlag for rapportar.

3.4.6. Førehandsvisning

Trass i at dei fleste redaktørane etter kvart har funne seg til rette med å sjå på artiklar som hierarkiske tre, er det sjølvstendig mogleg for redaktøren å sjå korleis artikkelen vil sjå ut i papirform eller i web-form. Redaktøren kan velje å kontinuerleg vise ein HTML-basert versjon av artikkelen, eller artikkeldelar, og kan også generere ut PDF-versjon av ein eller fleire artiklar.

3.5. Administrasjon

For eit prosjekt med over tjue redaktørar, er det naudsynt med administrasjon. Vi har lagt vinn på å legge så mykje som mogleg av administrasjonen inn i systemet på ein slik måte at administreringa vert ein del av dei daglege oppgåvene til redaktørane.

Redaksjonen i Norsk Ordbok er inndelt i grupper, med ein eller fleire gruppeleiarar. I tillegg er det oppretta nokre spesialgrupper innanfor redaksjonen. Desse spesialgruppene har særskilte oppgåver som til dømes å sjekke sitat eller samanhalde større delar av ordboka. I systemet er dette løyst ved at redaktørane er delt inn i redaksjonsgrupper, og i tillegg har roller.

Alle artiklar har til ein kvar tid ein redaktør. I tillegg til dette har alle artiklar ein status. Statussystemet er laga slik at ein artikkel med ein bestemt status berre kan flyttast over i eit avgrensa tal andre statusverdier. Statusen til artikkelen avgjer kven som har tilgang til å redigere han, og kven som kan flytte artikkelen mellom ulike statusar. Ved hjelp av dette systemet sikrar ein at ikkje fleire redaktørar redigerer same artikkel. Samstundes kan ein bruke systemet til å bygge opp ei produksjonsline der artikkelen må innom ein rad stasjonar før han vert publisert. Til dømes må ein artikkel innom ein gruppeleiar for kontroll, og alle artiklar må ha vore innom sitatkontroll før hovudredaktør tar tak i totalen.

3.6. Logging

Det er eit mål at redaksjonen til ein kvar tid har oversikt over kva som har skjedd med ein artikkel. Det er difor konstruert eit loggsystem knytt til statushandteringa. Kvar gong ein artikkel endrar status, vert det teken ein kopi av artikkelen med

statusinformasjon, og denne kopien vert liggande i databasen. Det er også mogleg å ta ein slik kopi kvar gong redaktøren ønskjer det. Dette gjer det mogleg å følgje utviklinga av artikkelen, både innhaldsmessig og også for å undersøkje kven som har gjort kva med artikkelen.

3.7. *Statistikk*

Basert på statusverdiar, dimensjoneringstal og loggar, vert det no manuelt laga produksjonsrapportar for at administrasjonen skal kunne følgje utviklinga av ordboka. Dette er viktig for å halde ein høg produksjon. Vi arbeidar no med eit generelt rapporteringssystem som skal gjere det lettare å utvikle dei rapportane som redaksjonen treng.

4. Problem

Ein del av problema vi har støtt på i utviklinga av systemet, er å strukturere ordboka, samstundes som ein skal opne for fleksibilitet. Ved å bruke eit relasjonsdatabasesystem i botnen, har vi lagt ganske sterke strukturingskrav på ordboka.

Vi har også hatt debattar om korleis redigeringa skal gå føre seg. I ein tidlegare versjon av systemet var trestrukturen meir gjennomgripande ved at det hovudsakleg berre var mogleg å redigere einskildnodar. I dagens versjon er det laga ein kombinasjon av ein trestruktur og eit vanleg skjema som eigentleg omsluttar fleire noder i treet. Den hybridløysinga som no er vald skapar ein del problem, og etter som redaksjonen vert meir van med å tenkje artikkelen som eit tre, vil det kanskje vere ønskjeleg å gå over til å kunne redigere nodane i treet meir direkte. I dette ligg det i så fall eit krav om å flytte fokus endå meir bort frå satsbiletet og meir i retning av strukturering av innhaldet.

5. Konklusjon

Redigeringsverktøyet slik det no ligg føre er eit godt verktøy til å fullføre det store NO 2014-prosjektet, sjølv om delar av systemet ikkje fungerer heilt slik vi kunne ønskje. Vidare – sjølv om det framleis står att å utvikle nokre modular av systemet – ser vi likevel at systemet har ført til høgare kvalitet, meir konsistens i ordboka, større effektivitet og enklare opplæring av nye redaktørar.

Litteratur

Extensible Stylesheet Language (XSL), Version 1.0, W3C Recommendation 15 October 2001, <http://www.w3.org/TR/xsl/>
 XSL Transformations (XSLT) Version 2.0, W3C Working Draft 15 September 2005, <http://www.w3.org/TR/xslt20/>

Web-sider: (alle sider sjekka 4. oktober 2005)

Cascading Style Sheets Home Page (CSS), <http://www.w3.org/Style/CSS/>

Eining for digital dokumentasjon, <http://www.edd.uio.no/>

Formatting Objects Processor (FOP), <http://xmlgraphics.apache.org/fop/>

Norsk Ordbok, <http://no2014.uio.no/>

RenderX, <http://www.renderx.com/>

The Enschedé Font Foundry, <http://www.teff.nl/>

The Extensible Stylesheet Language Family (XSL), <http://www.w3.org/Style/XSL/>

Ømålsordbogen og mejetærskergenerationen

Ømålsordbogen and the Combine Harvester Generation. The topic of this paper is the description of neologisms in a Danish dialect dictionary, *Ømålsordbogen* (The Dictionary of the Insular Dialects). *Ømålsordbogen* covers the period from 1750-1945 with 1850-1920 as its core period. The description of the neologisms in the dictionary, especially those reflecting agricultural developments in the late 19th and early 20th century, has become more thorough than originally planned regarding both definitions, encyclopaedic information and selection of entries. This growing interest in the later part of the period covered by the dictionary could be explained by the fact that this period has now become historical whereas in the 1920's, when the project began, these newer phenomena were contemporary to the editors and considered to be of less interest – or simply left unnoticed.

1. Indledning

I august 2004 var jeg sammen med to af de andre redaktører ved *Ømålsordbogen* på en ekskursion til Vestfyn, til en af ordbogens kontakter, landmanden Børge Jensen. Han havde nemlig stillet os i udsigt at vi kunne komme til at prøve at høste med *selvbinder* (eller *binder*, som man ofte sagde i det daglige). Jeg bruger datid, for selvbinderne blev afløst af mejetærskerne i 1960'erne. Selvbinderen er en forholdsvis kompliceret indretning der både skærer kornet af og binder det i neg, og den tilhører den nyere tids mekaniserede landbrug. Binderen begyndte dog allerede at blive udbredt i 1920'erne, og hermed hører den med til den periode som *Ømålsordbogen* beskriver. Jeg er ret sikker på at vi, når vi skal til at redigere *selvbinder*, vil være glade for at vi har oplevet hvordan sådan én fungerer. Sådan noget er det svært at læse sig til. Hvad der yderligere gjorde vores ekskursion nyttig, var at Børge Jensen ikke kun interesserer sig for landbrugshistorie og ældre landbrugsmaskiner, men også for dialekter, så vi fik også oplysninger om hvad de enkelte dele af selvbinderen kaldtes, oplysninger der er dårligt belagt i vores samlinger.

Da man for alvor begyndte at samle ind til *Ømålsordbogen* omkring 1930, var selvbinder og hele den landbrugskultur selvbinderen repræsenterer, nutid. Den forekom selvfølgelig og relativt uinteressant for redaktørerne. Interessen samlede sig i stedet om livet som det udfoldede sig i selvforsyningsens tid, hvor mændene mejede med mejetøj, og kvinderne bandt op i deres hvide forklæder. For os nuværende redaktører, der lever i mejetærskertiden, er 1920'erne derimod historie, og hermed helt anderledes interessant.

2. Ordbogens tid

Ømålsordbogen beskriver dialekterne på Sjælland, Lolland-Falster og Fyn med omliggende øer (som det også fremgår af ordbogens fulde titel, jf. litteraturlisten). Den omfatter en periode på ca. 200 år fra omkring 1750 til omkring 2. verdenskrig med tyngdepunkt i perioden 1850-1920. Afgrænsningen 1850 er betinget af kildemæssige omstændigheder. De oplysninger der findes om dialekterne før 1850, er få og tilfældige, og optegnelserne rummer ydermere stort set kun særdialektale ord. Grænsen omkring 2. verdenskrig er derimod betinget af ændringer i omverdenen, nemlig den mekaniseringsbølge i landbruget der introduceres med at traktoren bliver almindelig.

Men også inden for ordbogens hovedperiode sker der store forandringer. Landbruget omlægges fra overvejende at være baseret på korndyrkning til overvejende at være baseret på animalsk produktion af slagtesvin og mælk. Andelsbevægelsen vinder frem, og der kommer andelsmejerier næsten i hvert sogn til at behandle al den mælk som omlægningen af landbruget fører med sig. Der kommer nye arbejdsbesparende redskaber til og nye afgrøder med *sukkerroerne* som den mest betydningsfulde. Ofte tænker man på mekaniseringen af landbruget som noget der hører 1900-tallet til, men allerede i anden halvdel af 1800-tallet får man fx *hesteriven* og *rensemaskinen*, og inden døre får man *komfuret* og *kødhakmaskinen*. Og alle disse forandringer afspejles i ordforrådet – og i Ømålsordbogen. Materialet til sådanne ord er imidlertid ofte sparsomt fordi man som sagt fokuserede på tiden før moderniseringen af landbruget ved materialeindsamlingen.

3. Materialeindsamling

Allerede da jeg redigerede mit første ord, verbet *hakke*, blev jeg opmærksom på den skævhed der var ved den oprindelige materialeindsamling.

Verbet *hakke* har flere landbrugsfaglige betydninger, bl.a. bruges det i forbindelsen *hakke roer* i betydningen ‘udtynde og luge roer med et hakkejern’. Materialet på *hakke roer* bestod imidlertid med et par undtagelser udelukkende af udskrifter fra båndoptagelser og fra besvarelser fra en spørgeliste udsendt i 1969. Uden denne supplerings havde det ikke været muligt at redigere udtrykket. Og det samme gælder den øvrige roedyrkningsterminologi. Eftersom roedyrkingen blev indført før 1900, er det helt klart en forsømmelse at roedyrking ikke kom med i første omgang. Og optegnerne kunne bare have spurgt, for alle – eller så godt som alle – meddelere havde været med i roerne og ville have kunnet fortælle udførligt om emnet.

Men det er på den anden side også forståeligt at man især var optaget af det ældre lag. Det var nemlig på et relativt sent tidspunkt at man begyndte at optegne systematisk på Øerne. Ganske vist oprettede man allerede i 1908 Udvalg for Folkemaal, der fik som hovedopgave at udarbejde en ordbog over ømålene. På det tidspunkt var jysk og bornholmsk nemlig til en vis grad dækket ind med Espersens bornholmske

ordbog (Espersen 1908) og Feilbergs jyske (Feilberg 1886-1914), så nu var det afgjort ømålene der stod for tur. Men så kom 1. verdenskrig, og det blev først i slutningen af 20'erne at arbejdet reelt kom i gang. Nu hastede det hvis man skulle nå at "dokumentere det Kulturtrin, som det 19. Aarhundredes Dialekter er sprogligt Udtryk for", som Poul Andersen skriver i en redegørelse for planerne for den kommende ordbog over ømålene. Med noget var det allerede for sent:

At Fordringerne [til ordbogen, AG] nemlig ikke vil blive honoreret fuldt ud er rimeligt, allerede af den Grund at det Kulturtrin, som det 19. Aarhundredes Dialekter er sprogligt Udtryk for, ikke mere lader sig oplyse i alle Enkeltheder ad den mundtlige Traditions Vej, og det betyder, at hvad vi ikke faar med os nu, det vil stedse forblive uoplyst. Det er denne Kendsgerning, der bliver bestemmende for Dialektforskningens Krav til Indsamlingsarbejdet, og det er med Sorg og Bekymring, at Uff [dvs. Udvalg for Folkemaal, AG] ser, i hvor ringe Grad dets Kræfter slaar til, hvor lidt det magter at indsamle og oplyse, og hvor meget der derfor uafvendeligt forsvinder fra Dag til Dag. (Andersen 1939:39)

Ordbogen karakteriseres i sin fulde titel som en "sproglig-saglig" ordbog (jf. litteraturlisten). Hermed har redaktørerne villet pointere at ordbogen lægger stor vægt på også at skildre den materielle kultur og til dels også den åndelige, som ordene er indlejret i. *Saglig* har nemlig her den lidt ualmindelige betydning 'som vedrører encyklopædiske oplysninger'. Valget af netop ordet *saglig* afspejler at ordbogen har modtaget en afgørende inspiration fra "Wörter und Sachen"-retningen der blomstrede i 1920'erne (jf. Andersen 1939:24, Pedersen 1994:249).

Interessen for det saglige har præget både indsamlingsmetoder, tilrettelægning af redigeringen og ordbogens endelige udformning.

For at kunne indsamle materiale inden for et emne må man være godt inde i det, og redaktørerne delte derfor emnerne indbyrdes, således at én fx havde "pløjning", "tærskning" og "fiskeri", og en anden "høst" og "gården og dens indretning". De pågældende redaktører læste faglitteratur om emnet, interviewede kyndige dialekttalende og udarbejdede spørgelister inden for deres emne. Et kort uddrag af kornspørgelisten (stencileret) giver et godt indtryk af hvor omfattende opgaven var:

Mejning

1) Gloser for "meje", "Mejning", Hug, Skaar, Udtryk og Talemaader om Skaarets Bredde, Skaarkam, Stub (lang Stub v. Lejesæd).

2) Gloser for Høstkarl, Høstfolk, "være paa Høstarbejde". Udtryk for god eller daarlig Høstkarl.

På disse få linjer stilles mange, forskelligartede spørgsmål, men ikke desto mindre vedrører uddraget kun én arbejdsproces blandt mange inden for emnet kornhøst. Og kornhøsten er jo kun et blandt mange emner. Endelig skulle de samme spørgsmål stilles mange steder i ømålsområdet for at sikre den dialektgeografiske repræsentati-

vitet, samtidig med at der også skulle indsamles oplysninger om bøjning og udtale osv.

I midten af 1940'erne planlagde man endvidere at de enkelte redaktører skulle lave en sagredaktion af deres emne, dvs. redigere de særlige saglige betydninger og udarbejde et system af synonymhenvisninger med et hovedsynonym hvorunder de encyklopædiske oplysninger bringes, og hvorfra der henvises til de mere kortfattet redigerede undersynonymer (der igen har henvisning tilbage til hovedsynonymet). Og en hel del af emnerne blev faktisk forredigeret på denne måde, men desværre ikke alle, inden man – bl.a. på grund af presset fra bevilgende myndigheder – påbegyndte en alfabetisk betydningsredaktion.

I ordbogen afspejler den saglige interesse sig både i definitioner og citater, men kommer særligt til udtryk i de såkaldte leksikonafsnit, dvs. sammenhængende skilddringer af arbejdsprocesser, redskaber eller skikke. Afsnittene ligner artikler i et leksikon, men adskiller sig fra almindelige leksikonartikler ved at der gennemgående henvises til den folkelige terminologi. De er af meget varierende længde, men store emner som fx kornhøst kan fylde 4-5 spalter (jf. Pedersen 1994:250).

Hovedindsamlingsperioden var, som det vil være fremgået, 1930'erne og 40'erne, men man fortsatte med at supplere materialet sideløbende med sagredaktionen og den senere alfabetiske forredaktion. I 1975 begyndte den egentlige redigeringsfase, i 1992 udkom første bind, og det hidtil sidste bind, bind 7 af 16 planlagte, kom i 2004.

Det var først i begyndelsen af 1970'erne at artikelstrukturen blev endeligt fastlagt, så selvom projektet er næsten hundrede år gammelt, har nogle af de nuværende redaktører altså haft stor indflydelse på ordbogens format. Og vi har stadig mulighed for at gøre vores indflydelse gældende. Det er nemlig sådan at Ømålsordbogen i høj grad bygger på uskrevne regler, og det har jo en tendens til at medføre inkonsekvenser. Da nu ordbogen i forvejen er noget ujævnt redigeret, justerer vi lidt undervejs når vi bliver opmærksomme på noget som vi synes er uheldigt, bl.a. er vi i de senere år begyndt at lægge større vægt på det ordforråd der henviser til den senere del af ordbogens periode.

4. Beskrivelsen af det “nyere” ordforråd

4.1. Definitionerne

Det er uvist hvornår præcist det blev bestemt at ordbogen skulle gå til hen omkring 2. verdenskrig, men sådan har det i hvert fald været fra omkring 1970 (jf. et foreløbigt manuskript til en indledning, Bjerrum m.fl. ca. 1975). Om ord der betegner tekniske nyskabelser fra begyndelsen af det 20. århundrede, hedder det imidlertid i den publicerede indledning til ordbogen: “Sådant mere eller mindre perifert ordstof vil ofte blive behandlet summarisk eller blot blive medtaget af hensyn til en dialektal

udtaleform” (Ømålsordbogen, Tillægsbind:4). Sådan en summarisk behandling finder vi fx under *bil* og *cigaret*.

Jysk Ordbog har valgt at udelade definitionen ved fællesdanske ord eller betydninger, og i stedet bare skrive “= rigsm.” (dvs. = rigsmålet). Det gør vi ikke, dels fordi vi ønsker at Ømålsordbogen i princippet skal være en fuldstændig beskrivelse af sproget, dels fordi det undertiden kan være usikkert hvad der er rigsmålets betydning. Men undertiden havde “=rigsm.” næsten været at foretrække, fx er *bil* er forklaret med det næsten forældede ord ‘automobil’ (sådan forklarer Jysk Ordbog dog også *bil*). Eller endnu værre: Slår man op under *cigaret*, finder man ækvivalenten ‘tobakscigaret’, et knap nok eksisterende ord.

Netop disse løsninger er vi jo ikke så stolte af, men det principielle der ligger bag, fastholder vi. Det er ikke vores opgave at forklare hvad fx en bil er; undertiden kan vi nøjes med en antydende definition. Men hvis man sammenligner redaktionspraksis i de tidligste bind med de afsnit der er redigeret i de senere år, kan man konstatere en tendens til at redigere “nyere ord” mere fyldigt, især ord der vedrører landbokulturen.

Som eksempel på et ord vi i dag ville have redigeret mere fyldigt, kan nævnes *andel* med sammensætninger:

andel s m, dog > c S, c Æ – i sinster ofte [an'ndz-] vF(spor opt, y) – -1) den del af noget som tilhører ell tilfaldet én Ø(spor opt, y): [su 'mɑnɑ 'ɑndæjls 'hɑ's 'dɑ i mɑ'r'i:j's (mejeriet)] vF(Br); *samme Dag var jeg paa Boagde med Fattiges Andeel* M(Faro RStær.Dagb. 6/1 1812); hertil *andelsforetagende*, *-haver*, *-mejeri*, *-tærskeselskab*. -2) medvirken, delagtighed vF(Br, Øst): *jeg har ingen andel i, at han kom sådan ofsted* vF(Øst).

Det at *andel* får en særlig betydning i kraft af andelsbevægelsen, fremgår ikke af forklaringen; det er kun antydet ved citatet: *hvor mange andele har du i mejeriet?*. Sammensætningen *andelsbaver* kunne også have fortjent en artikel, muligvis også *andelsmejeri* og *-slagteri*, selvom de sidstnævnte ord næppe har været almindelige i det daglige; normalt sagde man bare mejeriet eller slagteriet. Der er dog en tillægsforklaring på at *andel* er redigeret så knapt: Det materiale redaktøren havde til sin rådighed dengang, var ikke så godt som det hun ville have haft nu. Det hænger sammen med at vi i de sidste 10-15 år har fået opbygget et korpus af båndoptagelser udskrevet i rigsmålsortografi, der er blevet et vigtigt supplement til optegnelserne på papir.

Lydoptagelserne er indsamlet i perioden 1934-1994, men de fleste interview er foretaget efter 1970'erne. Alle meddelelserne er imidlertid født inden for ordbogsperioden – de fleste er født 1870-1905 – og materialet bruges derfor næsten på lige fod med det ældre materiale, dvs. vi medtager stof som vi mener kunne have været sagt

før 2. verdenskrig. Og det har ordet *andel* jo givetvis, så belæggene på *andel* fra båndoptagelserne kan supplere materialet.

Høstativ betegner et redskab der er kommet i brug efter 1900, og er samtidig eksempel på et ord der er redigeret inden for de seneste år. *Høstativ* er forsynet med encyklopædiske oplysninger, og der er udarbejdet synonymhenvisningssystem, hvor *høstativ* er valgt som hovedsynonym:

høstativ s – indretning (indført i 1920'erne) til at tørre hø på (bestående af tre lægter der er sammenføjet foroven og forbundet med tværbjælker på den nederste tredjedel af lægterne) Ø(spor opt); jf *hæsje, horytter, rytter, stakrytter, stativ*.

Høstativ er ikke med i den oprindelige sagredaktion af emnet høhøst, selvom det er et af de emner der er gennemarbejdet grundigt. Materialet stammer nemlig, ligesom til *hakke roer*, næsten udelukkende fra en sen spørgeliste, denne gang fra 1984.

Høstativ er forsynet med udbredelsesangivelsen Ø(spor opt), dvs. "sporadisk optegnet i hele ømålsområdet (og muligvis almindeligt)" (jf. Ømålsordbogen, Til-lægsbind:22), men det har ikke noget at gøre med at ordet tilhører det 20. århundredes landbrug. Det er en formel man tit møder i ordbogen. "spor opt" siger noget om vilkårene for indsamling og dokumentation af talesprog. Til trods for det store indsamlingsarbejde der ligger bag, er vi ofte i den situation at vi ikke tør udtale os sikkert om udbredelsen.

4.2. Kronologiske markeringer

Angivelsen af udbredelse er naturligvis obligatorisk i en ordbog der omfatter mange dialekter, men undertiden er der også kronologiske markeringer. De kronologiske markeringer afspejler også den prioritering af stoffet før andelsbondekulturen der oprindeligt prægede projektet.

I udgangspunktet er Ømålsordbogen tænkt som en synkron ordbog, skriver Poul Andersen i redegørelsen for ordbogsplanerne (Andersen 1939). Der kan blive tale om at nogle ord markeres tidsmæssigt. For det første ord der ikke kendes længere i slutningen af 1800-tallet, for det andet yngre stof: "Der kan paa denne Maade paa nogle Punkter blive antydning af en Kronologi, men i det store og hele gaar Ordbogsplanen ikke ud over Fordringen: det Synchroniske Tværsnit af Sprog og Kultur" (Andersen 1939:20).

Poul Andersens udsagn har i det store hele stadig gyldighed. Det er kun en lille del af de oplysninger ordbogen bringer, der er forsynet med en form for kronologisk markering. Og med rette: Størsteparten af ordforrådet er givetvis blevet brugt i hele perioden.

Men behovet for at markere noget som værende nyt i perioden har været stigende i takt med at det nyere ordstof har fået en noget mere central placering. Desværre er

vores markeringer ikke fulgt med tiden, om jeg så må sige. Den almindeligste er nemlig “nyere”, jf. fx:

flæskesteg s – større, helstegt stykke svinekød (en nyere ret, i ældre tid var det ikke alm at stege større stykker kød) Ø(nu alm, LglO); jf *grise-*, *svinesteg*.

Og hvad betyder så nyere? Ja, det er ofte brugt om begreber og ord, der er indkommet i andelstiden, altså centralt i ordbogens kerneperiode (1850-1920), og som mange brugere allerede i dag vil opleve som noget der tilhører “de goe gamle dage”. I de senere år har vi forsøgt når det var muligt, at undgå “nyere” (eller “ældre”) og i stedet for anføre omtrentlige tidsangivelser, sådan som det er gjort ved *hårbørste* og *jomfru*:

hårbørste s – børste til at børste håret med (først alm br i slutn af 1800t) Ø(spor opt).

jomfru (også **jomfrue**) s f, > c S, c Æ .. -1) (indtil slutn af 1800t) om ugift kvinde af højere stand, fx om præstens, lærerens ell skovriderens datter .. S, M, L, F(spor opt) ..

Tidsangivelserne forekommer dog især i leksikonafsnittene, fx under ordet *juleleg*.

juleleg s – 1) *ungdomsgilde* (med dans), dels (L) om gilde i julen (ofte 2. ell 3. juledag, dels (F) om gilde i perioden fra jul til kyndelmisse ell fastelavn .. ; jf *julebal*, *-dans*, *-gilde*, *-lag*, *-musik*, *-stue*, *lege jul*; de unges gilder i julen (og tiden efter) var opr sammenkomster uden musik, hvor underholdningen var leg (jf SprKult og jf *julebisp*, *-buk*), men i begyndelsen af 1800t afløstes legene helt eller delvis af dans .., se videre u *ungdomsgilde*, *gilde* I.1; julegilder af denne type gik af brug i slutn af 1800t, men afløstes til dels af foreningsfester i forsamlingshus olgn; jf *julebal*, *-gilde*, *-træ*.

Juleleg er hovedsynonym i en synonymkreds med betydningen ‘sammenkomst for unge i juletiden’ (der igen indgår i synonymkredsen *ungdomsgilde*). Leksikonafsnittet gælder sagen som sådan, og ikke specielt ordet *juleleg*, og også materialet under de øvrige opslagsord er udnyttet til sagredaktionen. Leksikonafsnittet har, som det ses, givet mulighed for at beskrive faserne i udviklingen fra de gamle julelege i begyndelsen af 1800-tallet, over sammenkomsterne med dans til andelstidens fester, der ikke kan hedde *juleleg*, men *julebal*, *julegilde* og *juletræ*, som det fremgår af henvisningerne i afsnittets slutning.

At der i leksikonafsnittene også er omtale af nyere skikke, redskaber osv. er også noget der er blevet mere almindeligt i de senere år end før. Høstativet er fx tilsvarende nævnt i afsnittet om høhøst (under *hø*) og selvbinderen i afsnittet om kornhøst (under *høst*). Derimod er der fx ikke noget om den omfattende kagebagning der blev almindelig ved komfurets indførelse i afsnittet om bagning (under verbet *bage*), eller noget om de biavlsmetoder der blev indført i slutningen af det 19. århundrede i artiklen *bi*. Der er også et enkelt eksempel på et leksikonafsnit der stort

set kun beskriver nyere foreteelser, nemlig afsnittet under *kaffe*, hvor der gives en oversigt over de redskaber og de metoder for kaffebrygning der har været brugt siden kaffe blev almindeligt på landet efter 1850.

5. Lemmaselektion

Det er imidlertid ikke kun behandlingen af det nyere ordstof der har ændret sig, men også holdningen til hvilke ord der skal medtages.

Det har fra begyndelsen været meningen at ordbogen skulle omfatte hele ordforrådet og ikke særdialektale ord. Men undervejs har holdningen ændret sig til hvad dét omfatter. Man kan sige at sprogsynet er blevet mindre puristisk. I Poul Andersens artikel kommer han ind på afgrænsningen af det dialektale ordforråd over for "Skriftsproget og Rigestalesproget":

Dette Skel mellem hvad der tilhører Dialekten og hvad der ikke tilhører den, er det af mange grunde vanskeligt at drage. Hos alle yngre Øboer er Kendskabet til det i det hele nærstaaende Rigsmaal saa intimt, at Skellet mellem Dialekt og Rigsmaal er udvisket stærkt, og blandt vore Gamle er der forstaaeligt nok kun en mindre Elite, hos hvem man kan vente Resultater paa direkte Spørgsmaal om et Ords eller Udtryks Ægthed og Hjemlighed (Andersen 1939:21).

Han skriver videre at det sikreste er "den direkte Lytten til gammeldags Dialekttalende", men at metoden er alt for tidskrævende og dermed kostbar, og desuden, siger han, er der langt imellem disse.

Det vil altså sige at det ikke er nok at et ord indgår i en dialekttalendes aktive ordforråd, det skal altså også være en gammeldags én. Det mener vi er for snævert et kriterium for medtagelse. Det ændrer dog ikke ved at vi stadig har problemer med afgrænsningen over for rigsmålet.

Som eksempel på hvordan et sådant afgrænsningsproblem er løst, kan nævnes verber med præfikset *er-* (se tabel 1).

De verber der er anført i tabel 1, er de verber der er udskrevet til Ømålsordbogens seddelsamling, og altså dem som redaktøren har skullet tage stilling til.

Som det ses, er omtrent halvdelen af verberne medtaget i ordbogen, dog, bortset fra *ernære*, alle med markeringen "fra rm", der enten betyder at ordet skønnes at være indkommet i dialekterne på et forholdsvis sent tidspunkt, eller at ordet formentlig ikke står centralt i sprogbroen.

De ord der er udeladt, er alle kun repræsenteret med ét eller to belæg, der kun omfatter en lydform. De stammer fra ordsamlinger etableret af lokale dialektinteresserede, bl.a. vestfynboen Mikkil Kristiansens samling. Det er en samling man som ny redaktør meget hurtigt bliver kendt med, for det er en meget ordrig samling, men også en samling hvis kvalitet kan diskuteres. Som Poul Andersen skriver i sin nekrolog over Mikkil Kristiansen: "man undres tit over at M.K. tager ret meget

moderne Stof med; man finder Vendinger som man fristes til at kalde litterært Gods”, men, fortsætter han: “det er sikkert en urigtig bedømmelse”. Stoffet er ægte i *hans* dialekt: “.. hans vestfynske Mål er et intellektuelt præget, ejendomsstolt, vist også sprogligt reflekteret Bonde-Milieu’s Sprog, og på den baggrund må Samlingerne vurderes” (Andersen 1937:61). Formuleringerne er dog næppe ganske upåvirkede af at de optræder i en nekrolog, og ord der kun optræder hos Mikkel Kristiansen (og tilsvarende ordsamlinger) udelades normalt af ordbogen.

Medtaget i Ømålsordbogen	Udeladt
erfare Ø(spor opt, fra rm)	
	erholde
	erhverve
erindre Ø(spor opt, fra rm)	
	erkende
	erklære
	erkyndige
ernære Ø(spor opt)	
erstatte Ø(alm, fra rm)	

Tabel 1.

De verber med præfikset *er-* der er medtaget, er i modsætning til de udeladte alle belagt med autentiske citater eller i hvert fald citater der virker autentiske, fx under *ernære*: *han kan jo itte ernære en familie; de kan dårligt ernære sig ved stedet*. Citaterne er optegnet i lydskrift som så mange andre belæg i Ømålsordbogens samlinger, og lydskriften slører den omstændighed at der i mange tilfælde er tale om en slags redaktionelle eksempler, nemlig eksempler som en meddeler har lavet for at eksemplificere ordets brug enten spontant eller på opfordring af optegneren. Det er ikke i sig selv diskvalificerende for et citat, for mange var gode til det. Citaterne er ofte genereret med udgangspunkt i meddelerens egen socio-kulturelle kontekst, og kan betragtes som eksempler på en typisk replik; andre forsøg virker konstruerede (jf. Hovmark 2004:8f). At et ord er belagt med citater, garanterer altså ikke i sig selv at ordet faktisk bruges. Citaterne må vurderes et for et.

Med hensyn til de citater der stammer fra vores korpus af båndudskrifter, er problemstillingen en anden: Citaterne her er jo rent faktisk blevet sagt af dialekttalende. Men man må tage i betragtning at interviewene på båndoptagelserne repræsenterer en særlig form for samtale. Ganske vist påvirker optagelsessituationen især

fraseologi og fremstilling (se videre Hovmark 2004:12ff), men formentlig kun i ringe grad ordforrådet. Der er dog eksempler på ordvalg fremkaldt af optagelsessituationen, fx er *indretning* klart overrepræsenteret i båndoptagelserne fordi det er et praktisk ord at gribe til når man skal beskrive et redskab eller lignende, jf. fx:

(hvordan så en plejl ud?) *se der var sådan en rund indretning vi kaldte en håndvol* (Sydfyn, Hesselager Sogn).

Desuden må man være opmærksom på at mange af optagelserne er fra 1970'erne eller senere, og der indgår derfor også ord som de dialekttalende har taget i brug efter 2. verdenskrig. Ord som *fjernsyn* og *mejetærsker* og andre betegnelser på tekniske foreteelser er selvfølgelig helt uproblematisk at have med at gøre. Vanskeligere er det med fraseologiske udtryk, fx var *gå ind for noget* 'være gunstig stemt over for noget' i første omgang blevet redigeret under *ind*, men heldigvis blev vi i tide opmærksom på at det er tvivlsomt om udtrykket har været brugt dialektalt inden for perioden, eftersom det i 1920 i Ordbog over det danske Sprog er markeret som nyt og sjældent og antagelig lånt fra engelsk. Men også betegnelser på genstande kan være vanskelige at tage stilling til.

Det kan ikke udelukkes at vi undertiden tager for meget med. Samtidig er der nok tilfælde hvor vi tager for lidt med. Fx er ordet *kalender* udeladt. Den ældre dialektale betegnelse er givetvis *almanak*, der da også er medtaget i ordbogen. Efter al sandsynlighed har ordet *kalender* været brugt også på landet i begyndelsen af 1900-tallet. Materialet består imidlertid kun af nogle få lydsedler der som sagt ikke godtgør at ordet virkelig er blevet brugt. Og vi kan trods alt ikke redigere ud fra hvad vi tror vi ved. Der var så den mulighed at forhøre sig hos ældre landboere, men udbyttet ville måske ikke stå mål med arbejdet. Alle ville selvfølgelig kende ordet, men det ville i mange tilfælde være usikkert fra hvilket tidspunkt meddelelsen kendte det.

6. Afslutning

Principielt mener vi at Ømålsordbogen skal beskrive hele de dialekttalendes ordforråd inden for perioden. Lige så lidt som man kan undvære *okay* i en ordbog over moderne dansk, kan man undvære *bil* og *høstativ* i en stor dialektordbog. Men som jeg har gjort rede for, er det ikke altid muligt at leve op til de ideelle fordringer fordi materialet undertiden er utilstrækkeligt. Og når det kommer til stykket, er det nok heller ikke ord som *kalender* vi vil lægge mest energi i at efterhøre. I praksis vil ordbogen stadig være præget af at nogle ord i dialekterne anses for at være "finere" og "mere dialekt" end andre.

Litteratur

Ordbøger:

Espersen, I.C.S. 1908: *Bornholmsk Ordbog*. Med Indledning og Tillæg. Fotografisk optryk. København: Rosenkilde og Bagger 1975.

Feilberg, H.F. 1886-1914: *Bidrag til en ordbog over jyske almuesmål* 1-4. København: Thieles Bogtrykkeri.

Jysk Ordbog. <http://www.jyskordbog.dk>.

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. 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 1937: M. Kristiansen. 21. maj 1849 – 6. januar 1937. I: *Danske Folkemaal* 10, 65-67.

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.

Bjerrum, Marie m.fl. ca. 1975: Indledning til *Ømålsordbogen*. Utrykt, Afdeling for Dialektforskning's manuskriptsamling nr. 318.

Hovmark, Henrik 2004: Jagten på det autentiske citat. I: *Danske Talesprog* 5, 3-22.

Pedersen, Karen Margrethe 1994: Ordbog og encyklopædi. I: Anna Garde og Pia Jarvad (red.): *Nordiske Studier i Leksikografi* II. Rapport fra Konference om Leksikografi i Norden 11.-14. maj 1993. Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2. Oslo, 248-256.

Ett svenskt dialektlexikon i startgroparna

A Swedish dialect dictionary ready to start. The project *Swedish Dialect Dictionary* has been in progress at the Institute for Dialectology, Onomastics and Folklore Research in Uppsala since 2003. People interested in dialects constitute the target group and the dictionary will be published in book form. A timescale of ten years and a limit of 1200 pages are some preliminary constraints. The collections of dialect words at the Institute are considered a sufficient basis in spite of a slightly uneven geographical distribution. Selection criteria are discussed, among others, that the dialect words should be drawn from at least ten independent sources or derive from three provinces. Besides the usual items of information on pronunciation, meaning etc., the microstructure of the dictionary entries contains the geographical distribution of the occurrences. The need for increased resources if the project is to be completed within the time stated is emphasised.

1. Inledning

Våren 2003 fick vi i uppdrag av direktören vid Språk- och folkminnesinstitutet (SOFI) att utreda möjligheterna att framställa ett kortfattat dialektlexikon. Målgruppen för detta skulle vara den dialektintresserade allmänheten, däribland gymnasieelever, studenter och personer med intresse för hembygdens kultur och historia, släktforskning etc. Hittills har dessa användare varit hänvisade till Johan Ernst Rietz föräldrade *Svenskt dialekt-lexikon* från 1862-67 i den mån det varit bekant och funnits till hands. Alternativet har varit att ta kontakt med SOFI:s forskningsarkivarier för att under vägledning närma sig materialet i institutets omfattande dialektordssamlingar. – Vad har då hänt sedan uppdraget gavs och var står vi idag?

2. Bakgrund

Det kan vara på sin plats att ge en kort beskrivning av den arbetssituation som rådde då vi fick uppdraget. Under ett antal år hade vi båda arbetat som redaktörer inom det långsiktiga projektet *Ordbok över Sveriges dialekter* (OSD). Syftet med detta projekt var alltsedan starten 1954 att på vetenskaplig grund utarbeta en ordbok över svenska dialekter i Sverige. OSD skulle företrädesvis rikta sig till forskarsamhället och därmed fungera som ett komplement till *Svenska Akademiens ordbok* (SAOB). Efter en omorganisation av myndigheten SOFI 1998, som innebar att Dialektenheten vid f.d. Dialekt- och folkminnesarkivet i Uppsala och OSD slogs samman, kom OSD att bli ett av flera projekt inom den då nybildade Dialektavdelningen. En följd av sammanslagningen blev att den kraftigt decimerade personalen kunde ägna allt mindre tid åt redigeringen av OSD.

I det läget aktualiserades förslaget att satsa på ett kortfattat lexikon, som skulle finnas tillgängligt i ett hanterbart format inom rimlig tid. Projektet, med arbetsnam-

net *Svenskt dialektlexikon* (SDL), skulle planeras och utarbetas av oss båda, men på deltid, då vi även hade andra uppgifter att utföra inom avdelningens verksamhetsområden.

3. Omfång, tidsplan och publiceringsform

Redan från början uppmanades vi att försöka beräkna dels omfång, dels den tid det skulle ta att framställa ett lexikon – detta bl.a. med tanke på en projektansökan. De flesta som formulerat en sådan vet hur svårt det är att göra den typen av beräkningar. Likafullt använde vi inledningsvis rätt mycket tid åt att, utifrån det stora antal kortlådor som finns i SOFI:s arkiv, bestämma lexikonets omfång i sidor och lägga upp en tidsplan för genomförandet av projektet. Med förbehållet att det skulle behövas minst en redaktör till och helst ytterligare personal till Dialektavdelningen, föreslog vi preliminärt ett omfång av högst 1 200 sidor och en tidsram av 5-7 år. Efter senare uppgjorda, mera realistiska beräkningar har redigeringstiden förlängts till ca 10 år.

Svenskt ortnamnslexikon (SOL), utgivet av SOFI 2003, ansågs vara ett jämförbart projekt, men även om SOL:s framgång är eftersträvansvärd haltar jämförelsen. Detta projekt hade redan vid start en betydligt större redaktion till sitt förfogande, vilken också innefattade externa redaktörer. Ändå överskreds tidsplanen för SOL med råge.

När det gäller SDL:s omfång finns inga direkt jämförbara paralleller att utgå ifrån. Givetvis har vi studerat Rietz ordbok, den enda egentliga föregångaren. Vi har också gått igenom andra enbandsordböcker som *Nynorskordboka*, *Svensk ordbok* och många andra. Dock vill vi betona att det inte är frågan om att uppdatera en äldre ordbok. SDL ska inte heller vara en ersättare för OSD, utan det är helt och hållet ett eget projekt. Det enda som de båda ordboksprojekten har gemensamt är materialet. I det stadium vi befinner oss nu, skulle vi helst vilja avvakta med att precisera sidantal och tidsplan för det kortfattade lexikonet och i fortsättningen bara använda våra beräkningar som ett slags riktmärken tills vi producerat fler artiklar och bestämt en lämplig layout. Till stöd för vårt projekt har vi en expertgrupp som består av Staffan Fridell, professor i nordiska språk vid Uppsala Universitet och docent Birgit Eaker, förste redaktör vid SAOB. Båda har tidigare varit OSD-redaktörer.

Beträffande publiceringsformen är vår inriktning att lexikonet ska ges ut som bok i ett band. Lexikonets tänkta omfång talar för att det är en lämplig form. Detta hindrar naturligtvis inte att SDL med tiden också kommer att finnas åtkomligt på nätet. Fördelen med en nätversion är ju att den ökar möjligheterna att göra sökningar av olika slag, t.ex. regionala sådana.

4. Utgångsmaterial

Det material SDL kommer att bygga på är i huvudsak detsamma som utgör basen för OSD, nämligen OSD:s samlingar i sedesformat. De består av runt sju miljoner

blad och utgör huvuddelen av ordbokssamlingarna med dialektord vid SOFI i Uppsala. Insamlandet av dem påbörjades redan 1954 och har fortgått ända fram till nu. Parallellt med detta arbete har samlingarna hopsorterats med målet att uppnå en enda serie från *a* till *ö*. Dit har vi ännu inte kommit och i dagsläget fördelar sig materialet på åtta serier, det s.k. huvudregistret och sju underserier. Huvudregistret innehåller material i större eller mindre utsträckning från alla landskap, medan underserierna omfattar samlingar från ett eller några få landskap vardera. All sortering sker från underserierna in i huvudregistret, där ungefär hälften av samlingarna nu återfinns.

Materialets fördelning på landskap är inte helt jämn och det skulle också vara omöjligt att uppnå en fullständig jämlikhet. I vissa socknar och landsändar har folk helt enkelt varit mycket aktivare när det gäller att uppteckna och ge ut ordlistor än i andra och detta avspeglas i samlingarna. OSD:s arkivarier har emellertid kontinuerligt ägnat extra tid åt excerpering från missgynnade delar av landet och därigenom uppnått en tillräckligt god nivå för alla landskap, så att en ordbok ska kunna baseras på materialet. Vid redigeringen av SDL kommer även ett antal andra källor att ingå i utgångsmaterialet, såsom vissa tryckta ordböcker och annat arkivmaterial, i syfte att förbättra den geografiska fördelningen ytterligare.

Som ett alternativ till att sortera ihop samlingarna, har vi diskuterat att digitalisera dem för lagring i en databas. Sökmöjligheterna skulle därmed kraftigt utökas, inte minst vad gäller ord ingående i exempel, och samlingarna kunde göras tillgängliga för var och en på nätet. Men att skapa en databas skulle kräva en avsevärd insats i arbete och tid. Materialets storlek är en sak i sig, en annan att samlingarna till stor del utgörs av kopior av andra samlingar, där texten vanligen är handskriven och ibland kontrasterar dåligt mot bakgrunden. Uttalsuppgifterna anges med en särskild ljudskrift för svenska dialekter, det s.k. landsmålsalfabetet, eller med olika mer eller mindre individuellt utformade ljudskrifter. Sammantaget medför detta att en manuell inskrivning skulle krävas för att göra olika uppgifter sökbara. Vår bedömning i det här läget är därför att det inte är möjligt att skapa en databas innan SDL påbörjas, utan vi ser det som ett helt eget projekt att överväga längre fram.

5. Urvalskriterier

Den första verkliga stötestenen i planeringsarbetet med SDL har varit hur urvalet av lemmen ska göras. Ett strängt urval är nödvändigt, med tanke på utgångsmaterialets storlek. Vilka kriterier ska vi då utgå ifrån?

Av vår referensgrupp har vi fått rådet att basera urvalet på hur många gånger ett visst ord är belagt i samlingarna, framför allt för att själva valprocessen inte ska ta alltför lång tid. Bara ord som omfattas av minst tio belägg ur oberoende källor alternativt är belagda från minst tre landskap bör tas med. En sådan regel anser vi kan fungera som en utgångspunkt, men att den måste modifieras en del. Som nämnts

ovan varierar beläggstätheten över landet och särskilt hög är den i södra Sverige. Ett urval baserat på en strikt räkning av belägg skulle leda till att dessa landskap finge oproportionerligt stort utrymme i SDL. Regeln behöver därför viktas något, så att ord från sämre belagda landskap behandlas något mer generöst än ord från välbelagda delar av landet. Vi anser dessutom att man bör kunna tänja på en sådan regel, för att ge utrymme åt ord som kan vara av särskilt intresse för läsekretsen. Det kan gälla ord som finns även i fornvästnordiska eller i andra dialekter i Norden, växt- och djurnamn m.fl.

En annan aspekt i diskussionen om urvalskriterier är hur man ska förhålla sig till de ord som används i både dialekter och standardspråk. Det rör sig dels om arvord, dels om lånord från standardsvenskan som integrerats i dialekterna. Dessa dialektord kan dela alla betydelse med dagens svenska eller delvis skilja sig ifrån den genom att ha etablerat nya eller bibehållit gamla, i nutida svenska utdöda eller ovanliga, betydelse. Om utgångspunkten vore att SDL ska presentera det centrala och etablerade ordförrådet i dialekterna, skulle både särdialektala och med standardspråket gemensamma ord komma i fråga. Men av utrymmes- och tidsskäl kanske det ändå fordras att vi koncentrerar oss på bara särdialektala ord och betydelse i lexikonet. Det är ju dessa som användarna oftast kommer att ha anledning att slå upp. De ord som dialekterna delar med standardspråket utgör dessutom en förhållandevis stor andel av ordförrådet och överstiger i samlingarna inte sällan kriteriet på tio belägg alternativt tre landskap med bred marginal. De skulle alltså komma att kräva en god del av utrymmet i SDL. Vid en uteslutning av denna ordkategori skulle urvalsmetoden beskrivas i förordet och läsaren hänvisas till SOFI för upplysningar om ord som saknas i lexikonet.

En nackdel med denna typ av urval är att en felaktig uppfattning om att dialekterna bara består av en samling ”märkliga ord” kan vidmakthållas. Även om en förklaring till urvalet ges i förordet, är det troligen inte många användare som hittar dit, utan det dialektala ordförrådets sammansättning kommer att bedömas utifrån vad som faktiskt tas upp i lemmalistan. Ett alternativ till uteslutning kunde därför vara att behandla vissa av de med standardspråket gemensamma orden i mycket summariska artiklar där en explicit hänvisning till förordet ges. Denna variant kunde möjligen kombineras med ett högre krav på antal belägg och spridning över landskap än för de särdialektala orden. Det återstår dock att se om detta är lösningar som alls är möjliga med tanke på det begränsade utrymmet.

Ifall en redaktör är osäker på om ett visst ord finns i dagens standardsvenska, avgörs detta genom att *Nationalencyklopedins ordbok* (NEO) konsulteras. Ord som i den är markerade som provinsiala eller dialektala kommer att tas upp i SDL, eftersom en användare av NEO torde uppfatta dessa ord som möjliga att slå upp i SDL. Även en del andra ord som finns med i NEO kan komma att behandlas av oss, om vi tror att läsekretsen anser dem vara typiska dialektord, framför allt p.g.a. starkt

avvikande uttal. *Fäbus* som i dialekterna uttalas *ffjös*, *feus*, *fäx* m.m. är ett exempel på ett sådant ord.

6. Redigering

Redan på ett tidigt stadium påbörjade vi en provredigering. För att underlätta materialsituationen gjorde vi en del omflyttningar i samlingarna. Dessa har sedan lång tid varit placerade serie för serie från *a* till *ö*. Numera är de i stället uppställda bokstav för bokstav, vilket har förenklat proceduren att sammanställa allt material för ett visst ord. Vi bestämde oss också för att använda materialet till bokstaven *e* för en pilotstudie. Arbetet fördelades så att Eva arbetar med partiet *e* fram till *ek*, Karin skriver ett parti från och med *ek*. Efter åren som OSD-redaktörer kunde vi i början märka hur påverkade vi var av principer och skrivsätt som gällde för OSD. I arbetet med SDL prövar vi andra vägar. Ett parti har skickats för bedömning till vår expertgrupp, som tyckte att urvalet var alltför generöst och att artiklarna kunde kortas ned betydligt. Vi har tagit hänsyn till dessa synpunkter. Urvalskriterierna har stramats upp och en hel del tid ägnats åt att laborera med artiklarnas mikrostruktur.

Vi har också påbörjat en redigeringshandbok där principer och regler tas upp i detalj. Arbetet med den fortsätts parallellt med redigeringen. När principerna för redigering är fastlagda, avser vi att beräkna den totala tidsmängden som skulle krävas för att färdigställa SDL. Vi tänker då utgå från hur lång tid det tar att redigera t.ex. tio lådor och räkna om det till hela materialet med tillägg för efterarbete, marginaler för oförutsedda händelser osv. Hur många år det sedan kommer att röra sig om, avgörs av hur många personer det finns att fördela denna totala tidsmängd på. Inriktningen är att lexikonet ska vara klart för publicering inom en 10-årsperiod.

ed n. o. **ede** n. (ed: eï(d) o.d. JILpINb, er, air Nb; ede: ai'dje Vb, e'eid JI, e'er Nb)

1 av vatten (el. myrmark) på båda sidor omgiven landremsa, landtunga, udde e.d., ofta använd som gångställe för att dra el. bära båt från vatten till vatten SmBoDsiÖg, Sveal, Norrt. **2** landsträcka utmed älv förbi forsar (där man måste färdas till fots) ÅmLpINb.

efterfärd f. **1** efterbörd (hos kreatur) ÖIVgÖg, Sveal, Gst. **2** spår, lämningar VmGst. *Ja såg ätterfärd där räven ha tej (tagit) höna* Vm. **3** fast föda som äts vanl. efter soppa e.d. SmVg. **4** efterrätt, dessert HaVg. **5** den senare flyttningen till el. från fäbodarna HslHrj.

Figur 1. *Två preliminära artiklar från SDL-redigeringen*

7. Mikrostruktur

Varje artikel i SDL inleds av ett eller flera lemman i fetstil. Lemmana står i normaliserad form, vilket för de särdialektala ordens del innebär att de återges som de skulle ha sett ut i standardsvenska. Efter ordklass- eller genusangivelse följer ofta uttalsuppgifter placerade inom vinkelparentes. För uttalsnotationen används det vanliga alfabetet, kompletterat med ett mindre antal tecken hämtade ur landsmålsalfabetet. Samtliga tecken finns tillgängliga i den specialfont manuset skrivs med.

Vi har diskuterat uttalsparentesens utformning en hel del. I de tryckta OSD-häftena beskrivs uttalet mycket utförligt genom att ett stort antal varianter presenteras på ett starkt komprimerat sätt. Denna del av artiklarna har därför, generellt sett, varit den per rad mest tidskrävande att skriva. I SDL vill vi ha en parentes som är mer användarvänlig för den målgrupp vi vänder oss till, vilket kan innebära att antalet helt utskrivna former bör ökas och parentesen göras kortare – två åtgärder som tycks svåra att förena i praktiken. Ett sätt att minska antalet uttalsvarianter i SDL är att slå samman närliggande former och använda formuleringar som ”o.d.” oftare. Ett annat är att lemmanära uttal kan behandlas mera summariskt eller helt uteslutas. Det innebär att uttalsparentesen då lämnas därhän för de ord vars uttalsvarianter alltid ligger nära lemmat.

Ett stort antal hänvisningar blir nödvändiga att ange för att leda användaren till rätt ord. Den som söker efter ett dialektord har ofta bara hört det sägas och uttalet kan ibland ligga långt ifrån den normaliserade uppslagsform där ordet återfinns.

Geografisk utbredning ges framför allt för uttalsvarianter, betydelser och exempel, med angivelsen placerad direkt efter den uppgift den är adresserad till. Utbredningsangivelserna kan vara tämligen utrymmeskrävande, men vi anser att de inte kan avvaras när det är regionala varieteter som beskrivs. Landskap är den minsta enhet som anförs och det görs i form av förkortningar av landskapsnamnen, t.ex. Sk för Skåne och Dal för Dalarna. Landskapen anges alltid i en fast ordningsföljd från söder till norr. När en uppgift har större utbredning över landet, används också angivelser som ”Göta[and]”, ”Svea[and]” och ”Norra[and]”, ”täml. allm.” och ”allm.”. Till skillnad från OSD kommer beteckningarna Göta osv. att tas i bruk, även om det saknas belägg från ett mindre antal landskap i de aktuella områdena. Vid dessa större områden, men inte vid landskap, kan väderstreck (n, s, ö, v) sättas ut och längre uppräkningsdelar delas upp med kommatecken för att öka tydligheten. Vi har diskuterat om vissa ord kunde förses med utbredningskartor som illustrerar ordens polysemi eller uttal, men också detta är en fråga om utrymme.

Uttalsparentesen följs av betydelsebeskrivningen. För momentens ordningsföljd tillämpas den historiska principen, där den äldsta betydelsen placeras först och den yngsta sist. Denna princip kan dock frångås, om en annan ordning tycks vara mer givande för användaren. Betydelser som är svagt belagda eller av mer tillfällig karaktär förbigås. För att beskriva betydelserna används i första hand standardsvenska-

ska ekvivalenter. Men i och med att den begreppsvärld – i synnerhet den konkreta – som beskrivs i dialekterna i många stycken är föråldrad, kan vissa ekvivalenter i standardsvenskan te sig alltför ålderdomliga eller fackspråkliga för vår målgrupp. För att avgöra vilka ekvivalenter som kan anses möjliga att anföra, torde även i detta fall t.ex. NEO kunna fungera som rättesnöre. Alla ord som upptas i den, bör utan ytterligare förklaring kunna användas i SDL. I övriga fall får redaktören ta ställning till huruvida ett visst ord är möjligt att använda som enda betydelseuppgift eller måste ersättas av eller kompletteras med en förklaring.

Momenten avslutas med exempel – i de fall det finns lämpliga att anföra – vilka illustrerar ordets användning. De återges med ljudskrift, dock inte lika detaljerat som i uttalsparentesen, och översätts vid behov. De exempel vi anför är autentiska i så måtto att vi som redaktörer endast återger vad som uppges i källorna. Sedan är det en annan sak om det i varje fall rör sig om autentiskt talspråk, i den meningen att någon verkligen fällt yttrandet på just detta sätt i den aktuella socknen. Den vanliga har väl varit att det är upptecknaren som ansett att så här kan eller kunde man säga och skrivit ned detta. Ibland är emellertid våra belägg excerpter ur utskrifter av fonogram eller ur uppteckningar där ord och meningar antecknats under tiden som en meddelare talat.

8. Dialektavdelningen idag

En stor förändring sedan vi startade arbetet med SDL är att Dialektavdelningen från och med januari i år (2005) har en ny forskningschef, Margareta Svahn, som vill driva SDL som ett huvudprojekt i vilket avdelningens personal, för närvarande fyra personer, deltar på olika sätt och olika mycket.

En gång i månaden har vi lexikondag då vi diskuterar principer, problem el. likn. som har med SDL att göra. Som ett moment under våra lexikodagar ingår också ett par timmars samordning. För att alla ska bli hemmastadda i samlingarna har vi undervisat våra ”nykomlingar” i samordningens principer. Tillsammans gör vi vad vi kan för att arbeta oss fram i OSD:s samlingar med målet *en* serie i sikte.

9. Planer för framtiden

Det ingår i våra planer för framtiden att utforma en projektansökan. Som nämnts är det helt nödvändigt att knyta fler personer till Dialektavdelningen om SDL ska kunna genomföras inom rimlig tid. I en ansökan måste vi givetvis ge en tydlig beskrivning av hur vi tänkt oss att dialektlexikonet ska se ut och hur arbetet med det ska fungera. Övergripande frågor, som vi bör ha svaren på när den tiden kommer har vi redogjort för ovan och några svar har vi också kunnat leverera. I kort sammanfattning är frågorna och svaren så långt följande: Vilken är målgruppen för SDL? Lexikonet riktar sig till den dialektintresserade allmänheten. – I vilken form ska SDL publiceras? Lexikonet ska ges ut som bok, men det kan bli aktuellt med en

nätversion också. – Vilket omfång ska lexikonet ha och vilken tidsplan räknar vi med? Något entydigt svar på dessa båda frågor har vi inte kommit fram till ännu, men tills vidare tänker vi oss ett omfång av högst 1 200 sidor och vi räknar med att SDL ska finnas färdigt inom 10 år, under förutsättning att vi får en rejäl resursförstärkning. – Den allra svåraste frågan är, och det har vi också poängterat, hur ska urvalet göras? Den frågan är vi naturligtvis inte färdiga med än, men vi har börjat och tänker fortsätta att arbeta med en modifierad form av den tidigare nämnda tiobeläggsmodellen. – Till sist: hur löser vi frågan om artiklarnas mikrostruktur? Här närmar vi oss ett svar, vilket framgår av de exempel som visats ovan. Vi är helt medvetna om att mycket arbete återstår, men det är med större tillförsikt vi blickar framåt i år, än som var fallet våren för två år sedan.

Litteratur

- NEO = *Nationalencyklopedins ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. Höganäs: Bra böcker AB 1995.
- Nynorskordboka. Definisjons- og rettskrivingsordbok*. Oslo 1991.
- Rietz, Johan Ernst: *Svenskt dialekt-lexikon. Ordbok öfver svenska allmogespråket. 1-2*. Malmö, Köpenhamn, Leipzig & London. [1862-]1867. Register och rättelser av Erik Abrahamson. 1955. Uppsala & København. (Kgl. Gustav Adolfs Akad.).
- OSD = *Ordbok över Sveriges dialekter*. Uppsala: Språk- och folkminnesinstitutet.
- SAOB = *Ordbok över svenska språket*. Utg. av Svenska Akademien. 1-. 1898-. Lund.
- SDL = *Svenskt dialektlexikon*. Uppsala: Språk- och folkminnesinstitutet.
- SOL = *Svenskt ortnamnslexikon*. Språk- och folkminnesinstitutet. Uppsala 2003.
- Svensk ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. Tredje rev. uppl. Norstedts. Stockholm 1999.

Multilingval elektronisk idiomordbog

A multilingual electronic dictionary of idioms. In this article, I present a research project on multilingual lexicography and phraseology. The aim of the project is to analyse linguistic and lexicographical theories and evaluate practical solutions in order to provide a basis for the conceptual work on and the compilation of an electronic German ↔ Icelandic ↔ Danish ↔ German dictionary of idioms.

1. Indledning: Projektets baggrund

Projektet multilingval elektronisk idiomordbog bygger på resultaterne fra et forskningsprojekt med en tysk-islandsk fraseologisk ordbog som blev finansieret af *The Icelandic Centre for Research*. Projektet har to overordnede formål: (1) At udarbejde et teoretisk koncept for multilingvale elektroniske ordbøger og (2) at omsætte konceptet til praksis i en tysk ↔ islandsk ↔ dansk ↔ tysk database med idiommer (med mulighed for udvidelse med fraseologismer (i bred forstand) og etordsleksemer).

2. Teoretisk grundlag

2.1. Fraseologi

Et af de vigtigste grundlag for fraseografisk arbejde dannes i den ensprogede fraseologiske grundforskning. Det drejer sig om udførlige ensprogede korpusbaserede analyser og beskrivelser af fraseologismer. Først når denne slags beskrivelser foreligger, kan man lave kontrastive analyser af fraseologiske relationer og udarbejde kontrastive fraseologikorpora, der så danner grundlaget for ordbøger. I dette projekt er der tale om tysk ↔ islandske, dansk ↔ islandske og dansk ↔ tyske fraseologismer. Det efterfølgende er en kort beskrivelse af det fraseografiske grundlag i de tre sprog.

2.1.1. Tysk fraseologi

Inden for den tyske fraseologiforskning findes der allerede et bredt teoretisk grundlag, der vil indgå i mit projekt (sml. udførlig litteraturliste i Hallsteinsdóttir 2001 og 2006).

Af aktuelle forskningsprojekter der fokuserer på en korpusbaseret eller kontrastiv beskrivelse af tyske idiommer, kan f.eks. nævnes projekterne “Kollokationen im Wörterbuch” ved Berlin-Brandenburgische Akademie der Wissenschaften (sml. Fellbaum et al. under udg. og kollokationen.bbaw.de/index.html) og “Syntax und Semantik idiomatisierter verbaler Phraseologismen in einem abgestuft kompositionellen Grammatikmodell” ved universitetet i Jena (sml. Soehn/Römer under udg. og

www.personal.uni-jena.de/~xcr/website/forschung/IVP-Web/main.htm). Yderligere er idiommer integreret som opslagsord i korpusordbogen “Projekt Deutscher Wortschatz” (www.wortschatz.uni-leipzig.de) ved universitetet i Leipzig. I EU-forskningsprojektet EPHRAS (sml. Jesenšek under udg. og www.ephras.org) ved universitetet i Maribor indgår tysk som et af sprogene ved udviklingen af en multilingual lærer-cd med fraseologismer.

Med hensyn til grundlaget for fraseografisk arbejde er der dog stadigvæk behov for forskning på bestemte områder, f.eks. vedrørende beskrivelsen af semantiske og pragmatiske egenskaber, afdækning af hhv. onomasiologiske og konceptuelle strukturer og beskrivelse af kontrastive relationer til islandsk og dansk.

2.1.2. *Islandsk fraseologi*

Inden for ensproget islandsk fraseologi og fraseografi har Jón Hilmar Jónssons forskning en særstilling. Sammen med Jón G. Friðjónssons (Friðjónsson 1993) og Halldór Halldórssons (Halldórsson 1991) værker danner hans arbejde (Jónsson 2002) et grundlag for beskrivelse af islandsk fraseologi i kontrastiv sammenhæng.

2.1.3. *Islandsk ↔ tysk fraseologi*

Bortset fra mine forskningsprojekter (Hallsteinsdóttir 2001, Hallsteinsdóttir under udg. a, b og c) og nogle korte artikler findes der kun en større teoretisk afhandling inden for kontrastiv tysk-islandsk fraseologi (Sverrisdóttir 1987). Her mangler altså stort set et grundlag for fraseografisk arbejde.

2.1.4. *Dansk fraseologi*

Inden for den danske fraseologi kan man konstatere øget forskningsinteresse i løbet af de sidste år, bl.a. med grundlæggelsen af Dansk Selskab for Fraseologi (DANFRAS). Der findes nyere oversigter over dansk fraseologiforskning og dansk fraseografi (f.eks. Farø 2004a+b og Farø under udg.), større idiomordbøger er udkommet og en elektronisk idiomordbog er blevet publiceret på internettet (www.idiomordbogen.dk) (sml. review article Farø 2004b og Almind et al. under forb.). Yderligere arbejdes der på et stort online-opslagsværk over det danske sprog, hvori beskrivelse af idiommer indgår (sml. Henrik Lorentzens og Lars Trap-Jensens artikel i denne publikation og www.ordnet.dk). Men alligevel er der stadig mange uløste problemer i forbindelse med systematiseringen og beskrivelsen af dansk fraseologi, eller, med Ken Farøs egne ord: “Der venter en opgave forude med at systematisere den fraseologiske forskning og beskrivelse i Danmark.” (Farø 2004b, 234).

2.1.5. *Tysk ↔ dansk fraseologi*

I den tysk-danske fraseologi findes nogle få nyere teoretiske og kontrastive undersøgelser (f.eks. Farø 2004a, Farø under udg. og Farø under forb.), der bl.a.

beskæftiger sig med ækvivalensproblemer og fraseografiske løsninger. Den danske og den dansk-tyske fraseologiforskning danner et forholdsvis solidt grundlag for videre forskning.

2.1.6. Dansk ↔ islandsk fraseologi

Selv om der i Island er en lang tradition for at undervise i dansk som fremmedsprog både i gymnasierne og i folkeskolerne, har der hidtil ikke været lagt nogen stor vægt på kontrastiv pædagogisk-leksikografisk forskning i disse to sprog. Bortset fra Auður Hauksdóttirs (Islands Universitet) nyere (p.t. ikke publicerede) undersøgelser i islandsk-dansk fraseologi findes der hverken teoretiske fraseologiske afhandlinger, større kontrastive arbejder eller fraseologiske ordbøger. Mit projekt er derfor et pilotprojekt i forbindelse med beskrivelsen af dansk-islandsk fraseologi.

2.2. Fraseografi

Mangler i fraseologiske ordbøger og forslag til udbedring af disse har været et hyppigt diskuteret tema i den tysksprogede fraseologiforskning inden for de sidste 20 år. Denne diskussion danner en bred teoretisk basis for fraseografisk arbejde. Diskussionen tog sin begyndelse med Werner Kollers forslag til en ideel tosproget idiomatisk lærerordbog (Koller 1987); dette forslag er delvis fulgt i trykte ordbøger, bl.a. findes onomasiologisk-tematisk opbygning (organisation efter betydning/koncepter) og øvelser i Regina Hesskys og Stefan Ettingers tyske onomasiologiske idiomordbog (Hessky/Ettinger 1997).

Først med muligheden for at bruge computere som leksikografisk medium er der kommet et arbejdsgrundlag for at realisere mange af de teoretiske forbedringsforslag. Det bliver muligt at bryde den todimensionale, lineære struktur trykte bøger har, op. Alligevel må man anse mange elektroniske ordbøger for at være biprodukter af trykte ordbøger. De opbygges efter koncepter for trykte ordbøger og har de samme begrænsninger, de udnytter altså ikke de muligheder computere byder på som medium (f.eks. ubegrænset plads, multimedia, hypertext, ikke-lineær opbygning).

Den teoretiske diskussion inden for elektronisk fraseografi bærer i stor grad præg af at forskningsområdet er under udvikling, og at der mangler teoretisk forarbejde i fraseologien og i leksikografien, f.eks. i forbindelse med ækvivalensproblematikken og spørgsmålet om grundlaget for lemmaselektion (sml. udførligt om fraseografiens problemstillinger i Hallsteinsdóttir 2006).

2.3. Metaleksikografisk arbejde

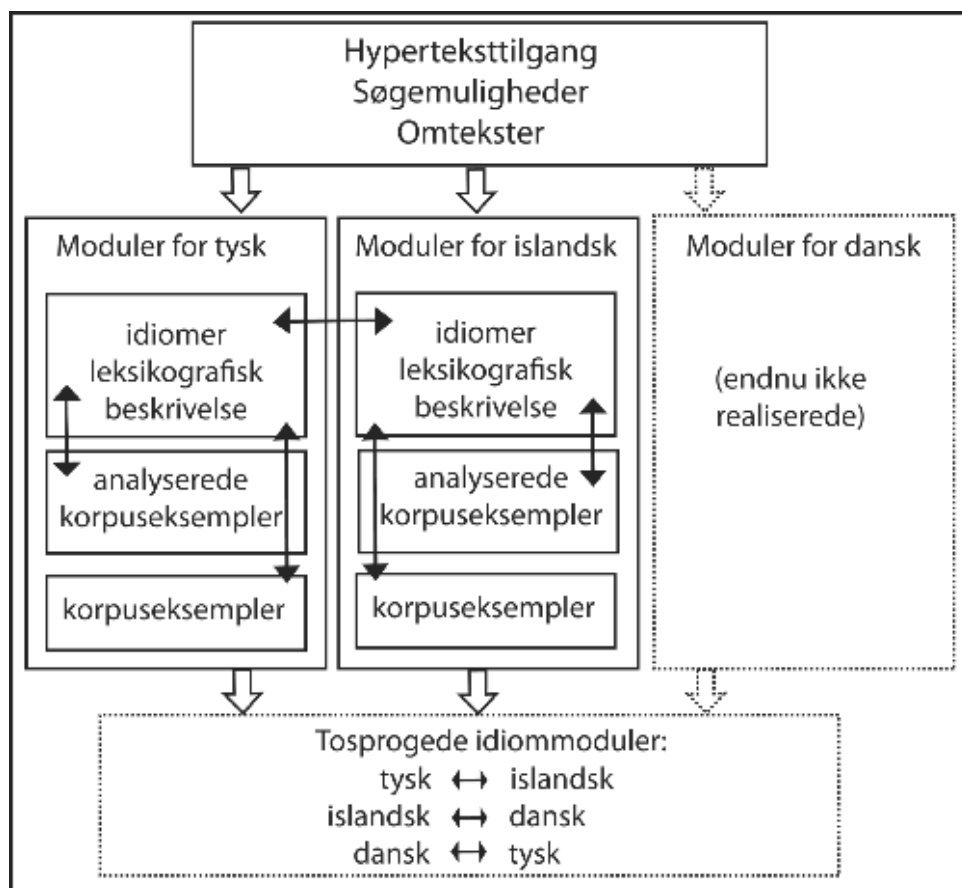
En stor del af arbejdet ved det aktuelle projekt består i metaleksikografiske overvejelser i forbindelse med udformningen af et leksikografisk koncept for en multilingval database, bl.a.:

1. design af multilingval databasestruktur,
2. design af ordbogens informationsenheder i forhold til ordbogens funktioner (i forbindelse med brugergrupper og ordbogsbrugernes behov) og
3. løsninger til ækvivalensproblematikken i multilingvale ordbøger.

2.3.1. Design af multilingval leksikografisk databasestruktur

Ved valget af databasesoftware skulle følgende kriterier kunne opfyldes:

1. Der skal være mulighed for databasemoduler for hvert sprog, dvs. den overordnede struktur skal være modulær (sml. figur 1), bl.a. med hensyn til senere udvidelse med flere sprog.



Figur 1. Overordnet leksikografisk struktur

2. Inden for modulerne skal strukturen være en ikke-lineær hypertextstruktur med en hierarkisk opbygning. Informationer inden for modulerne skal forbindes med hyperlinks.
3. Databaseprogrammet skal byde på forskellige muligheder for publicering, varierende tilgangsstruktur og layout, bl.a. forskellige søge- og rapportmuligheder.
4. De ikke-leksikografiske kriterier, brugervenlighed (der skal bruges minimal tid på tilegnelse af programmet) og pris (projektet har begrænset budget), spillede ligeledes en stor rolle da valget faldt på databasesoftwarens askSam (www.asksam.com).

2.3.2. Design af databasens informationsenheder

Design af ordbogens informationsenheder udarbejdes i forhold til ordbogens funktioner i forbindelse med brugergrupper og ordbogsbrugernes behov (se om ordbogsfunktioner i Bergenholtz 1997 og Tarp 1998). Selv om der findes en del teoretiske afhandlinger om problemet, er brug af ordbøger ud fra min erfaring tit meget individuel, for ikke at sige tilfældig, idet der hos de fleste brugere er tale om en selv lært kompetence. Ordbogsdidaktik er mere eller mindre ukendt i skolen, dvs. der finder nærmest ingen systematisk undervisning i ordbogsbrug sted. Alligevel prøver jeg at give et bud på hvordan man kunne strukturere de leksikografiske informationer i forhold til potentiel brug (sml. også Hallsteinsdóttir under udg. a og c).

Informationsstrukturen er opbygget efter ordbogsbrugernes potentielle behov for informationer for sprogreception (høre, læse), sprogproduktion (tale, skrive) og oversættelse. Der er tale om et informationshierarki som udfoldes ved hjælp af hyperlinks:

Niveau 1: Sprogreception.

Niveau 2: Sprogproduktion.

Niveau 3: Oversættelse af sprog.

I databasen bliver der yderligere taget højde for informationer til indlæring og korrektur, og den indeholder ligeledes en del metaleksikografiske informationer. Efterfølgende gennemgås databasens informationsenheder summarisk.

Databasen indeholder en del ensprogede informationer, såvel for sprogreception som sprogproduktion. Der er tale om:

- 1) En minimalform, f.eks. *grünes Licht*, som fungerer som metaleksikografisk identifikation af idiomet og har funktion af bookmark for hyperlinks.
- 2) De typiske syntaktiske strukturer, som på den ene side er vigtige med henblik

på identifikation af idiomer, især når formen eller dele af konstruktionen forekommer i andre kontekster (man kunne nok også sige: i ordret brug), og som på den anden side er uundværlige ved sprogproduktion:

- (a) *grünes Licht für jemanden/etwas*
 - (b) *jemand/etwas gibt (jemandem/etwas) grünes Licht (für/zu etwas)*
 - (c) *jemand/etwas erhält grünes Licht (für/zu etwas) (von jemandem) (mit etwas)*
 - (d) *jemand/etwas hat grünes Licht*
 - (e) *etwas bedeutet grünes Licht für jemanden/zu etwas*
- 3) Hyperlinks til analyserede korpuseksempler med en udførlig beskrivelse af grammatiske regler og restriktioner og eksempler på negation. Yderligere er der hyperlinks til andre idiomer med samme eller lignende form: *jemand rechnet mit grünem Licht*.
 - 4) Idiomets betydning(er) beskrives ud fra de analyserede korpuseksempler, og synonyme og antonyme idiomer angives. Desuden forklares idiomets etymologi og pragmatiske restriktioner.

De ensprogede informationer suppleres med tosprogede informationer som fungerer som hjælpemiddel fortrinsvis ved sprogreception og oversættelse. Disse er:

- 1) ordret oversættelse til islandsk/dansk,
- 2) betydningsangivelse på islandsk/dansk,
- 3) angivelse af ækvivalente islandske/danske idiomer og
- 4) en forklarende kommentar til ækvivalensrelationen.

Metaleksikografiske informationer i databasen bruges på den ene side ved struktureringen af de forskellige tilgangs- og søgemuligheder (alfabetisk liste med de enkelte komponenter, søgning efter de enkelte ord i idiomer, de enkelte komponenters semantiske felter, angivelse af koncepter til konstruktion af onomasiologisk tilgang til idiomernes betydning), på den anden side giver de mulighed for videregående forskning på grundlag af projektets resultater. Som eksempler på den slags metaleksikografiske informationer nævnes angivelse af frekvens i det tyske korpus *Deutscher Wortschatz* (www.wortschatz.uni-leipzig.de) og bekendthedsgrad hos tyske sprogbrugere (sml. Hallsteinsdóttir/Šajánková/Quasthoff, under forb.).

2.3.3. Ækvivalensproblematikken

Selv om der findes rigelig teoretisk litteratur om idiomer og ækvivalens, er en adækvat gengivelse af ækvivalens i en multilingval idiomatisk database stadigvæk en kompliceret problematik at redegøre for (sml. en udførlig teoretisk diskussion og empirisk undersøgelse af idiomer og ækvivalens i Farø under forb.). Ækvivalens beskrives i dette projekt altid kun mellem to sprog ad gangen. Der er ingen

begrundelse for bestemmelse af multilingval ækvivalens idet der ikke findes nogen potentiel brugssituation til denne.

Der er overordnet tale om to ækvivalenstyper:

- 1) *Idiom* ↔ *idiom(er)*. Denne type kræver beskrivelse af semantiske, pragmatisk og strukturelle forskelle mellem funktionelt ækvivalente idiommer. Hovedvægten i dette projekt ligger følgelig på situativ/funktionel ækvivalens (pragmatik og semantik, semantiske relationer). Idiomformen er kun relevant ved "tematiseret" brug (Farø under forb.). Sproglige (semantiske, strukturelle, etc.) relationer og/eller metaforik inddrages i begrænset omfang, først og fremmest med henblik på "falske venner".
- 2) *Idiom* ↔ *etordsleksem(er)*. Denne type betragtes i traditionel fraseologiforskning som nul-ækvivalens. Nul-ækvivalens er dog udelukkende en metodisk konsekvens af en teoretisk afgrænsning mellem idiommer og ikke-idiommer. Ægte funktionel nul-ækvivalens er med stor sandsynlighed meget sjælden. I første omgang er lemmataene i databasen dog af metodiske årsager begrænset til idiommer (med mulighed for udvidelse med andre sproglige enheder senere).

Idiommer indgår desuden i relationer til andre sproglige enheder (f.eks. synonymi, antonymi, orddannelse) som burde inddrages i den leksikografiske beskrivelse. Denne form for beskrivelse kræver dog at sproget betragtes i sin helhed. En konsekvent fortsættelse heraf ville være at sætte spørgsmålstegn ved "rene" idiomordbøger, ikke for at gøre mit eget arbejde overflødig, men for at foreslå integration af idiommer i omfattende leksikografiske opslagsværker (leksika, tesaurusser). Integration af idiommer (og andre flerordsforbindelser) i leksikografiske opslagsværker og beskrivelse af intrasproglige relationer stiller den ensprogede leksikografi over for store udfordringer med hensyn til formulering af betydnings-, synonym- og antonymangivelser. I multilingval leksikografi med beskrivelse af intrasproglige og intersproglige relationer er udfordringerne endnu større idet de intrasproglige relationer ikke nødvendigvis er symmetrisk *intersprogligt* set, dvs. ens i to sprog.

3. Sammenfatning

Beskrivelse af fraseologismer er et højaktuelt forskningstema. En del af resultaterne i dette projekt forventes at blive en konkret beskrivelse af dansk, tysk og islandsk fraseologi og kontrastive analyser i en eksemplarisk database der f.eks. kan bruges som basis for lærebøger eller i andre undervisningsmaterialer. Databasen skal desuden, med forudgående kontrastive analyser af de pågældende sprog, kunne udvides med andre sprog. En brugbar teoretisk konception for multilingvale ordbøger danner forudsætningen for det leksikografiske arbejde. Denne konception åbner muligheder for fraseografiske arbejder i andre sprog.

Det forventes ligeledes at projektets resultater vil være et vigtigt bidrag til teoretisk viden om fraseologi som sprogligt fænomen. Endvidere forventes det at projektets resultater bliver til et grundlag for (og opmuntring til) yderligere fraseologiforskning, især i kontrastiv og leksikografisk sammenhæng.

Litteratur

Ordbøger:

- Friðjónsson, Jón G. 1993: *Mergur Málsins. Íslensk orðatiltæki. Uppruni, saga og notkun*. Reykjavík: Örn og Örlygur.
- Halldórsson, Halldór 1991: *Íslenskt Orðtakasafn*. Reykjavík: Almenna bókafélagið.
- Jónsson, Jón H. 2002: *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandskrá*. Reykjavík: JPV Útgáfa.
- Hessky, Regina/Ettinger, Stefan 1997: *Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene*. Tübingen: Gunter Narr Verlag.

Anden litteratur:

- Almind, Richard/Bergenholtz, Henning/Vrang, Vibeke under forb.: Theoretical and Computational Solutions for Phraseological Lexicography. I: *Linguistik Online*.
- K. Farø og E. Hallsteinsdóttir (udg.): *Neue theoretische und methodische Ansätze der Phraseologieforschung*. <http://linguistik-online.com>.
- Bergenholtz, Henning 1997: Polyfunktionale ordbøger. I: *LexicoNordica* 4, 15-28.
- Farø, Ken 2004a: Am Anfang. Zur Erforschung der dänisch-deutschen Idiomatik. I: J. Wirrer og Cs. Földes (udg.): *EUROPHRAS 2002. Phraseologismen als Gegenstand sprach- und kulturwissenschaftlicher Forschung*. Baltmannsweiler: Schneider Verlag Hohengehren, 105-117.
- Farø, Ken 2004b: Idiomor på nettet: Den danske idiomordbog og fraseografien (<http://www.idiomordbogen.dk>). I: *Hermes, Journal of Linguistics* no. 32, 201-235.
- Farø, Ken under udg.: Danish Phraseography. I: H. Burger, D. Dobrovolskij, P. Kühn og N.R. Norrick (udg.): *Phraseology. Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (HSK). Berlin/New York: Walter de Gruyter.
- Farø, Ken under forb.: *Idiomatizität – Ikonizität – Arbitrarität: Beitrag zu einer Theorie phraseologischer Äquivalenz*. Ph.d.-afhandling ved Københavns Universitet.
- Fellbaum, Christiane/Kramer, Undine/Neumann, Gerald under udg.: Korpusbasierte lexikografische Erfassung und linguistische Analyse deutscher Idiome. I: A.H. Buhofer (udg.): *EUROPHRAS 04*.
- Hallsteinsdóttir, Erla 2001: *Das Verstehen idiomatischer Phraseologismen in der Fremdsprache Deutsch*. Hamburg: Verlag Dr. Kovač. http://www.verlagdrkovac.de/0435_volltext.htm.
- Hallsteinsdóttir, Erla 2006: Phraseographie. I: *Hermes* 36.
- Hallsteinsdóttir, Erla under udg. a: A bilingual electronic dictionary of idioms. I: H. Gottlieb og J.E. Mogensen (udg.): *Dictionaries, Lexicographical Options and User Needs. Selected*

- Papers from the 12th International Symposium on Lexicography*. Amsterdam: John Benjamins.
- Hallsteinsdóttir, Erla under udg. b: Vom Wörterbuch zum Text zum Lexikon. I: U. Fix, G. Lerchner, M. Schröder og H. Wellmann (udg.): *Zwischen Lexikon und Text – lexikalische, stilistische und textlinguistische Aspekte*. Leipzig: Sächsische Akademie der Wissenschaften zu Leipzig.
- Hallsteinsdóttir, Erla under udg. c: Konzeption und Erstellung einer computergestützten zweisprachigen Phraseologiesammlung Isländisch – Deutsch. I: A.H. Buhofer (udg.): *EUROPHRAS 04*.
- Hallsteinsdóttir, Erla/Šajánková, Monika/Quasthoff, Uwe under forb.: Phraseologisches Optimum für Deutsch als Fremdsprache. Ein Vorschlag auf der Basis von Frequenz- und Geläufigkeitsuntersuchungen. I: *Linguistik Online*. K. Farø og E. Hallsteinsdóttir (udg.): *Neue theoretische und methodische Ansätze der Phraseologieforschung*. <http://linguistik-online.com>
- Jesenšek, Vida under udg.: Zum Problem der Äquivalenz in mehrsprachigen phraseologischen Wörterbüchern. I: A.H. Buhofer (udg.): *EUROPHRAS 04*.
- Koller, Werner 1987: Überlegungen zu einem Phraseologie-Wörterbuch für Fremdsprachenunterricht und Übersetzungspraxis. I: H. Burger og R. Zett (udg.): *Aktuelle Probleme der Phraseologie*. Frankfurt a. M.: Peter Lang, 109-120.
- Soehn, Jan-Philipp/Römer, Christine under udg.: Wann ist ein Idiom ein Idiom? Eine Analyse von Phraseologismen ohne freie Lesart. I: A.H. Buhofer (udg.): *EUROPHRAS 04*.
- Sverrisdóttir, Oddný G. 1987: *Land in Sicht. Die kontrastive Untersuchung deutscher und isländischer Redensarten aus der Seemannssprache*. Frankfurt a. M.: Peter Lang.
- Tarp, Sven 1998: Leksikografien på egne ben. Fordelingsstrukturer og byggede i et brugerorienteret perspektiv. I: *Hermes, Journal of Linguistics* 21, 121-137.

Internetadresser:

- <http://corpora.informatik.uni-leipzig.de/>
<http://kollokationen.bbaw.de/index.html>
<http://www.asksam.com>
<http://www.dendanskenetordbog.dk>
<http://www.ephras.org>
<http://www.idiomordbogen.dk>
<http://www.ordanet.is>
<http://www.ordnet.dk>
<http://www.personal.uni-jena.de/~xcr/website/forschung/IVP-Web/main.htm>
<http://www.wortschatz.uni-leipzig.de>

Muligheder for samarbejde mellem ordbøger der arbejder inden for det nordiske sprogområde

Plans for co-operation between Nordic-language dictionary projects. The West-Nordic language area covers Norway, the Faeroes and Iceland. The languages of these three countries all develop from a common source, usually called Old Norse (Danish: 'norrønt'). The article describes the way in which dictionaries of the modern forms of these languages can benefit from being able to refer to corresponding words both in the sister languages and in Old Norse. *Ordbog over det norrøne prosasprog // A Dictionary of Old Norse Prose* makes considerable use of lexicographical material from related medieval and modern languages when undertaking semantic analysis. An initiative is under way to enable dictionary projects that cover all medieval Nordic languages to co-operate on an interlinked presentation of their lexical data and some background materials on the World Wide Web.

1. Vestnordiske ordbøger og nabosprog

Det vestnordiske sprogområde dækker de moderne sprog der tales i Norge, på Færøerne og på Island. Fælles for disse sprog er at de har et historisk udgangspunkt, nemlig det norrøne sprog. Leksikografiske beskrivelser af de moderne sprog varetages af *Norsk Ordbok* i Norge, *Føroysk orðabók* på Færøerne og *Orðabók Háskólans* på Island. *Ordbog over det norrøne prosasprog // A Dictionary of Old Norse Prose* (ONP) i København beskæftiger sig, som navnet siger, med det norrøne sprog.

Der er tradition for at ordbøger over moderne sprog giver henvisninger til ældre, historiske former for enkelte ord. Det giver ordforrådet en historisk dimension og er oplysende for den interesserede læser af ordbogen. Derimod er der kun sjældent henvisninger fra et moderne sprog til et andet. Inden for det nordiske sprogområde ser vi således i ordbøger ofte henvisninger til ældre former af et ord, som f.eks. fra moderne norsk til norrønt, men vi ser sjældent en henvisning til et ords form i et andet moderne sprog, som f.eks. mellem moderne norsk og islandsk eller færøsk, selvom en sådan henvisning ofte kan være relevant og oplysende. Ordbøgerne ser i almindelighed sig selv som nationale ordbøger med sig selv som udgangspunkt og mål, og for langt de fleste brugere af ordbøgerne er det vel også tilstrækkeligt.

Ordbog over det norrøne prosasprog har til opgave at beskrive det gamle sprog i det vestnordiske sprogområde, altså gammelnorsk (indtil ca. 1370) og islandsk (indtil 1540). Vi beskriver det sprog som ordbøgerne over de moderne sprog henviser til. Redaktørerne på ONP ser sagen fra en anden side. Vi kan se det ordforråd vi beskriver, udvikle sig netop til alle de moderne sprog under et. Fra vores synsvinkel

er grænserne mellem moderne norsk, færøsk og islandsk mere et spørgsmål om dialektgeografi end om nationale sprog.

2. ONP og de moderne sprog

Redaktørerne ved ONP løber jævnligt ind i problemer når vi skal udarbejde betydningsbeskrivelser. Det skyldes bl.a. at vi ikke kan gribe til ældre stadier af sproget, for vores materiale er det ældste. Mange ord er kun meget sparsomt belagt, og uden en god kontekst kan det være vanskeligt at fastlægge betydningen af et enkelt ord. Redaktionen er nødt til at gribe til de nyere repræsentanter for sprogene, især yngre norsk og islandsk.

Et eksempel. Ordet **belgssonr** optræder kun én gang i ONP's materiale, nemlig i et norsk diplom fra 1360 i formen *bægsson*. Det er indlysende at førsteledet er det maskuline substantiv *belgr*, hvor ONP har definition 5 '(menneskets) "bælg", krop, vom, udposning // skin (of a person), body, belly, distension/bulge'. Betydningen af *belg-* i *belgssonr* ligger inden for dette område, men er dog ikke ganske klar. Her kommer *Norsk Ordbok* os til hjælp. Under opslagsordet **belg** i definition 3 b) findes betydningen 'sviv., ein som er tjukk og feit og har stor mage'. Denne definition passer fint med ONP's belæg, og redaktionen har udmøntet den i ONP's definition '(pejor.) "søn af tykvom" // (pejor.) "son of big-belly"', med behørig henvisning til *Norsk Ordbok*.

Det har altid været et håb for redaktørerne ved ONP at det en dag vil være muligt at få etableret et samarbejde med andre ordbøger inden for det nordiske sprogområde. Håbet og drømmen har specielt drejet sig om et samarbejde der består i at både brugere af ordbøgerne og redaktionerne har fri adgang til udveksling af ordbogsdata og baggrundsmateriale. Tanken om at redaktioner og brugere frit og let kan få alle oplysninger om et ord og dets former i geografisk og historisk spredning præsenteret direkte, er meget tillokkende.

3. Et fælles ordbogsinitiativ

Nu hvor mange ordbøger der beskæftiger sig med de nordiske sprog, er blevet tilgængelige på internettet, er dette håb og disse drømme ved at kunne tage form. Den 29. og 30. april 2005 blev der afholdt et møde i København på Den arnamagnæanske samling primært for middelalderordbøger i hele Norden, altså også danske og svenske. Mødet var arrangeret af Menota (Medieval Nordic Text Archive).

Formålet med mødet var at sondere mulighederne for et samarbejde mellem de ordbøger der beskæftiger sig med de nordiske middelaldersprog, med henblik på om det vil være muligt at oprette et fælles format for de involverede ordbøger.

Hensigten er at det skal være muligt for netbrugere at kunne håndtere ordbøgerne som en helhed. Det skal være muligt at kunne slå op på et ord i f.eks. norrønt og derudfra kunne få oplysninger om ordets former i andre nordiske sprog, f.eks. i

mellemnorsk, nynorsk og moderne islandsk. Men også i andre nordiske middelalder-sprog som f.eks. gammeldansk og fornsvensk.

Ordbøgerne har vidt forskellige redaktionelle edb-systemer og redskaber, der ikke kan eller skal modificeres. Det endelige mål er en integration af materialet på internettet fra de forskellige ordbøger, således at ordbogsdata og baggrundsmateriale set med brugernes øjne kommer til at udgøre en enhed. Hvorledes det skal se ud og fungere, er under overvejelse. Indtil et fælles koncept er udarbejdet, er et væsentligt skridt undervejs at opbygge et forholdsvist simpelt linksystem mellem ordbøgerne, således at brugerne i det mindste bliver mindet om at et ord er udbredt over hele det nordiske sprogområde og gennem hele de nordiske sprogs attesterede historie. For at simplificere konceptets udvikling er det et skridt på vejen at kun middelalderordbøgerne i første omgang indarbejdes i planerne. Hensigten er senere at inddrage de store ordbøger over de nationale sprog i Norden.

Der blev på mødet nedsat to arbejdsgrupper. Den ene skal varetage den overordnede planlægning og ideudvikling, mens den anden skal søge at finde frem til et fælles format hvori de forskellige ordbøgers data kan præsenteres på internettet, og på anden måde søge at løse de tekniske problemer der er forbundet med realiseringen af ideerne. De enkelte ordbøgers interne udviklingsværktøjer bliver ikke berørt af dette initiativ.

Herudover er det hensigten at – i første omgang – ordbogsredaktionerne på de involverede ordbøger skal have let og direkte elektronisk adgang til de andre ordbøgers ressourcer som f.eks. elektroniske tekster og bibliografiske data.

Arbejdsprocessen fortsætter, og der blev afholdt møde i arbejdsgruppen d. 10. september 2005. Man kan følge med i projektets udvikling på Menotas website <http://www.menota.org>.

Geografiske hol i ei nasjonal ordbok – om geografisk representasjon av kjeldene i Norsk Ordbok

Geographical holes in a national dictionary – about geographical representation of the sources in Norsk Ordbok. Norsk Ordbok covers both the written language 'nynorsk' and the spoken language in Norway. As a national dictionary, there is a demand for geographical representation when it comes to the spoken language. Nevertheless this survey shows that the spoken language in the northern part of Northern Norway and in most cities, is very poorly documented. The main reason for this seems to be a heritage from an earlier neogrammatical influenced linguistic tradition which was characterised by studies with Old Norse as one comparative component. Little attention was paid to dialects which were a product of language contact, which is the case for both the dialects in the northern part of Northern Norway and in the cities until late in the 1970's. Norsk Ordbok was organised as a project i 2002 and there is now some effort put in balancing the source material with a view to geographical representation.

1. Innleiing

For ein som ikkje er kjend med den norske språksituasjonen, kan kanskje tittelen på denne artikkelen verke overraskande. I mange land er normalsituasjonen at eit nasjonalt ordboksverk beskriv skriftspråket som igjen ligg nært opp til standardtalemålet i det økonomiske og politiske sentrumet, oftast hovudstadsområdet. I eit slikt perspektiv er det merkeleg å snakke om geografiske hol, fordi alle område utanfor dette sentrumet på sett og vis representerer eit einaste stort hol. Norsk Ordbok (NO) er ikkje eit slikt ordboksverk. Tvert imot skal verket – når det gjeld talespråket – dekkje varietetane i heile landet. Derfor kan det likevel ha noko for seg å undersøkje den geografiske fordelinga av kjeldene i NO.

NO har undertittelen *Ordbok over det norske folkemålet og det nynorske skriftmålet*. Verket har altså eit todelt mål. Det skal gje ei uttømmende vitskapleg framstilling av ordtilfanget i dei *norske dialektane* frå 1600-talet til i dag og av det *nynorske skriftmålet*. For den som kjenner den norske språkhistoria og veit at det nynorske skriftmålet byggjer direkte på talemåla rundt om i landet, er denne kombinasjonen ikkje så overraskande, men han gjer NO unik, iallfall i europeisk samanheng. Ordbøker som beskriv eit talemål (dialektordbøker), er ikkje noko ukjent fenomen i andre språksamfunn, men at talemålsmaterialet får plass ved sida av skriftmålet i eitt og same verk, er uvanleg. Dette faktumet seier òg noko om statusen talemåla har i Noreg.

Med "det norske folkemålet" må ein forstå norsk talemål, altså dei norske dialektane. Samtidig er ordboka òg ei nasjonal ordbok, det vil mellom anna seie at ho

skal dekkje det norske talespråket i *heile* landet. Dette kravet til geografisk representativitet blir understreka i fleire presentasjonar av ordboka. Til dømes går det fram i ei tidleg utgreiing om det første faglege opplegget for ordboka at ”maalet for Norsk Ordbok maa – sjølv sagt – vera aa faa samlarar i arbeid i kvar einaste norsk bygd og by, helst fleire paa kvar stad” (Skard 1932:21).

2. Geografiske fordeling av målførekjeldene

2.1. Oversyn

For å få eit bilete av den geografiske fordelinga av målførekjeldene i NO har eg gått inn i seteldatabasen og søkt fram setlane fylke for fylke og så fordelt setlane etter dagens kommuneinndeling for å kunne visualisere fordelinga ved hjelp av eit kartprogram. Setelarkivet til NO inneheld totalt om lag 3,2 millionar setlar (som også er digitaliserte og kan hentast fram som faksimilar). Oppteljinga viser at av desse er det om lag 660 000 målføresetlar, det vil seie lappar med dialektopplysningar som kan heimfestast til eit geografisk område. Resten av setlane er litterære setlar med ekserpt frå skriftlege kjelder. I tillegg kjem materialet frå Trønderordboka, som utgjer om lag 150 000 setlar fordelte på trøndelagsfylka og Møre og Romsdal.

Å lage eit oversyn over geografisk fordeling er i seg sjølv eit omstendeleg arbeid fordi NO byggjer på kommuneinndelinga slik ho var rundt 1950, og ho skil seg frå inndelinga i dag fordi mange kommunar er slått saman etter den tid. Dessutan er mange setlar dessverre feil merkte med fleire enn eitt kommunenamn, men dette er langt på veg korrigert ved stikkprøver av setlar frå kvar geografisk eining og utluking av systematiske feil som oppstår fordi ein heimelsperson er registrert med galne geografiske opplysningar i høve til det som står på setlane. Ein del setlar er merkte med geografiske einingar større enn kommune (på landområde som omfattar delar av eitt eller fleire fylke), og slike setlar er fordelte på kommune etter opplysningar på setelen. Ei tredje feilkjelde er at ein (god) del litterære setlar også er merkte med geografisk eining (oftast med heimkommunen til personen som har ekserpert boka) utan at ekserpisten har ført på målføreopplysningar på setelen. I tillegg inngår ei rekkje både publiserte og upubliserte ordsamlingar frå ulike kommunar i materialgrunnlaget til NO. Setlar frå desse ordsamlingane er i hovudsak ikke merkte med områdenamn, og kjem såleis ikke med når ein søker på geografiske område. Det vil med andre ord seie at NO rår over noko meir informasjon om målføra enn det som kjem fram på dei 660 000 målføresetlane i basen. Ei anna kjelde til informasjon om målføra som ikke er teken med i dette oversynet, er ein synopsis (Norsk dialektatlas) som (i hovudsak) inneheld informasjon om bøyning og uttale av ein del utvalde ord heilt ned på kommunenivå.

Desse forholda har gjort at det har vore eit tidkrevjande og vanskeleg arbeid å få ei påliteleg oversikt over den geografiske spreinga på kjeldene til NO, men sjølv om det framleis finst feil i materialet kartet byggjer på, gjev det etter mitt syn eit

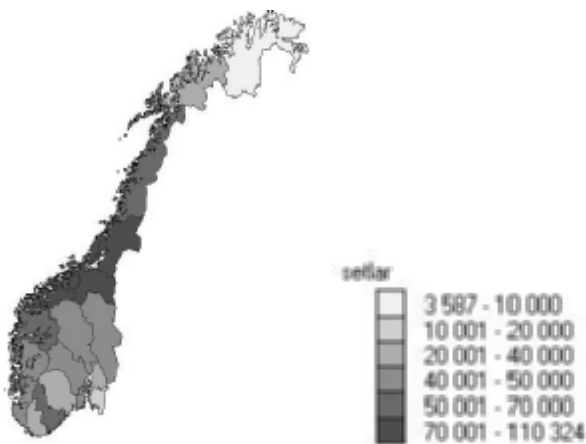
tilstrekkeleg godt bilete av situasjonen.

Det optimale ville vere å presentere tala i eit kart med den kommuneinndelinga som gjaldt rundt 1950, men eit slikt kartprogram er ikkje tilgjengeleg. Ein konsekvens av dette er særleg at tala for tidlegare reine bykommunar blir noko tilslørte, fordi bykommunane i dag oftast omfattar område utanfor bygrensa.

2.2. *Kva viser materialet?*

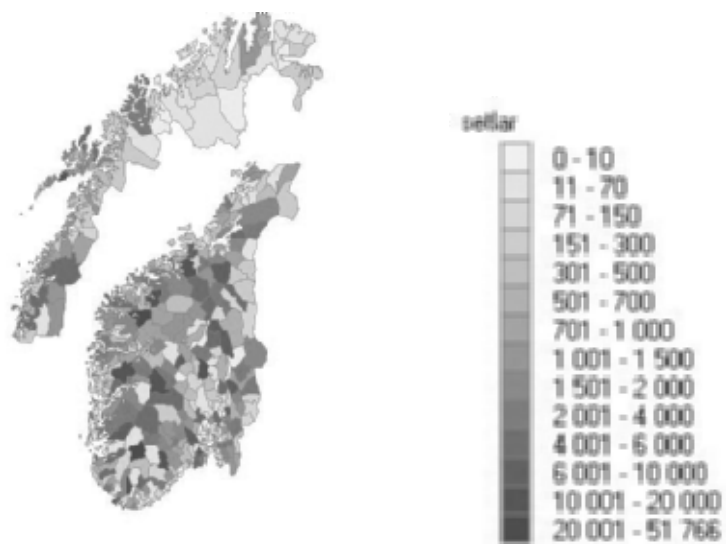
Oversikta i figur 1 viser at det nordlegaste fylket, Finnmark, har svært få kjelder. Dette er ikkje overraskande. Andre har peikt på den same mangelen tidlegare, mellom anna Jenstad (2005:224), som hevdar at av dei rundt 1000 frivillige medarbeidarane som har sendt inn målføretilfang til NO, er det truleg berre ein eller to aktive frå det nordlegaste fylket i landet. Talmaterialet underbyggjer denne påstanden, og samtidig får ein talfesta korleis underlagsmaterialet til NO fordeler seg geografisk. Ein kan òg merke seg det låge talet for hovudstadsfylket Oslo.

Grunnlagsmaterialet viser at Finnmark utmerkar seg ved å ha berre 3 587 setlar (0,43 %) (fordelte på 19 kommunar), mens gjennomsnittstalet for alle 19 fylke i landet er 43 932 setlar (5,26 %). Denne tendensen kjem òg til syne i kommuneoversikta (jf. figur 2), og i tillegg kan ein der sjå at det finst få setlar frå kommunane i Nord-Troms. Nord-Troms og Finnmark utgjer dermed eit samanhengande område med svært lite målføretilfang.



Figur 1. *Målføresetlar fordelt etter fylke¹*
(Kartgrunnlag: Statens kartverk)

¹ Oversikta inkluderer òg setlar frå Trønderordboka (om lag 150 000 setlar jamt fordelt på fylka Nord-Trøndelag, Sør-Trøndelag og Møre og Romsdal).



Figur 2. Målføresetlar fordelt etter kommune (Kartgrunnlag: Statens kartverk)

Eit avgjerande spørsmål når ein ser på landet under eitt, er om fordelinga er heilt arbitrær, eller om det kan vere ideologiske eller andre årsaker til ubalansen. Ein kan utan tvil slå fast at det er eit visst innslag av slump med omsyn til spreininga på kjeldematerialet som både kan ha årsak i sjølve prosessen for å rekruttere informantar og manglande interesse for prosjektet i enkelte kommunar.² Det finst fleire andre kommunar i landet som har få eller ingen setlar, men altså ingen større samanhengande område, som tilfellet er lengst nord i landet. Med unntak av at ein del bykommunar har få setlar, til dømes langs sørlandskysten, er det vanskeleg å sjå noko system blant dei andre kommunane som har få setlar. Desse lyse flekkene på kartet kan vanskeleg forklarast som anna enn tilfeldigheter. Derimot kan den gjennomgåande tendensen med få setlar i Nord-Troms og Finnmark knapt vere tilfeldig.

3. Moglege årsaker til den geografiske ubalansen

NO må sjåast i samanheng med og kanskje også som eit framhald av ordbøkene til Ivar Aasen (*Ordbog over det norske Folkesprog* (1850) og *Norsk Ordbog* (1873)) og Hans Ross (*Norsk Ordbog. Tillæg til Norsk Ordbog af Ivar Aasen* (1895)). Desse verka, og ikkje minst den lingvistiske tradisjonen som var rådande i samtida, spelar ei viktig rolle i denne samanhengen. Som dialektordbok har NO hittil vore sterkt

² Det finst lite informasjon med omsyn til rekrutteringa av informantar til Norsk Ordbok.

påverka av den tradisjonen som prega norsk målføregransking frå midten av 1800-talet til langt inn på 1970-talet (Vikør 1997:8), det vil seie den tradisjonelle dialektologien med si djupe forankring i junggrammatikken. Jahr (1996) lanserer omgrepet ”nynorsk språkforsking” og i det legg han

den forskningsaktiviteten som med forankring i målreisingsideologien til Ivar Aasen tok sikte på å beskrive norsk språk historisk, dialektologisk og som skriftnorm, uten at det blei tatt noe som helst hensyn til at det også fantes en muntlig og skriftlig riksmål/bokmåls-tradisjon i landet. Denne tradisjonen representerte, som et klart resultat av språkkontakt i dansketida, ”dansk” språk i Norge, og skulle derfor overhodet ikke tas hensyn til i beskrivelsen av norsk (Jahr 1996:89).

Ein la vekt på den beskrivande, historiske og samanliknande dimensjonen og knytte synsmåtane uløseleg til språkhistoria. Blant mange forskarar i nordisk språkvitskap i Noreg gav denne teoretiske forankringa seg utslag i ei leiting etter samanhengen mellom gammalnorsk og dialektane i samtida, mellom dialektane og til dels også etter samanhengen mellom dialektane og nynorsk skriftmål (Jahr 1996:97).

Ei slik forståingsramme må sjåast i samband med nasjonsbyggingstanken og den sterke ideologiske funksjonen som norsk dialektologi har hatt frå Aasens tid og framover. For ein ny nasjon blei det å få eit *eige* skriftspråk eit svært viktig prosjekt. Ein sjølvstendig nasjon måtte ha eit eige sjølvstendig skriftspråk; språket blei dermed eitt av dei viktigaste nasjonalsymbola. Etter unionsoppløysninga i 1814 var skriftspråket plutselig utanlandsk, og i nasjonalromantisk ånd byrja Aasen arbeidet med å knyte band tilbake gjennom det som har blitt kalla 400-årsnatta (eller den dansk-norske parentesen i norsk historie!) til tida før unionen. Aasen viste gjennom arbeidet sitt at språket som blei talt av bøndene omkring i landet, var utvikla direkte frå gammalnorsk uavhengig av dansk. I programskriftet ”Om vort Skriftsprog” omtaler Aasen dette som ”denne kostelige Skat fra Fortiden, som vore Forfædre gjennem alle sine Trængsler have troligen bevaret og overladt til os som en hellig Arv” (Aasen 1909:3). Med andre ord leverte Aasen språklege prov på at dansketida verkeleg var ein parentes i norsk historie (Jahr 1989:14). Sjølv om nasjonsbygging og nasjonalromantikk her er løfta fram som viktige fødselshjelpearar for Aasens prosjekt, vil det ikkje seie at det ikkje finst andre føresetnader og forklaringar. Aasens arbeid må òg sjåast i samanheng med opplysningsidealet frå 1700-talet som igjen heng saman med ønske og vilje til ei sterkare demokratisering utover på 1800-talet (jf. Apelseth 1996).

Områda lengst nord i Noreg var busett av samar lenge før noko germansktalande folkeslag slo seg ned så langt nord. Norsk busetnad går truleg ikkje lenger tilbake enn til 1200-talet. I tillegg til norsk- og samisktalande folkegrupper var det på 1700- og 1800-talet stor finsk innvandring til Nord-Troms og Finnmark. Med andre ord har dette samfunnet vore fleirspråkleg lenge før omgrepet blei aktualisert med

innvandring frå ikkje-vestlege land for nokre tiår sidan. Dette forholdet gjer at talemålsgrunnlaget i Nord-Troms og Finnmark er eit anna enn i resten av landet. Ved gjensidig kontakt mellom desse språka har det norske talemålet i dette området teke opp i seg drag frå dei andre involverte språka både når det gjeld fonologi, morfologi og syntaks.

Dermed var det ikkje i Nord-Troms og Finnmark ein venta å finne dei banda som kunne dokumentere linjene tilbake til gammalnorsk tid. Dette synet på nordnorsk talemål kan sporast fleire stader hos Aasen. I eit brev til vitenskapselskapet i Trondheim i 1845 uttalte han m.a.: ”Thi hvad Dialekten i Nordlandene angaar, da maa det antages at denne er af mindre Vigtighed” (1957:121). Men i 1851 reiste Aasen så langt nord som til Tromsø, og i eit brev til C.R. Unger skriv han dette:

der [blev] dog endda et ikke ubetydeligt Bidrag til Ordbogen; [...] De fleste av disse Ord ere fra Lofoten [...] Nordenfor Lofoten er der, saavidt jeg nu kan skjønne, intet at udrette i Sprogveien, da Indbyggerne for en meget stor Deel bestaae af indflyttede Folk, og Sproget i det Hele har megen lighed med det Trondhjemske (Aasen 1957:209).

Det ser ut til at Aasen tydelegvis modererer oppfatninga si av nordnorsk mål etter to reiser i landsdelen (han besøkte Helgeland i 1846), og i eit brev frå 1861 (Aasen 1957:386) gjev han uttrykk for at han ville reise nordover igjen, men denne reisa blei det ikkje noko av.

Også Ross deler Aasens syn på dei nordnorske dialektane og viser i fortalen i ordboka si til Aasen når han uttaler at ”Egnene nordenfor Trondhjem synes at have lidet eller fattigt Ordforraad, idet de mangle en Deel af de Ord, som ere nogenlunde almindelige i andre Dele af landet (Norsk Gramatik 356³)” (Ross 1971:XI). Dermed slår han alt mål nord for Trondheim saman med ”det Trondhjemske”.

Ein kan nok seie at Aasen og Ross hadde avgrensa kunnskapar om nordnorsk mål, ikkje minst om nordleg nordnorsk, men i lys av forskingsprogrammet hans var det heller ikkje nødvendig å skaffe meir informasjon om språket lenger nord. Ein slik innsats ville frå Aasens ståstad ikkje svare seg fordi han der neppe ville finne band tilbake til gammalnorsk.

Likevel kan ein merke seg det Aasen skriv i fortalen til *Ordbog over det norske Folkesprog* (1850): ”der er saaledes adskillige Egne, hvori der endnu kunde være meget at samle” (s. 4), og vidare: ”Samlingen i det Hele kan altsaa ikke endnu ansees som fuldstændig” (s. 5).

Og det skulde være ønskeligt, om Flere vilde bidrage hertil, da man her vel kan sige, at Een kan ikke gjøre Alt. Efter nogen Tids Forløb kunde der saaledes udgives en Samling af Tillæg til denne Bog, og hertil burde man da have saa meget som muligt samlet fra forskjellige Egne og især fra de Distrikter som hidtil ere komne mindst i Betragtning (s. 15).

³ Denne tilvisinga som Ross gjev til Aasen, stemmer ikkje.

Ein legg merke til at Aasen sjølv såg veikskapane ved ordboka, særleg når det gjeld den geografiske balanseringa, og oppmodinga hans om at andre må ta opp arbeidet etter han.

Både Aasen og Ross ser behovet for at ordboksverket må bearbeidast og oppdaterast dersom det skal kunne forsvare sin plass som (i den grad det er mogleg) uttømmmande diakron og synkron ordbok over det norske språket slik det kjem til uttrykk i alle delane av landet. Det må likevel understrekast at også i nordnorsk er det ord som høyrer til "det nationale Ordforraad" som Aasen er oppteken av, sjølv om han i nokon grad har teke med "endeel fremmede Ord, som ikke godt kunde udelates, da deres Brug i Landet er mere almindelig og tildeels ogsaa noget gammel" (Aasen 1873:7). I eit slikt perspektiv blir det stabile språket til eldre informantar med einsarta ordformer frå dei dialektområda i landet som ligg nærast gammalnorsk, mest interessant (Karlsen 1998:33). Derfor kan t.d. Ross fint kvalitetsvurdere dialektane slik han gjer i fortalen til ordboka si. Han snakkar om "gode Dialekter" (s. XVI) (t.d. store delar av Agder) og "ganske gode Maal" (s. XVIII) (t.d. Sunnmøre og Nordfjord). Sjølv om ikkje Ross går så langt, ville døme på mindre gode mål vere dialektar som ikkje i så stor grad kviler på det gammalnorske opphavet.

Dei norske dialektane i Nord-Troms og Finnmark kan sjåast som dei yngste og mest nyutvikla av dei norske dialektane (Bull 1994:236), men som mange språkkontaktprodukt, har desse dialektane lett for å bli stigmatiserte, ikkje minst fordi det er norsk som har hatt høgast sosial status av dei tre involverte språka. Målføre som er resultat av språkkontakt, blei heller ikkje vurderte som interessante (nok) innanfor Aasens språkprogram, trass i unntaket han gjorde for ein viss type lånord og sjølv om han i nokon monn reviderte synet sitt på dei nordnorske dialektane. Nyare forskningstradisjonar, særleg sosiolingvistikken (og med den også språkkontaktstudiet), der språkvariasjon og språkendring (ofte blant yngre språkbrukarar) blir sjølve forskingsobjektet, gjorde det interessant å studere nordleg nordnorsk :

De norske dialektene som historisk sett relativt nylig var utvikla i tidligere samisk- og kvensk/finsk-talende bygder, og som tidligere var blitt fullstendig ignorert av dialektologene, blei nå vurderte som spesielt viktige og interessante (Jahr 1996:95).

Men innanfor leksikografien har forskingsobjektet ikkje endra seg vesentleg. Vikør (1997:8) understrekar at denne tendensen nok også gjeld for NO, og dette arbeidet underbyggjer denne påstanden. Synsmåtane som prega den tradisjonelle dialektologien, har blitt ført vidare heilt fram til moderne tid først og fremst ved at ein ikkje på eit tidlegare tidspunkt har intensivert innsatsen for å balansere kjeldene betre med omsyn til geografi.

4. Kva kan gå tapt dersom hola ikkje blir tetta?

4.1. Ord av norsk opphav

4.1.1. Ord som berre er registrerte i (nordleg) nordnorsk

Etter at redaksjonen dei siste åra har intensivert arbeidet med å dekkje dialektane i Nord-Troms og Finnmark betre, er det i dag mogleg å finne mange døme på ord som berre er registrerte i denne landsdelen:

mellomberr adj 'glipe med bar hud (mellom skjorte og bukse)' (Nordland, Troms)
månestav m 'atterskin frå månen i vatn med form som ei loddrett stripe' (Skjervøy)
svartseglar m 'svart seglskute som seglar i lufta, og varslar ulykke' (Lebesby)
utskjersfjøre f 'serskilt lågt lågvatn (fjøre)' (Lofoten)

4.1.2. Allereie registrerte ord får utvida tyding eller utbreiingsområde

surka f ('sats (for produksjon av heim Brent)') (Vardø)

Ordet er ikkje tidlegare belagt i denne tydinga i arkiva til Norsk Ordbok. (Det næraste ein kjem er tydinga 'sur væske' frå Malangen i Midt-Troms).

jårr n ('tullprat; surr av mange stemmer samstundes')

Ordet var tidlegare berre registrert i Midt-Troms, men etter at nytt materiale frå nordleg nordnorsk er lagt inn, kan utbreiingsområdet utvidast til også å omfatte Vardø i Finnmark.

klenning m (> klining) 'oppbløytt og samanrulla flatbrød med lever'

Den vanlege tydinga av dette ordet er 'flatbrød eller lefse påklint smør o.a.', men har i delar av Nord-Noreg (Senja, Vesterålen) fått ei meir spesialisert tyding.

Ein opplagt fare ved at talemålsmateriale frå den nordlegaste delen av Nord-Noreg er sterkt underrepresentert, er at ein ikkje får noko inntrykk av kva for ord av gamalnorsk opphav som er vanleg i desse talemåla.

Det er karakteristisk at fleire av dei nemnde orda er henta frå sjø- og kystverksemd. Dette kan ein sjå i samanheng med at fiske og fangst tradisjonelt har vore dei viktigaste næringsvegane i Nord-Noreg, mens bondekulturen, som var berar av språkarven innanfor Aasens prosjekt, ikkje har stått så sterkt i nord. Ei vektlegging av bondekulturen vil med andre ord også slå negativt ut for målføretilfang frå Nord-Noreg.

4.2. Lån og interferens frå samisk og finsk

Ein del ord frå samisk og finsk er komne inn i det norske ordtilfanget og blir i dag oppfatta som ein meir eller mindre naturleg del av standardspråket og finst som oppslagsord i mange ordbøker. Døme på slike ord frå samisk er *joik* ('song eller resitasjon til melodiskala'), *kommag* ('sommarfottøy utan hæl av hårlaust reinskinn'), *pesk* ('kufte av reinskinn'). Velkjende lånord frå finsk er til dømes *komse*

(‘transportabel vogge’), *sauna* (‘badstove’) og *sisu* (‘innsatsvilje, tæl’). Alle desse orda er belagde i arkivet til Norsk Ordbok og er, eller vil bli, redigerte.

I tillegg til desse orda har Norsk Ordbok med nokre ord av samisk opphav som nok ikkje er like vanlege i allmennspråket. Dette er to døme frå bind 3, og det finst truleg fleire:

gaim(i) m [frå samisk *gaibmi*] namnebror (Malangen, Karlsøy): *han gaimi* namnebror min (Malangen) / *Emil gaim* namnebror til ein meir kjend Emil (Karlsøy).

gaisse m [frå samisk *gais(s)a*] høg, kvass fjelltopp, serl i namnet på fjellrekka *Gaissane* i Finnmark.

Med andre ord finst det spreidde døme på lånord frå finsk-ugrisk i NO utover dei som er blitt vanlege i standardordbøker.

Eit anna nokså vanleg finsk lånord er *bjekso* (‘høg lærstøvel med opphøgd tupp framme’), men det er ikkje teke med i Norsk Ordbok, sjølv om det er registrert i setelarkivet.

Uttrykket *gå (i) gaiiken* (‘bli øydelagt, gå i stykke’) frå finsk *gaiikki* er notert frå Sør-Varanger, men ordet er registrert så seint at det ikkje blei redigert i bind 3. (Uttrykket er vanleg i Nord-Troms og Finnmark.)

Språkkontakt kan òg gje seg utslag i interferens på syntaktisk plan. Det vanlege ordstillingsmønsteret i norsk er at det finitte verbet kjem som ledd nummer to i spørjesetningar innleidd av kv-ord (V2-regelen). Fleire dialektar i Nord-Troms og Finnmark kan oppvise brot på dette mønsteret ved at det finitte verbet kjem på tredje plass i spørjesetningar innleidd av kv-ord. Desse døma frå Tana (i Finnmark) illustrerer nettopp dette fenomenet: *kor du ska?* (‘kor/kvar skal du?’), *ka klokka e?* (‘kva er klokka?’). Sjølv om denne ordstillinga er mogleg også i andre norske dialektar, har fleire forskarar forklart denne ordstillinga med interferens frå samisk og/eller finsk (der slik ordstilling er fullt mogleg), iallfall når det gjeld nordleg nordnorsk (jf. t.d. Nilsen 1996). Dette er døme på eit lingvistisk interessant fenomen som kan relaterast til kontakten mellom norsk, samisk og finsk, og det ville vere svært synd om det ikkje kjem med døme på slik språkbruk i Norsk Ordbok.

4.3. *Russenorsk*

Finnmark har felles grense med Russland, og fram til revolusjonen i 1917 var det i fleire hundreår livleg handelssamkvem med Nordvest-Russland, den såkalla pomorhandelen. Russiske fartøy gjesta hamner langs kysten av Finnmark og Troms for å selje mjøl og kjøpe fisk. Behovet for eit kommunikasjonsmiddel førte til utviklinga av pidginspråket russenorsk, og det finst framleis reminisensar av denne kontakten i språket. Det er rapportert ein del lånord frå russenorsk i norsken heilt fram til vår tid som til dømes *prennek* (russeno *prennik* ‘småkake’) (kjent frå Sør-

Senja, Andørja, Lenvik), *spræke* ('prate, slarve, snakke om laust og fast') (øg subst: "slå av ein spræk") (kjent frå Lenvik) og *kludde* (russeno *kludi* 'jule opp') (kjent frå Lenvik og Senja) (Broch og Jahr 1984:88ff). Ingen av desse orda er registrerte i arkivet til Norsk Ordbok.

Substantivet *balldúska* f. ("kveite") (russeno *balduska*) er kjent frå fleire stader i Nord-Noreg (Nesna, Senja, Hillesøy). Ordet er òg komme med i NO under den normerte forma *ballduske* med tydinga 'godnamn på kveite'. Men ordet har fått galen etymologi ved at det er vist til ball < norr. *bqllr*. Ein av informantane gjev også opp spesialtydinga 'tøvete jente, mindre evnerik kvinne' (frå Hillesøy), utan at denne tydinga er komen med i NO. Broch og Jahr (1984:88) meiner at russisk *boltuska* subst ('pratsam kvinne') må liggje til grunn for denne siste tydinga.

Verbet *kråle* (russeno *kralom* 'stele') er registrert i denne tydinga med fleire setlar i arkivet til NO frå fleire stader i Nord-Noreg (men er enno ikkje redigert). Det er interessant at denne tydinga også er med i Carl Richard Ungers *Norske Dialekt-Ord udtagne af en Reisendes Optegnelser* (1844). Unger har tydinga frå så langt sør som Inderøy i Nord-Trøndelag.

Eit anna lånord frå russenorsk er adjektivet *bålsa* ('svær'). Broch og Jahr (1984) har ordet frå Vesterålen. I NO finst adjektivet *bolsut* med tydinga 'tjukk, klumput, lite nett'. Grunnlaget for denne ordartikkelen er ein setel frå Trondenes. Etymologien er ikkje gjeven opp i NO, men truleg er det snakk om eit lån frå russenorsk.

Maskulinet *starúrk* (russeno *starik* 'gamling') er kjent frå Borge i Lofoten og Levanger i Nord-Trøndelag. Broch (1984) meiner sisteleddet er overført frå *skurk* (m) i same tyding. I arkivet til NO er ordet *starruk* (m) registrert i Rennebu og Meldal i Sør-Trøndelag med tydinga 'gammal (ugift) mann'. På setelen frå Meldal er det vist til *Ordbog over det danske Sprog* som har ordet *starut* (m) med tydinga 'merkeleg, sjuskete eller upåliteleg person'. Etymologien er usikker ifølgje ODS, men det er vist til polsk *staruch* ('gammal mann'). Det er nok meir rimeleg at ordet er lånt inn i nordnorsk og trøndsk gjennom russenorsk enn at det er kome gjennom dansk (jf. tydingsendringa og slektskapen mellom polsk og russisk).

Eit anna døme på språkkontakt med Russland er substantivet *katank(a)* m ('filtstøvlar, luggar') som er registrert i arkivet til NO fleire stader frå Troms og Finnmark. Ordet finst i ei rad variantar i nordrussiske dialektar og er lånt inn i forma *kátanka* f (*Språknytt* 1987 nr. 3:16). Om det kjem direkte frå russisk eller om det har komme gjennom russenorsk, er noko usikkert. Ordet vil venteleg bli redigert.

Etter denne gjennomgangen legg ein merke til at ingen av dei russenorske lånorda som er komne med i NO, har (riktig) etymologi.

4.4. Kjeldemateriale frå bymål

Oversikta i tabell 1 viser at tala for dei reine bykommunane (slik grensene var rundt 1950) gjennomgåande er svært låge. Gjennomsnittstalet for alle bykommunane i

landet er 394 setlar, mens gjennomsnittstalet for landkommunane er 1203 setlar. Når det gjeld bykommunane, dreg dei 5 byane med flest setlar, Stavanger, Oslo, Bergen, Kristiansand og Arendal, gjennomsnittet monaleg opp med mange setlar. Tek ein omsyn til folketalet i byane (sjølv rundt 1950), blir misforholdet mellom bykommunar og landkommunar endå klarare.

Bykommune	Setlar ⁴
Stavanger	9371
Oslo	5003
Bergen	4340
Kristiansand	2675
Arendal	2062
Trondheim	411 ⁵
Tromsø	248
Sarpsborg	193
Porsgrunn	157
Skien	39
Sandnes	35
Larvik	24
Fredrikstad	8
Ålesund	5
Bodø	5
Drammen	3
Haugesund	2
Sandefjord	1
Moss	1
Halden	1
Tønsberg	0
Hamar	0

Tabell 1. *Oversikta viser talet på setlar i byar med meir enn 20 000 innbyggjarar (per 1.1.2004)*

Det klaraste dømet på denne veikskapen ved materialet er dei få belegga frå hovudstadskommunen Oslo (som òg er eige fylke). Kommunen har om lag 530 000 innbyggjarar, eller rundt 12 prosent av folketalet i landet, men berre vel 5 000 setlar (0,60 %) i arkivet viser døme på språkbruk frå hovudstaden. Innvandringa dei siste tiåra frå ikkje-vestlege land har, særleg i hovudstaden, ført til eit nytt kontaktpro-

⁴ Denne oversikta inkluderer òg setlar frå målføresamlingar i basen som ikkje er merkte med geografisk område.

⁵ Dette talet inkluderer òg grunnlagsmaterialet for Trønderordboka, dei såkalla trønderetlane.

dukt, den såkalla kebabnorsken. Denne (eller desse) varietet(-ane) veit vi enno ikkje så mykje om, men det er samla ein del leksikalsk materiale, t.d. Aasheim (1995) og Østbye (2005).

Også dette forholdet må sjåast i lys av det språksynet som rådde innanfor den tradisjonelle dialektologien. Målet i byane blei ikkje oppfatta som målføre av Aasen og samtida hans. Dei var for påverka av riksmål/bokmål til at dei blei vurderte som interessante for språkforskarar som leita etter samanhengen med gammalnorsk. Aasen karakteriserer måla i byane som ein ”mellemart af Norsk og Dansk” (1957: 111) (Den same merkelappen får også språket til ”conditionerede Folk paa Landet”). At det i byane ofte fanst både eit ”folkemål” og ”eit høgare talemål”, var det lite medvit om i Aasens samtid. Denne distinksjonen kom først klart fram med bymålsstudiane utover på 1900-talet (jf. Larsen (1907), Larsen og Stoltz (1912) og Berntsen og Larsen (1925)). Tendensen når det gjeld bymåla, ser ut til å vere den same som for nordleg nordnorsk. Bymåla og nordleg nordnorsk er begge eit produkt av språkkontakt, og dei har på mange måtar lidd felles skjebne. Sjølv om bymål i større grad enn nordleg nordnorsk blei emne for talemålsforskinga utover på 1900-talet, ser det ikkje ut til at det auka fokuset på bymål, har slått like sterkt gjennom i leksikografi-en. I alle fall må ein karakterisere NO som tradisjonell med omsyn til dette.

5. Kva har blitt gjort dei siste åra for å betre situasjonen?

Særleg etter at Norsk Ordbok blei gjort om til prosjektet Norsk Ordbok 2014 i 2002, har det skjedd ein del når det gjeld å styrkje målføre kjeldene frå den nordlegaste delen av Noreg. Resultatet for Nord-Troms og Finnmark hadde nok vore enda lågare om oppteljinga hadde skjedd for tre år sidan eller tidlegare, for ein god del av dei setlane som finst i dag, er nyregistreringar som har kome til etter 2002.⁶

Generelt er det no eit større fokus på å få kjeldemateriale som er betre geografisk balansert, samtidig som enkelte redaktørar har skaffa til veges og lagt inn materiale frå dei nordlegaste fylka i basen til NO. Den språklege interessa blant lekfolk er stor, og privatpersonar, lag og foreiningar rundt om i landet gjev jamleg ut lokale målføresamlingar. Mange av desse kjem inn til redaksjonen i NO, og i arbeidet med å leggje desse samlingane inn i basen til NO, har prosjektet prioritert samlingar frå Nord-Noreg.

Ei rad nyttilsetjingar dei siste åra har ført til ei samansetjing av redaksjonen med betre geografisk spreining enn tilfellet har vore tidlegare. Her, som ofte elles, viser det seg at redaktørar kjenner eit spesielt ansvar for det distriktet der dei høyrer heime. Ein positiv ringverknad av dette har vore større fokus på nordleg nordnorsk.

Sist, men ikkje minst, rår redaksjonen i dag over elektroniske dataverktøy som

⁶ Det var registrert 284 setlar frå Finnmark før nyregistreringa kom i gang.

gjer det mogleg å få eit oversyn over kjeldematerialet også med omsyn til heimfesting. Sjølv om eit slikt oversyn per i dag ikkje kan søkjast fram utan vidare, er det i alle høve enklare i dag enn det ville vore før materialet blei elektronisk tilgjengeleg.

Når det gjeld kjeldetilfanget frå byane, er det ikkje sett inn spesielle tiltak for å betre stoda. Det er sjølvsagt synd, men på mange måtar er dette ein naturleg konsekvens av dei rammene prosjektet NO 2014 arbeider innanfor. Dersom redaksjonen skal innfri produksjonskrava slik at verket blir ferdig til 2014, er det få ressursar som kan setjast av til å utvide kjeldematerialet ordboka byggjer på.

Denne kritiske gjennomgangen av setelarkivet til NO viser at kjeldematerialet ikkje omfattar ”kvar einaste norske bygd og by”. No kan ein nok seie at dette var eit svært pretensiosøst mål som det skulle mykje til for å nå. At det finst hol i eit så omfattande prosjekt som NO kviler på, er i seg sjølv ikkje kritikkverdig, men den systematiske underrepresentasjonen når det gjeld kjelder frå bymåla og måla i Nord-Troms og Finnmark, må likevel karakteriserast som uheldig viss målet er å beskrive norsk talespråk slik det kjem til uttrykk i alle delar av landet.

Trass i den geografiske ubalansen i kjeldene til NO som det er peikt på her, må det likevel understrekast at NO forsvarer plassen sin som det mest altomfattande ordboksverket i Noreg. Ved å kartleggje språkleg praksis og sjå talemålet i samanheng med skriftspråket over tid er NO både eit vitskapleg og kulturhistorisk unikt prosjekt som er med på å fremje ei demokratisk haldning til språk og kultur.

Litteratur

Ordbøker:

Ordbog over det danske sprog. XXI. bind. København: Gyldendalske boghandel. Nordisk forlag 1943.

Norsk Ordbok. Bind 3 flusker-gigla. Oslo: Det Norske Samlaget 1994.

Ross, Hans 1895: *Norsk Ordbog. Tilleg til ”Norsk Ordbog” af Ivar Aasen*. [Ny utgåve 1971. Oslo/Bergen/Tromsø: Universitetsforlaget.]

Trønderordboka. Trondheim: Tapir akademisk forlag 2002.

Unger, Carl Richard 1844: *Norske Dialekt-Ord uddragne af en Reisendes Optegnelser*. [Oslo].

Østbye, Andreas E. 2005: *Kebabnorsk ordbok*. Oslo: Gyldendal.

Aasen, Ivar 1850: *Ordbog over det norske Folkesprog*. [Ny utgåve ved Kristoffer Kruken og Terje Aarset 2000. Oslo: Det Norske Samlaget.]

Aasen, Ivar 1873: *Norsk Ordbog*. [Ny utgåve ved Kristoffer Kruken og Terje Aarset 2003. Oslo: Det Norske Samlaget.]

Annan litteratur:

Apelseth, Arne 1996: Ein europear så god som nokon. I: *Syn og Segn 1/96*. Oslo: Det Norske Samlaget, 33-47.

- Berntsen, M. og Amund B. Larsen 1925: *Stavanger bymål*. Oslo: Utgjeven av Bymålslaget. I kommisjon hos Aschehoug.
- Broch, Ingvild og Ernst Håkon Jahr 1984: *Russenorsk – et pidginspråk i Norge*. Tromsøstudier i språkvitenskap III. Oslo: Novus.
- Bull, Tove 1994: Nordnorsk språkvariasjon. I: *Nordnorsk kulturhistorie 2*. Oslo: Gyldendal Norsk Forlag, 228-239.
- Jahr, Ernst Håkon 1989: *Utsyn over norsk språkhistorie etter 1814*. Oslo: Novus.
- Jahr, Ernst Håkon 1996: Nynorsk språkforskning – en historisk oversikt. I: Carol Henriksen et al. (eds.). *Studies in the development of linguistics in Denmark, Finland, Iceland, Norway and Sweden*. Oslo: Novus, 84-101.
- Jenstad, Tor Erik 2005: *Norsk Ordbok* som dialektordbok. I: Ruth Vatvedt Fjeld og Dagfinn Worren (red.): *Nordiske studiar i leksikografi 7. Rapport frå Konferanse om leksikografi i Norden. Volda 20.-24. mai 2003*. Skrifter utgjevne av Nordisk foreining for leksikografi. Skrift nr. 8, 221-227.
- Karlsen, Knut E. 1998: Hallfrid Christiansen – offer for eit manglande paradigmeskifte i nordisk språkvitenskap i Noreg? I: Knut E. Karlsen (red.): *Mål og middel i nordisk språkvitenskap* (HBO-rapport 4/1998), 29-44.
- Larsen, Amund. B. 1907: *Kristiania bymål: vulgærsproget med henblik på den utvungne dagligtale*. Kristiania: Cammermeyer.
- Larsen, Amund. B. og Gerhard Stoltz 1912: *Bergens bymål*. Kristiania: Utgjeven av Bymålslaget. I kommisjon hos Aschehoug.
- Nilsen, Hilde 1996: *Koffer dæm sir det?: spørresetninger i nordreisadialekten*. Hovudoppgåve i nordisk språk. Universitetet i Tromsø.
- Skard, Sigmund 1932: *Norsk Ordbok. Historie – plan – arbeidsskipnad*. Oslo: Det Norske Samlaget.
- Skjekkeland, Martin 1997: Ivar Aasen i møte med talemålet på Agder. I: Arnold Dalen (red.): *Gransking av norsk mål i hundre år etter Ivar Aasen*. Det Kongelige Norske Videnskabers Selskab. Skrifter 3. Trondheim: Tapir Forlag, 105-117.
- Språknytt* 1987: Meldingsblad for Norsk språkråd. Nr. 3.
- Vikør, Lars S. 1997: Innleiing. I: Lars S. Vikør: *Norsk Ordbok – nynorskens leksikografiske kanon? Rapport frå eit seminar på Blindern 31. mai 1996*. Universitetet i Oslo, 5-12.
- Aasen, Ivar 1909: Om vort Skriftsprog. I: *Syn og Segn*. Oslo: Det Norske Samlaget, 1-5.
- Aasen, Ivar 1957: *Brev og dagbøker*. Oslo: Det Norske Samlaget.
- Aasheim, Stine 1995: *”Kebab-norsk”: framandspråkleg påverknad på ungdomsspråket i Oslo*. Hovudoppgåve i nordisk. Universitetet i Oslo.

De første skridt til en islandsk fremmed- og/eller låneordbog

Towards an Icelandic dictionary of foreign words and/or loanwords. In this paper I start by discussing the differences between the terms “framandorð” or “aðkomuorð” (foreign word) and “tökuorð” (loanword). I then give a brief account of language purism in Iceland and its influence on the vocabulary, referring in particular to examples from articles by Halldór Laxness and Guðmundur Finnbogason. The last part of the paper is devoted to some guiding principles for composing an Icelandic dictionary of foreign words. Five rules, that a word must fulfill to be considered as adapted, are discussed. The conclusion is that the terms “tökuorð”, including both foreign words and loan words, and “tökuorðabók”, dictionary of foreign words and loanwords, are the best suited for Icelandic.

1. Indledning

Jeg har valgt at tale om de første skridt til en islandsk fremmed- og/eller låneordbog. Dette tema er noget jeg har spekuleret på i lang tid, men først nu er jeg for alvor begyndt at samle materiale til en sådan bog. Men hvad betyder at samle materiale i denne forbindelse? Hvilke ord og ordforbindelser skal med i bogen, hvilke ord skal udelades, og hvorfor? Skal man angive ordenes alder i sproget, og sidst, men ikke mindst, kan man skelne mellem låneord og fremmedord? Jeg vil prøve på at besvare disse spørgsmål selvom jeg i virkeligheden endnu ikke har klare svar på dem alle.

2. Låne- og fremmedord på Island

Temaet låne- og fremmedord har man ikke skrevet meget om på Island ud fra et lingvistisk synspunkt, men ganske mange har bestemte meninger om hvad der bør undgås i sproget. En pensioneret professor gav mig følgende svar da jeg spurgte ham om man ikke burde samle materiale til en islandsk fremmedordbog: Nej, det skal man ikke, for hvis ordene står i en ordbog, mener man at de også kan bruges, og det ønsker vi jo ikke.

I et land som Island, hvor sprogpurismen er kommet til udtryk i hvert fald siden det sekstende århundrede og har været dominerende siden romantikkens tid, har ganske mange bestemte meninger om hvad det er som skal defineres som et ord eller en ordforbindelse af fremmed oprindelse i islandsk. De fleste er ikke vant til at skelne mellem låneord og fremmedord. Alle ord af fremmed oprindelse er for almenheden *tökuorð*, d.v.s. ord optaget fra et andet sprog uanset hvor tilpassede de er. Lad os se lidt nærmere på termerne. I en artikel af Jón Hilmar Jónsson (1980:62) står der f.eks.:

Det er vigtigt at nevne at islandsk ikke skiller mellem fremmedord og låneord, begrepet *tökuorð* omfatter både assimilerede og uassimilerede ord. Inndelingen i fremmedord og låneord synes for øvrigt å være tvilsom, ikke minst fordi den legger altfor ensidig vekt på formelle kriterier uten å ta hensyn til aspekter som gjelder bruken av de enkelte ord.

I *Íslensk-dönsk orðabók* (Islandsk-dansk ordbog) fra 1920-24 har man dog allerede gjort forskel på de to begreber. Den danske oversættelse af opslagsordet *tökuorð* er 'låneord', mens der findes et andet opslagsord, *aðkomuorð*, som har definitionen 'fremmedord'. Så vidt jeg ved, har man her for første gang forsøgt at skelne mellem disse to begreber, sikkert efter dansk forbillede.

Institut for Leksikografi (www.lexis.hi.is) har ikke mange citater med ordene *tökuorð* og *aðkomuorð*. De ældste citater med *tökuorð* i instituttets samlinger stammer fra begyndelsen af det tyvende århundrede, mens de ældste eksempler på *aðkomuorð* er lidt ældre eller fra slutningen af det nittende århundrede. Ud fra citaterne kunne man slutte at *aðkomuorð* i begyndelsen havde en mere negativ klang end *tökuorð*, men eksemplerne er få og giver måske ikke et rigtigt billede af bruken. I nyere tid er lingvister begyndt at bruge *aðkomuorð* i betydningen 'fremmedord' og skelne mellem disse ord og *tökuorð*. Termen *framandorð* for fremmedord er derimod ganske ny og så vidt jeg ved, blev ordet *framandorðabók* første gang brugt i *Nordisk leksikografisk ordbok* 1997 som et islandsk ord for fremmedordbog. I den tredje og nyeste udgave af *Íslensk orðabók* (Islandsk ordbog) fra 2002 er *framandorð* for første gang taget med som opslagsord. Definitionen er (i min oversættelse): 'et ord af udenlandsk oprindelse i et sprog, uafhængigt af dets anerkendelse'. Definitionen af *tökuorð* er derimod: 'et ord som er optaget i et sprog fra et andet', og i parentes står der '(og har fået en vis anerkendelse)', d.v.s. hverken *framandorð* eller *tökuorð* er efter definitionen helt anerkendte ord. *Aðkomuorð* er ikke et opslagsord i ordbogen og var det heller ikke i de første to udgaver.

3. Sprogpurismens rolle

Men nu lidt om hvad der skal med, og hvad der skal udelades i den bog jeg påtænker. Som jeg allerede har nævnt, har sprogpurismen spillet en stor rolle på Island i lang tid, og det er grunden til at offentlige medier har forsøgt at undgå fremmedord og i mange tilfælde også nogenlunde tilpassede låneord. Man kan dog spore en vis ændring i den senere tid, ikke mindst i medierne. Der er skrevet mange artikler om den negative danske indflydelse på sproget, og i dem kan man ofte finde eksempler på ord som man efter forfatterens mening skal lade være med at bruge. Disse artikler kan være til stor hjælp ved indsamling af fremmed- og låneord som man ikke kender på forhånd. Jeg vil kun nævne to artikler som gælder dansk indflydelse, og som hver for sig viser situationen i den første halvdel af det tyvende århundrede.

Forfatteren Halldór Laxness var stærkt imod dansk indflydelse på det islandske

sprog og skrev i 1939 en artikel derom under titlen “Ill danska” (Dårligt dansk). Han kritiserede islændingenes kundskaber i dansk, som han mente var grunden til at de ikke kunne undgå at bruge ord eller ordforbindelser stammende fra dansk, og at denne indflydelse var stadig voksende og ikke havde været så stærk siden den tid “da man i det 18. århundrede brugte danismer af forfængelighed” (1939:33). Halldór Laxness henviste til en liste over danismer som han havde samlet fra aviserne i Reykjavík. På listen er der mange ord og ordforbindelser som stadig er yderst almindelige i sproget. Jeg nævner kun få eksempler:

finna út úr einhverju (finde ud af noget)
 ekki nóg með það (ikke nok med det)
 ganga út frá (gå ud fra)
 ganga inn á (gå ind på)
 það gengur út á (det går ud på)
 reikna með (regne med)
 bera sig að (bære sig ad)
 afgerandi (afgørende)
 svo mikið sem (så meget som)
 koma inn á (komme ind på)
 vel að merkja (vel at mærke)
 hafa eitthvað á tilfinningunni (have noget på fornemmelsen)
 svo að segja (så at sige)
 slá striki á (slå en streg over)
 þvert á mót (tværtimod)
 yfirbjóða (overbyde)
 óviðkomandi (uvedkommende)

I alle tilfælde har vi her at gøre med låneoversættelser. Jeg har søgt efter alle disse ord og ordforbindelser i Halldór Laxness' værker og kun fundet de tre som står nederst på listen, så han har selv undgået at bruge de fleste af dem. Grunden til at jeg nævner denne liste her, er at jeg har tænkt meget på i hvor høj grad man skal tage ordforbindelser som disse med i en låne- og fremmedordbog i dag. Mange purister er endnu meget imod dem, regner dem som en del af den fremmede indflydelse og mener at de sætter en plet på det islandske sprog. Det er en meget svær opgave at finde frem til den slags ordforbindelser og kræver at man har meget gode kundskaber i dansk. Det er lettere at identificere de senere årtiers engelske fremmedord og udtryk. I samtaler med folk, f.eks. af min generation, om en fremmedordbog nævner man netop ordforbindelser som disse og mener at de absolut skal med fordi vi hele tiden i skolen fik formaninger om at undgå den slags og fik store røde streger i vores stile hvis vi brugte dem. Men som sagt, jeg er i tvivl og vil gerne høre mine kollegers mening om sagen.

En anden puristisk artikel stammer fra professor og landsbibliotekar Guðmundur

Finnbogason, som var en kendt mand på Island der skrev artikler som man lagde mærke til. Den aktuelle artikel har titlen *Hreint mál* (Rent sprog) og indeholder en beskrivelse af et middelklassehjem i Reykjavík omkring 1920 (1928:145-155). Forfatteren gik fra det ene værelse til det andet og beskrev det han så, med de ord han hævdede at man brugte i sådanne hjem. Han nævnte f.eks. ganske mange ord for al slags køkkenting som han påpegede var danske. Blandt dem er *kastarhola*, *dörslag* og *eggjapískari*, *sigti*, *propptrekkjari*, *viskustykki*, *karklútur*, *uppvöskunarbusti*, *skrúbbur*, *kústur* og *fægiskúffa*. Han nævnte også forskellige madvarer som man kunne finde i spisekammeret, ordene *polegg*, *medisterpylsa*, *karbonaði* og *plokkfiskur*, men også adskillige verber som blev og i mange tilfælde endnu bliver brugt ved madlavning og rengøring, f.eks. *sigta*, *píska*, *plokka*, *hakka*, *kokka*, *speða*, *skræla*, *pússa* og *skrúbbu*. Næsten alle disse ord er levende i sproget i dag, men gælder endnu for mange som danismer. Mange af dem findes i Institut for Leksikografis samlinger, nogle er belagt allerede for århundreder siden, f.eks. *kastarhola*, *dörslag*, *sigti* og verberne *sigta*, *hakka*, *kokka* og *pússa*, andre findes kun i citater fra Guðmundur Finnbogasons artikel. Grunden dertil er den ret store forskel på skriftsproget og talesproget som gør det vanskeligt at finde frem til ord som faktisk blev og måske endnu bliver brugt selvom man ikke finder dem i ordbøger eller på tryk andre steder. Jeg syntes det var vigtigt at finde ud af den faktiske brug af disse ord, og benyttede lejligheden til at spørge om alle fremmedordene i Guðmundur Finnbogasons artikel i instituttets radioudsendelser og fik en masse svar. Det ser ud til at de fleste ord lever deres gode liv i sproget. Mange af mine hjemmelmænd sendte mig lister med ord som de huskede fra deres barndom forskellige steder i landet, og på denne måde er jeg nået frem til en god del af de danske låneord som brugtes i hjemmet, og som man kæmpede imod gennem næsten hele det tyvende århundrede.

I det jeg allerede har sagt, har jeg kun nævnt dansk indflydelse, men selvfølgelig har engelsk sat sine spor i sproget på Island som i de øvrige nordiske lande. Dets indflydelse startede under den anden verdenskrig og vokser stadig, mens indflydelsen fra dansk næsten er forsvundet. Samfundet er også blevet mere mangfoldigt med tilkomsten af indvandrere fra forskellige lande. Deres indflydelse på ordforrådet er først og fremmest mærkbar i forbindelse med mad og madvarer. Men deres tilstedeværelse har også øget brugen af engelske glosser i kommunikationen mellem islændinge og andre som heller ikke har engelsk som modersmål.

Jeg spurgte om man i bogen skulle medtage angivelse af ordenes alder i sproget. I øjeblikket er jeg meget i tvivl. Jeg har tænkt mig at medtage ordforråd af fremmed oprindelse indkommet omkring 1900 og frem til i dag. Undertiden kan man nogenlunde finde frem til en datering, især hvis ordene forekommer i annoncer for nye ting som er lige blevet importeret, mens man i andre tilfælde ikke har oplysninger som man reelt kan stole på. Som jeg allerede har påpeget, forekommer mange af disse ord ikke i de ældre ordbøger, og de findes ikke i Institut for Leksikografis

samlinger. I den sidste tid er man dog på Landsbiblioteket gået i gang med at affotografere alle tidsskrifter og aviser frem til 1920 og *Morgunblaðið*, landets ældste avis, fra 1913 og frem til i dag, og jeg gør mig forhåbninger om at disse kilder kan være en hjælp ved dateringsarbejdet.

Man kan sagtens sige at der er rigeligt stof til en låne- og fremmedordbog selvom det ikke er let at finde frem til en god del af ordene som kun lever i talesproget. Men er det muligt at skelne mellem låne- og fremmedord i islandsk, og hvorledes skal man bære sig ad? Jeg har overvejet den mulighed i hver ordbogsartikel at angive om ordet skal betragtes som låneord eller som fremmedord. Nu er jeg mere i tvivl end jeg var i begyndelsen. Grænsen mellem disse to termer er ofte ganske uklar. Kan der opstilles enkle regler for valget mellem de to kategorier?

4. De fem regler

I min artikel fra den leksikografiske konference i Göteborg (2001:132) nævnte jeg Baldur Jónssons forslag til de krav som man kan stille til et ord af fremmed oprindelse hvis det skal accepteres som en del af det islandske ordforråd (1997:16). Jeg vil nævne de fem regler her igen og se lidt nærmere på dem bagefter. Baldurs kriterier er:

1. Trykket skal være på første stavelse.
2. Et fremmedord skal følge islandske fonologiske regler.
3. Et fremmedord skal følge islandske fonotaktiske regler.
4. Et fremmedord skal tilhøre en bøjningsklasse der allerede eksisterer i sproget, følge dens bøjningsregler og andre specielle regler.
5. Et fremmedord skal skrives med islandske bogstaver således at der er regelbunden overensstemmelse mellem skrivemåde og udtale.

Hvis et ord opfylder disse betingelser, kan det efter hans mening betragtes som et *låneord*, mens andre ord som ikke opfylder alle disse betingelser, men alligevel bruges i sproget, kun delvis er assimilerede og derfor falder i gruppen *fremmedord*. Lad os se lidt på disse kriterier.

Vi behøver ikke at bruge tid på det første kriterium. Det er en gældende regel i islandsk at tryk ligger på første stavelse. Ligger det et andet sted i ordet, så drejer det sig om et uassimileret fremmedord.

Et låneord må kun bestå af fonemer som eksisterer i sproget. Således kan ord med *y* udtalt med rundet vokal, som for længst er forsvundet i islandsk, ikke regnes som tilpasset. Hvis udtalen er med urundet *i* eller *í*, så hører det hjemme i gruppen assimilerede fremmedord hvis det opfylder de andre kriterier. Som eksempel kan jeg nævne *system* og *systematískur*, udtalt med rundet vokal, som ikke-tilpassede ord på grund af udtalen, og *sístem* eller *sistem* og *sístematískur*, som er tilpassede.

Regelen om fonologisk assimilation gælder ikke kun danske ord, men også ord optaget fra engelsk. Som eksempel kan nævnes at visse engelske konsonanter ikke findes i det islandske fonologiske system. Lukkelyde bliver f.eks. palataliserede hvis de står foran en palatal vokal i islandsk. Ordene *gír* 'gear', *gin* 'gin' og *gel* 'gel' udtales med palataliseret *g* og kan derfor regnes assimilerede, mens *geim* 'party' og *keip* 'kort slag af stof eller pelsværk' udtales uden palatalisering. Det samme gælder et *g* foran diftongen *æ*. Regelen er at det udtales med palatalisering, f.eks. *gær* 'i går', *gæta* 'passe på'. I ordet *gæi*, eng. 'guy' udtales *g* uden palatalisering, og ordet er derfor ikke assimileret.

Af Baldur Jónssons artikel fremgår det at definitionen af en islandsk fonotaktisk regel ikke er selvklar. Som jeg har påpeget før (2001:133), ville man udelukke en mængde låneord som er optaget i sproget før 1500, hvis man kun ser på arveordenes opbygning. Men i gamle assimilerede fremmedord har man allerede indført fonologiske strukturer som ikke fandtes i arveordene. Dette gælder f.eks. ord hvor en lang konsonant følger et *ú*:

- ú + bb: *skrúbba* (substantiv, verbum) 'skrubbe'
- ú + dd: *húdd*, eng. 'hood'
- ú + ff: *skúffa* 'skuffe'
- ú + kk: *dúkka* (substantiv, verbum) 'dukke', *húkka*, eng. 'hook'
- ú + pp: *grúppa* 'gruppe'
- ú + tt: *tútta* 'sut', *prúttu* 'prutte om noget', *púttu* 'putte'

Selvom disse ord er blevet accepteret som assimilerede fremmedord, føles de dog af mange som fremmede, især hvis de stammer fra engelsk, som *hood*, eller findes både i dansk og i engelsk, som *gruppe* og *group*.

Bøjningen er sjælden et problem. Enten falder et fremmedord i en bøjningsklasse og kan accepteres som en del af ordforrådet hvis det også opfylder de andre kriterier, eller det gør det ikke og betragtes som et uassimileret fremmedord. Fremmede substantiver og verber falder normalt let i en bøjningsklasse. Det samme gælder stort set kun de adjektiver som bøjes som perfektum participium, f.eks.

- forkelaður* – *forkeluð* – *forkelað* 'forkølet'
- fornermaður* – *fornermuð* – *fornermað* 'fornærmet'
- húkkadur* – *húkkuð* – *húkkað* eng. 'hooked'

samt adjektiver med endelsen *-ískur*, f.eks.

- partískur* – *partísk* – *partískt* 'partisk'
- pólítískur* – *pólítísk* – *pólítískt* 'politisk'
- sýstematískur* – *sýstematísk* – *sýstematískt* 'systematisk'

Andre adjektiver er normalt ikke assimilerede med hensyn til bøjning, f.eks. *smart*, *bræt* engelsk 'bright', *kúl* 'cool', *röff* 'rough', *töff* 'tough' og *næs* 'nice', som har samme form i alle køn, d.v.s. han er *kúl*, hun er *kúl*, barnet er *kúl*. En af de få undtagelser er *penn* 'pæn' – *pen* – *pent*. Som eksempel på hvort svært det har været for et ord at blive accepteret i islandsk hvis det ikke vil eller kan følge sprogets regler, er verbet *ske* 'ske'. Udtalen er i orden, *k* er palataliseret foran en palataliseret vokal, det bøjes regelmæssigt, *skeði* i præteritum og *skeð* i participium. Men det ender ikke i infinitiv på *-a* som andre islandske verber. Ordet dukker op i sproget i det 16. århundrede og har levet et relativt godt liv hele tiden trods purismens forsøg på at rydde det af vejen. I den sidste udgave af *Íslensk orðabók* (Islandsk ordbog) fra 2002 står der som forklaring i parentes "optaget i islandsk i det 15. århundrede, men anses endnu for at være upassende i formelt sprog".

Så er vi kommet til den sidste regel, som jo er ganske svær. Kun en lille del af fremmedordene har været optaget i ordbøger, man har undgået dem på tryk, de har ikke været behandlet i islandske retskrivningsregler og slet ikke i lærebøger. Derfor findes der i mange tilfælde ingen tradition for hvordan ordene skal skrives. Lad os se lidt på nogle ord fra Guðmundur Finnbogason's artikel. Jeg har nævnt ordet *kastarhola* med *h* i sidste sammensætningsled. Jeg har eksempler på *kastarhola*, *kastarola* uden *h*, *kastarolla* med dobbelt *-ll-*, *kastarhol* i neutrum, *kasserolla* med *e* i stedet for *a* i første sammensætningsled og *kassarolla*. Når man betragter de former som jeg har fundet, ser det ud til at to danske former ligger bag de islandske. Den ene er *kasserolle* og den anden *kasterolle*, som begge to var levende i det danske sprog omkring den tid hvor de islandske ord blev indlånt. Eksemplerne med *h-*, *-hol*, *-hola* er så et islandsk forsøg på at forklare den senere del af ordet. Desuden er *h* normalt stumt i begyndelsen af andet sammensætningsled. I dette tilfælde burde der måske være to opslagsord i ordbogen med henvisning fra det ene til det andet.

Da jeg spurgte om *dörslag* i radioudsendelsen, fik jeg også oplyst forskellige former, og de fleste tilstod at de ikke havde set ordet på tryk. Ordformerne var: *dörslag*, *dorslag* med *-r* og *doslag*, *duslag* eller *döslag* uden *-r*. I dette tilfælde vil der kun blive ét opslagsord med sideformer, og ordet sættes i gruppen uassimileret fremmedord på grund af skrivemåden.

Et af de ord Guðmundur nævnte, var *etasjer*, som ganske mange husker, men de fleste i formerne *ediser*, *edisjer* eller *ediseri*. Ordet kommer fra dansk *etager*, som igen har det fra fransk *étagère*. Oprindelsen findes i ordet *etage*; som oftest er det brugt om en etage i et hus, men de islandske ordformer blev øjensynlig ikke sat i forbindelse med det ord og er uassimilerede fremmedord. Etage i et hus blev nemlig til *tasía* fordi det ikke passede til systemet at ordet begynder på tryksvagt *e-*. *Tasía* i et hus blev brugt før i tiden, men høres ikke nu. Mange af mine hjemmelmænd kendte ordet *tasíupvottur* og fortalte mig at det var levende i sproget, især når et barn skal vaskes. Selv kendte jeg det godt fra min barndom, men havde ikke hørt det

længe. Det blev lavet efter det danske ord *etagevask*. De nævnte også *tasíudiskur*, som på samme måde har sin oprindelse i dansk *etagetallerken* ('kageopsats'). *Tasía* ville jeg regne som et fremmedord som andre ord på *-ía*, *melódía*, *útópía*, *harmónía*, ikke som et assimileret låneord.

Men én ting må man ikke glemme: Et ord kan opfylde alle Baldur Jónssons kriterier og alligevel ikke regnes som accepteret låneord. Reglerne er nemlig for lingvister. Kriteriet "føles fremmed" spiller en meget stor rolle hos almenheden. Ordet *blók* f.eks., som er indlånt fra engelsk *bloke*, opfylder alle kriterier, får blandt andet omlyd i plural, *blækur*, men "føles fremmed" og er derfor ikke accepteret som låneord, det er et fremmedord. Det samme gælder verberne *fíla* og *frika*, som opfylder alle kriterierne, er ganske almindelige i talesproget, men får betegnelsen "slang" i Íslensk orðabók og er ikke accepteret som en del af ordforrådet af ganske mange fordi de "føles fremmede".

5. Hvad skal med i en låne- og fremmedordbog?

Da jeg oprindeligt begyndte at tænke på en ordbog over den del af ordforrådet som er af fremmed oprindelse, havde jeg i sinde at bruge danske og tyske ordbøger som forbillede og kun tage fremmedord med, d.v.s. ord som føles fremmede såvel af lingvister som af almenheden. Så ville kun en del af de eksempler som jeg har nævnt, komme med i bogen. Men nu er jeg igen kommet i tvivl. Sådant en bog ville være den første af sin art på islandsk, og man ville forvente at kunne finde oplysninger om et ords oprindelse uden hensyn til dets klassifikation som enten låneord eller fremmedord. *Tökuorðabók*, i den gamle betydning af *tökuorð* som jeg før har nævnt, d.v.s. alle ord af fremmed oprindelse, ville måske være den rigtige titel på sådan en bog. I øjeblikket har jeg i sinde at gå den vej, begynde omkring år 1900 og skelne mellem låneord og fremmedord i selve ordbogsartiklerne i databasen på grundlag af lingvistiske kriterier. Så ville de fleste få hvad de ønsker.

Litteratur

Ordbøger:

Árnason, Mördur (red.) 2002: *Íslensk orðabók*. 3. udgave. Reykjavík: Edda.

Blöndal, Sigfús (red.) 1920-1924: *Íslensk-dönsk orðabók*. Reykjavík: I kommission hos Verslun Þórarins B. Þorlákssonar og Aschehoug & Co. (W. Nygaard).

Nordisk leksikografisk ordbok. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 4. 1997. Oslo: Universitetsforlaget.

Anden litteratur:

Finnbogason, Guðmundur 1920: Hreint mál. I: *Skírnir* 102, 145-155.

Jónsson, Baldur 1997: Zur formalen Anpassung von Fremdwörtern im Isländischen. I:

Skandinavistik 27, 15-23.

Jónsson, Jón Hilmar 1980: Om skrivemåte og bøyning av fremmedord i islandsk. I: *Språk i Norden* 1980, 61-67.

Kvaran, Guðrún 2001: Tanker om en islandsk fremmedordbog. I: *Nordiska studier i lexikografi*. Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 6. Göteborg, 125-136.

Laxness, Halldór 1939: Ill danska. I: *Tímarit Máls og menningar*, 33-34.

Varför förändras (inte) ordböcker? Exemplet Florinus

Why do/don't dictionaries change? The Florinus example. This article discusses changes occurring in a minor thematic dictionary, compiled by the Finnish priest Henrik Florinus and first published in 1678 in Turku under the title *Nomenclatura Rerum brevissima Latino-Sveco-Finmonica*. In the following hundred years, this dictionary was reissued in at least fifteen editions, the last in 1771 in Stockholm under the title *Vocabularium Latino-Svecum*. The article identifies and discusses the different reasons for the changes (modernisation, product development, space requirements, mistakes, supply of printing types, and modifications marking a new edition) as well as for the absence of certain changes (labour savings, the circumstance that this type of dictionary no longer was viewed as modern, and the authoritativeness of dictionaries).

1. Florinus ordbok

År 1678 utgavs i Åbo en oktavvolym med följande tidstypiskt omständliga titelsida:

NOMENCLATURA Rerum brevissima, Latino-Sveco-Finmonica, In usum Juventutis patriæ, ordine naturæ in certos Titulos digesta, & publici juris facta, Studio & sumptibus HENRICI M. FLORINI Pastoris Ecclesiæ Pæmarensis Austro Finn. ABOÆ. Excusa à JOHANNE Winter/ ANNO 1678.

Det rör sig alltså om en latinsk-svensk-finsk ordbok i det mindre formatet ("brevissima"), avsedd för fäderneslandets ungdom ("Juventutis patriæ") och inte bara utarbetad utan även förlagd och finansierad ("sumptibus") av Henrik Florinus, kyrkoherde i Pëmar i södra Finland. Av titelsidan framgår också att den är tematiskt uppställd ("ordine naturæ") och indelad i kapitel efter ämnesområden. Benämningen "NOMENCLATURA Rerum" antyder ett annat av ordbokens kännetecken: den innehåller inga verb utan förtecknar så gott som enbart substantiv och (till en mindre del) adjektiv.

Det som framför allt gör denna ordbok intressant är att den – i mer eller mindre (o)förändrat skick – fortsatte att utkomma i nära 100 år, sista gången 1771:

H. F. VOCABULARIUM LATINO-SVECUM, In Usum Juventutis, Ordine Naturæ, in certos Titulos digestum, & publici juris factum. En kårt Orde-Bok, På Latin och Swenska, Til Ungdomens Nytto, efter en naturlig Ordning, uti wißa Titlar författad, och af Trycket utgången; Å nyo uplagd och förbättrad. STOCKHOLM, Tryckt på Direct. LARS SALVII kostnad, 1771.¹

¹ På sedvanligt sätt återges här antikvan med kursiv och frakturen med rak stil.

Mellan 1678 och 1771 utkom ytterligare 14 utgåvor av ordboken. En schematisk presentation av samtliga utgåvor ges i tabell 1.²

år	utgivningsort (utgivare)	språk	paginerade sidor	antal arti- klar	(huvudsaklig) sättningsförlaga
1678	Åbo	lat/sv/fi	2-148	3 163	–
1683	Åbo	lat/sv/fi	2-148	3 177	1678
1695	Sthlm (Hök)	lat/sv/ty/fi	4-239	3 176	1683
1706	Sthlm (Hök)	lat/sv	2-116	3 174	1695
1708a	Sthlm (Hök)	lat/sv/ty/fi	4-239	3 177	1695
1708b	Sthlm (Hök)	lat/sv/ty/fi	4-239	3 177	1708a
1720	Strängnäs	lat/sv	2-116	3 173	1706
1722	Sthlm (Horrn)	lat/sv	2-108	3 180	1708a
1728	Lund	lat/sv	2-112	3 176	1708a
1733a	Sthlm (Horrn)	lat/sv/ty/fi	4-217	3 180	1722
1733b	Sthlm (Horrn)	lat/sv	2-108	3 180	1733a
1740	Sthlm (Horrn)	lat/sv	2-108	3 180	1722
1754	Sthlm (Salvius)	lat/sv	2-108	3 180	1740
1755	Göteborg	lat/sv	2-112	3 180	1733b
1770	Göteborg	lat/sv	2-112	3 180	1755
1771	Sthlm (Salvius)	lat/sv	2-108	3 180	1754

Tabell 1. *De bevarade utgåvorna av Florinus ordbok*

Att jag utan större betänkligheter tillåtit mig att sammanföra dessa 16 ordboksutgåvor under samlingsbeteckningen *Florinus ordbok* har givetvis sin grund i de stora likheterna (och de uppenbara beroendeförhållandena) dem emellan. Ordbokens grundläggande struktur med den tematiska uppställningen och indelningen i 22 kapitel förblir helt oförändrad, och detsamma gäller med vissa smärre undantag också för antalet artiklar (jfr tabell 1) och deras inbördes ordning liksom för uppsättningen av latinska ingångsord. Likaså bibehålls karaktären av skolordbok, avsedd för latinstuderande ynglingar med svenska (eller i de två tidigaste utgåvorna: finska) som modersmål. Visserligen gjorde utgivaren 1695 vissa – om än inte särskilt

² I samtliga utom två fall har jag använt mig av de exemplar av ordboken som finns på Uppsala universitetsbibliotek eller Kungliga biblioteket i Stockholm. Undantagen är 1678 som har utgetts i faksimil (se litteraturlistan) och 1722 som bara tycks finnas på Åbo Akademis bibliotek. – När det gäller sättningsförlagorna är uppgifterna med nödvändighet behäftade med en viss osäkerhet; det *kan* ju finnas ytterligare utgåvor som inte är bevarade. – För en närmare presentation av de olika utgåvorna och deras inbördes beroendeförhållanden hänvisas till Larsson under utgivning.

långtgående och helhjärtade – försök att utvidga ordbokens användningsområden genom att tillfoga tyskan och ställa upp vokabulärerna spaltvis (med latinet och svenskan på versosidorna och tyskan och finskan på rectosidorna); därigenom blev det ju i princip möjligt att använda ordboken också i andra riktningar än utifrån latinet, något som också antyds i det nyskrivna svenska förordet. Dock fick detta uppenbarligen inget större genomslag. I de följande fyrspråkiga utgåvorna 1708 och 1733 har ambitionen förbleknat, och i den tvåspråkiga variant som med tiden blir allennarådande rör det sig entydigt om en ordbok i latin för svenskar och ingenting annat.

Samtidigt kan man konstatera att Florinus ordbok i andra avseenden genomgår nog så betydelsefulla förändringar under sin nära sekellånga utgivningshistoria: språkens antal varierar, ordboken får (med början 1695) en ny huvudtitel *Vocabularium* i stället för det tidigare *Nomenclatura rerum brevissima*, artiklarnas typografiska arrangemang blir betydligt striktare och enhetligare, och – sist men inte minst – sker en hel del förändringar i den svenska vokabulären. Och i det följande kommer jag att lite närmare diskutera hur orsakssammanhangen ser ut: Vad ligger bakom förändringarna eller avsaknaden av förändringar? I huvudsak har jag utgått från den svenska vokabulären i de åtta utgåvor som kan sägas utgöra ordbokens huvudfåra: de två utgåvorna i Åbo 1678 och 1683 samt de sex Stockholmsutgåvorna 1695, 1708a, 1722, 1740, 1754 och 1771 (jfr kolumnen med sättningsförslag i tabell 1).

2. Förändringarna i den svenska vokabulären

För att få ett mått på förändringarna i den svenska vokabulären har jag summerat antalet svenska artikelfält som skiljer sig från den ena utgåvan till den andra (se tabell 2). (Vid sammanräkningarna har jag med två undantag när inkluderat samtliga grafematiska förändringar. Undantagen är dels vissa slag av förkortningar (i form av nasalstreck och underförstådda ordled samt förkortningen *it.* för *item*), dels sådana uppenbara tryckfel som korrigerats i följande utgåva. Inte heller har jag räknat in skillnader i fråga om versal och gemen, sär- och sammanskrivning samt interpunktion.)³

³ Det bör framhållas att de siffror som presenteras här och i det följande inte gör anspråk på fullständig exakthet. Dels kan klassificeringarna av förändringarna i enstaka fall vara diskutabla, dels är det knappast möjligt att helt undvika förbiseenden vid genomgången av ordböckerna och vid sammanräkningarna. De felmarginaler man har att räkna med är dock knappast större än de variationer som (till följd av t.ex. uppdragna och förstörda typer) torde finnas mellan olika exemplar av en och samma utgåva av ordboken.

	1678 ↓ 1683	1683 ↓ 1695	1695 ↓ 1708a	1708a ↓ 1722	1722 ↓ 1740	1740 ↓ 1754	1754 ↓ 1771
S:a	207	1457	216	360	314	848	155

Tabell 2. *Antalet förändrade artikelfält i den svenska vokabulären*

Ser man till det totala antalet förändringar kan man alltså urskilja två förändringstoppar, framför allt 1695 då närapå hälften av artiklarnas svenska artikelfält förändras, men även 1754 då drygt en fjärdedel berörs. I övrigt är andelen förändrade artikelfält däremot förhållandevis liten.

Minst lika intressant som antalet förändringar är frågan vilka slags förändringar det handlar om. I tabell 3 har jag därför särskilt (I) sådana förändringar som enbart berör de i artikelfälten ingående ordens form, dvs. ortografiska och morfologiska förändringar (för termerna, se Santesson 1986:24), och (II) sådana förändringar som på ett eller annat sätt påverkar betydelsen och som jag valt att kalla lexikala. (Till de senare har jag räknat inte bara förändringar i fråga om rotmorfem utan också grammatiska morfem som artiklar och pluraländelser. Dessutom har jag inkluderat förändringar när det gäller förekomsten av de metalexikografiska angivelserna *item* och *eller*, liksom också de fåtaliga ändringarna i artikelföljden – i en tematisk ordbok kan ju även artiklarnas ordningsföljd sägas vara betydelsebärande.)

	1678 ↓ 1683	1683 ↓ 1695	1695 ↓ 1708a	1708a ↓ 1722	1722 ↓ 1740	1740 ↓ 1754	1754 ↓ 1771
I.	170	1415	204	335	312	836	155
II.	37	42	12	25	2	12	–
S:a	207	1457	216	360	314	848	155

Tabell 3. *Antalet förändrade artikelfält uppdelat i kategorierna I. formella förändringar, II. lexikala förändringar*

I tabell 3 är det som synes en delvis annorlunda bild som framträder. När det gäller lexikala förändringar dominerar 1683 och 1695, och likaså finns ett förhållandevis stort antal 1722. Däremot finns det inte särskilt många 1754; i denna utgåva handlar det framför allt om ortografiska förändringar.

Också inom gruppen lexikala förändringar går det att urskilja olika kategorier. I tabell 4 har jag särskilt (I) strykningar, där ett eller flera morfem har strukits, (II) utbyten, där ett eller flera morfem har ersatts av ett eller flera andra, och (III) tillägg, där ett eller flera morfem har tillfogats.

	1678 ↓ 1683	1683 ↓ 1695	1695 ↓ 1708a	1708a ↓ 1722	1722 ↓ 1740	1740 ↓ 1754	1754 ↓ 1771
I.	2	33	4	1	–	2	–
II.	12	6	4	11	1	5	–
III.	23	3	4	13	1	5	–
S:a	37	42	12	25	2	12	–

Tabell 4. *Antalet lexikala förändringar uppdelade i kategorierna I. strykningar, II. utbyten, III. tillägg*

Mest frapperande i tabell 4 är att det 1683 är tilläggen som dominerar medan den stora merparten av de lexikala förändringarna 1695 däremot utgörs av strykningar. Påfallande är också det ringa antalet strykningar från 1722 och framåt.

Ytterligare ett sätt att belysa förändringsmönstren är att dela upp kategorin tillägg i olika underkategorier beroende på vilken strukturell nivå som berörs. I tabell 5 har jag därför skilt mellan (I) helt nya artiklar, (II) uppdelning av artiklar med mer än ett latinskt ingångsord, och (III) tillägg inom de svenska artikelfälten. (När det gäller kategorierna strykningar och utbyten handlar det med ett litet undantag enbart om förändringar på artikelfältsnivå).

	1678 ↓ 1683	1683 ↓ 1695	1695 ↓ 1708a	1708a ↓ 1722	1722 ↓ 1740	1740 ↓ 1754	1754 ↓ 1771
I.	14	–	–	–	–	–	–
II.	–	–	1	3	–	–	–
III.	9	3	3	10	1	5	–
S:a	23	3	4	13	1	5	–

Tabell 5. *Antalet tillägg uppdelade i kategorierna I. helt nya artiklar, II. uppdelning av artiklar, III. tillägg inom artikelfälten*

Mest iögonfallande i tabell 5 är att helt nya artiklar bara förekommer 1683. De små tillökningarna av antalet artiklar 1708 och 1722 handlar bara om att artiklar delats upp, och längre fram förekommer över huvud taget inga nya artiklar; tilläggen håller sig helt inom ramen för de svenska artikelfälten.

3. Varför förändras ordboken?

Bakom förändringarna i Florinus ordbok kan man – mer eller mindre ovedersägligt – urskilja åtminstone sex olika orsaksfaktorer: tidsanpassning, produktutveckling, utrymmesbesparingar, misstag och missförstånd, tillgången på typer samt självständighetsmarkeringar.

3.1. Tidsanpassning

Att tidsanpassning eller modernisering är (eller åtminstone kunde förväntas vara) ett viktigt förändringsincitament är givet: i och med att den samhälleliga och språkliga verkligheten – en verklighet som ordboken både är en del av och en avspeglning av – förändras över tiden måste också ordboken förändras för att inte bli otidsenlig. De mest uppenbara fallen av tidsanpassning är de många ortografiska förändringarna i 1695 och 1754 års utgåvor. Mest iögonfallande 1695 är att morfemfinalt <dh> och <gh> i stort sett försvunnit medan vokalgeminering blivit undantag snarare än regel (jfr Larsson under utgivning). Till de mest frekventa förändringarna 1754 hör att vokalbetecknande <v> ersatts av <u>, <i> som konsonantgrafem av <j>, <z> av <s>, <ch> av <g>, <k> eller <ck>, och <ck> i ställning efter konsonant av <k>. Likaså är särmarkeringarna av lång vokal genom geminering eller <h>-inskott i det närmaste försvunna (se Larsson 2005:181). (Jfr Santesson 1986 passim. Att förändringarna 1754 låg i tiden visas om inte annat av att i stort sett samma ortografiska förändringar också uppträder i 1755 års utgåva från Göteborg, trots att denna av allt att döma tillkom oberoende av 1754 års ordbok.)

Till tidsanpassningen kan likaså föras många av de förändringar på det morfologiska planet som möter 1722 och 1754. Några exempel är 1708 års **Seande** (s. 22), **Banno** (s. 88), **Arbet** (s. 126), **Spatzerning** (s. 182), **Then lyckelig igenkommen är** (s. 188) och **Kåld** (s. 218) som 1722 blivit respektive **Seende** (s. 10), **Banna** (s. 38), **Arbete** (s. 56), **Spatzering** (s. 80), **Then lyckeligen igenkommen är** (s. 83) och **Kall** (s. 98), samt 1740 års **Sängkläder** (s. 17), **Handtog** (s. 20), **Försumelig** (s. 30), **Furoträ** (s. 64), **Ett onaturligit Foster** (s. 69) och **Tarmalopp** (s. 90) som 1754 ändrats till **Sängkläder**, **Handtag**, **Försumlig**, **Furoträ**, **Et onaturligt Foster** respektive **Tarmlopp**. På den lexikala nivån tycks moderniseringar däremot vara mycket sällsynta. I stort sett begränsar de sig till några enstaka fall 1722. Två exempel är **Triumph/ en härlig Inridning efter behållen seger** för latinets *Triumphus* (1708a:200) där den nya utgåvan har **effter erhållen seger** (s. 88), samt artikeln **Peninsula**, där de tidigare utgåvornas **Nääs/ thet nämsta omflutit** (1708a:134) 1722 getts lydelsen **Nääs/ thet mäst är omflutit** (s. 59).

3.2. Produktutveckling

Oavsett och oberoende av förändringarna i språket och omvärlden är det förstås alltid möjligt att på olika sätt utveckla och förbättra ordboken och därigenom

(förhoppningsvis) förläna den en ökad attraktionskraft på marknaden. Till denna produktutveckling hör de lexikala förändringar i form av tillägg (såväl på artikelnivå som inom de svenska artikelfälten) som införs 1683, den enda nyutgåva där Florinus själv är involverad. Likaså kan många av de lexikala förändringarna 1722 föras till denna rubrik: så t.ex. de tre artiklar som delats upp (ett exempel är *Arbustum, i, lucus, i, saltus, us. Lund* (1708a:144) som 1722 motsvaras av dels *Arbustum, i, lucus, i. Lund*, dels *Saltus, us. Tiock skog* (s. 64)), liksom de tre tidigare hänvisningsartiklar som nu försetts med svenska ekvivalenter (t.ex. *Quadriga, æ, &c.* (1708a:186) med den nyinsatta betydelseförklaringen **En med 2. par hästar före** (s. 82)). I fall som dessa handlar det ju inte om några av tiden framtvingade förändringar utan om försök till produktförbättringar.

Produktutvecklingen framträder också på ett mera övergripande plan. Framför allt gäller det för många av de förändringar som möter 1695: så t.ex. att de latinska böjningsändelserna regelmässigt är utsatta (tidigare förekom det bara undantagsvis), att den tyska vokabulären har tillfogats, liksom att titelsidorna, registret och kapitelrubrikerna ges på samtliga de fyra språken. Detsamma är fallet med introduktionen av den tvåspråkiga varianten 1706 och (inom huvudfåran) 1722. Någonstans i gränsområdet mellan produktutveckling och tidsanpassning ligger också den utveckling mot allt större enhetlighet och konsekvens i fråga om artiklarnas arrangemang som kännetecknar ordboken: den spaltvisa uppställningen 1695, den strikta spalthållningen 1708 (1695 tillät sig sättaren att vid brist på utrymme göra avkall på de raka vänstermarginalerna i svenskan och finskan) och de enhetliga typgraderna 1722 (i de tidigare utgåvorna laborerade sättarna med två eller tre olika typgrader i var och en av vokabulärerna).

3.3. Utrymmesbesparingar

Även om tidsanpassning och produktutveckling säkerligen utgör de främsta förändringsincitamenten finns det ett inte obetydligt antal förändringar som pockar på andra förklaringar. Viktigt i sammanhanget var utan tvekan det tillgängliga utrymmet. Kostnaden för papperet utgjorde en dryg post i tryckeriernas budget (av de löpande kostnaderna stod enligt Gaskell (1974:177) papperet för ”often about 75 per cent of the total in the sixteenth century, dropping to about 50 per cent by the eighteenth”), och att hålla nere denna kostnad genom att spara på utrymmet var därmed en nog så betydelsefull faktor vid bokutgivningen. Även om dessa utrymmesbesparingar är mer eller mindre skönjbara i så gott som samtliga utgåvor märks de tydligast 1695: i och med att tyskan tillkom, att latinets böjningsändelser sattes ut och att vokabulärerna ställdes upp spaltvis blev det uppenbarligen nödvändigt att på sina ställen komprimera de svenska artikelfälten.

Utrymmesbesparingarna kommer till uttryck på såväl den formella som den lexikala nivån. När det gäller de ortografiska förändringarna kan de visserligen vara

svåra att skilja från tidsanpassningen: moderniseringen av stavningen innebar ju ofta en reducering av antalet grafem (som t.ex. när 1683 års **Klookheet/ försichtigheet** (s. 45) och **Lustigh/ gladh** (s. 141) 1695 blivit **Klokhet/ försichtighet** (s. 76) respektive **Lustig/ glad** (s. 228)). I vissa fall är det emellertid svårt att förklara stavningsförändringarna på annat sätt än som utrymmesbesparingar. Ett exempel – där det av allt att döma är typernas relativa bredd som fällt avgörandet – är 1683 års **Qwittans/ qwittobref** (s. 73) där den senare ekvivalenten 1695 har fått den mindre utrymmeskrävande stavningen **quittobref** (s. 116). Ett intressant fall är likaså 1683 års **Commodus, aptus, idoneus, Beqväm/ tienligh** (s. 144) som 1695 blivit **Commodus, a, um, aptus, a, um, idoneus, a, um. Beqwē/ tienlig** (s. 232). Det mesta talar för att ändringen från <ä> till <e> berodde på att det var ont om plats: något *ä* med nasalstreck fanns ju inte att tillgå, men genom att i stället välja ett *ē* kunde man undkomma den spatiösa *m*-typen. (För att utbytet är betingat av utrymmesskäl talar också att ändringarna 1695 i de allra flesta fall går i den motsatta riktningen från <e> till <ä>.)

Framför allt märks dock de utrymmesbesparande åtgärderna på det lexikala planet. Som framgått av tabell 4 utgörs det stora flertalet (33 av 42) av de lexikala förändringarna 1695 av strykningar. Och av allt att döma har huvuddelen av dessa sin förklaring i brist på utrymme. Några exempel är 1683 års **Contumelia, opprobrium, Försmädelse/ spott/ wahnähra/ smäleek** (s. 54) mot 1695 års **Försmädelse/ spott/ wanähra** (s. 88), **Colluvies, En församblad orenligheet/ en hoop arga bofwar** (s. 58) mot **Samlad orenlighet/ en hop bofwar** (s. 92), **Chorus, En hoop Folck/ som siunger eller dantzar/ item en hoop** (s. 78) mot **En hop Folck/ som siunger eller dantzar** (s. 124) samt **Theatrum, Skådeplatz/ ther man seer Comædier och andra speel** (s. 78) mot **Skådeplatz/ ther man ser allehanda Spel** (s. 124). Utan dessa förkortningar av uttryckssätten skulle sättaren haft mycket svårt att inrymma artiklarna på det utrymme som stod till buds.

3.4. Misstag och missförstånd

En annan – inte helt oviktig – orsak till förändringarna i Florinus ordbok skulle kunna rubriceras som misstag och missförstånd. Ett exempel av det mera banala slaget är 1740 års **Brudesång/ bröllopskrifft** för latinets *Epithalamium* (s. 43) som 1754 och 1771 (kanske under inflytande av den närmast föregående artikelns **Ekta Säng** för *Torus*) kommit att bli **Brudesäng, Bröllopskrift**. (Givetvis skulle ett fall som detta helt enkelt kunna avfärdas som tryckfel, men i och med att förändringen vidareförs i nästa utgåva kan den ju sägas ha blivit en del av ordbokens lexikala innehåll.) Av större intresse är sådana missgrepp och missförstånd som i senare utgåvor för vidare till andra lösningar. Ett exempel är 1678 års **Hypothecarius. Pantägande** (s. 73) som 1771 motsvaras av **Hypothecarius, i. Panttagare** (s. 51). Sannolikt går förändringen tillbaka på att sättaren 1695 av misstag fick ett *a* i sin

hand i stället för ett *ä*, varför den svenska ekvivalenten kom att bli **Pantagande** (s. 116). Medan 1708 års utgåva nöjde sig med att bättra på stavningen till **Panttagande** föredrog 1722 års ordbok att i stället byta ut efterleden till **Panttagare** (s. 51).

Ett annat exempel är 1678 års *Sicarius, Mördare/ som bär Mordwärja* (s. 119) mot 1771 års *Mördare, som bär Mordgewär* (s. 85). Även om ändringen vid första anblicken kanske kan se ut som en modernisering förefaller det onekligen lite märkligt med tanke på den nära anknytningen till den närmast föregående artikeln *Sica*, där en och samma betydelseförklaring **Fördoldt swärd/ lönwärja** återfinns i samtliga utgåvor av ordboken. Och antagligen står förklaringen att finna i 1708 års utgåva där bristen på utrymme fick sättaren att då och då ta till icke konventionaliserade förkortningar; till dessa hörde *Mördare/ som bär Mordwär.* (s. 194), en förkortning som 1722 tydligt kom att utläsas som **Mordgewär**.

3.5. Tillgången på typer

En orsak till förändringar på den formella nivån som, även om den knappast låter sig påvisas helt entydigt, nog inte bör underskattas är tillgången på typer. Att tryckeriernas typförråd inte alltid ville räcka till är ett ofta omvittnat faktum, och fullt möjligt är att vissa av de till synes godtyckliga ortografiska förändringarna i ordboksutgåvorna har en sådan bakgrund. Frånsett utgåvorna 1754 och 1771 (liksom de två Göteborgsutgåvorna), som med sin enhetligare och konsekventare ortografi pekar framåt mot en ny tid med betydligt fastare stavningsregler, ger ordböckernas svenska vokabulärer prov på ett avsevärt mått av fri variation på det ortografiska området (jfr Santesson 1986:280 passim); och i och med denna fria variation hade sättarna ett betydande manöverutrymme för att hantera eventuella brister i typförrådet. När t.ex. utgåvan från 1740 på en och samma sida (s. 43) uppvisar fyra ändringar från <ft> till <fft> (däribland **Brudesång/ bröllopskrift** mot 1722 års **Brudesång/ Bröllopskrift**), samtidigt som det på en och samma sida i följande ark (s. 52) finns fyra ändringar i motsatt riktning (däribland **Motskrift/ genskrift** mot den äldre utgåvans **Motskrift/ genskrift**), skulle förklaringen kunna vara den relativa tillgången på de två typerna *f* och *ff*; för 1740 års sättare torde nog enkelt och dubbeltecknat <f> i kontexter som dessa hur som helst ha framstått som helt utbytbara.

3.6. Självständighetsmarkeringar

Kanske kan en orsak till förändringarna i Florinus ordbok också vara att utgivaren ville markera sin självständighet gentemot föregångaren. I vilken mån sådana självständighetsmarkeringar verkligen spelade någon roll i sammanhanget är givetvis vanskligt att avgöra. Men kanske kunde det vid de stridigheter om utgivningsrättigheter som var så vanliga under detta tidevarv (se t.ex. Schück 1923 passim) vara av värde att kunna peka på att nyutgåvan var något mer än ett eftertryck. Och likaså

kunde kanske den förpliktigande formulering som möter på titelsidorna i samtliga sex utgåvor från Horrns och Salvius tryckerier – ”Å nyo uplagd och förbättrad” – ha sporrat utgivaren att här och där tillfoga någon synonym i de svenska artikelfälten; det är annars inte helt lätt att hitta någon förklaring till att t.ex. 1740 års *Miseratio, onis, commiseratio, onis, misericordia, æ, condolentia, æ. Medynckan/ Förbarmelse/ Barmhertighet/ Medlidande* (s. 42) 1754 har utökats med ännu en svensk ekvivalent till *Medynkan, Förbarmelse, Barmhertighet, Medlidande, Mißkunsamhet*.

4. Och varför inte?

De ovan redovisade förändringarna till trots är det som faller i ögonen vid en jämförelse mellan de 16 utgåvorna av Florinus ordbok snarare likheterna än skillnaderna: som redan framgått förblir utgåvorna i mångt och mycket identiska. Och den svenska vokabulären utgör inget undantag därvidlag. Vad man kan konstatera utifrån den statistik som presenterats ovan är ju framför allt två saker. För det första är förändringarna i ordbokens svenska vokabulär i det stora hela ganska måttfulla. Av de sammanlagt 3 180 artiklarna 1771 återfinns 1 118 – dvs. mer än en tredjedel – i samma gestalt 1678. Därtill kommer att skillnaderna i de allra flesta fall enbart rör ortografin; antalet artikelfält som förändrats lexikalt är bara 120. Det mesta är sig alltså likt, och med tanke på hur stora förändringarna i både samhället och språket är under dessa nära 100 år – utgåvorna sträcker sig ju från det karolinska till det gustavianska tidevarvet, från Samuel Columbus tid till Abraham Sahlstedts – kan Florinus ordbok förefalla märkligt statisk och oföränderlig; man kunde ha förväntat sig mer av förändringar.

För det andra finns det (tvärtemot vad den första summeringen i tabell 1 gav för handen) en tilltagande förändringsobenägenhet. Förändringarna i makrostrukturen, när det gäller uppsättningen av artiklar, liksom tilläggen i de svenska artikelfälten möter till allra största delen redan i den andra utgåvan 1683, och i de tre sista utgåvorna i huvudfåran, 1740, 1754 och 1771, rör det sig mestadels bara om förändringar i stavningen.

En minst lika intressant fråga blir alltså varför ordboken *inte* förändras mer än den gör. I huvudsak torde det röra sig om en samverkan mellan tre olika faktorer: det handlade om arbetsbesparing, att Florinus ordbok inte var någon framtidsprodukt, samt den inneboende auktoritet som ordböcker besitter.

4.1. Arbetsbesparing

Som framgått ovan hade de tre först nämnda – och säkerligen viktigaste – orsakerna till förändring det gemensamt att de var ekonomiskt betingade: det handlade om att antingen sälja mer genom att marknadsanpassa produkten eller hålla nere produktionskostnaderna genom att minska pappersåtgången. Och med all sannolikhet var

den viktigaste orsaken till att ordboken inte förändrades, till att så mycket förblev vid det gamla, också av ekonomisk natur. Det var frågan om en arbetsbesparing: att inte omarbete och revidera innebar mindre arbetsinsatser och därmed lägre kostnader. Man kunde använda en tidigare utgåva som sättningsförlaga utan att det krävde något mera omfattande redigeringsarbete. Och om det bara rörde sig om ändringar på den formella nivån behövdes kanske inte någon förberedande redigering över huvud taget, utan arbetet kunde mer eller mindre i sin helhet överlåtas åt sättningsförlagaren: mycket möjligt är att det var han som applicerade de för tidsanpassningen nödvändiga korrigeringsfiltren (för termen, se Santesson 1986:23). Givetvis underlättades också produktionen om man i möjligaste mån kunde bevara förlagans uppställning och sidbrytning, något som i sin tur förutsatte att förändringarna inte var av alltför genomgripande natur.

4.2. Ingen framtidsprodukt

En annan orsak till att Florinus ordbok i så liten utsträckning förändrades var antagligen det faktum att den med tiden i allt mindre utsträckning framstod som en framtidsprodukt. Det handlade om ett på flera sätt föråldrat koncept. För det första hade den tematiska uppställningen på 1700-talet sin storhetstid bakom sig; det var den alfabetiska ordningen som hade framtiden för sig (jfr Hüllen 1999:444). För det andra kan man skönja en förändring av den latinsk-svenska lexikografins inriktning under 1700-talet. Belysande är en jämförelse mellan Florinus ordböcker och den 1790 utkomna *Lexicon Latino-Svecanum*: medan Florinus har en samtidsinriktning i samma tradition som t.ex. Lincopensen och Comenius latinläroböcker – där det i första rummet var frågan om att lära sig använda latinet som kommunikationsspråk – är 1790 års ordbok på ett annat sätt inriktad på antikens språk och samhälle. Dels visar det sig i de svenska betydelseförklaringarna. Några exempel är latinets *Lora*, *Accubitalia* och *Gulo* där Florinus svenska ekvivalenter är de samtidsförankrade *Spijsööl* (1678:19), *Sängekläder* (s. 23) och *Järf* (s. 99) medan 1790 års ordbok har respektive *Efter-win, som präßas af utpräßade drufwor, Kläder, hwarmed Sofforne öfwerhöljdes, på hwilka man låg till bords* och *Fråßare*. Dels framgår det av lemmaurvalet: sådana beteckningar för ideologiska och tekniska landvinningar som *Idololatria*, *Affguderij* (s. 38), *Typographus*, *Booktryckiare* (s. 45) och *Bombarda*, *Byssa* (s. 120) har över huvud taget inga motsvarigheter 1790.

Florinus ordbok torde alltså ha representerat ett koncept på utgående. Och i och med detta måste incitamenten för att lägga ner tid och arbete på en grundligare revidering ha varit ganska små. I stället valde man att (frånsett den förhållandevis lättvindiga ortografiska tidsanpassningen) fortsätta i de gamla hjulspåren så länge produkten gick att sälja – betecknande är att man knappast kan tala om någon produktutveckling efter 1722.

4.3. Ordböckers auktoritet

Vid sidan av dessa två orsaker till att så lite förändrades kan man ana ytterligare en: ordböckers inneboende auktoritet. Som McArthur (1986:46) skriver har det alltsedan de medeltida klosterskrivarnas tid funnits ”an element of *reverence* in our civilization to works of *reference*”; och att så mycket gammalt hänger med från den ena utgåvan till den andra kan nog också ha att göra med att det just handlade om en ordbok. (Att Lars Salvius – utgivaren av 1754 och 1771 års ordböcker – annars inte var främmande för att grundligt omarbete texter finns det flera exempel på; se Santesson 1986:297 och Johannisson 1961.)

Signifikativt är inte minst att antalet strykningar är så få i de senare utgåvorna (se tabell 4). Trots att 1722 års ordbok fortfarande uppvisar en tydlig ambition att utveckla produkten – och trots att den enhetliga typgraden måste ha gjort det ännu besvärligare att disponera och få plats med texten – innehåller den bara ett enda exempel på strykningar. Det gäller 1708 års *FUNDUS, i. Fast egendom/ såsom Åker/ Äng/ etc. item Botn* (s. 130) som 1722 – av allt att döma av utrymmesskäl – förkortats till *Fast egendom/ såsom Åker/ item Botn* (s. 58) (att man i detta enda fall tillät sig en strykning kan nog hänga samman med att det strukna morfemet stod att finna i den närmast följande artikeln *Pratum, i. Äng*). Och vad gäller de två strykningarna 1754 är bakgrunden lite speciell. Det handlar om de två artiklar som 1678 lyder *Morbus, siukdō/ siuka/ kranckheet* och *Æger, ægrotus, morbidus, Siuk/ siukligh/ kranck* (s. 124) men där 1754 års ordbok bara har två ekvivalenter, *Sjukdom, sjuka* respektive *Sjuk, sjuklig* (s. 89). Redan 1708 hade här den förra artikelns ekvivalenter genom en kombination av förkortning och sättningsfel blivit *Siukdom/ Siuka/ Kanckh.* (s. 200), en lydelse som återkommer i oförändrat skick 1722 och 1740. Och måhända är det detta som ligger bakom strykningarna: den som vid utgivningen 1754 ställdes inför detta märkliga *Kanckh.* kan nog ha funnit för gott att utelämna det helt och hållet, för att sedan när han kom fram till adjektivet *Kranck* åtta artiklar längre ner i konsekvensens namn stryka även detta.

Det förefaller alltså som om man med tiden blev allt ovilligare – man frestas att skriva: inte vågade – att avlägsna några betydelsebärande element ur texten. Och kanske hänger det samman med ordbokstextens auktoritet. Att här och där stryka några av synonymerna i de svenska artikelfälten kunde annars ha varit ett förhållandevis enkelt sätt att råda bot på flera av de problem som vidhäftade Florinus ordbok. Ett är förstås den brist på utrymme som så ofta gör sig påmind i utgåvorna. Ett annat problem är att vissa av ordbokens ekvivalenter under 1700-talets senare hälft nog kan ha framstått som mer eller mindre obsoleta. Det faktum att t.ex. *hickan* (H 885), *glödkara* (G 625), *begravningsståt* (i betydelsen ’likprocession’ (B 736)), *högstol* (i betydelsen ’kateder, pulpet’ (H 2278)), *lustighet* (i betydelsen ’raskhet, hurtighet’ (L 1240)), *nystestol* (N 899) och *spetskenekt* (S 9541) samtliga har Florinus 1695 som sista belegg i SAOB antyder att dessa ord eller betydelser inte var helt och fullt

gångbara 1771. Trots detta – och trots att de ur funktionell synvinkel lätt hade kunnat undvaras – återfinns de ännu i den sista utgåvan av ordboken i sammanställningarna **Hickning**, **Hickan** (s. 10), **Glöd-kara**, **Eld-raka** (s. 20), **Begravnings-ståt**, **Lik-Process** (s. 26), **Högstol**, **Lärarens Stol** (s. 29), **Lustighet**, **Snällhet**, **Wighet** (s. 34), **Nystfot**, **Nyststol**, **Hjulråk** (s. 77) samt **Pikner**, **Spetskneckt** (s. 85).⁴

Det finns ytterligare en aspekt av strykningarna som har med ordbokens funktion att göra. I och med att det var frågan om en receptionsordbok – eller kanske hellre: en passiv översättningsordbok (se NLO s. 214) – för latinstuderande svenskar kan sådana uppradningar av synonymer som **Faarsyster/ Faster** och **hynda/ tijkä/ täfwa** för latinets *Amīta* (1678:64) och *Canicula* (s. 97) förefalla tämligen överflödiga; men ännu i 1771 års ordbok återfinns alltså samma ekvivalenter **Farsyster**, **Faster** (s. 45) respektive **Hynda**, **Tik**, **Täfwa** (s. 69) i de svenska artikelfälten (vad gäller *farsyster* kunde kanske redan det faktum att SAOB:s sista belägg är från ca 1750 (F 358) ha motiverat en strykning).

Litteratur

Ordböcker:

Florinus, Henrik 1678: *Nomenclatura rerum brevissima Latino-Sveco-Finnonica, in usum juventutis* [...]. Åbo. Faksimilutgåva med efterskrift av Riitta Suhonen, 1976. Helsingfors.

[För de övriga utgåvorna av Florinus ordbok hänvisas till tabell 1.]

Lindblom, J. A. 1790: *Lexicon Latino-Svecanum, Latinsk och Svensk Ord-Bok* [...]. Uppsala.

NLO = *Nordisk leksikografisk ordbok*. Skrifter utgitt av Nordisk forening for leksikografi.

Skrift nr. 4. Oslo 1997.

SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1-. Lund 1898-.

Annan litteratur:

Gaskell, Philip 1974: *A New Introduction to Bibliography*. Oxford.

Hüllen, Werner 1999: *English Dictionaries 800-1700. The Topical Tradition*. Oxford.

Johannisson, Ture 1961: Nils Matson Kiöping och Lars Salvius. I: *Nils Matson Kiöpings Resa. Parallelltexter ur andra och tredje upplagorna. Med efterskrift och anmärkningar utgivna av Ture Johannisson*. (Svenska texter 6.) Stockholm, 125-142.

Larsson, Lennart 2005: Mönstrens makt. Om särmarkeringarna av lång vokal i Florinus ordböcker. I: *Studier i svensk språkhistoria* 8. Utg. av Cecilia Falk och Lars-Olof Delsing. (Lundastudier i nordisk språkvetenskap A:63.) Lund, 179-187.

Larsson, Lennart, under utgivning: Florinus ordbok. I: *Språk och stil* 16, 2006.

McArthur, Tom 1986: *Worlds of Reference. Lexicography, learning and language from the clay tablet to the computer*. Cambridge.

⁴ Att SAOB:s sistabelägg från Florinus regelmässigt tidfästas till 1695 beror uppenbarligen på att det i stort sett bara är 1678 och 1695 års utgåvor som excerperats.

Santesson, Lillemor 1986: *Tryckt hos Salvius. En undersökning om språkvården på ett 1700-talstryckeri med särskild hänsyn till ortografi och morfologi.* (Lundastudier i nordisk språkvetenskap A:37.) Lund.

Schück, Henrik 1923: *Den svenska förlagsbokhandelns historia 2.* Stockholm.

Genveje i den elektroniske bilingvale ordbog illustreret ved hjælp af *man* → *you, one, we, they* og syntaktiske løsninger

Short cuts in an electronic bilingual dictionary illustrated by *man* → *you, one, we, they* and syntactic solutions. Generally speaking, advanced language students in Denmark fail to tackle the non-existence of a full equivalent of the Danish indefinite pronoun *man* competently when expressing themselves in English – they heavily overuse the partial equivalent *you*. German and French have similar pronouns: *man* and *on*, but their dictionaries, just like Danish-produced English grammars, make a better point of warning and of offering solutions than do most Danish-English active dictionaries. On the basis of the combined material from dictionaries and grammars, this paper provides a suggestion for a flow chart for the Danish lemma *man*, intended for an electronic, dynamic, active Danish-English dictionary with links to external sources such as corpora and word lists.

1. Baggrund

Når et lemma i en bilingval aktiv ordbog har mange gengivelsesmuligheder i målsproget, kan artiklen blive omfattende, og selv for den rutinerede bruger kan det være trælsomt at træffe det rette valg, især hvis der er mange dimensioner at tage hensyn til.

Det danske indefinitte pronomen *man* er i relation til engelsk et eksempel på et sådant lemma. Dets eneste syntaktiske rolle er som subjekt, det anvendes på alle sprogniveauer og rangerer som nr. 29 i frekvens for det danske sprog ifølge 'Sproglige hitlister' i Korpus 2000 (tallet er dog ifølge Korpus 2000-siden indtil videre kun hentet fra Korpus 90). *Man* modsvares af et leksikalsk hul i engelsk ('lucka', Svensén 2004:332), men er et hyperonym med ikke alene en række delvise, (hyponyme) ækvivalenter i engelsk, idet der foreligger divergens fra kilde- til målsprog (op.cit.:317), men også en række oversættelsesmuligheder, hvor lemmaet ikke gengives som egentlig ækvivalent, men kommer til udtryk gennem syntaksen. Problemet diskuteres som et eksempel på 'impersonal constructions in translation' ud fra den professionelle oversætters synsvinkel af Smith (1998). Et samlet billede af mulighederne ses i figur 5. Summen af udtryksmuligheder på målsproget overstiger således summen af 'betydelser' på målsproget i den forstand, hvori det diskuteres af Svensén (2004:317). Valget af oversættelse kan derfor bero i mange hensyn, og ofte kan brugeren således ikke være sikker på at have truffet det mest hensigtsmæssige valg, før stort set hele artiklen er set igennem, en problematik der har ført til omarbejdelse af præpositionsartiklerne i Blinkenberg & Høybye: *Fransk-dansk ordbog*, 3. udg., 1997 (Rasmussen 1998:134).

Det spørgsmål, der stilles her, lidt provokeret af Bergenholtz & Tarp, der efterlyser mere orientering om løsninger i elektroniske ordbøger (2005), er derfor, om ikke den elektroniske ordbog i et sådant kompliceret tilfælde kan give brugeren bedre service end den traditionelle papirordbog, dels i kraft af dynamisk indretning og dels i kraft af links til tekstkorpora (ibid.), men også til mange andre typer af interne og eksterne links (Svensén 2004:523-526). Det skal der her gives et bud på i form af et rutediagram. Men forinden ses der på behandlingen af *man* i en række traditionelle ordbøger og grammatikker, for det er der, inspirationen til indhold og struktur af diagrammet er hentet, i lighed med traditionelle ordbøger (Rasmussen 1998).

1.1. *Problemets dobbelte karakter*

Leksemet *man*'s rige oversættelsesmuligheder fremgår af figur 1, hvor den største dansk-engelske ordbog, Vinterberg & Bodelsen (V&B98), der figurerer som ordbogskolonne nr. 6, har 18 muligheder. Tilføjelse af bilingvale ordbøger med andre kildesprog end dansk giver yderligere to muligheder (se figur 1). Endnu to muligheder kommer til, hvis man går til grammatikkerne, som det fremgår af figur 2. Og på det spekulative plan kan man forestille sig, at mulighederne ikke er udtømt med disse registreringer.

Den ene side af *man*-problemet er således ækvivalentmulighedernes mangfoldighed, men den anden og lige så vigtige side er, at mange danskere overforbruger ækvivalenten *you*, når de formulerer sig på engelsk. Dette overforbrug synes der at være en god forklaring på, hvis man ser på figur 1's ti dansk-engelske ordbøger, udgivet i perioden fra 1975 til i dag, hvor *you* har højeste prioritet i langt de fleste, markeret med '1'. Ser man derimod på fremstillingen i engelsk-grammatikker skrevet for danskere, er der al mulig grund til at forvente andre gengivelser end *you* prioriteret, som det fremgår af figur 2, og som det vil blive diskuteret nærmere i afsnit 3.2. Overforbruget af *you* har muligvis også en anden forklaring, nemlig at der i de senere år er sket en stigende anvendelse i dansk af pronomenet *du* på bekostning af *man* (Allan et al. 1998, Buhl 1998:254 & Sørensen 1995:165), hvilket her illustreres med et enkelt eksempel udtalt af læge Anette Storr-Paulsen i Danmarks Radio i et interview om det problem, at teenagere ofte får utilstrækkelig søvn:

Og er *man* træt i hverdagen, kan *du* ikke fungere (8-radioavisen, 25.04.2005)

Vore dages danskere har måske således den fejlopfattelse, at pronomenet *you* altid generelt har samme elastiske mulighed for referentielt indholdspotentiale (Pedersen 2005:307), nemlig til både 1., 2. og 3. person, som pronomenet *man* har. Hvor tvivlsomt det kan være altid at forlade sig på oversættelse til *you*, illustrerer nedenstående eksempel, som er anvendt i undervisningen i økonomisk sprog på Handelshøjskolen

i København, og som er hentet fra et interview med den tidligere formand for bestyrelsen i Tele Danmark Communications, Knud Heinesen, om 'privatisering' af statsmonopoler i Danmark. Knud Heinesen siger:

Jeg mener, at det [privatisering] har været et gode, og det har samtidig været nødvendigt. Hvis [1] *man* ser på, hvad der har drevet den udviklingså må [2] *man* sige at det er inspireret af EU, hvor [3] *man* skal have fri konkurrence på et fælles marked. (Berlingske Tidende, 22. september 2002)

hvor det første *man* uden problemer kan gengives med *you*, men hvor en gengivelse af det andet *man* næsten uundgåeligt ville virke docerende. Intervieweren, og læseren, kunne føle sig talt ned til, måske endda manipuleret med – altså næppe nogen loyal oversættelse. Men Heinesen vil antagelig gerne invitere journalist og læser med i sit synspunkt, altså et solidariseringsiltag fra hans side, og derfor ville han nok godtage ækvivalenten *we*.

Ved det tredje *man* er fokus især på dem, der er aktører på EU's markeder, men først og fremmest på selve konkurrencen. Derfor bør valget nok falde på passiv, altså en gengivelse uden deiktisk udpegning.

2. Gengivelsesmuligheder

2.1. Gengivelsesmuligheder for *man* i ordbøgerne

Et nærmere blik på ti dansk-engelske ordbøger fra de sidste 30 år i figur 1, ordbogskolonner 1-10, viser, at de gengivelsesforslag i den enkelte ordbog, der rangerer højest, gennemgående er egentlige ækvivalenter, anført i fed skrift. Ordbøgerne er opstillet fra venstre mod højre i udgivelses-kronologisk orden. Derpå følger tre tysk-engelske og endelig otte fransk-engelske ordbøger, ligeledes for hvert sprog i kronologisk orden. For disse ordbøger drejer det sig naturligvis om gengivelsesmuligheder for henholdsvis det tyske indefinitte pronomen *man* og det franske *on*, dog uden at der er set på, hvorvidt disse to pronomener eventuelt ikke ganske svarer til det danske *man*.

Tallene i kolonnerne angiver nummer i rangfølgen i den enkelte ordbog. Således angiver det øverste tal i den første talkolonne, nemlig tallet 2 ud for muligheden 'you', at ordbogen Syn75 (Næsted: *Engelsk-Dansk Synonym-Ordbog*), har ækvivalenten 'you' som nr. 2. Et kig hen gennem de danske kolonner viser, at den placering deler Næsted med kun to andre dansk-engelske ordbøger ud af de ti undersøgte. Næsteds højst prioriterede mulighed fremgår at være ækvivalenten 'one', en placering han deler med de samme to andre ordbøger, og med to tysk-engelske og tre fransk-engelske.

Den rækkefølge af ækvivalenter, altså kolonnen yderst til venstre, som er tertium comparationis for sammenligning af samtlige ordbøger og grammatikker, er den

rækkefølge, der figurerer i den største af de dansk-engelske ordbøger, nemlig V&B98.

Ord- bøger Mulig- heder	Syn75	GR-U90	Rou91	Gj&Ga92	GR95	V&B98	GylMin01	Pol02	Gad stri03	IT	Pen 90	OxDu99	Har 01	HaNe72	Cas80	Web 80	Rob 83	HaOx94	Larous95	HaPe 01	OxUn01
you	2	1	2	1	1	1	1	1	1	2	2	2	2		4		5	3	4	5	1
one	1	5	1	2	6	2		2	2	1	1	1		1	1	1	6	10	5	4	6
we/I		4			4	3	3	3			5			3	3	4	11	5	11	3	2
they	3	2	3	3	2	4	2	4	4		4			4	2	2	3	12	1	1	3
people/men		3		4	3	5			3	5	3	9	3	2	5	3	10		3	2	4
folkel: body, girl, chap, man, fellow	5				5	6								6					8		
Passiv	4	6	4	5	7	7	4		5	4	6	4	1	7		5	1	1	2		7
it-konstr		7			8	8	5			3	7	8		8	8	6	7	2	6		
Bydemåde		8	5		9	9			6		8	5						9			
udbrud: really!		9			10	10															
verbalsubst						11												7			
Verbalskift= sige→høre						12															
this is+inf						13															
there	2		6			14					3		5			8	8	10			
any-/some- /nobody/-one						15					6		9	6		4	11	9	6	5	
vb→subst						16										2		7			
finit rel → infin						17															
kendt vending			7			18	6	5			10		11	7		9	4	12			
Præcisering													10			12	6				
Tidsangiv.														9							

Figur 1. *Dansk-, tysk- og fransk-engelske ordbøger*

Et blik ned over skemaet som helhed viser, at mulighedernes udfyldning og rækkefølge varierer meget, måske fordi ordbøgerne er meget forskellige i omfang. Dette er ikke undersøgt, og det er heller ikke formålet med denne undersøgelse. Det, der er interessant her, er at få mange dokumenterede muligheder ind i billedet, og det er også grunden til, at andre ordbøger end dansk-engelske er inddraget. Yderligere inddragelse af ordbøger kunne måske have givet et rigere billede, men udvalget af tysk- og fransk-engelske ordbøger fra et dansk udgangspunkt har vist sig at være meget beskedent og er derfor stort set det, der findes på Handelshøjskolen i

København. Som det fremgår af figur 1, har tilføjelsen af de 3+8 tyske og franske ordbøger øget listen på atten muligheder med to, og det antages derfor, at yderligere ordbøger ikke ville øge antallet af muligheder væsentligt.

Hvad der især forekommer interessant i skemaets brogede billede, er to ting, nemlig for det første, at muligheden *you* rangerer meget højt i de danske ordbøger, medens omskrivning med passiv rangerer på 4.-7. pladsen. Omskrivning med passiv rangerer derimod højt i flere af de ikke-danske ordbøger. Og for det andet, at bortset fra V&B98, som har mange sofistikerede muligheder med, så er sofistikerede muligheder til stede i større udstrækning i de ikke-danske ordbøger.

Grunden til, at disse to ting er interessante, er, at på disse to områder er der god overensstemmelse mellem de ikke-danske ordbøger og engelsk-grammatikkerne udgivet i Danmark, se figur 2.

2.2. Gengivelsesmuligheder for man i grammatikkerne

	Gramma- tikker	Stel.&Sø.66	Tho,Ni&Ve81	Hersk&Ped85	Vesterg85	Johans&Stev90	Bruun&Vei94	Swan95	Olsen97	Buhl98	Hju&Schw98	Lassen&Tarp00	Borg00	Bendtsen00	Olsen03	Hju&Johan04
	Mulig- heder															
0	Advarsel	(x)		(x)	x	x	x	(x)	x	x	(x)	x	x	x	x	x
1	<i>you</i>	+	1	2	2	2	5	2	3	5	3	1	11	2	3	1
2	<i>one</i>	+	3	1	4	3	4	1	2	4	6	2	9	3	2	2
3	<i>we/I</i>	+		5	1	4	6		6	9	2	5	10	4		3
4	<i>they</i>	+	2	4	3	5	2	3	4	7	4	3	12	5	4	4
5	<i>people/ men</i>			3	5	6	3		5	8	5	4	13	6		5
6	folkeligt: <i>body, girl, chap, man, fellow</i>										9					
7	passiv 1) simp vend 2) He is believed to	+	4	6	6	1	1		1	++	1	6	1	1	1	6
8	it-konstr				7					2	10	8	2	7		
9	Bydemåde									3	10		4			
10	udbrud: really!															
11	Verbalsubst															
12	Verbalskift+ Rolleskift															
13	this is+inf															

14	There					7					7	3	7		
15	any-/some- /nobody/-one														
16	vb→subst the hope is														
17	fin.rel →infi- nitiv														
18	Kendt vending														
19	præcisering							6				8	8		
20	Participium	+			+							5			
21	-able/-ible											6			
22	Tidsangivelse					8						7			

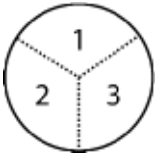
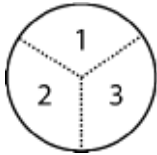
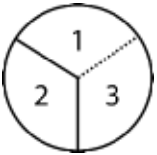
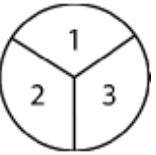
Figur 2. *Engelsk-grammatikker udgivet for danskere + Swan: Practical English Usage*

Divergenserne mellem ordbøger og grammatikker manifesterer sig foruden i rangordenen ved, at en del muligheder aldrig er med i de danske engelsk-grammatikker. Antagelig af indlysende grunde – en grammatik er ikke en ordbog. Iøjnefaldende ligheder er derimod, at ækvivalentadskillende informationer på deiktisk grundlag figurerer såvel i grammatikkerne som i ordbøgerne. Disse informationer skal der herefter ses nærmere på i de to typer opslagsværker, først i ordbøgerne, figur 3.

3. Betydningsdifferentiering

3.1. Betydningsdifferentiering i ordbøgerne

Et vigtigt spørgsmål i lyset af divergensen (kapitel 1) fra kilde- til målsproget er, hvor stort et persondeiktisk felt de enkelte pronominale ækvivalentmuligheder dækker.

Ækvivalenter Ordbøger	<i>you</i> 	<i>one</i> 	<i>we</i> 	<i>they</i> 
Syn75	indbefatter den tiltalte	hvor den talende først og fremmest har sig selv i tankerne		især når man hentyder til folk på et bestemt sted
GR-U90	den tiltalte medregnet	i mere formelt sprog = en selv, folk	den talende medregnet, den tiltalte ikke medregnet	den talende og den tiltalte ikke medregnet
Rou91	andre mennesker	inkl. en selv	*	andre mennesker
GJ&Ga92	når den talende kan indbefattes	den talende medregnet	*	den talende og den tiltalte ikke medregnet
GR95	den tiltalte medregnet	F	den talende medregnet, den tiltalte ikke medregnet	den talende og den tiltalte ikke medregnet
V&B98	især med den tiltalte indbefattet	især med den talende indbefattet	[anført som syn til 'one']	den talende og den tiltalte ikke indbefattet
GylMin01		*		
Pol02			*	
Gadstri03	inklusive den man taler med	inklusive en selv		Andre mennesker end en selv og den/dem man taler med
IT				

Figur 3. Ordbøgernes betydningsdifferentiatorer for udvalgte ækvivalenter. (* Angiver at ækvivalenten mangler, og der er således heller ikke betydningsangivelse).

3.1.1. Ækvivalenten *you*

Her er ordbøgernes svar vanskelige at forholde sig til, idet det er usikkert hvem der kan være indbefattet foruden den, man taler til (2. person i cirklen). Rent kombinatorisk er der mange muligheder, hvis man ser på ental og flertal for sig. Men kan brugeren forlade sig på, at de alle kan dækkes af *you*? Et udtømmende svar kræver en større undersøgelse.

3.1.2. Ækvivalenten *one*

Her gælder samme usikkerhed som for ækvivalenten *you*, blot med et andet kernefokus, nemlig den talende selv.

3.1.3. *Ækvivalenten we*

Tre ordbøger har ikke denne mulighed med, men V&B98, som har den med under betydningsdifferentiatoren for *one* og altså sidestiller disse to, har imidlertid ingen oplysning om formelt sprog. Og virker det i øvrigt logisk, at *we* ikke skal kunne omfatte ‘tiltalte’ (2. person)? Et målsprogseksempel, der harmonerer godt med betydningsdifferentiatoren, figurerer i tre af Gyldendals ordbøger: ‘that is how we do it in Denmark’. Men i Politikens eksempel under *one*, uden betydningsdifferentiering; ‘hvad skal ~ gøre? = what is one to do’, kunne *man* i passende kontekst sagtens gengives med *we*. Eksempler kan således bekræfte en betydningsdifferentiator, men ikke nødvendigvis afklare, hvad der måtte være usikkert.

3.1.4. *Ækvivalenten they*

Her er der ingen tvivl, der er total udelukkelse af 1. og 2. person, og det forvirrende ‘indbefattet’ optræder da heller ikke.

3.1.5. *Konklusion*

På baggrund af ordbøgerne alene må konklusionen være, at hvis brugeren ønsker at omfatte potentielt hele det deiktiske persongalleri og undgå at bruge formelt sprog, så må *you* forekomme at være løsningen. Og der sker som sagt også et overforbrug af *you*. Men det burde der jo ikke gøre, for grammatikkerne advarer grundigt mod det.

3.2. *Betydningsdifferentiering i grammatikkerne*

Her skal nu ses lidt nærmere på, hvordan grammatikkerne griber problemet an, idet der her alene fokuseres på prioritering af passivløsningerne, figur 4, kolonne 3, og på de advarsler, der gives, om at *man* oversættelsesmæssigt udgør et problem. Den altovervejende anvisning i grammatikkerne ses at være en engelsk passivkonstruktion, kolonne 4.

Forfatter	Udg.-år	Passivs rang	Advarsel i den enkelte grammatik
Poul Steller & Knud Sørensen	66		“Flere af de personlige pronomener bruges generisk (=‘man’); det gælder <i>you</i> , <i>we</i> og <i>they</i> : ... – Mærk dog, at i denne anvendelse har disse pronomener stadig bevaret noget af deres centrale betydninger, således at <i>you</i> bruges, når man ønsker at inddrage den, man taler til ... Note. Det, der oftest svarer til dansk ‘man’, er en engelsk passivkonstruktion ” (side 201)
Thomsen, Nielsen & Vester	81	4	““man” oversættes slet ikke: sætningen laves om til lideform ” (side 60)

Herskind & Gravers Pedersen	84	6	“103 Det danske ord man kan gengives på forskellige måder: 103a <i>one</i> bruges, når den talende inkluderer sig selv... 103b <i>you</i> ... 103e passivkonstruktionen er hyppig og ofte den bedste oversættelse af “man”” (s. 44-45)
Vestergaard	85	6	I note til afsnit om reference: “Engelsk har ikke noget generisk personligt pronomen svarende til det danske <i>man</i> . “Generisk person” kan udtrykkes ved (1) <i>we</i> ... I øvrigt kan dansk <i>man</i> hyppigt gengives ved <i>people</i> , passivkonstruktion eller upersonlig konstruktion.” (side 153)
Johansson, Stevens, Almqvist & Wiksell	90	1	“Oversættelse af dansk “man” A passiv konstruktion ”
Bruun, Veirup Systime	94	1	“Engelsk har ikke et ord, der svarer til det danske “man”. Man kan gengive det på flere forskellige måder, men man skal passe virkelig godt på at vælge rigtigt: a) Passivkonstruktion – en meget almindelig form”
Swan (britisk opslagsværk)	95	?	“ <i>One</i> is not used to generalise about people who could not include the speaker; <i>you</i> is not used to generalise about people who could not include the hearer” (side 394)
Flemming Olsen	97	1	“Der er ingen enkelt gløse for det danske ord <i>man</i> . Det danske ord er ubestemt i den forstand, at det bruges uden at den talende kender eller ønsker at komme nærmere ind på identiteten af den eller dem, der har sagt eller gjort et eller andet. ... Den hyppigste gengivelse for <i>man</i> er passiv omskrivning ” (side 162)
Ole Buhl	98	?	““ Man ” er uhyre udbredt i det danske sprog. Det skyldes formentlig, at sprogbrugeren ved at bruge “man” slipper for at tage stilling til, hvem “man” egentlig er. ... Oversættelsen af “man” volder ofte problemer. ... “ Man ” er med andre ord ganske ubestemt . Det er derfor ganske naturligt, at den bedste oversættelse af det danske “man” ofte er en engelsk passivkonstruktion ...” (side 252-256)
Hjulmand & Schwarz	98	1	“The most neutral translation of Danish MAN is the passive without an agent. This construction is of course only possible when the active clause has a DO/IO” (143)
Lassen & Tarp	00	6	“På engelsk afhænger valget af, hvem der refereres til. Det, der oftest svarer til dansk “man”, er engelsk passivkonstruktion”
Torben Borg	00	1	“Engelsk har ikke et præcist enkeltord, der svarer til “man”, og vælger derfor i mange tilfælde at formulere sætningen i passiv (uden agensled)” side 137. Herefter inddeles problematikken under overskrifterne: “ Hvis uklart, hvem “man” er ” og “ Hvis klart hvem “man” er ”, hvor passiv figurerer under førstnævnte.
Knud B. Bentsen, Ole Bønnerup & Frede S. Pedersen	00	1	“111 Engelsk har ikke et enkelt ord, der svarer til det danske <i>man</i> . Der er bl.a. følgende muligheder for at gengive det danske <i>man</i> : 111a Brug af passiv ” (side 120)
Flemming Olsen	03	1	Afsnittet “DET DANSKE MAN” indledes med “ Passiv omskrivning ” (side 77)
Johansson, ovs. & bearb.: Lise-Lotte Hjulmand	04	6	Som sjette mulighed under afsnittet om MAN henvises til afsnittet om passiv “som ofte er en god måde at oversætte det danske <i>man</i> ” (side 77)

Figur 4. Grammatikkernes anbefalinger af passiv

Samtidig ses betoningen af passiv som mulighed at være rykket op i rangfølgen. Flere grammatikere gør desuden opmærksom på, at gengivelse med pronomener reducerer anonymiteten i kildesprogets *man*. Disse to forhold er elementer i strukturen i det afsluttende rutediagram, figur 6, hvor Borgs grammatik (2000) med spørgsmålet om, hvorvidt det er uklart eller klart, hvem *man* er, har været en stor inspiration til at udarbejde diagrammet, og til i det hele taget at arbejde med muligheden for en dynamisk mikrostruktur.

3.3. Det samlede billede af betydningsdifferentieringen

Det billede, der samlet tegner sig af gengivelsesmulighederne på grundlag af de undersøgte ordbøger og grammatikere, ses i figur 5, hvor de tekstelementer i betydningsangivelseskolonnen, der figurerer i kursiv, er gengivelser eller synteser af de betydningsdifferentieringer, der optræder i ordbøgerne. De øvrige elementer er udledt af gengivelsernes karakter, inspireret af grammatikkerne.

	Mulighed	Betydningsangivelse	Dansk eksempel	Engelsk oversættelse
1	you	<i>især med den tiltalte indbefattet, uformelt</i>	Man får valuta for pengene der	You get a lot for your money there
2	one	<i>især med den talende indbefattet, formelt</i>	Hvad skal man gøre?	What is one to do?
3	we/I	<i>den talende medregnet, Solidaritet. Den tiltalte ikke medregnet</i>	Sådan gør man her i huset	That is how we do it in this house
4	they	<i>den talende og den tiltalte ikke indbefattet</i>	Man siger hun er rig	They say that she is rich
5	people/ men/ the world	= ovenfor for they	Man siger hun er rig	People say that she is rich
6	folkeligt: body, girl, chap, man, fellow	Folkeligt uformelt sprog, daglig tale	(Og hvad skal man så gøre)	What is a body to do
7	passiv	a) <i>simpel vending:</i> objekt til subjekt b) subjektsrokade: subjekt i objektssætning → subjekt i <i>passiv</i> sætn	a) Man fangede ham a) Man så ham klatre over muren a) Man fortæller os b) Man mener han er død b) Man vil vide at han har sagt	a) He was caught a) He was seen to climb over the wall a) We are told ... b) He is believed to be dead b) He is understood to have said
8	it-konstr kataforisk	kræver objekt i kilde-sp. passivvending med it som foreløbigt subj. cl. skift fra vb → adj.	Man mener at ... Man er klar over at	passiv: It is held that ... vb → adj: It is obvious that

9	bydemåde	<i>direktiv tekst</i> <i>MAN uoversat</i>	Man bedes ringe to gange	Please ring twice
10	udbrud: really!	MAN uoversat	Ser man det?	Is that a fact? Indeed!
11	verbal-subst	MAN's finitte verbum ændres til <i>ing-form</i>	som MAN siger	as the saying goes
12	verbal-skift	rolleskift: obj. → subj. ved verbalskift: f.eks. sige → høre	Man har sagt mig On va encore augmenter l'essence	I have heard Petrol is going up again
13	this is+inf		Man skal herved meddele	this is to inform you
14	there	<i>THERE som foreløbigt subjekt = skift fra verbum til substantiv</i>	a) Man kan ikke vide det b) Man lo og snakkede	a) There is no way of knowing b) There were laughter and talking
15	any-/some-/nobody/-one			From the way he talked anyone would have thought I was a common thief (V&B)
16	vb>subst the hope is	hovedsætningens finitte verbum substantiveres og bliver subjekt	(Lige nu håber man at få mindst én ekspert i faget til at undervise i det.)	the present hope is to get at least one expert on the subject to teach it
17	finit rel-sætning → infinitiv		(Det er noget man kan være stolt af)	That is a feat to be proud of
18	kendt vending	Liste i Korpus 2000	(Hvordan har man det i dag?)	How are we today?
19	præcisering	<i>konteksten afslører identiteten af MAN</i>	Dans aucun pays on ne semble pouvoir arrêter l'inflation	No country seems able to stop inflation
20	participium	finit verbal i en betingelses-sætning → participium	Hvis man ser det fra den vinkel Hvis man ser generelt på det	Seen from this point of view Taking it by and large (Borg)
21	finit relativ-sætning → -able	<i>hele rel-sætningen gengives med verbalet som efterstillet adjektiv med -able -ible</i>	- man kan forestille sig - den eneste man kunne se	The worst outcome imaginable The only person visible
22	tidsangivelse → subjekt	<i>de hyppigste verber i målsproget er see og find</i>	Der kan ikke herske tvivl om at man i nær fremtid vil opleve store fremskridt inden for informationsteknologien	No doubt the near future will see a lot of progress within information technology

Figur 5. Præsentation af betydningsinddeling

Det samlede differentieringsbillede for *man* er således materialet til det følgende rutediagram, der tænkes at indgå bag brugergrænsefladens mikrostruktur i en

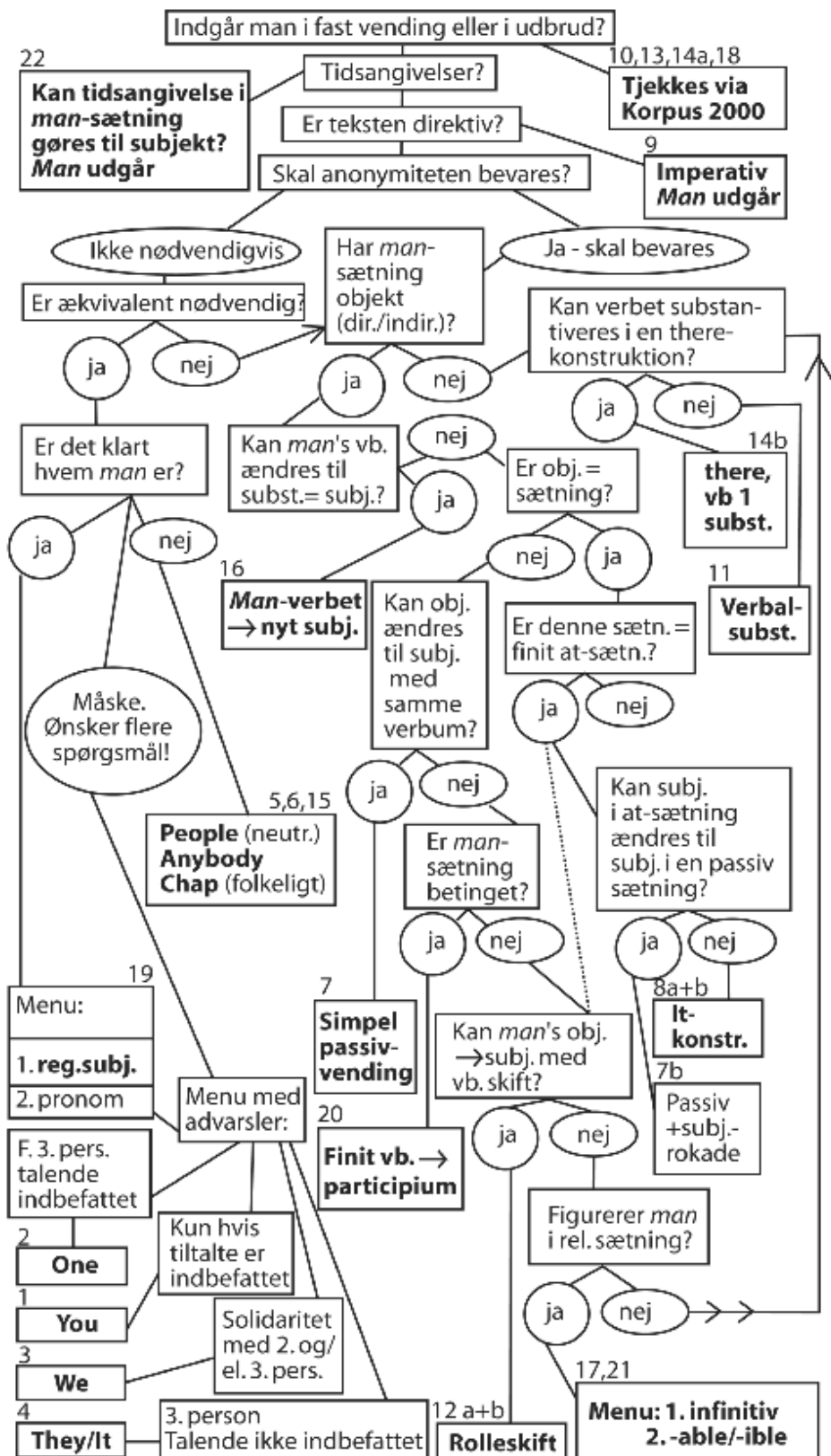
konkret artikel i en dynamisk elektronisk aktiv ordbog for lemmaet *man*, idet ordbogen tænkes at have flere brugerniveauer (Simonsen 2001:202), hvor det foreliggende diagram er beregnet på det højeste niveau.

4. Rutediagram

Diagrammet er begrænset til alene at rumme de løsninger, som figurerer i ordbøgerne og grammatikkerne, se figurerne 1-3, og dermed de iboende syntaktiske afsæt i kildesproget. Rækkefølgen i diagrammet i figur 6 er omvendt i forhold til f.eks. V&B98, hvor 'kendte/faste vendinger' figurerer til sidst. I en dynamisk ordbog, hvor det tænkes at være muligt at linke til eksterne ressourcer, eksempelvis et korpus med en voksende liste over faste vendinger, som det allerede findes for *man* i forbindelse med Korpus 2000, vil det være at gøre grin med brugeren, hvis der skal svares på spørgsmål, som slet ikke har relevans for de faste vendinger. Dem skal brugeren kunne få oversættelse for med det samme ud fra en passende tekststreng omkring *man*. En bilingval liste vil kunne oprettes med udgangspunkt i en allerede etableret liste i Korpus 2000 over kendte vendinger med *man*.

Brugeren skal derefter tage stilling til spørgsmål om tidsangivelse og om direktiv tekst, og derpå til det vigtige spørgsmål om anonymitet (Borg 2000:39). Hvis brugeren ønsker størst mulig anonymitet, kan det kun opnås ved at give afkald på en egentlig ækvivalent, for eksempel ved at vælge en passivkonstruktion. Og netop passivkonstruktionen er interessant, ikke kun fordi den anbefales så kraftigt i grammatikkerne, men også fordi det analyseværktøj, der er knyttet til Korpus 2000, gør det muligt at afgøre, om *man* har objekt, som jo er betingelsen for simpel passivvending, løsning nr. 7 (figur 1, 2 og 5). Og endelig også fordi en manuel optælling af *man* ifølge sætningsanalyse af Korpus 2000 (Bick, uden dato) viser, at ud af de første 100 har de 69 objekt, noget der understøtter, at passivkonstruktion anbefales så konsekvent. Det, vi så ikke ved, er, om ækvivalerende engelske transitiv verber altid stiller sig til rådighed for passivvendinger, altså hvordan deres 'leksikalske profiler' og 'konstruktionspotentiale' ser ud (Pedersen 2005:303), men dette kan antagelig søges oplyst via links til Wordsketch (Kilgarriff & Rundell 2002:813) i tilslutning til British National Corpus.

Som det fremgår af tabellerne i figur 1-5 og af diagrammet, er der mange andre syntaktiske muligheder. Ved opbygningen af diagrammet har strategien været at gå fra de mere overordnede spørgsmål som anonymitet og egentlig ækvivalent, via analyse til inddragelse af ordniveauet som f.eks. muligheden for at skifte fra verbet 'fortælle' til verbet 'høre' og ændre de øvrige elementer tilsvarende, altså en vekselvirkning som den, Svensén er inde på (2004:337), men hvor formelle kriterier er lagt til grund frem for semantiske (Rasmussen 1998), dog ikke i abstrakte formler, men med 'mere frie formuleringer' (Lorentzen & Trap-Jensen 2005:263), som de figurerer i figur 5 og i diagrammet.



Rutediagrammet (forrige side): Hvert rektangel i diagrammet med fed skrift er et facit og er forsynet med et tal, der refererer til tallene 1-22 i tabellerne i figur 1, 2 og 5 og er altså en gengivelsesmulighed.

Forsøger man at oversætte eksemplerne i figur 5 ét for ét via diagrammet, vil mange af dem kunne få flere facitter, som følge af at brugeren skal træffe valg, og vel også som følge af at oversættelse har mange muligheder. Forsøger man at finde eksempler, der ikke vil blive oversat, skal der ikke stor kreativitet til for at konstruere sådanne. Her har målet alene været at opstille et diagram for de muligheder, som eksisterende opslagsværker peger på.

Således har pladsen i denne sammenhæng heller ikke tilladt at afprøve diagrammet på et større materiale. Opstillingen af det foreliggende diagram har i sig selv været et betydeligt puslespil.

Litteratur

Ordbøger og korpora:

- Cas80 = *Cassell's French Dictionary: French-English English-French*. New York: Macmillan Publishing.
- Gadstri03 = Eva Jørgensen: *Gads sribede engelsk kombi*. København: G.E.C. Gads Forlag.
- GR95 = Jens Axelsen: *Gyldendals Røde Dansk Engelsk*. København: Gyldendal.
- GR-U90 = Jens Axelsen: *Gyldendals Røde Dansk Engelsk – Undervisning*. København: Gyldendal.
- Gj&Ga1992 = *Engelsk Skoleordbog*. Tilrettelægning: Knud Udbye. København: Gjellerup og Gad.
- GylMin01 = *Engelsk-Dansk/Dansk-Engelsk*. Gyldendals Miniordbøger. København: Gyldendal.
- HaOx94 = *The Oxford Hachette French Dictionary, French-English English-French*. Oxford: Oxford University Press.
- Har01 = *Harrap Compact English – English German/German English*. Compact German Dictionary. Edinburgh: Harrap.
- HarNe72 = *Harraps New Standard French and English Dictionary*. London: Harrap.
- HaPe01 = *Harraps Petit Dictionnaire. Anglais-Français. Français-Anglais*.
- IT = *Ordbogen.com – Danmarks største ordbog. Dansk-engelsk, engelsk-dansk*.
www.ordbogen.com.
- Korpus 2000 = <http://korpus.dsl.dk/korpus2000>
- FranceLarou95 = *Grand Dictionnaire Anglais-Français Français-Anglais*. Faye Carney. Paris: Larousse.
- OxDu99 = *The Oxford-Duden German Dictionary (German-English/English-German)*. Oxford: Oxford University Press.
- OxUno01 = *Oxford French Dictionary and Grammar*. Oxford: Oxford University Press.
- Pen90 = *The Penguin German-English Dictionary*. Compiled by Timothy Buch. London: Penguin Books.

- Pol02 = *Politikens Engelsk-Dansk Dansk-Engelsk*. København: Politikens Forlag.
 RobCol83 = *Dictionnaire Français-Anglais*. B.T.S. Atkins, Alain Duval & Rosemary C. Milne. London: Le Robert & Collins.
 Rou91 = *Danish Dictionary: English-Danish, Danish-English*. Editor: Anna Garde. Additional material by W. Glyn Jones. First published 1991 by G.E.C. Gads Forlag, Copenhagen. London: Republished 1995 by Routledge.
 Syn75 = Henning Næsted: *Engelsk-Dansk Synonymordbog*. København: Grafisk Forlag.
 V&B98 = Herman Vinterberg/C.A. Bodelsen: *Dansk-Engelsk Ordbog*. 4. udgave ved Viggo Hjørnager Pedersen. København: Gyldendal.
 Web80 = *Webster's French English Dictionary*. Castle Books.

Grammatikker:

- Bendtsen, Knud Børge, Ole Bønnerup & Frede Salling Pedersen 2000: *Engelsk Multimedie Grammatik*. Århus, Danmark: System A/S.
 Borg, Torben 2000: *Engelsk Grammatik – med komparative aspekter*. Rev. udgave, 2. oplag. Aalborg, Danmark: Institut for Sprog og Internationale Kulturstudier.
 Bruun, Poul, Lisbeth Veirup & forlaget system a/s 1994: *English Grammar Guide*. Herning, Danmark: system a/s.
 Buhl, Ole 1998: *Pragmatisk Engelsk Erhvervsgrammatik*. Holstebro, Danmark: Forlaget Ventus.
 Herskind, Aase & Uffe Gravers Pedersen 1984: *Engelsk grammatik med synonymer*. Lyngby, Danmark: Kaleidoscope Aps.
 Hjulmand, Lise-Lotte 2004: *Base – Engelsk basisgrammatik af Christer Johansson*. København: L&R Uddannelse A/S.
 Hjulmand, Lise-Lotte & Helge Schwarz 1998: *A Contrastive Grammar of English For Danish Students*. Frederiksberg, København: Samfundslitteratur.
 Johansson, Bertil, Michael Stevens & Almqvist & Wiksell Läromedel AB 1990: *Gads Grammatik Engelsk*. København: Gjellerup & Gad.
 Lassen, Inger & Gertrud Tarp 2000: *Sprogrevision og oversættelseskritik*. Publikation fra Institut for Sprog og Internationale Kulturstudier, Aalborg Universitet – vol. 26.
 Olsen, Flemming 1997: *Basics – engelsk grammatik med øvemateriale*. København: Munksgaard.
 Olsen, Flemming 2003: *Kursus i engelsk grammatik – Grundbog*. København: Gyldendalske Boghandel, Nordisk Forlag A/S.
 Steller, Poul & Knud Sørensen 1966: *Engelsk Grammatik*. København: Munksgaard.
 Swan, Michael 1995: *Practical English Usage*. Oxford: Oxford University Press.
 Thomsen, Jan, Lars Ole Nielsen & Hans Henrik Vester 1981: *Den lille fejlstøvsuger*. Vejle, Danmark: Musagetes Aps.
 Vestergaard, Torben 1985: *Engelsk Grammatik*. København: Schønberg.

Anden litteratur:

- Allan, Robin, Philip Holmes & Tom Lundskær-Nielsen 1998: *Danish: A Comprehensive Grammar*. Reprinted with corrections. London and New York: Routledge.
 Bergenholtz, Henning & Sven Tarp 2005: *Electronic dictionaries: Old and new lexicographic solutions*. I: *Hermes* 34, 7-9.
 Bick, Eckhard (uden dato): *Constraint Grammar-baseret DanPars-tagger*, der er online

- tilgængelig igennem VISL-projektet ved Syddansk Universitet.
<http://corp.hum.sdu.dk/cgi-bin/tgrepeye.exp.cgi>.
- Kilgarriff, Adam & Michael Rundell 2002: Lexical Profiling and its Lexicographic Applications – a Case Study. I: *Proceedings of the Tenth EURALEX International Congress, 13-17 august 2002, Copenhagen Volume II*. Copenhagen: Center for Sprogteknologi, 807-818.
- Lorentzen, Henrik & Lars Trap-Jensen 2005: Grammatiske oplysninger i *Den Danske Ordbog*. I: Ruth Vatvedt Fjeld & Dagfinn Worren (red.): *Nordiske Studier i Leksikografi 7, Rapport frå Konferanse om leksikografi i Norden*. Oslo, 252-266.
- Pedersen, Bolette Sandford 2005: Datamatisk Leksikografi i Norden – status og visioner. I: Ruth Vatvedt Fjeld & Dagfinn Worren (red.): *Nordiske Studier i Leksikografi 7, Rapport frå Konferanse om leksikografi i Norden*. Oslo, 302-314.
- Rasmussen, Jens 1998: Artikelstruktur for præpositioner. I: Anna Garde, Pia Jarvad & K.T. Thomsen: *Elefant – se også myg. Festskrift til Jens Axelsen*. København: Gyldendal. 133-141.
- 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.
- Smith, Ross 1998: *Problems with impersonal constructions in translation – a case study*. I: *The Linguist*, Vol. 37, No. 5, 135-136.
- Svensén, Bo 2004: *Handbok i lexikografi – Ordböcker och ordboksarbete i teori och praktik*. Andra upplagen. Stockholm: Norstedts Akademiska Förlag.
- Sørensen, Knud 1995: *Engelsk i dansk – Er det et must?* København: Munksgaard.

Ordbogseksterne henvisninger – fra belæg til funktioner

External References in Dictionaries – From Attestation to Functions. While there has been a lot of metalexicographical interest in the use of internal cross references in dictionaries, the use of references to texts and sources outside the dictionary has been largely ignored. The purpose of this article is threefold: Firstly to examine the general practice of external references in printed and Internet dictionaries; secondly to develop the necessary theoretical principles, and finally to establish a set of guidelines for best practice. It is argued that external references should always be made according to the planned dictionary functions in order to provide problem-solving dictionary solutions to specific user problems in either communication or knowledge related user situations. The proposition defended here is that the selected external references should follow the same functional orientation as the dictionary articles from which they are derived.

1. Indledning: Se kilden – ja, men hvorfor?

Ønsker man at vurdere nytteværdien af den leksikografiske praksis, som er kendt under termen ‘ordbogsekstern henvisning’ eller ‘kildehenvisning’, er der ikke megen hjælp at hente i Nordisk Leksikografisk Ordbok (NLO). Ordbogsfunktioner forbigås i tavshed. Definitionen forklarer ikke, hvordan henvisningerne skal hjælpe brugerne, men fokuserer på omdirigeringen af opslaget:

ordbogsekstern henvisning

henvisning som leder brugeren fra et sted i ordboken til et sted udenfor ordboken – Med dette menes en henvisning til en annen ordbok, en håndbok etc., f.eks. i en juridisk ordbok en henvisning til en bestemt lovsamling.

↔ ordboksintern henvisning; →henvisningsstruktur

Eksempel 1. *Artiklen ‘ordbogsekstern henvisning’ i NLO*

Et andet problem rejses af definitionen af de beslægtede ‘kildehenvisninger’, hvor der synes at være tale om ‘kildeangivelser’ frem for egentlige henvisninger. Mange af disse kildehenvisninger antager dog karakter af ordbogseksterne henvisninger, for så vidt oplysningerne om kilden klart er identificerede, eller for så vidt ordbogsforfatteren direkte opfordrer brugeren til at slå op i disse kilder. Kildeangivelserne kan derved også betragtes som potentielle ordbogseksterne henvisninger og vil blive behandlet på lige fod med disse.

kildehenvisning

angivelse av den teksten og det tekststedet som et autentisk eksempel er hentet fra

Eksempel 2. *Uddrag af artiklen ‘kildehenvisning’ i NLO*

Ovennævnte definitioner rejser en række metaleksikografiske problemer. Det første problem er at forklare, hvorfor brugeren skal ledes til et sted uden for ordbogen, og at gøre rede for, hvornår der er god mening i at gøre det. Det andet problem ligger i grænsedragningen mellem ordbogsintern og ordbogsekstern. Online ordbøger, der benytter sig af aktive links, udviser disse grænser og skaber metaleksikografisk tvivl (jf. Laursen & Duvå 2003). Vi vil fokusere på ordbogsfunktionerne. De funktioner, som de ordbogseksterne henvisninger er eksponenter for, forudsætter:

1. at der er et udvidet leksikografisk informationsbehov til stede;
2. at den leksikografiske information er relevant;
3. at ordbogsfunktionerne opfyldes på trods af de leksikografiske omkostninger, der er forbundet med omdirigeringsprocessen.

2. Funktionelle præmisser

2.1. Tre argumenter for etablering af henvisningsløsninger

Plads, tid samt funktionel fleksibilitet udgør de tre argumenter, der retfærdiggør leksikografens etablering af kildehenvisninger som leksikografisk løsning. Plads, fordi der ved henvisninger til kilder kan etableres komprimerede løsninger, der åbner tilsvarende op for spalteplass og inddragelse af et større antal datatyper andre steder i papirordbogen. Tid, fordi udarbejdelse af meget specifikke, kommunikations- og vidensorienterede løsninger er en tidskrævende opgave, der vil gå ud over den samlede tid, der anvendes til udarbejdelse af de øvrige løsninger alle andre steder i ordbogen. Funktionel fleksibilitet, fordi leksikografen kan etablere niveauopdelte løsninger, der tilfredsstiller den intenderede, heterogene brugergruppes behov. Forudsat leksikografen har tænkt dem rigtigt ind i henhold til ordbogskonceptet, antager ordbogseksterne kildehenvisninger karakter af polyvalente ordbogsløsninger.

2.2. Leksikografisk ekspansion

De ordbogseksterne henvisninger er både implicitte og eksplicitte henvisninger, der har til formål at lede brugeren ud af ordbogen og hen til eksterne kilder, som brugeren ifølge leksikografen med fordel kan vælge at slå op i, såfremt leksikografen med rette må formode, at et sådant opslag vil være medvirkende til at frembringe en bedre løsning på det leksikografiske problem, der oprindeligt lå til grund for opslaget, og i det omfang den leksikografiske løsning, som er sikret via kerneangivelserne i ordbogsartiklen, måtte forekomme at være utilstrækkelig eller mangelfuld i den fastlagte brugersituation. Henvisningen er udtryk for en komprimeret leksikografisk løsning, der først falder sammen med den samlede ordbogsfunktion i det øjeblik, der ved opslag i kilden opnås en leksikografisk ekspansion.

2.3. Todelte ordbogsfunktioner

Kildehenvisningerne er den leksikografiske signalering af et ordbogsinternt løsningsunderskud, som det i bestemte brugersituationer anbefales brugeren at søge dækket ved sikring af eksternt adgang til udvalgte, ikke-ordbogsdeterminerede datatyper, der i sammenhæng med kerneangivelser i udgangsartiklen opbygger en sammenhængende leksikografisk løsning. De ordbogseksterne henvisninger er en to-trins løsning, hvor det funktionelle afsæt sker fra selve kerneangivelsen. Den samlede funktion etableres først i det øjeblik, brugeren vælger at følge ordbogens anvisning.

2.4. Behovet for en leksikografisk cost-benefit-analyse

Brugerens søgning på – og udnyttelse af – de eksterne kilder er forbundet med højere leksikografiske informationsomkostninger, som ifølge Nielsen og Mourier 2005 omfatter både søge- og forståelsesomkostninger. Kilderne kan være svære at finde og svære at forstå, specielt på baggrund af den ordbogsartikel, der henvises fra. Anvendelsen af kildehenvisningerne som leksikografisk løsning bør derfor nøje afvejes, da det leksikografiske udbytte bør overstige de leksikografiske informationsomkostninger, der fordeler sig mellem søgeomkostninger – vanskeliggørelse af søgeprocessen pga. forskellige dataformater – og forståelsesomkostninger – højere krav til forståelsen af det udvidede leksikografiske informationsbehov og af den leksikografiske relevans.

3. Leksikografisk informationsbehov og leksikografisk relevans

Sikringen af ordbogsfunktionerne i forbindelse med kildehenvisningsløsningerne inddrager to leksikografiske parametre:

- a. **Udvidet leksikografisk informationsbehov:** et behov, der på forhånd er blevet identificeret og registreret af leksikografen ud fra en brugerprofilanalyse og ud fra en begrundet viden om, at den udvidede løsning, som kildehenvisningen er udtryk for, bedst passer til den pågældende brugersituation og til det pågældende brugerproblem i betragtning af ordbogens koncept.
- b. **Leksikografisk relevans:** Relevansen af informationen for den intendede brugergruppe sikres ved tilvejebringelse af en kildehenvisningsløsning, der funktionelt bygger på en homogen indbyrdes relation mellem de relaterede datatyper.

I det følgende skal fire ordbøger konfronteres med ovenstående principper.

3.1. Den Danske Ordbog (DDO)

DDO inddrager talrige kildehenvisninger, der har karakter af korpusbelæg, men savner klare ordbogsfunktioner. Bergenholtz & Vrang 2005 anfører, at inddragelsen

af kilderne er et udslag af den filologiske tradition. Forklaringen kunne dog også være, at ordbogen er udarbejdet på grundlag af et korpus, og at redaktionen derfor må have følt behov for at skabe respekt om ordbogens empiriske grundlag. Mere præcist anføres det i præsentationen af ordbogen på DSL's hjemmeside (www.dsl.dk/ddo_indhold.html), at "Ordbogen viser desuden typiske eksempler på opslagsordenes anvendelse, og ordenes brug og betydning illustreres ved autentiske citater fra aviser, romaner, tv-udsendelser m.m." Kigger vi på et tilfældigt udvalgt eksempel ud fra ordbogsbrugerens synsvinkel, er det svært at få øje på egentlige ordbogsfunktioner. Det leksikografiske informationsbehov mangler. Ved lemmaet 'brun' finder brugerne en række henvisninger, som er vanskeligt gennemskuelige. Her er behovet for kildeinformation tilsyneladende meget stort. Er 'brun' et ord, der har specielle litterære kvaliteter? Næppe. Henvisningerne har til formål at attestere, at man må godt sige og skrive, at man kan have 'brune øjne', eje en 'brun hest', bryste sig af 'brune lår og brune ben', gå med en 'jakke i brunt', være 'brun og lækker' og bruge 'brun som eyeliner', fordi nogle journalister og forfattere har sprogliggjort samme eksistentielle erkendelser.

brun adj.

-t, -e; -ere, -est; (...)

af en (mørk) farve som fx chokolade, kaffe og kakao • farven opstår ved blanding af gul, rød og sort □ ~ sovs, ~ konvolut, *brune pletter*, *brune øjne*, *brune krøller*, *brune nuancer*, ~ frakke, *brune sko* □ *Begge mændene var høje og mørke, med brune øjne og et smil som Cary Grant* SøndBT89, *En lille brun hest rejser sig og begynder at græsse* MeMads92 • med hud der er farvet lys brun pga. pigmentdannelse efter at have været udsat for sol el. anden ultraviolet stråling; SYN solbrændt; ANT bleg, hvid a □ *brune lår*, *brune arme*, ~ ryg, *brune ben* □ *Om sommeren cykler hun rundt i kornblå kjole og har brune ben* BerlT90, *Lotte var i bikini, brun og lækker* ErAmdr87b • (som sb.) brun farve; noget som er brunt □ *Settet har en streng klassisk jakke i brunt til en superkort, knaldrød silkenederdel* BT91, *Som eyeliner bruger jeg brun* FrbergBl91

Eksempel 3. *Artiklen 'brun' i DDO*

Der er her tale om autentiske citater, der illustrerer og attesterer sprogbrugen i modsætning til de konstruerede sætningseksempler, som mange leksikografer vægrer sig ved at bruge. Kildehenvisningerne virker her som metaleksikografiske indikatorer, der skal certificere belæg. Set ud fra ordbogsbrugerens krav til ordbogens pålidelighed kan certificering i princippet forsvares, men kun i princippet. Problemet er, at denne certificering ikke behøver at blive ekspliciteret. Brugeren har ikke praktisk mulighed for at tjekke de angivne sætningseksempler i deres rette kildeomgivelser. En informationssøgning i Søndags BT89 vil kræve uforholdsmæssig mange ressourcer i forhold til det ringe udbytte. Idet redaktionen i ordbogens forord oplyser brugerne om, at samtlige citater er hentet fra et dansk korpus, inklusive en

kort præsentation af principperne for dette korpus, kunne man have nøjedes med at præcisere, at der derved implicit var sikret belæg for autenticiteten af sprogbrugen i de bragte citater. Derved havde man sparet værdifuld angivelsesplads og undgået at belaste artikelstrukturen unødigt. Kostbar leksikografisk tid ville man også have sparet, idet inddragelse af autentiske eksempler også volder mange leksikografiske problemer, der gør det til en leksikografisk tidsrøver (jf. Lorentzen 2001).

3.2. Regnskabsordbogen på papir – dansk-engelsk – hjælp til at skaffe sig viden

Dansk-engelsk Regnskabsordbog (RO) har til formål at hjælpe den intenderede brugergruppe med de specifikke problemer, som danske og engelske regnskabstekster afføder, og har logisk nok fastlagt følgende kommunikationsorienterede funktioner: oversættelse dansk > engelsk samt produktion, korrektur og reception på begge sprog. Brugergruppen omfatter danske eksperter (revisorer), semiekspertter (oversættere, sprog- og økonomistuderende), som udgør den primære målgruppe samt lægfolk (læsere af regnskabstekster). Desuden er der i konceptet, ud over den nødvendige baggrundsviden, der i forvejen formidles via definitionerne, indbygget en række udvidede vidensorienterede funktioner efter princippet “hjælp til at skaffe sig viden” (RO:16), hvor det oplyses, at den bedste hjælp til at løse disse udvidede vidensorienterede problemer fås ved at gå til kilderne. Der henvises i nedenstående artikel under lemmaet ‘hensat forpligtelse’ til et bilag i årsregnskabsloven, til online-udgaven af ordbogen samt til et bestemt afsnit i regnskabsvejledning 17.

hensat forpligtelse

En hensat forpligtelse er en forpligtelse, der er uvis med hensyn til størrelse eller forfaldstid, og som vedrører regnskabsåret eller tidligere regnskabsår.

provision

▲ *andre hensatte forpligtelser* other provisions ▲ *årets forbrugte hensatte forpligtelser* provisions used during the year ▲ *hensatte forpligtelser til omstrukturering* provisions for restructuring costs ▲ *indregne en hensat forpligtelse* recognise a provision ▲ *måle en hensat forpligtelse* measure a provision ▲ *regulere en allerede indregnet forpligtelse* adjust a provision already recognised ▲ *tilbageføre hensatte forpligtelser* reverse a provision
KILDE årsregnskabsloven bilag 1. C nr. 7 (www.regnskabsordbogen.dk), Regnskabsvejledning 17, afsnit 5 => hensættelse

Eksempel 4. Artiklen ‘hensat forpligtelse’ i RO

Et kontrolopslag i loven på hjemmesiden ‘retsinfo’ (<http://www.retsinfo.dk>) afslører, at definitionen netop er kopieret fra lovens bilag 1. Det må formodes, at der er tale om en kildeangivelse, der alene skal gøre brugeren opmærksom på, at ordbogsdefinitionen er lig med den officielle definition. Derved løser kildehenvisningen ikke noget udvidet vidensorienteret problem, mens det i denne forbindelse er uvist, hvorfor man ikke har valgt at henvise til lovens § 47, der netop bringer

yderligere, relevante oplysninger om emnet. Den interesserede bruger kan dog også skaffe sig viden om de hensatte forpligtelser ved at læse den bestemte vejledning om emnet, som er udarbejdet af den danske forening af statsautoriserede revisorer (FSR). Har brugeren ikke vejledningen stående på sin hylde, skal brugeren først slå op i kildefortegnelsen på side 17 for at finde frem til de bibliografiske oplysninger om denne publikation. Her kunne man have valgt at henviser til danske online kommentarer til regnskabsvejledningerne, der udgives af internationale revisionsfirmaer. Ordbogen indeholder dog en række kildehenvisninger til de internationale regnskabsstandarder (IAS), der bringer relevante oplysninger til de brugere, der måtte have behov for mere viden i forbindelse med den faglige sikring af de engelske tekster, som de skal oversætte eller producere. Samlet set benytter RO sig af en række funktionsbestemte kildehenvisninger, hvor den specifikke kombination af kommunikations- og vidensorienterede funktioner kan være ganske nyttig i givne brugersituationer, mens andre henvisninger, der udelukkende fungerer som identificerende kildeangivelser, savner et klart leksikografisk løsningspotentiale. Det overordnede problem er dog, at der ikke skelnes mellem de forskellige henvisningstyper, og at brugeren derfor kan ledes hen til kilder, der hverken opfylder et reelt udvidet leksikografisk behov eller bringer relevante oplysninger til løsning af brugerproblemet.

3.3. *Juridisk ordbog fransk-dansk (JOFD)*

Denne ordbog – som i øvrigt er anmeldt i Leroyer & Rasmussen 2004 – rummer talrige kildehenvisninger til franske og danske retskilder og til EU-kilder med mere. Ordbogens primære funktioner er oversættelse L2 > L1, mens de sekundære funktioner er reception L2 > L1 af franske juridiske tekster. Den intenderede brugergruppe omfatter danske jurister og danske oversættere. Nedenstående henvisninger ved lemmaet ‘attestation’, hvor der henvises til den franske retsplejelov og til EU-konkurrenceret, er repræsentative for ordbogens henvisningspraksis, der mere har karakter af belægssikring end gennemtænkte funktioner:

attestation *f* 1. attest, erklæring; 2. (*retspl*) skriftlig vidneerklæring (jf. NCPC art. 200 ff.)
 ~ **négative** (*EU-konk.r*) negativattest (attest fra Kommissionen, hvorved denne fastslår, at der ikke findes at være anledning til at skride ind over for en aftale, samordnet praksis el.lign., jf. art. 2 i forordning nr. 17/62)
 ~ **sur l'honneur** erklæring på tro og love

Eksempel 5. *Artiklen ‘attestation’ i JOFD*

Som der ovenfor er blevet argumenteret for, bør kildehenvisningerne implementeres i henhold til ordbogskonceptet. På denne baggrund er det vanskeligt at finde frem til brugersituationer, hvori kildernes nytteværdi vil tilgodese de intenderede brugergruppers behov og løse deres problemer. For brugergruppen danske jurister er

henvisningen til den franske retskilde næppe relevant med henblik på reception af franske tekster. Hypotetisk set vil henvisningen være nyttig for de få danske jurister, der er rigtig gode til fransk, og som måtte ønske at tilegne sig mere viden. At tjekke anvendelsen af opslagsordet som foreslået af ordbogsforfatterne er ligeledes en fagsproglig vidensorienteret problemløsning, som ikke kan være relevant for danske jurister. Det er også vanskeligt at bestemme nytteværdien af henvisningen til den danske version af EU-forordningen, da der ved EU-tekster er tale om fuldkomne paralleltekster, der fjerner receptionsproblematikken. For brugergruppen danske oversættere, der måtte have problemer i oversættelsens receptionsfase, er kildehenvisningen til retsplejeloven heller ikke relevant, da receptionsproblemerne kun kommer på tale, for så vidt de berører valget af den rette ækvivalent, og det er ikke tilfældet her. Også her er funktionerne vidensorienterede, idet der henvises til paragraffer, hvor oversætteren vil kunne finde yderligere baggrundsviden i form af franske definitioner og/eller forklaringer til opslagsordet. Oversætterne har især behov for henvisninger til kilder, der bringer oversættelsesløsninger eller illustrerer oversættelsesproblemstillinger, men netop disse mangler.

Samlet set er JOFD's kildehenvisninger udtryk for vidensorienterede funktioner, som ikke er medregnet i ordbogskonceptet. Kombinationen af kommunikationsorienterede og vidensorienterede funktioner peger til gengæld i retningen af et lærerkoncept for studerende, en ordbog, der i forbindelse med undervisning i juridisk oversættelse vil kunne bruges til at tilegne sig viden om faget. Denne funktionelle kombination er dog heller ikke fastlagt i ordbogskonceptet. Man kunne dog have selekteret og opbygget kildehenvisninger, der understøtter ordbogens primære oversættelsesfunktioner. Fx kan oversættelsen til dansk af den komplekse juridiske præposition *sans préjudice de* volde mange vanskeligheder, idet kun den rette kontekst kan hjælpe til at afgøre, hvordan den rette betydning skal udtrykkes på dansk. Angivelse af kontekstuelle oversættelsesløsninger vil dog være pladskrævende, og en henvisning til et korpus af udvalgte parallelcitater vil være en langt bedre løsning på problemet. Citaterne kunne nemt integreres i den tekstbase, som ordbogen i sin cd-rom-udgave er understøttet med.

préjudice

[...]

sans ~ de (*lovtekn*) jf. dog (hvis i slutningen af sætningen), med forbehold af (hvis i starten af sætning), uden at det dog berører/for så vidt andet ikke er bestemt i • ex: *sans ~ des dispositions de l'article sus-mentionné* med forbehold af ovennævnte artikel • *se også *réserve*
→ se flere oversættelseksemples på CD

Eksempel 6. *Ændringsforslag til artiklen 'préjudice' i JOFD*

3.4. Regnskabsordbogen på internettet – Dansk-dansk og Dansk-engelsk

Her skal behandles et repræsentativt eksempel på denne online-ordbogs henvis-

ningspraksis under lemmaet ‘skat af årets egenkapitalbevægelser’. Ligesom i papirudgaven af ordbogen henvises der til regnskabsvejledningerne, dog ikke i form af et aktivt link. Kildehenvisningen er af samme art som henvisningerne i papirordbogen. Problemet er, at denne type kildehenvisning henvender sig til eksperterne, der er aktive brugere af vejledningerne i deres arbejde, men som egentlig ikke har et udvidet leksikografisk behov ved dette opslag, fordi det leksikografiske problem mangler (informationen i kilderne er relevant, men brugerne har og kender den i forvejen). Oversætterne til gengæld, der vil opleve et udvidet behov for en vidensorienteret kildehenvisningsløsning under deres arbejde, vil ikke kunne finde hjælp i den dansk-engelske online-version, da denne, i modsætning til den danske version, har fravalgt links til *www*-kilder som leksikografisk løsning.

skat af årets egenkapitalbevægelser

< ingen ubestemt artikel; skatten af årets egenkapitalbevægelser, ingen pluralis >

betydning

Skat af årets egenkapitalbevægelser er den andel af årets skat, der vedrører posteringer direkte på egenkapitalen.

eksempler

Skat af årets egenkapitalbevægelser posteres direkte på egenkapitalen.

se også

skat af årets resultat

kilde

Regnskabsvejledning 14, afsnit 7

Eksempel 7. Artiklen ‘skat af årets egenkapitalbevægelser’ i *RO online, Dansk-dansk*

Samlet set er kildehenvisningerne i *RO dansk-dansk online* udtryk for leksikografiske ekspertløsninger til udvidede, vidensorienterede problemer, der kan opstå ved behov for opdatering af viden relateret til danske regnskabstermer. For semiekspertterne antager kildehenvisningerne snarere karakter af rene kildeangivelser, der bringer metainformationer om ordbogens dataangivelser i form af specifikke dataidentifikations- og certificeringsoplysninger. Hverken på dansk eller mellem dansk og engelsk tilbyder *RO online* til gengæld hjælp i form af kildehenvisningsløsninger, der skal opfylde de kommunikationsorienterede funktioner, selvom hjælpen til tekstproblemerne ligger centralt i ordbogskonceptet. Man må derfor undre sig over, hvorfor der fra papirudgaven af regnskabsordbogen *Dansk-engelsk*, hvor teksthjælpen netop er i centrum af de fastlagte ordbogsfunktioner, henvises til *www.regnskabsordbogen.dk*, der i sin danske version netop savner disse tværproglige, kommunikationsorienterede funktioner. Problemet ligger med andre ord ikke så meget i fastlæggelsen af kildehenvisningernes funktionsorientering som i deres funktionelle relation til ordbogens koncept, først og fremmest ordbogens primære, intenderede målgruppe og ordbogens primære funktioner.

4. Fire typer løsninger

Vi er nu i stand til at kortlægge kildehenvisningsløsningernes indbyrdes, funktionelle relationer, idet vi vil skelne mellem 4 typer løsninger:

4.1. Funktionsmaksimerede 1&2 – *viden/viden & kommunikation/kommunikation*

Ved de funktionsmaksimerede kildehenvisningsløsninger (**Fmaks.**) har vi den situation, hvor ordbogsartiklens angivelsesfunktion er af samme art som kildeangivelsesfunktionen. Den samlede ordbogsfunktion sikres ved sammenlægning af de to funktioner. Kombinerer vi ordbogens kommunikation, dvs. fastlæggelsen af funktionen, og handlingsforløbet (jf. Bergenholtz 2005), dvs. det kombinerede opslag i ordbogen og opslag ved kilden, har vi ved vidensorienterede ordbogsfunktioner en sammenhængende relation, som er lineær. Omdirigeringen integreres i det funktionelle handlingsforløb. Ved sammenlægning af den primære vidensfunktion i angivelsen og den associerede vidensfunktion i kilden opbygges en sammenhængende vidensorienteret ordbogsfunktion. Samme relationer har vi ved kommunikationsorienterede kildeløsninger. Omdirigeringen integreres ligeledes i handlingsforløbet. Ved sammenlægning af den primære kommunikationsorienterede funktion i angivelsen og den associerede kommunikationsorienterede funktion i kilden opbygges ligeledes en sammenhængende kommunikationsorienteret ordbogsfunktion. I begge tilfælde er de funktionsmaksimerede kildeløsninger kendetegnet ved leksikografens kortlægning af et udvidet leksikografisk informationsbehov, der af leksikografen søges løst ved kombination af interne ordbogsdata (iD^1) og associerede, eksterne kildedata af samme type (eD^1). Det kombinerede opslag opfylder det udvidede leksikografiske informationsbehov ($+LiB$) og bringer samtidig relevante leksikografiske informationer ($+LR$), idet den samlede funktion (**F**) etableres ved sammenlægning af hovedfunktionen HF^1 og delfunktionen df^1 ved både iD^1 og eD^1 . Vi kan for de funktionsmaksimerede kildeløsninger opstille det i en formel¹ som i figur 1.

$$\mathbf{Fmaks.}^{+LiB+LR} \leftarrow ((iD^1 \Rightarrow HF^1) \cup (\sigma eD^1 \Rightarrow df^1))$$

Figur 1. Funktionsmaksimerede løsninger

¹ I formlerne er følgende logiske symboler anvendt:

() sæt af ordnede elementer

\leftarrow resultat, ny relation

^{1,2} datatype

\Rightarrow medfører

\cup fællesmængde, sammenlægning

\cap foreningsmængde, intersektion

\notin er ikke element i

σ basal operation, selekt ved opslag

4.2. Funktionsbegrænsede 1 – kommunikation/viden

De funktionsbegrænsede kildehenvisningsløsninger (**Fb.**) er kendetegnet ved kombinationen af ordbogsfunktioner, der i angivelsen og i kilden har modsatrettede orienteringer og beforder forskellige handlingsforløb. Ved kombinationen af kommunikationsorienterede angivelsesfunktioner og vidensorienterede kildefunktioner er de vidensorienterede funktioner ved kilden hægtet af. Omdirigeringen forstyrrer handlingsforløbet. Brugeren kan gå tilbage til sin tekst, men har muligvis spildt sin tid.

4.3. Funktionsbegrænsede 2 – viden/kommunikation

Samme relationer gælder for den omvendte kombination, hvor de kommunikationsorienterede funktioner ved kilden er hægtet af; omdirigeringen forstyrrer handlingsforløbet, da selve ordbogsopslaget er tekstafhængigt.

De funktionsbegrænsede henvisninger er mere afhængige af hypotetiske brugersituationer i forhold til ordbogens koncept og har et mere hypotetisk leksikografisk potentiale. Opslag ved kilden kan være begrundet i et udvidet leksikografisk informationsbehov, mens den i kilden fundne information leksikografisk set er irrelevant. Tilsvarende kan opslag ved kilden sikre en leksikografisk relevansopfattelse, som ikke er betinget af et reelt udvidet leksikografisk informationsbehov i brugersituationen. Informationen er med andre ord relevant, men brugeren har ikke behov for at få formidlet denne information, da brugeren har den i forvejen. De funktionsbegrænsede løsninger bidrager ikke til at etablere en sammenhængende leksikografisk løsning. På denne baggrund kan der opstilles to formler – **Fb.1** og **Fb.2** – hvor der blot er byttet om på fortegnene for **I** og **R** (henholdsvis **-LiB + LR** og **+ LiB - LR**) og på datatyperne 1 og 2 (figur 2).

$$\mathbf{Fb. 1}^{-LiB + LR} \leftarrow ((iD^1 \Rightarrow HF^1) \cap (\sigma eD^2 \Rightarrow df^2))$$

$$\mathbf{Fb. 2}^{+LiB - LR} \leftarrow ((iD^2 \Rightarrow HF^2) \cap (\sigma eD^1 \Rightarrow df^1))$$

Figur 2. Funktionsbegrænsede løsninger 1 & 2

4.4. Funktionsminimerede

De funktionsminimerede kildehenvisningsløsninger (**Fmin.**) er både ved kommunikations- og vidensorienterede funktioner afskåret fra den leksikografiske kommunikationssituation og kendetegnes ved fraværet af funktionsfastlæggelsen ved selektionen af kilden. Der er ingen forbindelse mellem ordbog og kilde, idet opslag ved kilden kun meget vanskeligt kan effektueres og heller ikke er begrundet i et udvidet leksikografisk informationsbehov. Kildens indhold er ligeledes leksikografisk irrelevant. Løsningen bidrager ikke til at etablere en effektiv leksikografisk løsning (figur 3).

$$F_{\min.} \text{ } (^{-}LiB - LR) \leftarrow (\sigma \text{ } iD^1 \Rightarrow HF^1) \notin (eD^2)$$

Figur 3. Funktionsminimerede løsninger

5. De fire kildeløsninger og deres leksikografiske potentiale

5.1. Forudsætningerne for funktionsmaksimering

Forudsætningerne for funktionsmaksimering af kildehenvisningsløsningerne er:

- at brugen af kilden understøtter samme funktionsorientering, som er fastlagt i udgangsartiklen – kommunikation eller viden
- at den nye relation er i overensstemmelse med ordbogens koncept, først og fremmest ordbogens primære ordbogsfunktioner
- at de leksikografiske omkostninger står i rimeligt forhold til det faktiske leksikografiske udbytte

5.2. Leksikografisk potentialekort

Bruger vi Sepstrups (2002) model og anbringer de forskellige typer kildeløsninger i relation til det udvidede leksikografiske informationsbehov og til relevansen af den leksikografiske information, kan vi mappe et informationspotentialekort. De fire typer kildeløsninger anbringes i de respektive potentialefelter. De funktionsmaksimerede henvisninger findes ved sammenlægning af **+I+R**-relationer, de funktionsminimerede ved **-I-R**, og de funktionsbegrænsede ved intersektion af **+R-I**- og **-I+R**-relationer, som vist i figur 4.

Leksikografisk potentialekort	+ leksikografisk relevans	÷ leksikografisk relevans
+ udvidet leksikografisk informationsbehov	<i>Højt leksikografisk potentiale</i> Det leksikografiske udbytte af kildeløsningen er højt → funktionsmaksimerede 1 & 2	<i>Begrænset leksikografisk potentiale</i> Det leksikografiske udbytte af kildeløsningen er begrænset → funktionsbegrænsede 1
- udvidet leksikografisk informationsbehov	<i>Begrænset leksikografisk potentiale</i> Det leksikografiske udbytte af kildeløsningen er begrænset → funktionsbegrænsede 2	<i>Lille leksikografisk potentiale</i> Det leksikografiske udbytte af kildeløsningen er lavt → funktionsminimerede

Figur 4. Leksikografisk potentialekort for kildehenvisningsløsningerne

6. Retningslinjer for kildehenvisningsløsninger

Vi er nu i stand til at opstille retningslinjer for etablering af kildehenvisningsløsninger.

6.1. Udarbejdelse af brugerinstruktioner

- Der gives ud fra en beskrivelse af de intenderede brugergrupper og brugersituationer typiske eksempler på det leksikografiske løsningspotentiale, der ligger i brugen af kildehenvisningsløsninger.
- Der gøres opmærksom på, at opnåelsen af kildehenvisningsløsningernes fulde leksikografiske løsningspotentiale er betinget af tilstedeværelsen af de anførte problemer og behov i de pågældende brugersituationer.

6.2. Udarbejdelse af metaleksikografiske redegørelser

- Der gøres rede for de metaleksikografiske overvejelser, der har ført til selektionen af kildehenvisningsløsninger set i relation til de udvidede informationsbehov, der udspringer af de berørte ordbogsartikler, og til ordbogskonceptet.
- Der gøres rede for de brugerdifferentieringskriterier, som henvisningsløsningerne er udtryk for, herunder hvilke brugere, hvilke situationer og hvilke typer problemer kildehenvisningsløsningerne bedst vil kunne imødekomme.
- Der gøres rede for de funktionelle kriterier, der har ligget til grund for kildeudvælgelsen, herunder en kort kildekritisk kommentar til de enkelte kilders leksikografiske potentiale.

6.3. Etablering af kildehenvisningsløsninger i papirordbøger

- Der foretages leksikografiske cost-benefit-analyser, herunder en samlet vurdering af det udvidede informationsbehov og af den leksikografiske relevans.
- Der etableres fortrinsvis funktionsmaksimerede kildehenvisningsløsninger.
- Der etableres kildehenvisningsløsninger, der klart identificerer kilden, og som er fuldt integreret i ordbogen i form af en udtømmende kildefortegnelse samt specifikke instruktioner i brug af kildeløsningerne i ordbogens brugervejledning.

6.4. Etablering af kildehenvisningsløsninger i internetordbøger

- Der anvendes samme regler som for papirordbøger.
- Der etableres helst dybe links til de relevante kilder.
- Der skelnes grafisk mellem links af typen 'læs mere' og af typen 'se flere eksempler'.

7. Konklusion: Se kilden, ja, men kun fordi det er en god ordbogsløsning

Ovenstående gennemgang af ordbogseksterne henvisninger som funktionelle leksikografiske løsninger har på ny bekræftet, at ordbogsfunktioner og deres succeskriterier (dvs. fastlæggelse og opfyldelse af gennemtænkte, optimale funktioner) forbliver alfa og omega i de mange metodologiske valg, leksikografen bør træffe. Skal ordbogen, herunder den spirende generation af online-ordbøger, ikke ende i henvisningskaos på grund af de fristende links, skal ordbogsfunktioner ses og respekteres som et menneskeligt, leksikografisk interface, der er indlejret midt i den

leksikografiske kommunikation, og som er styret af en intelligent og forpligtende dialog mellem leksikograf og bruger. Implementeringen af funktionelle kildehenvisningsløsninger frem for simple kildebelæg er heller ingen undtagelse fra denne grundlov. Se kilden, ja – men kun fordi det er en god ordbogsløsning (fordi den set i relation til ordbogens koncept og til den aktuelle brugersituation er i stand til at bringe en sammenhængende funktionel ordbogsløsning med højt leksikografisk potentiale) – er det simple svar til det simple spørgsmål, vi indledte med.

Litteratur

Ordbøger:

- DDO = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. København: Gyldendal 2003-2005.
- JOFD = Christensen, Dorthe & Fich, Thomas: *Juridisk ordbog fransk-dansk*. København: Gads Forlag 2004.
- NLO = Bergenholtz, Henning/Cantell, Ilse/Fjeld, Ruth Vatvedt/Gundersen, Dag/Jónsson, Jón Hilmar/Svensén, Bo: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget 1997.
- RO = Bergenholtz, Henning/Mourier, Lise/Nielsen, Sandro: *Regnskabsordbogen Dansk-Engelsk*. København: Forlaget Thomson 2004-2005.
- RO online = Bergenholtz, Henning/Mourier, Lise/Nielsen, Sandro: *Den Danske Regnskabsordbog, Dansk-Dansk* <http://netdob.asb.dk/iasdk/index.html> og *Dansk-Engelsk* <http://netdob.asb.dk/iasdkgb/index.html>. 2004-2005.

Anden litteratur:

- Bergenholtz, Henning 2005: Den usynlige produktions- og korrekturordbog. I: *LexicoNordica* 12, 19-38.
- Bergenholtz, Henning & Vrang, Vibeke 2005: Den Danske Ordbog bind 2 (E-H) og 3 (I-L) – en ordbog for folket eller for akademikere? I: *LexicoNordica* 12, 169-187.
- Laursen, Anne Lise & Duvå, Grete 2003: IT-baseret fagleksikografi: nye perspektiver. I: Zakaris Svabo Hansen & Anfinnur Johansen (red.): *Rapport fra Konference om leksikografi i Norden*. Tórshavn: NFL og Føroyamálsdeild Fróðskaparseturs Føroya, 191-202.
- Leroyer, Patrick & Rasmussen, Kirsten Wølch 2004: Anmeldelse af Christensen, Dorthe & Fich, Thomas: *Juridisk ordbog fransk-dansk*. I: *Sprint*, 2004, 2. København: CBS – Fakultetet for Sprog, Kommunikation og Kultur, 93-103.
- Lorentzen, Henrik 2001: Jagten på det gode citat. Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus. I: Martin Gellerstam et al. (eds.): *Nordiske studier i leksikografi 5. Rapport från Konferens om Lexikografi i Norden*, Göteborg 26.-29. maj 1999. Göteborg: Nordisk forening for leksikografi, 202-216.
- Nielsen, Sandro & Mourier, Lise 2005: Internet Accounting Dictionaries: Present Solutions and Future Opportunities. I: *Hermes, Journal of Linguistics* 34. Århus: Handelshøjskolen i Århus, 83-116.
- Sepstrup, Preben 2002: *Tilretteleggelse af information*. Århus: Systime.

ordnet.dk – et nyt sprogligt opslagsværk på internettet

ordnet.dk – a new lookup tool on the Internet. We give an account of the architecture and functionalities of a web-based application that gives joint access to three existing resources and seeks to combine them in a new way. The project includes digitisation of the historical 28-volume dictionary *Ordbog over det danske Sprog* (Dictionary of the Danish Language) as well as the integration with its five supplementary volumes. The modern *Den Danske Ordbog* (The Danish Dictionary) will continuously be supplied with new material and adapted to a concept for a digital dictionary which will among other things introduce new features such as onomasiological queries and larger integration with the *Korpus 2000* component. The corpus module is supplied with additional material, and a revision of both the interface and functionalities of the module is foreseen as part of the project which runs until the end of 2009.

1. Introduktion

I denne artikel vil vi redegøre for et opslagsværk som er under udarbejdelse i Det Danske Sprog- og Litteraturselskab (DSL). Det drejer sig om et elektronisk opslagsværk som skal koble tre eksisterende værker sammen på hjemmesiden www.ordnet.dk, nemlig *Ordbog over det danske Sprog* (ODS), *Den Danske Ordbog* (DDO) og *Korpus 2000*. Projektet løber i årene 2004–2009 og finansieres af Carlsbergfondet og Kulturministeriet med støtte fra Det Elektroniske Forskningsbibliotek.

Projektet indebærer at ODS først digitaliseres og derefter sammenflettes med *Supplementet til ODS*, hvis femte og sidste bind udkom i efteråret 2005.

DDO foreligger allerede som digitalt manuskript, og det er derfor her mulighederne for at eksperimentere med det elektroniske medie kommer til at foregå. En af fordelene ved netpublicering er muligheden for at koble flere værker sammen til ét og springe imellem dem. Den mulighed vil vi udnytte ved at knytte ordbogsdelen og korpusteksterne sammen, både sådan at korpusmaterialet udnyttes mere effektivt i den beskrivelse man får i ordbogsartiklerne, og sådan at brugeren får mulighed for at lave sine egne undersøgelser med udgangspunkt i det ord han/hun lige har slået op. Derudover skal både ordbog og korpus forsynes med flere artikler og tekster så basen holder sig opdateret med den sproglige udvikling.

2. Den Danske Ordbog på nettet – en prototype

Som et første skridt i retning af at lægge DDO på internettet har vi udarbejdet en intern prototype til en elektronisk version af ordbogens data. Og det er med overlæg at vi taler om ordbogens data, og ikke bare ordbogen, for det er afgørende for os at versionen på nettet ikke blot er en tro kopi af den trykte bog vist på en computerskærm. De første generationer af elektroniske ordbøger bestod netop blot af de

trykte ordbogsspalter forsynet med visse søgefunktioner, oftest begrænset til søgning på opslagsord og fritekstsøgning i den fulde tekst. Vi vil gerne gå et par skridt videre og ud over de nævnte tilbyde mere avancerede søgninger, fx søgninger der ikke tager udgangspunkt i et bestemt opslagsord eller udtryk (semasiologisk søgning), men i et bestemt betydningsindhold (onomasiologisk eller begrebsorienteret søgning). Det er vigtigt at understrege at der er tale om en prototype der lægger vægt på mulighederne og ikke på præsentationen, så derfor er det ikke den endelige grænseflade der ses på de følgende figurer.

Som ved de fleste opslagsværker og andre tjenester på nettet, er der et indtastningsfelt hvor man indtaster det ord man vil slå op, fx substantivet *dans*. Resultatet af denne søgning er en liste med ord der begynder med opslagsordet, altså en højre-trunkeret søgning, der også inkluderer opslagsord som fx *dansemyg* og *dansk*. Det er også muligt at lave en venstretrunkeret søgning, nemlig på strengen **dans*, hvilket ud over rigtige andetledssammensætninger som fx *krigsdans* og *kædedans* også giver *impedans* og det for korpusleksikografer nok så vigtige ord *konkordans*. Der er i dette tilfælde altså tale om rent tegnorienterede søgninger.

Hvis man klikker på ordet *dans* i resultatlisten, får man et skærmbillede der minder om en bog, men alligevel er lidt anderledes (se figur 1). Noget af det nye er at man i begyndelsen af artiklen får vist et resumé over ordets forskellige betydninger, for *dans'* vedkommende 3 betydninger. Til højre for dette ses en liste over de faste udtryk som *dans* indgår i. På den måde får man hurtigere end i en trykt bog et overblik og kan relativt let gå til det ønskede sted med et enkelt klik. Fordelene er naturligvis endnu større ved opslagsord med mange betydninger og mange faste udtryk.

Figur 1. Artiklen *dans* i ordnet-prototypen

Definitionerne i Den Danske Ordbog følger i stor udstrækning den klassiske aristoteliske definitions måde med *genus proximum* og et eller flere *differentiae specificae*. Betydningsresuméerne er automatisk genereret ud fra dette *genus proximum*, som blev udvalgt af redaktørerne i redigeringsprocessen og skrevet i et særligt, usynligt element der oprindeligt var tænkt som et eksperiment, men nu viser sig at være yderst anvendeligt, ikke bare til betydningsresuméer. Disse resuméer bygger enten på *genus proximum* alene eller *genus proximum* i kombination med et par af de omgivende ord. Hvis dette system skal fungere hensigtsmæssigt i alle tilfælde, kræver det en manuel redigering, men som et foreløbigt bud på hvordan man hurtigt kan præsentere betydninger, synes vi det er ganske godt. I de store engelsksprogede *learner's dictionaries* som fx Macmillan og Longman har man allerede denne resuméfacilitet i de elektroniske udgaver, især for lange artiklers vedkommende, men det har tydeligvis også krævet menneskelig redigering da der ikke altid er én til én-overensstemmelse mellem tekststrengene i definitionerne og i resuméerne.

De faste udtryk fremgår som nævnt af en liste der som i den trykte bog er alfabetiseret efter første ord i udtrykket. Alle udtrykkene er klikbare, dvs. at man fx kan klikke på udtrykket *latinamerikanske danse* og blive ledt derhen i strukturen. Når man har læst forklaring m.m., har man mulighed for at klikke sig tilbage til resuméet i artiklens begyndelse.

I den trykte udgave af DDO er der ofte anvendt traditionelle strategier for at spare plads, fx brug af bindestreger eller tilde til at erstatte opslagsordet. I det digitale medium er pladshensynet ikke længere afgørende, så derfor har vi allerede i prototypen valgt at lade alle sammensætninger være skrevet helt ud, sammenlign sammensætningerne i tabel 1.

Trykt DDO	Prototype
danseform, -kunst, -undervisning; folke-, kind-, kæde-, mave-, par-, regn-, sportsdans	danseform, dansekunst, danseundervisning; folkedans, kinddans, kadedans, mavedans, pardans, regndans, sportsdans

Tabel 1. Udsnit af artiklen *dans* (sammensætninger)

Vi har ligeledes droppet tilden som erstatning for opslagsordet i ordforbindelser (kollokationer), sammenlign eksemplerne i tabel 2.

Trykt DDO	Prototype
moderne ~, klassisk ~; sang og ~, musik og ~, spille (op) til ~	moderne dans, klassisk dans; sang og dans, musik og dans, spille (op) til dans

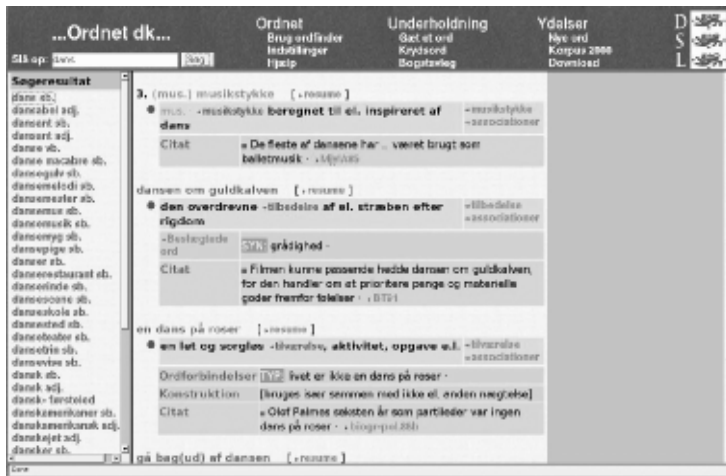
Tabel 2. Udsnit af artiklen *dans* (kollokationer)

Forøgelsen i pladsforbrug er beskeden, mens læsbarheden til gengæld vinder betragteligt ved afskaffelsen af bindestreg og tilde. I den endelige netversion vil man også kunne skrive sigler som SYN, ANT, JF helt ud og tilføje nye elementbetegnelser såsom *ordforbindelse*, *konstruktion* og *citater*.

Forkortelsesstrategien har også spillet en rolle ved udformningen af kildeangivelserne til citaterne (ca. 100.000 i alt), idet der i den trykte ordbog ved alle citater er anført en kilde, men i forkortet og komprimeret form. Idéen har været at forkortelsen skulle være så gennemskuelig som muligt (med mulighed for at slå forkortelsen op i ordbogens bind 6), men i prototypen er kildeangivelserne i stedet gjort klikbare så man ved et klik med musen får vist et lille vindue med mere detaljerede bibliografiske oplysninger. Det vil være en stor hjælp for mange (danske) brugere der måske nok kan afkode *BerlT90* (avisen Berlingske Tidende 1990), men vil have svært ved at regne ud hvad der gemmer sig bag fx *LDJørg86* eller *biogr.-pol.86b*.¹

De faciliteter og de ændringer i forhold til den trykte bog som vi har nævnt indtil nu, anser vi for forholdsvis almindelige og forventelige. I det følgende vil vi omtale nogle mere specielle funktioner som vedrører den tidligere nævnte genus proximum-opmærkning og sammenkoblingen med korpustekster.

Vi vil igen se på artiklen **dans**, nærmere betegnet betydning 3, der er defineret som ‘musikstykke beregnet til el. inspireret af dans’, se figur 2.



Figur 2. Udsnit af artiklen *dans* (betydning 3)

Der er her knyttet to søgemuligheder til ordet *musikstykke*, der altså er genus proximum i denne definition. Klikker man på *musikstykke* i selve definitionen, bliver dette ord slået op i ordbogen. Hvis man klikker på *musikstykke* til højre for

¹ Henholdsvis Lone Diana Jørgensen: *Svanedans*, Gyldendal 1986 og Jørgen Poulsen: *Olof Palme*, Samleren 1986

definitionen, får man i venstre side vist en liste over alle ord der i ordbogen er defineret som ‘musikstykke’, dvs. har dette ord som genus proximum. Denne liste er genereret automatisk og sorteret efter hvilke ord der typisk optræder i definitioner hvor *musikstykke* er genus proximum. Som kriterium for typiskhed er brugt den velkendte mutual information-statistik, som først blev introduceret af Church og Hanks (1989), og som groft sagt lister ord efter hvor stor den indbyrdes tiltrækningskraft er. I dette tilfælde giver det os en liste hvor ord der betegner musikstykker, er ordnet efter hvilke karakteristiske træk (*differentiae specificae*) der indgår i deres definition. På denne måde får vi musikstykker der hører til danse, muntre musikstykker, musikstykker inddelt i satser, lyriske musikstykker, musikstykker for klaver osv. (se tabel 3).

Ord der betyder ‘musikstykke’

Træk: **dansen**
 cha-cha-cha sb.
 polonæse sb.
 sarabande sb.
 samba sb.
 rumba sb.
 tarantel sb. bet. 2
 bolero sb.
 menuet sb.
 mazurka sb.
 mambo sb.
 hopsa sb.
 gigue sb.
 gavotte sb.
 tango sb.
 fandango sb.
 polka sb.
 czardas sb.

Træk: **muntert**
 capriccio sb.
 scherzo sb.
 humoreske sb.
 polka sb.
 bagatel sb.
 rondo sb.

Træk: **satser**
 divertimento sb.
 symfoni sb.
 suite sb. bet. 2
 serenade sb.
 sonate sb.
 klavertrio sb.

Træk: **sangstemmer**
 septet sb. bet. 2
 kvintet sb. bet. 2
 sekstet sb. bet. 2
 kvartet sb. bet. 2

Træk: **lyrisk**
 impromptu sb.
 romance sb. bet. 2
 nocturne sb.
 ballade sb. bet. 3

Træk: **rytme**
 cha-cha-cha sb.
 tarantel sb. bet. 2
 mazurka sb.
 menuet sb.
 tango sb.
 polonæse sb.
 sarabande sb.
 samba sb.
 rumba sb.
 mambo sb.
 vals sb.
 bolero sb.
 gavotte sb.
 gigue sb.
 hopsa sb.
 polka sb.
 fandango sb.
 czardas sb.
 sørgemarch sb.
 march sb. bet. 4

Træk: **klaver**
 klaverstykke sb.
 impromptu sb.
 prælude sb.
 novелlette sb. bet. 2
 klavertrio sb.
 ballade sb. bet. 3
 toccata sb.

Træk: **instrumenter**
 trio sb. bet. 2
 oktet sb. bet. 2
 septet sb. bet. 2
 kvintet sb. bet. 2
 sekstet sb. bet. 2
 kvartet sb. bet. 2
 koncert sb. bet. 2

Træk: **orkester**
 ouverture sb.
 kammerkoncert sb. bet. 2
 dødsmesse sb.
 symfoni sb.
 ouverture sb. bet. 2
 passion sb. bet. 2
 kantate sb.
 solokonzert sb. bet. 2

Træk: **takt**
 galop sb. bet. 2
 siciliano sb.
 sørgemarch sb.
 march sb. bet. 4
 chaconne sb.
 passacaglia sb.

Tabel 3. Uddrag af listen over ord med musikstykke som genus proximum

Det er klart at der her er tale om et eksperiment: Listen er genereret og sorteret automatisk og ikke gennemgået af et menneske; fx er det ikke indlysende at korværker som dødsmesse, passion og kantate kun er placeret under *orkester* og ikke også under *kor*, som slet ikke optræder som signifikant træk. Vi synes dog alligevel den rummer perspektiver, og ved manuel tilretning vil resultatet blive langt bedre (jf. Asmussen 2004, Trap-Jensen 2004). Et næste skridt i arbejdet vil netop være at indbygge mulighed for begrebsorienterede søgninger, hvor udgangspunktet er begrebet og ikke ordet. Hvis man eksempelvis er interesseret i at finde ord for mørke øltyper, vil man kunne indtaste et genus proximum (*øl*) og et eller flere specifikke træk (*mørkt*), og man vil hurtigt få foreslået kandidater som *stout*, *lagerøl*, *porter*, *maltøl* og *nisseøl*.

En yderligere facilitet i prototypen er koblingen med korpus. Til højre for ordbogsartiklen findes nemlig en mulighed for at se mere om ordet i korpus. Det giver foreløbig en konkordans fra Korpus 2000 (<http://korpus.dsl.dk>): et korpus med i øjeblikket ca. 56 millioner løbende ord fordelt på to delkorporer fra omkring 1990 og omkring år 2000. I prototypen er de to resurser adskilt fra hinanden, men vi ønsker at skabe en større integration mellem korpus og ordbog, en friere navigering mellem de to resurser. Fx kan en konkordans, fremkaldt af et ordbogsopslag, udløse ønsket om at slå op i ordbogen et nyt sted. I de definitioner eller eksempler man finder der, ser man måske ord eller udtryk der kan give anledning til en ny korpus-søgning osv. Korpus 2000's interface tilbyder i dag visse grundlæggende korpus-funktioner som fx sortering på nøgleordet selv eller på dets venstre eller højre kontekst og lister over ord der er hyppige naboer til opslagsordet. I et nyt korpusinterface, som er under udvikling i DSL, regner vi med at tilbyde mere avancerede funktioner som fx lister der på grundlag af en syntaktisk opmærkning af korpus giver overblik over et ords typiske syntaktiske og leksikalske medspillere² og mulighed for søgning i brugerdefinerede delkorporer som fx avistekster før 1995 eller talesprog fra radio og tv.

3. Omfang, lemmaudvælgelse og redigeringsprincipper

Overgangen fra papir til skærm ændrer på flere punkter betingelserne for redigering, og det vil betyde at de elektroniske artikler efterhånden vil ændre sig i forhold til bogværkets artikler. Præsentationen af oplysninger vil ændre sig: Ud over selve designet, som i sagens natur vil tage sig anderledes ud i netversionen, har vi allerede set at pladsøkonomi ikke er så vigtig ved netpublicering, og mange forkortelser, brug af tilde og bindestreger og andre pladsbesparende foranstaltninger vil derfor blive gjort overflødige. Vi har også tidligere set at betydningsresuméer i begyndelsen af

² Det der af Kilgariff et al. (2004) kaldes 'word sketches'.

artiklerne er nyttige til at skaffe overblik over en elektronisk artikel. I netversionen vil man også få mulighed for med teleskopløsninger enten at skjule visse oplysningstyper for at få større overblik eller udfolde oplysningerne, alt efter éns behov. Og endelig forestiller vi os at nogle oplysningstyper skal præsenteres på en anden måde end i bogværket. Det gælder især bøjningsoplysninger og syntaktiske oplysninger, som også af sprogteknologiske grunde bør gives en anden struktur i den elektroniske udgave (se Trap-Jensen, under udgivelse). Det samme gælder til en vis grad de etymologiske oplysninger.

På indholdssiden vil artiklerne også ændre sig, dels ved at det bliver muligt at bringe flere autentiske eksempler end det stramt redigerede udvalg som man finder i bogværket, dels ved at nye funktioner bliver tilbudt som kun er mulige i det elektroniske medie. Koblingen til korpus har været nævnt som noget vi prioriterer højt. Korpus skal udnyttes ved at brugeren ud over den generelle mulighed for at slå det tilsvarende ord op i korpus også skal have adgang til nogle specifikke oplysninger baseret på korpusundersøgelser der knytter sig til den artikel brugeren befinder sig i. Fx kan det være interessant at se fordelingen i korpus af dobbeltformer eller alternative bøjningsformer, eller at se autentiske eksempler på de syntaktiske mønstre der gives i artiklerne. Begge dele kan lade sig gøre fordi korpusteksterne er blevet opmærket både morfologisk og syntaktisk.

Det andet område vi prioriterer højt, er muligheden for at lave begrebsorienterede søgninger. I gennemgangen af prototypen viste vi hvordan ordbogsmanuskriptet i forvejen indeholder mange af de forbindelser mellem ordene som man er interesseret i at få oplyst når man søger efter begrebet snarere end betegnelsen, en såkaldt onomasiologisk søgning. Ud over over-/underbegrebsrelationen bestemt ved genus proximum kan man finde oplysning om synonymi, antonymi, del-helhed og flere andre relationer. I prototypen kan man på baggrund af genus proximum-opmærkningen finde mange interessante begrebsområder, men det er et problem at systemet af begrebsområder ikke har været behandlet ud fra en konsistent, ontologisk synsvinkel. Ordbogsartiklerne er nemlig først og fremmest skrevet til mennesker og ikke til en computer. En artikel kan derfor fungere fint som oplysning til en menneskelig bruger, men hvis den forsynder sig mod det ontologiske hierarki, vil computeren ikke registrere ordet som en del af det pågældende begrebsområde. I DDO er en *hare* fx defineret som 'et gråbrunt pattedyr med hvid bug, to store fortænder i over- og undermunden, lange ører og kraftige bagben', hvilket fungerer fint for en menneskelig bruger, men udelukker ordet fra en søgning på begrebet gnaver. Og når en *benvarmer* er defineret som 'en slags ulden strømpe uden fod til at trække på benet', er det så underbegreb til strømpe eller til beklædningsgenstand? Den slags problemer kommer man hurtigt ud i hvis de begrebsorienterede søgninger foretages direkte på artiklernes genus proximum-opmærkning. Vi har derfor indset at der kræves temmelig meget redigering af ordstoffet, og har derfor indledt et samarbejde

med Center for Sprogteknologi ved Københavns Universitet om at lave et decideret leksikalsk-semantic ordnet for dansk på baggrund af artiklerne i DDO, efter modellen fra de internationale *wordnets*. Det danske projekt kalder vi DanNet (se www.wordnet.dk), og tanken er at det skal kunne integreres i *ordnet.dk* og få de begrebsmæssige søgninger til at fungere hensigtsmæssigt – foruden at det naturligvis vil få en række selvstændige, sprogteknologiske anvendelser ligesom søsterprojekterne for en række andre sprog.

Hvor mange ordbogsartikler man vil kunne slå op i *ordnet.dk*, ved vi ikke helt sikkert endnu. Under indtryk af den kritik der har været rettet mod den trykte DDO,³ vil vi dog godt oplyse at det er planen at både denne ordbogs ca. 60.000 hovedopslagsord og de ca. 40.000 underopslagsord (primært sammensætninger og afledninger) skal kunne slås op som selvstændige artikler. Derudover vil der som sagt blive nyskrevet et antal helt nye artikler, især om nye ord der kommer til i sproget, men ideelt set vil det også være nyttigt at supplere ordstoffet med andre typer artikler. Det kan være ord som blev sorteret fra under redigeringen af DDO (faglige eller forholdsvis sjældne ord), eller ord fra tiden mellem ODS' afslutning og DDO's primære periode (dvs. fra ca. 1950 til ca. 1980) som ikke er repræsenteret i DDO's korpus. Der er også en vis usikkerhed om lemmabestanden i ODS. Den første netudgave indeholder 180.000 opslagsord, men hertil skal muligvis lægges en række sammensætninger der gemmer sig nede i artiklerne – sammensætninger som vi overvejer at gøre søgbare – og i hvert fald de nye artikler der er indeholdt i supplementsbindene. Og endelig ved vi endnu ikke hvor stor fællesmængden mellem de to ordbøger er.

For Korpus 2000's vedkommende er planen som nævnt ovenfor at udarbejde en ny brugergrænseflade hvor flere nye og forbedrede faciliteter stilles til rådighed, herunder ikke mindst udnyttelse af teksternes syntaktiske opmærkning. Det er også tanken at man skal kunne søge mere fleksibelt i korpus, fx ved at kunne vælge at søge i samtlige tekster eller kun relevante udsnit. Endelig vil mængden af tekster løbende blive forøget så også tekstbasen holdes opdateret. Det er det der på længere sigt vil gøre det muligt at foretage diakrone undersøgelser af sprogets udvikling, og redaktionsinternt vil det være et nyttigt redskab til at generere automatiske lister over nye lemmakandidater.

4. Ordbog over det danske Sprog på internettet

Som nævnt i indledningen drejer en væsentlig del af ordnet-projektet sig om at gøre Ordbog over det danske Sprog (ODS) tilgængelig i et elektronisk format, ligesom andre tilsvarende store nationalordbøger er blevet digitale, fx Svenska Akademiens ordbok (SAOB) i Sverige, Oxford English Dictionary (OED) i den engelsktalende

³ Jf. bl.a. Bergenholtz og Vrang 2004.

verden og Grimms Deutsches Wörterbuch i Tyskland. Norsk Ordbok og Norsk Riksmålordbok er også trådt ind i den digitale tidsalder, jf. Guttu 2005 og Grønvik og Tvedt 2006.

I den indledende fase skulle det besluttes om bogværket skulle scannes eller indtastes. Begge modeller er blevet brugt af andre store ordbogsprojekter, fx scanning af SAOB og indtastning af OED, men vi besluttede at lægge os op ad den model der blev brugt i forbindelse med Grimms Deutsches Wörterbuch. Den går ud på at manuskriptet indtastes i to uafhængige versioner, der siden sammenlignes elektronisk. På den måde kommer man så godt som samtlige tastefejl til livs uden en lang og sej korrekturlæsning, der ellers ville være nødvendig ved scanning.

Denne del af arbejdet foregår i samarbejde med en afdeling ved universitetet i Trier ved navn Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften (se <http://germazope.uni-trier.de/Projects/KoZe2>). Først er bogen tastet ind fysisk, hvilket er foregået i et firma i Kina der har specialiseret sig i denne type opgaver og også stod for indtastningen af Grimm. Efter indtastningen er de to parallelle versioner så sammenlignet automatisk i Trier, idet uoverensstemmelserne er blevet bearbejdet manuelt, dvs. at en person med kompetence i germanske sprog har slået op i ODS for at se hvad der skulle stå.

Denne fase af arbejdet er nu afsluttet. Vi har modtaget filer hvor alle 28 bind er indtastet og den automatiske sammenligning er foretaget. Teksten til ODS foreligger altså nu i et rent tekstformat spækket med typografiske koder, og i figur 3 kan man få et indtryk af hvordan det ser ud.

```
$0040.12 <P>__<A+1>#F+herske-syg,#F-</A+1> #/+adj. meget, alt for be-
$0040.13 g#.^orlig efter at herske (jf.#/- -k#.^ar, -lysten#/+).
$0040.14 vAph.(1759).#/- Emilie var herskesyg og #.on-
$0040.15 skede en mere glimrende Stilling i Sta-
$0040.16 rostens Huus.#/+Hauch.I.265.#/- Det er en Ytring
$0040.17 af Tyranni og herskesygt Sindelag. #/+PM#.oll.
$0040.18 II.388.#/- det at hj#.^alpe #/+(er)#/- ikke . . at v#.^are
$0040.19 den Herskesygeste men den Taalmodigste.
$0040.20 #/+Kierk.XIII.533.#/- <A+1>#F+-syge,#F-</A+1> en. #/+det at v#.^ore
$0040.21 herskesyg; magtbrynde (jf.#/- -lyst#/+). Moth.H
$0040.22 173.#/- Alexander Magnus . . havde og store
$0040.23 Fejl. Hans Herske-Syge alleene . . kand
$0040.24 tiene til Beviis derpaa. #/+Holb.Ep.III.168.
$0040.25 Bagges.L.I.279. Brandes.VIII.151.#/-</P>
```

Figur 3. Artiklerne *herskesyg* og *herskesyge* i indtastet format

Det drejer sig om artiklerne *herskesyg* og *herskesyge*, og som det fremgår, kan det godt være vanskeligt at læse for mennesker. Til gengæld er det typografiske billede

repræsenteret fuldt ud: Hver linje i bogen indledes med en kode der oplyser om spaltenummer og linjenummer; alle typografiske markeringer såsom skriftstørrelse, fed, kursiv, spatiering er gengivet, ligesom alle specialtegn, dansk æ, ø og å samt de små symboler der brugtes dengang i form af ankre, noder, fugle, tandhjul osv., har hver sin specifikke kode.

Der er endnu ikke foretaget nogen systematisk kontrol af indtastningen, men alting tyder på at der er meget få indtastningsfejl, så det er vi meget tilfredse med. Der er dog ting i den trykte bog som har været vanskelige at afkode både for de kinesiske indtastere og ved den efterfølgende sammenligning i Trier. Disse tvivlstilfælde er markeret i filerne, og der vil være henved 2.000 steder hvor vi skal ind og vurdere og rette individuelt.

Der er bl.a. tale om steder hvor vi som indfødte danskere umiddelbart kan se hvad der skal stå, mens det ikke har været muligt for en ikkedansker, fx tilfælde hvor mellemrum mellem ord er blevet meget små eller nærmest væk. En bestemt kildeangivelse er indtastet på disse to måder i filerne:

```
Ris paa150Maader
Rispa150Maader
```

for sådan ser det faktisk ud i bogen, men en person med modersmålskompetence ser straks at det rigtige ser sådan ud:

```
Ris paa 150 Maader
```

I andre tilfælde er der "flueklatter" i satsen, hvor indtasterne ikke har vidst om klatten hørte med til bogstavet, fx et n med flueklat over, der mest ligner et n med tilde. I andre tilfælde igen er satsen helt udvisket, og man kan som dansker måske rekonstruere den, men ellers må man søge tilbage til ordbogssedlerne hvis der er tale om et citat. Et eksempel på dette er et indtastet:

```
Løvemank{??}{?A}
```

som bør være

```
Løvemanken
```

Den næste fase i arbejdet skal foregå hos DSL, og det er den strukturelle fortolkning af de typografisk opmærkede data. Det er den fase vi skal til at begynde på, og vi ved endnu ikke hvilke problemer vi står over for, men vores forventning er at vi kan nå meget langt ved hjælp af de typografiske signaler der er givet videre fra bogformatet.

Vi regner med at gå trinvis frem sådan at vi kan offentliggøre en foreløbig version i efteråret 2005, en version hvor i hvert fald opslagsordene er opmærket, og hvor præsentationen er gjort mere overskuelig med flere linjeskift og brug af farver. Det er selvfølgelig primitivt og slet ikke endemålet, men for mange brugere af ODS vil det være et kæmpe fremskridt.

I en senere fase vil vi prøve at komme dybere ned i strukturen og opnå en finere opmærkning af indholdet i de forskellige oplysningstyper. En anden hovedopgave, som bestemt ikke må undervurderes, er sammenfletningen af ODS og de 5 supplementsbind. Fordelene ved supplementsbindene er at de allerede foreligger i elektronisk form, omend i et andet format end hovedværket; til gengæld er oplysningerne i supplementet ganske heterogene – der kan være tale om elementer der skal ind på alle niveauer i hovedværkets struktur: helt nye artikler, tilføjelse af betydninger eller citater, ændring eller sletning af eksisterende oplysninger. I første omgang kunne man tænke sig at supplementets artikler og andre oplysninger blot bringes efter hovedværkets, men i samme skærbillede (ligesom man har gjort i OED med de nye oplysninger der endnu ikke er integreret i de eksisterende artikler). Senere vil man, formentlig manuelt, skulle ind og placere oplysningerne det rigtige sted i strukturen.

5. *ordnet.dk* og fremtiden

Med *ordnet.dk* er Det Danske Sprog- og Litteraturselskab for første gang gået ind i at publicere ordbøger elektronisk. Men det behøver ikke slutte med ODS og DDO. Selskabet råder over flere resurser både mht. ordbøger og tekster: En ordbog over gammeldansk er under udarbejdelse, Holberg-Ordbog I-V udkom 1981-1988, og et elektronisk renæssanceprojekt der indledes fra 2006, omfatter foruden tekstudgivelser på både latin og dansk også begyndelsen til en ordbog over ældre nydansk. På tekstsiden råder DSL i kraft af projektet *Studér Middelalder på Nettet* (<http://smn.dsl.dk>) over en række centrale danske middelaldertekster i digital form. I det elektroniske *Arkiv for Dansk Litteratur* (www.adl.dk) findes desuden et meget stort antal tekster af klassiske danske forfattere, baseret på DSL-udgaver, som ville være oplagt at bruge som tekstkorpus for en overvejende litterær ordbog som ODS. Selvom det altså stadig er i afdelingen for visioner og ønsketænkning, finder vi det ikke helt urealistisk at forestille sig *ordnet.dk* udvidet til engang i fremtiden at omfatte både tekstudgaver og leksikografiske beskrivelser af dansk sprog lige fra runerne til i dag.

Litteratur

Ordbøger:

- DDO = *Den Danske Ordbog. Bind 1-6*. Udgivet af Det Danske Sprog- og Litteraturselskab. Hovedredaktører: Ebba Hjorth og Kjeld Kristensen. København: Gyldendal 2003-2005.
- ODS = *Ordbog over det danske Sprog. Bind 1-28*. Udgivet af Det Danske Sprog- og Litteraturselskab. Grundlagt af Verner Dahlerup. Ledende redaktører: H. Juul-Jensen og Jørgen Glahder. København: Gyldendalske Boghandel, Nordisk Forlag 1918-1956.

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.
- Bergenholtz, Henning/Vibeke Vrang 2004: Den Danske Ordbog imponerer og skuffer. I: *Hermes* 33, 149-178.
- Church, Kenneth og Hanks, Patrick 1989: Word association norms, mutual information and lexicography. I: *ACL Proceedings, 27th Annual Meeting, Vancouver*. 76-83.
- Grønvik, Oddrun og Tvedt, Lars Jørgen 2006: Norsk Ordbok 2014 – presentation av eit komplekst leksikografisk verktøy. I denne rapport.
- Guttu, Tor 2005: Videreføringen av *Norsk Riksmålordbok* – om redigeringsarbeidet. I: Fjeld, Ruth V. og Worren, Dagfinn (red.): *Nordiske Studiar i Leksikografi 7. Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24. mai 2003*. Oslo. 166-174.
- Kilgariff, Adam; Rychly, Pavel; Smrz, Pavel; Tugwell, David 2004: The Sketch Engine. I: Williams, Geoffrey and Vessier, Sandra: *Proceedings of the 11th EURALEX International Congress*. Lorient. 105-115.
- Trap-Jensen, Lars 2004: Et net af ord – *ordnet.dk*. I: *Mål & Mæle* 4, 20-26.
- Trap-Jensen, Lars, under udgivelse: Making Dictionaries for Paper or Screen: Implications for Conceptual Design. I: *Proceedings of the 12th EURALEX International Congress*. Torino 2006.

Websider:

- ADL = *Arkiv for Dansk Litteratur*. Det Kongelige Bibliotek og Det Danske Sprog- og Litteraturselskab. København. www.adl.dk
- DanNet. Center for Sprogteknologi og Det Danske Sprog- og Litteraturselskab 2005-2008. København. www.wordnet.dk
- Korpus 2000*. Det Danske Sprog- og Litteraturselskab. København. <http://korpus.dsl.dk>
- Studér Middelalder på Nettet*. Det Danske Sprog- og Litteraturselskab. København. <http://smn.dsl.dk>

Är ordboken *Svenskt språkbruk* en deskriptiv ordbok?

Is the dictionary *Svenskt språkbruk* a descriptive dictionary? The dictionary *Svenskt språkbruk. Ordbok över konstruktioner och fraser* [Swedish Usage: A Dictionary of Constructions and Phrases] is characterised as a descriptive dictionary, i.e. one in which the phraseological material is presented as it appears in the data source, which consists of available Swedish concordances and text databases. When working with concordances and text databases, you are easily led to believe that you have captured the real usage because you are dealing with authentic material. This is of course partly an illusion, since every selection and treatment of data is necessarily based on some analytic framework. In this article, I discuss the problems with the notion that as a linguist one works descriptively in an unprejudiced way, whereas in reality in many phases of the work one is influenced, consciously or not, by various kinds of norms. In this discussion, I refer to the linguistic classification on which the organisation of this dictionary is based.

1. Inledning

Det har hållits många föredrag om ordböcker och normer, så ämnet ordböcker och normer är om inte uttömt så i alla fall inte obearbetat. Att jag ändå tyckte att jag skulle ta upp det här ämnet berodde på att det finns en del finlandssvenskt material i ordboken *Svenskt språkbruk. Ordbok över konstruktioner och fraser* (SSB). Syftet med att också ta med en del konstruktioner och fraser som enbart kan förekomma i den svenska som talas i Finland var att få några finlandssvenska uttryck kodifierade och förklarade för en större svensk läsekrets, i synnerhet för dem som också ibland läser finlandssvensk skönlitteratur. Men det har kommit en del kritik från finlandssvenskt håll mot urvalet av dessa finlandssvenska uttryck. Det visade sig nämligen att det finns sådant i SSB som de finlandssvenska språkvårdarna inte ansett sig kunna godkänna för användning. Nina Martola redovisar detta i sin recension av ordboken och påpekar för de finlandssvenska läsarna ”För oss finlandssvenska användare gäller det att komma ihåg att SSB är en deskriptiv ordbok och inte en normativ” (Martola 2003:25). Denna reaktion aktualiserade problemet med att ordboksförfattarnas intentioner inte nödvändigtvis överensstämmer med ordboks-användarnas förväntningar. Ordboksförfattarna strävar vanligtvis efter att vara deskriptiva, dvs. att beskriva bruket så som det föreligger i tillgängliga källor. Men kan ordböcker verkligen vara rent deskriptiva? Uppläggningsen av en ordbok, klassificeringen av språkliga enheter och urvalet kan väl inte ske utan att normer är inblandade? Syftet med den här artikeln är att försöka närma mig svaren på dessa frågor.

2. Ordboken *Svenskt språkbruk*

Svenskt språkbruk är en tämligen nyskriven fraseologisk ordbok utarbetad av Svenska språknämnden. Föregångaren, *Svensk handordbok* från 1966, som också kom till på initiativ av Svenska språknämnden, är en heltäckande ordbok över konstruktioner och fraser i svenskan. När det behövdes en ny mera samtida fraseologisk ordbok bestämde man att den också skulle vara heltäckande. Därför innehåller SSB alla de förväntade slagen av fraskategorier. SSB innehåller både produktiva element och lexikaliserade. De produktiva inslagen gör att ordboken kan klassificeras som en produktionsordbok, avsedd som hjälpmedel vid produktion av text och de lexikaliserade inslagen gör att den kan ses som en receptionsordbok, avsedd att ge upplysningar om frasernas betydelse och funktion. SSB är alltså en sammansatt ordbok, som är tänkt att vara ordboksanvändarna behjälplig på många olika sätt.

3. Vad är språknormer?

Det finns förhållandevis mycket skrivet om *normer*, även om språknormer, men i många fall har begreppet norm smält samman med närliggande begrepp och företeelser. Normer är regler, rättesnören som reglerar språkbruket. De påverkar alltså bruket men är skilda från bruket och ska inte sammanblandas med det. Enligt Renate Bartsch är normerna den normaliserande faktorn i språket (Bartsch 1987:171). Normerna är med andra ord den faktor som skapar enhetlighet i språket. En typ av språklig enhetlighet är den varietet som brukar gå under beteckningen standard-språk. Det är det språk som språkbrukare oftast förväntar sig att finna i ordböcker. Men ordböcker innehåller i regel mer än så, så också SSB. Problemet är att man i regel måste räkna med att allt som står i ordböcker är acceptabelt, men att allt inte är acceptabelt i alla sammanhang. SSB innehåller en del talspråkligt material och också en del innovationer, som t.ex. vissa nya alternativa prepositioner och strykningar av formord. Det betyder att många ordboksanvändare kan hitta språkliga varieteter som de inte tänkt sig eller normförändringar som de inte ännu hunnit anamma. Standardspråk kan enligt Olle Josephson ”måhända beskrivas som föreställningen att det finns något som alla tycker är rätt” (Josephson 2004:158). Termen standardspråk är inte på något sätt en entydig term. Den kan bl.a. uppfattas som liktydig med ”den språkform som uttryckligen standardmärkts, eller kodifierats”, alltså den språkform som finns i ordböcker av typen Svenska Akademiens ordlista. Standardspråk kan också uppfattas som ”den varietet av ett språk som kan brukas i och är accepterad i flest situationer” (Josephson 2004:159). Man kan säga att standardspråk är en språkform som på olika sätt är avgränsbar, även om gränserna för vad standardspråk är kan variera. Renate Bartsch anser att normer är konservativa. Individer envisas med att hålla fast vid sina normer för att inte förlora orienteringen i tillvaron (Bartsch 1978:173). Man måste räkna med att det tar tid innan normförändringar hunnit tränga in i det som olika individer uppfattar som standardspråket.

3.1. Svenskt språkbruk och normerna

SSB betecknas i inledningen till ordboken som en *deskriptiv ordbok*. Med *deskriptiv* avses här att alla konstruktioner och fraser med exempel presenteras så som de kom till uttryck i källmaterialet. Som källmaterial användes konkordanserna i Språkbanken i Göteborg, textdatabasen Presstext och Google. Gångse ordböcker användes också som källa. När man arbetar med konkordanser och databaser är det lätt hänt att man vaggas in i föreställningen att man nått det verkliga språkbruket eftersom man arbetar med autentiskt språkmaterial. Detta är givetvis till en del en chimär. Det är visserligen det autentiska språkbruket som finns lagrat i konkordanserna och databaserna, men för att få fram ett faktiskt och presenterbart språkmaterial för en ordbok måste detta material på olika sätt filtreras genom ordboks författarnas språkvetenskapliga och lexikografiska kunnande. Den här processen sker inte utan normer. Det som presenteras i ordboken måste vara prototypiskt i någon mening, dvs. det måste kunna utgöra mönster för ordboks användarna. För att få fram dessa mönster behöver ordboks författarna utnyttja lingvistiska begreppssystem. Dessa begreppssystem är givetvis styrande. För en ordbok som SSB spelar grammatik i vid bemärkelse en mycket stor roll, vilket modellen i figur 2 är en illustration till. Grammatiken, som ju består av språkliga regelsystem, måste självklart ses som en normativ företeelse. Per Linell tar också upp detta samband i sin bok *The Written Bias in Linguistics*. Han anser att det språkvetenskapliga arbetet är överfullt med normativa aspekter (Linell 2005:162). Dessutom står man som ordboks författare ständigt inför valsituationer. Då påverkas man väl medvetet eller omedvetet av vedertagna språknormer för att kunna sovra bland alternativen och ta ställning till dem? Hur gör man då? Är det så att man väljer det alternativ som passar bäst in i det etablerade språkssystemet? Eller väljer man det alternativ som har det högsta frekvenstalet i databaserna, för att man utgår ifrån att det alternativet sannolikt är det mest representativa för bruket? Eller är det kanske så att man helt omedvetet stöder sig på sina egna privata normer vid val av alternativ?

3.2. Bruklighetsmarkörerna och normerna

De konventionella *bruklighetsmarkörerna* är de medel med vilka ordboks författaren kan vägleda sina läsare i vilka sammanhang ett uttryck går att använda. Bruklighetsmarkörerna syftar till att ange stil, attityd, användbarhet och regional variation. Se figur 1. De är av tradition förhållandevis få. Vissa av dem är till sitt innehåll mycket vida. En sådan markör är stilbeteckningen *vard.* för vardaglig stil, som kan avse allt från ett lite ledigare skriftspråkligt uttryckssätt till någonting som bara förekommer i talspråket i okonventionella situationer. Dessutom är gränsdragningarna mycket subjektiva mellan vad som är vardagligt och vad som är neutralt. Det är i regel ordboks författarens egna normer som avgör var gränsen går mellan det markerade fallet med bruklighetsmarkören *vard.* och det omarkerade fallet utan bruklighetsmarkör.

Man måste utgå ifrån att de stilistiska normerna varierar ansenligt mellan olika språkbrukare och att den hjälp man kan få som ordboksanvändare därför är tämligen blygsam i det här fallet. Det finns en annan typ av bruklighetsmarkörer i SSB som är tänkt att vara vägledande. Det är förkortningarna *el.* och *äv.* Dessa står mellan alternativa konstruktioner och fraser. Förkortningen *el.* betyder att båda alternativen är likvärdiga och att ingetdera alternativet är mer brukbart än det andra. Förkortningen *äv.* betyder att det första alternativet är mera brukligt än det senare. Det senare alternativet är medtaget för att det finns säkra belägg på det även om dessa belägg i regel är mycket färre. Det kan röra sig om något nytt som ännu inte hunnit etableras i större skala eller något äldre fall som fortfarande är aktuellt i språket och som det därför inte finns skäl att utesluta. Den här typen av bruklighetsmarkörer har en tydligt normerande funktion. De syftar till att skicka godtagbara alternativ. Med hjälp av dem viktas presentationen av konstruktioner och fraser.

Stil: *högt.* – högtidligt, *vard.* – vardagligt, *åld.* – ålderdomligt

Attityd: *iron.* – ironiskt; *neds.* – nedsättande, *skämts.* – skämtsamt; *vulg.* – vulgärt

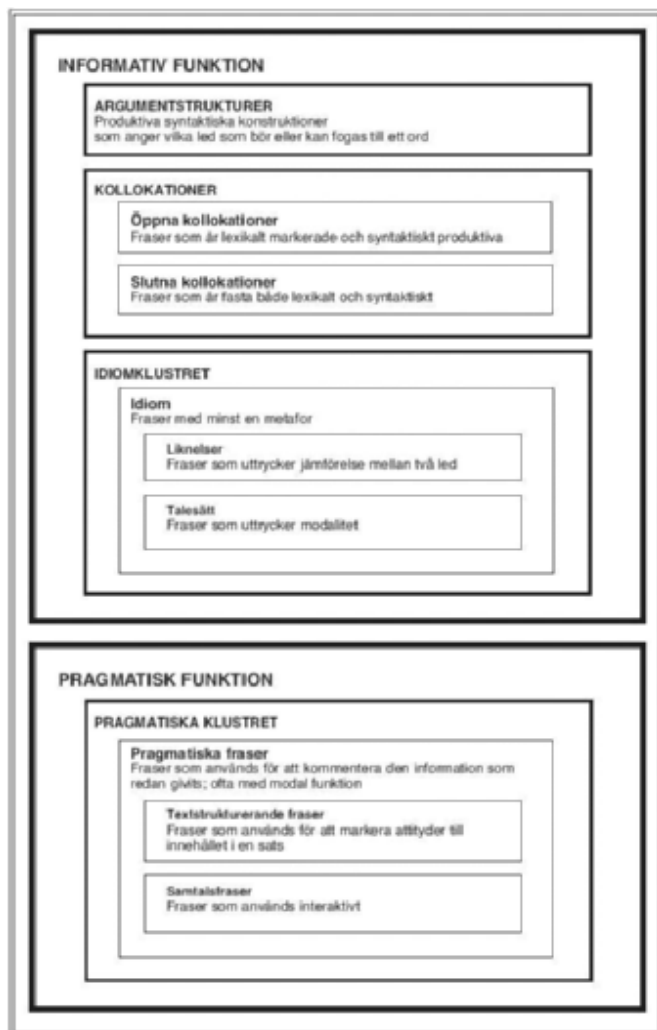
Användbarhet: *el.* – eller, *äv.* – även

Regional variation: *finl.* – finlandssvenska

Figur 1. *Bruklighetsmarkörer i SSB*

4. *Svenskt språkbruk och begreppssystemen*

För uppläggningsen av ordboken skapades en modell som listade de fraskategorier som skulle ingå i ordboken. Se figur 2. Den behövdes för klassificeringen av konstruktioner och fraser och för dispositionen av ordboksartiklarna. Man ska inte se den här modellen som en teoretisk modell i den meningen att den skulle vara en schematisk figur för en sammanhängande fraseologisk teori. Den syftar enbart till att visa vilka fraseologiska kategorier som förekommer i ordboken, hur vi definierar dem och i vilken ordning de uppträder i ordboksartiklarna. Definitionerna är inte på något sätt uttömmande. Syftet med dem är att på ett enkelt sätt ange vad som skiljer den ena fraskategorin från den andra. De aktuella fraskategorierna är följande: *argumentstrukturer* (dvs. syntaktiska mönsterkonstruktioner), *kollokationer* med underkategorierna *öppna kollokationer* och *slutna kollokationer*, alla *idiomkategorier* med de speciella underkategorierna *liknelser* och *talesätt* och *pragmatiska fraser* med de speciella underkategorierna *textstrukturerande fraser* och *samtalsfraser*. Av definitionerna i modellen framgår att alla fraskategorier utom fraskategorierna i det pragmatiska klustret har definierats strukturellt. Det som förenar dessa fraskategorier är att de har funktionen att vara informationsbärande delar i en satsstruktur, därav termen *informativ funktion* som övergripande beteckning. Fraskategorierna i det pragmatiska klustret har däremot definierats enligt hur de används, därav termen



Figur 2. Modell för kategorisering av fraser

pragmatisk funktion som övergripande beteckning. Det fanns ytterligare en lista på ett stort antal konstruktionstyper och frastyper i kodform, som inte är redovisade här. De blev med tiden sammanlagt drygt 200 och behövdes för eliciteringen av konstruktioner och fraser ur källmaterialet. Jag ska inte här gå närmare in på de språkvetenskapliga begrepp och teorier som legat till grund för modellen utan jag nöjer mig med att bara nämna dem, eftersom de varit normativt styrande och samtidigt mycket betydelsefulla för ordbokens utformning. För själva eliciteringen av fraserna utgör enkel frasstruktur, kollokationsbegreppet och Fillmores teorier om sambandet mellan syntax och semantik grunden. För kategoriseringen av fraserna

har begreppen lexikalisering och produktivitet spelat en stor roll. För funktionsbegreppet, som också har stor betydelse för kategoriseringen av fraser, användes främst Hallidays teorier. Vidare inhämtades en hel del insikter om metaforer, idiom och pragmatik.

Argumentstrukturer:

säkra ngt; avgå med ngt, seger över ngn/ngt; försöka göra ngt äv. försöka att göra ngt

Öppna kollokationer:

säkra segern, avgå med segern; hård vind, sträng kyla

Slutna kollokationer:

nautisk mil; falla i sömn

Idiom:

göra ngt i grevens tid, kasta yxan i sjön, dra åt svängremmen

Liknelser:

ngn är fattig som en kyrkråtta

Talesätt:

hoppet är det sista som överger människan, [man ska inte] kasta ut barnet med badvattnet

Pragmatiska fraser:

går det så går det, affärer är affärer

Textstrukturerande fraser:

först och främst, i själva verket, när allt kommer omkring

Samtalsfraser:

tusen tack! förlåt att jag besvärar, vad kan jag stå till tjänst med? god jul!

Figur 3. Exempel på konstruktioner och fraser

Argumentstrukturerna består av formaliserade grammatiska konstruktioner. De anger vilka slags enheter som uppslagsordet kan kombineras med. Dessa och också de öppna kollokationerna är produktiva. Argumentstrukturerna erbjuder en ram som språkbrukaren får vara kreativ inom medan de öppna kollokationerna visar på tänkbara ordkombinationer. Argumentstrukturerna har alltid ofyllda led som t.ex. *ngt* som i konstruktionerna *säkra ngt* och *avgå med ngt*. De öppna kollokationer har motsvarande led fyllda som i fraserna *säkra segern* och *avgå med segern*. De öppna kollokationerna ska bara ses som exempel på typiska markerade ordkombinationer. Det finns gott om öppna kollokationer som inte tagits med i ordboken. Slutna kollokationer är lexikaliserade och där är ingen produktivitet möjlig. De slutna kollokationerna inrymmer termer, som t.ex. *nautisk mil* och lexemartade fraser, som t.ex. *falla i sömn* med en motsvarighet i det enkla verbet *somna*. Idiomen i den här modellen består av fraser som innehåller åtminstone en metafor. Ett exempel på ett konventionellt idiom är t.ex. idiomet *kasta yxan i sjön* med betydelsen 'att ge upp allt'. Att 'ge upp' jämför man med att man kastar sitt viktiga verktyg, yxan, i sjön

varifrån man inte kan få upp det igen. En annan idiomartad fras är liknelserna som t.ex. frasen *ngn är fattig som en kyrkerätta* som används 'om ngn som har mycket litet pengar'. Ytterligare en idiomartad fraskategori är de fraser som traditionellt kallas talesätt. Typiskt för dem är att de har modal funktion, dvs. syftet med dem är att tala om hur man ska förhålla sig till något. Modaliteten kan vara explicit, men den kan också vara implicit som t.ex. i frasen *hoppet är det sista som överger människan*. De s.k. pragmatiska fraserna är inte lika systematiskt beskrivbara som de övriga fraserna. De är strukturellt mycket olika och kan närmast karakteriseras som ett slags klichéartade stereotyper. De tillför inte texten eller yttrandet någon ny information i egentlig mening utan används för att kommentera den information som redan getts. En sådan fras är t.ex. *affärer är affärer*, som kan ses som en överslätande kommentar om att det enbart är det affärsmässiga som gäller. Bland de pragmatiska fraserna finns också fraser som uttrycker många olika interaktiva funktioner som hälsningsfraser, fraser med vilka man uttrycker olika grader av tacksamhet, ursäkt o.dyl.

5. Markering av normerna

En stor del av materialet i SSB kan sägas vara konventionellt skriftspråkligt standardspråk, vilket ordboksanvändarna i praktiken förmodligen har mest nytta av. Om detta råder det sällan delade meningar. Däremot kan innovationer vara mera diskutabla. De innebär normförändringar, nya regler som gör att standardspråket omformas. Detta berör framför allt de grammatiska konstruktionerna, som i modellen går under beteckningen argumentstrukturer, även om andra fraskategorier som öppna kollokationer och idiom också kan komma ifråga. För argumentstrukturernas del kan som normförändringar t.ex. gälla nya användningar av prepositioner, strykningar av formord och intransitiva verb som börjat användas transitivt och tvärtom. Det kan ur normsynpunkt vara intressant att jämföra med hur samma konstruktion presenteras i SSB och i den äldre fraseologiska ordboken Svensk handordbok. I denna ordbok får man t.ex. i ordboksartikeln *bild* följande information om val av preposition när betydelsen är, enligt formuleringen i ordboken, "avbildning":

en ~ av (ej: på) kungen.

Att direkt avråda från en konstruktion som t.ex. (ej: på) är ett rent preskriptivt presentationsätt, som inte förekommer i SSB. Det kan ytterligare vara värt att uppmärksamma att Svensk handordbok skrevs under en tid då det ännu inte var en självklarhet att lyfta upp konstruktioner och fraser till generell frasnivå. Allt presenteras därför i form av korta exempelmeningar som exemplet ovan med "kungen". Det gäller sedan för användarna att själva göra generaliseringarna och

plocka fram mönstren. I SSB presenteras konstruktion med *bild* + PP på följande sätt:

bild av ngn/ngt el. *bild på ngn/ngt* han målade för några år sedan en ~ av (som föreställer) Humphrey Bogart; på hennes vägg sitter en ~ på idolen;

Drygt tre decennier efter att Svensk handordbok kom ut har prepositionen *på* till *bild* vunnit insteg i språket. Det finns fortfarande en stilskillnad mellan *av* och *på*, men den är minimal. Man kan numera fritt använda båda alternativen, vilket också bruklighetsmarkören *el.* anger här. I skriftspråket är *av* fortfarande vanligare än *på*, därav placeringen av alternativen. Det händer också att verbpartiklar stryks. Konstruktionen *hoppa över ngt* i betydelsen 'strunta i, avstå ifrån' kan även konstrueras så som anges i SSB:

hoppa ngt – Ska ni gå på bio i kväll? – Nej, vi ~r det.

Det partikellösa alternativet är som exemplet antyder talspråkligt och det är numera vanligt i språkbruket. Det här fallet är tämligen nytt och står enbart i SSB. Ett annat fall, som också enbart står i SSB, är *att*-strykning efter verbet *kommer* som presenteras på följande sätt:

ngn kommer att göra ngt el. *ngn kommer göra ngt* vi vet inte hur kommunen kommer att använda pengarna; hon kommer säkert köpa en ny dator till hösten.

Det här är ett fall där den *att*-lösa formen ökat väldigt i användning, vilket av reaktionerna att döma inte har gällt alla kategorier språkbrukare. Ökningen av det *att*-lösa alternativet har varit så stor att alternativen kunnat markeras med bruklighetsmarkören *el.*, vilket betyder att båda alternativen är fullt acceptabla. Ett tredje fall som tidmässigt inte heller kunnat kodifieras i en ordbok tidigare är t.ex. den ändrade transitiviteten hos vissa verb som t.ex. hos verbet *accelerera*. Det är ett verb som tidigare bara kunde användas intransitivt, men som numera också kan användas transitivt. I Handordboken finns enbart den intransitiva formen: bilen, motorn ~r snabbt. SSB har givetvis också samma konstruktion med viss semantisk utvidgning:

ngt accelererar bilen ~de och försvann ur synhåll; äv. om att ngt sker snabbare försäljningen ~de (ökade) i mars och har fortsatt även i april.

Det nya i SSB är att det också finns en transitiv användning med kausativ betydelse 'få att öka':

accelerera ngt för att ~ elektroner behöver man separera dem från kärnan; äv. om att påskynda ngt regeringens tvingades ~ reformeringen av utbildningssystemet.

Tendensen att intransitiva verb transiveras verkar öka. De fyra fallen ovan representerar alla tydliga normförändringar. Tidigare fasta grammatisklexikala regler har lösts upp och kompletterats med nya grammatisklexikala regler. Det gamla finns alltså kvar och det nya uppstår som ett alternativ som ordboks författarna måste ta ställning till. Lars Trap-Jensen har också skrivit om detta problem med lexikografiska val mellan olika alternativ. Enligt honom uppstår problemet när språkbruket ”i mange tilfælde giver et andet svar end den eksternt givne norm, og redaktøren må derfor enten foretage et valg eller prøve at finde en løsning der tilgodeser begge aspekter” (Trap-Jensen 2002:64). I det här kapitlet har jag försökt visa hur detta problem har lösts i SSB. I allmänhet har problemet lösts så att bägge aspekterna har tillgodosetts.

6. Det deskriptiva och normerna

Det framgår av de föregående kapitlen att normer reglerar och påverkar. De är normaliserande och de kan ha en konservativ inverkan på språkbruket. Det finns olika slag av normer som är av betydelse för tillkomsten av en ordbok. Som ordboks författare får man räkna med att man är styrd av normer som man är medveten om och också av normer som man inte är medveten om. Medvetet handskas man med de normer som styr själva språkssystemet, grammatiken och lexikonet. Medvetet utnyttjar man också generella lingvistiska begreppssystem. Hur dessa har utnyttjats i SSB framgår av kapitel 5. När normer utnyttjas omedvetet är det framför allt fråga om det egna privata normsystemet som kommer till uttryck. Ett lexikografiskt sådant fall är t.ex. bruklighetsmarkören *vard.* i kapitel 3.2. Ulf Teleman har tidigt beskrivit detta samband, även om det i hans fall gäller allmän språkbeskrivning och inte lexikografi (Teleman 1979:19). Hur går detta med ordboks författarnas beroende av normer ihop med att de normalt deklarerar att deras ordböcker är deskriptiva? Ordet deskriptiv är karakteriserande och syftar på sambandet mellan framställnings-sätt och bl.a. normer. En ordbok anses deskriptiv om den presenterar språkbruket utan preskriptiva kommentarer. Enligt Bo Svensén avses med deskriptiv lexikografiskt ”lexikografens sätt att förhålla sig till det språkliga utgångsmaterialet” och vidare ”att det skall råda överensstämmelse mellan detta material och ordbokens beskrivning av det” (Svensén 2004:29). Kristina Nikula för in ytterligare en annan aspekt på deskriptivitet lexikografiskt. Hon kommer i sin analys fram till att begreppen deskriptiv och normativ är företeelser på olika plan. Det normativa träder in först när man ska utnyttja de resultat som man fått fram på deskriptiv väg, även om hon lite längre fram i texten konstaterar att ”deskriptiva utsagor om språket” inte är ”fria från värderingar” (Nikula 1992:46). Man kan alltså dra den slutsatsen att det

förmodligen inte existerar någon deskriptivitet utan normer och att deskriptiva ordböcker i egentlig mening inte förekommer. Ändå finns det inte något skäl att frånga termen deskriptiv ordbok. Termen har sitt berättigande så länge man är medveten om vad den innebär. En ordbok som SSB är närmast lätt normativ med en klar deskriptiv intention att kodifiera de konstruktioner och fraser som används i svenskan.

Litteratur

Ordböcker:

SSB = *Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Stockholm: Norstedts ordbok 2003.

Svensk handordbok. Konstruktioner och fraseologi (1966). Stockholm: Esselte Studium 1989.

Annan litteratur:

Bartsch, Renate 1987: *Norms of Language*. London: Longman Linguistic Library.

Josephson, Olle 2004: *Ju: ifrågasatta självklarheter om svenskan, engelskan och alla andra språk*. Stockholm: Norstedts ordbok.

Linell, Per 2005: *The Written Language Bias in Linguistics*. London: Routledge.

Martola, Nina 2003: Svenskt språkbruk – ordbok över konstruktioner och fraser. I: *Språkbruk* 4, 20-25.

Nikula, Kristina 1992: Deskriptiva ordböcker – finns dom? I: *Nordiske studier i leksikografi* 1. Oslo, 43-53.

Svensén, Bo 2004: *Handbok i lexikografi*. Stockholm: Norstedts Akademiska förlag.

Teleman, Ulf 1979: *Språkrätt*. Lund: Liber Läromedel.

Trap-Jensen, Lars 2002: Normering og deskription i Den Danske Ordbog – mere eller mindre? I: *LexicoNordica* 9. Oslo, 63-78.

Om felaktiga särskrivningar i svenskan och norskan och om ordboksbaseade och statistiska program för att hitta dem

On decomposed compounds in Swedish and Norwegian and Dictionary-based and statistical programs to detect them. In written Swedish and Norwegian, as well as in Danish, there is a tendency to decompose compounds, e.g. *barn bok* instead of, correctly, *barnbok* ('children's book'). The article deals with several aspects of this phenomenon, in particular the possibilities of detecting them by means of dictionary-based and statistical programs. It is claimed that statistical (and non-linguistic) programs should, at most, be regarded as complements to dictionary-based ones. Intelligent programs must also be based on morphological analysis of the (potential) compounds. As for Norwegian, an extensive corpus investigation indicates that decomposition of compounds is much more common in informal texts than in formal ones.

1. Inledning

I svenskan och norskan, liksom även i danskan, tycks på senare år ett lexikalt skrivfel ha brett ut sig mer och mer. Fenomenet vi tänker på kan på svenska kallas *felaktig särskrivning* och exemplifieras av meningar som *vi gick på en skogs promenad* (>*skogspromenad*) och *hon skriver mest barn böcker* (>*barnböcker*). Man kan möta det i de flesta informella stilarter, ibland t.o.m. i mer formella texter. Det har tilldragit sig stor uppmärksamhet både i svensk och norsk språkvårdsdebatt och blivit föremål för åtskilliga mindre specialstudier. Det har också givit upphov till en (numera nedlagd) svensk webbsida där man bl.a. samlat särskilt roliga fall av särskrivningar, t.ex. *bar* (=naken) *disk* (>*bardisk*) och *kassa* (=dåliga) *apparater* (>*kassaapparater*). Det finns också dataprogram som försöker upptäcka felaktiga särskrivningar och leda skribenten till rätta. De mer intelligenta av dessa program måste nödvändigtvis utnyttja uppgifter från elektroniska lexikon. Problemet med felaktiga särskrivningar har därför stor lexikografisk relevans.

Vår artikel är uppdelad i en svensk och en norsk del, men mycket av det som sägs om svenska är relevant för norska (och danska), och vice versa. I den svenska delen ges först en bakgrund och en kort diskussion om orsakerna till (den eventuella ökningen av) felaktig särskrivning. Sedan presenteras två ganska färska svenska studier som ger en uppfattning om felets vanlighet. Därefter vidtar en diskussion kring möjligheterna att med hjälp av elektroniska lexikon upptäcka de viktigaste typerna, utan att man övergenererar felmarkeringarna, samt en presentation av det troligen intelligentaste – och ordboksbaseade – svenska särskrivningsprogrammet ("Granska"). Slutligen nämns ett enklare, statistiskt baserat program som många Googleanvändare har stött på. En jämförande undersökning av Granskas och

Googles prestanda avslutar den svenska delen.

Den norska delen redogör för en undersökning av en onormerad korpus från Internet som kontrollerats mot sammansättningar från Tanums stora rättskrivningsordbok. Denna sökning gav en stor korpus av potentiellt felaktiga särskrivningar. Detta material har kategoriserats efter morfologiska och etymologiska kriterier. Därefter följer en diskussion av möjliga förklaringar till de olika feltyperna som kanske kan bidra till bättre insikt i orsakerna till den nya tendens i norska. Såvitt vi känner till finns det inget särskrivningsprogram för norska, jämförbart med det svenska. Det borde vara en utmaning för norska lingvister och språkteknologer att skapa ett sådant program, gärna med utgångspunkt i de svenska erfarenheterna och i en grundligare analys av de registrerade särskrivningsfelen.

2. Felaktig särskrivning i svenskan

2.1. Bakgrund

Felaktig särskrivning (härefter förenklat *särskrivning*) är ingalunda något nytt i svenskan. En känd språkvårdare, Nils Linder, klagar över det i en språkhandledning från 1886. De första studierna över fenomenet dyker upp på 1970-talet, och många tycks ta för givet att det stadigt har ökat i omfattning sedan dess. Det är emellertid inte säkert att det förhåller sig på det sättet (jfr Daher 2004).

Den mest populära och närliggande förklaringen är att vi har att göra med ännu ett fall av engelsk påverkan. En annan förklaring är det moderna bruket i svenska barnböcker att skriva sammansättningar med bindestreck (*barn-bok*). En observation som Hallencreutz (2001) gör är att långa sammansättningar är mer utsatta för särskrivning än förhållandevis korta. Man kan möjligen tänka sig att långa sammansättningar blir allt vanligare i takt med att samhället blir alltmer komplicerat. En oväntad källa till felaktiga särskrivningar är faktiskt Words rättskrivningsprogram, som gärna vill dela upp långa sammansättningar det inte kan tolka i mindre, begripliga delar. Ordet *felmarkeringarna* tidigare i artikeln ovan ville Wordprogrammet t.ex. nödvändigtvis ändra till det felaktiga *fel markeringarna*.

De viktigaste empiriska studierna över svenska särskrivningar är Hallencreutz (2001) och Hallencreutz (2003). Hallencreutz har bl.a. undersökt ett stort antal svenska grundskoleuppsatser från 1989 som sammanlagt innehåller ca 5000 sammansättningar. Av dessa är knappt 500, eller 9,6 procent, särskrivna. I en senare studie, Daher (2004), undersöks gymnasistuppsatser från 2004 med sammanlagt ca 6000 sammansättningar, och det visade sig att ca 300, eller 5,1 procent, var särskrivna. Den senare gruppen gjorde alltså bara hälften så många fel som den förra, men eftersom den utgjordes av mer avancerade skribenter är det svårt att dra några slutsatser beträffande ev. ökning eller minskning av särskrivningsfrekvensen i svenskan.

2.2. Automatisk upptäckt av svenska särskrivningar

Det är ganska självklart att elektroniska fullformslexikon med ordklassinformation kan ge värdefull information när det gäller automatisk upptäckt av särskrivningar. Det enklaste fallet föreligger när den särskrivna förleden inte kan förekomma som självständigt ord. Meningen *han studerade kvinno frågan* innehåller ett icke-existerande ”ord”, *kvinno*. När särskrivningsprogrammet inte hittar det i ordboken kan det i förväg ha fått instruktionen att pröva att sätta ihop det med nästa ord. I bästa fall finns sammansättningen *kvinnofrågan* i ordboken och problemet är då löst. Men även om sammansättningen saknas kan programmet godkänna den, under förutsättning att förleden *kvinno-* återfinns i ordboken och *frågan* identifieras som ett substantiv.

Men i de flesta fall är problemet mer komplicerat än så, genom att förleden är ett möjligt ord. I den här korta artikeln nöjer vi oss med att se på de två viktigaste sammansättningstyperna, nämligen substantiv + substantiv (*barnbok*) och adjektiv + substantiv (*vitsippa*). Dessa båda typer (grupp 1 och 2) utgör tillsammans en stor majoritet av sammansättningarna i nästan vilka texter som helst. Två viktiga undergrupper i grupp 1 är (a) sammansättningar där förleden är ett substantiv i grundform (*barn-bok*) och (b) sammansättningar där förleden är ett substantiv i genitiv (*skogs-promenad*). I fortsättningen studerar vi enbart exempel från grupperna 1a, 1b och 2.

Den minst problematiska av dessa grupper för ett särskrivningsprogram är grupp 1a. Det har att göra med ett speciellt fenomen i svensk (och norsk/dansk) syntax, nämligen att substantiv i grundform ytterst sällan följs av substantiv (i vilken form som helst). Ett bra särskrivningsprogram kan därför säga att *barn böcker* med minst 99 procents sannolikhet är en felskrivning av *barnböcker*, även om sammansättningen inte skulle finnas i programmets lexikon. Någon enstaka gång råkar ”ordsträngen” *barn böcker* förekomma i meningar *hon brukar skänka fattiga barn böcker* och då gör programmet troligen en felaktig felmarkering, men det måste man acceptera.

Mer problematisk är grupp 1b. Substantiv i genitiv följs typiskt av andra substantiv, och ordpar som *skogs promenad* och *livs innehåll* kan därför inte utan vidare felmarkeras (i varje fall inte utan avancerad semantisk och/eller syntaktisk analys, jfr *hans livs innehåll var ...*). Däremot kan ibland vissa kongruensförhållanden utnyttjas. Ett substantiv i genitiv kan t.ex. aldrig följas av ett substantiv i bestämd form, och därför kan ett bra särskrivningsprogram lugnt ändra *skogs promenaden* till *skogspromenaden*.

Liknande problem möter vi i grupp 2. Ett adjektiv följs typiskt av ett substantiv, och *en vacker vit sippa* är en lika rimlig fras som *en vacker vitsippa*. Men även här kan kongruensförhållanden utnyttjas. Ordsträngen *vit sippa* (jfr *den vita sippa*) är otänkbar och ändras riskfritt av programmet till *vitsippa*.

Överväganden av denna typ verkar ligga till grund för Granska-programmet

(<http://skrutten.nada.kth.se>). Programmet föreslår ändringar av nästan alla grupp 1a-fall men i allmänhet inte av särskrivningar som tillhör grupp 1b och grupp 2 (utom när kongruensförhållanden gör detta möjligt). Det är uppenbart att programmet utnyttjar ett elektroniskt lexikon med ordklass- och böjningsuppgifter och en begränsad syntaktisk analysmodul (för närsyntax, särskilt kongruens). För att komma längre i grupperna 1b och 2 skulle programmet bl.a. behöva ett lexikon med mer avancerad semantisk information (frasen *en skogs promenad* är ytterst osannolik av semantiska skäl, frasen *en skogs rikedomar* däremot fullt möjlig). Det bör också påpekas, att Granska-programmet givetvis tar hänsyn till homografi och därför i många fall avstår från att ge rättningsförslag. Det komiska *bar disk* (jfr ovan) lämnas t.ex. okommenterat eftersom *bar* kan vara både substantiv, verb och adjektiv (jfr *han bar disk till köket*).

En annan möjlighet att förbättra prestanda för intelligenta särskrivningsprogram av Granska-typ kan vara att utnyttja ordfrekvenser. Vi kan först se på ett program som ingår i Googles sökmotor och som enbart använder sig av statistik. Om man slår upp ”barn bok” i Google (svenska sidor) möts man av en vänlig fråga ”Menade du *barnbok*?”. Det betyder inte att programmet (som ”Granska”) ”vet” att *barn* är ett substantiv i grundform och att det följs av ett substantiv *bok* (jfr grupp 1a) utan bara att det har räknat antalet belägg av *barnbok* och antalet belägg av *barn bok* på nätet och funnit att det förra antalet är mycket större. Det är alltså fråga om en ytterst osofistikerad analys som i motsats till analyserna i Granska-programmet inte utnyttjar lexikografiska och syntaktiska hjälpmedel. Speciellt behöver (och kan!) Google-programmet inte bekymra sig om homografiproblemet utan kan frejdigt fråga den som söker efter ”bar disk” ”menade du *bardisk*?”

Trots att Googleprogrammet är primitivt kommer det därför att ge många korrekta ändringsförslag som Granska-programmet inte kan ge. Men vilket av programmen presterar bäst på en större, slumpmässigt vald mängd av felaktiga särskrivningar? För att få ett svar på detta testade jag de båda programmen på alla sammansättningar som enbart innehöll substantiv och/eller adjektiv i Dahers (2004) material, allt som allt 157 stycken. Resultatet blev att Granska klarade 111 (71 procent) och Google 80 (51 procent), alltså en klar seger för Granska. Bland särskrivningar som Google men inte Granska klarade märktes *drog missbrukare* (*drog* är homograft), *allergi framkallande*, *ensam stående* och *funktions hindrade* (*hindrade* är homograft). Omvänt klarade Granska men inte Google t.ex. *barn kontakter*, *diktsamlings boken*, *fantasi förmåga* och *fantasi lek*. Det är inte svårt att förstå varför; det är här fråga om ovanliga sammansättningar med bara enstaka, eller inga, belägg på nätet. En ohjälplig svaghet hos Googleprogrammet är att det är helt beroende av att antalet belägg av de aktuella sammansättningarna är stort.

Inte desto mindre skulle en statistisk komponent kunna förbättra Granska-programmets prestationsförmåga en del. I vår jämförande undersökning klarade

Googleprogrammet 19 av de 46 särskrivningar som Granska inte klarade. Tillsammans klarar de båda programmen således 130 av 157 särskrivningar, eller 84 procent. Det tycks alltså innebära en mycket betydande förbättring av det lingvistiskt baserade Granskaprogrammets prestanda. Förutsättningen för att det är fråga om en verklig förbättring är givetvis att Googleprogrammet inte ger många felaktiga felmarkeringar, och så tycks faktiskt inte vara fallet (även om detta bör undersökas närmare).

Det är dock ingen tvekan om att den viktigaste rollen i ett intelligent särskrivningsprogram spelas av ett bra elektroniskt fullformslexikon med böjningsuppgifter – ett inte ovälkommet besked till lexikografer!

3. Feilaktig særskrivning i norsk

3.1. Vandalisme og mishandling eller naturlig språkutvikling?

Som i svensk er det registrert en økende tendens til særskrivning av komposita i moderne norsk. Særskrivningen skjer trolig i total ubevissthet for kanskje de aller fleste som gjør disse feilene, men det er tydeligvis til stor gremmelse for normbevisste språkbrukere. Det er interessant å registrere at en ikke-lingvistisk faggruppe som astronomer har organisert en forening mot orddeling, med egen hjemmeside. Til skrekk og advarsel samles det inn feildelte ord som de mener gir grunnlag for feiltolkning (<http://folk.uio.no/tfredvik/amo/hvaeramo.html>). Foreningens mål er klart begrunnet på hjemmesiden:

De siste åra har en helt ny måte å mishandle det norske språket på blitt mer og mer vanlig, nemlig orddeling, eller *særskrivning* som det egentlig heter. Ord som *spisestue* blir til *spise stue*, *lammelår* blir til *lamme lår* og *hostesaft* blir til *hoste saft*. Vi i AMO mener slik språkvandalisme må bekjempes med alle lovlige midler, siden særskrivning kan gjøre en tekst tung å lese, kan føre til misforståelser og er et skritt nærmere en enda sterkere anglisering av språket vårt.

Tanken bak foreningens argumentasjon må være at de feildelte eksemplene kan bli tolket som fraser, *spise stue* som verbal pluss objekt, *lamme lår* som adjektiv og substantiv og *sjokolade biter* som subjekt og verbal. Sannsynligheten for slike misforståelser er liten, da konteksten vil gjøre de tolkningene umulige. Likevel blir fenomenet omtalt som vandalisme av språket og at det gjør tekster tunglest. Tunglesthet forutsetter i det minste at leserne oppfatter disse feilene, noe de vel ikke gjør, de som selv deler ord på denne måten. De mange feilene tyder heller på at språkfølelsen er i endring med hensyn til sammenskriving av komposita i norsk og andre liknende språk. Påvirkning fra engelsk er sikkert en delforklaring, men det er grunn til å anta at det ikke er hele forklaringen.

3.2. *Bakgrunn*

Det er ikke mye vi faktisk vet om særskriving av komposita i moderne norsk. De fleste undersøkelsene av fenomenet hittil er gjort med elevtekster som grunnlag. Den største undersøkelsen er hovedoppgaven til Roar Walmsness (2000). Laila Sakshaug (1992) har brukt reklametekster som materiale i sin hovedoppgave, og selv om det er lite og noe tilfeldig innsamlet, er det et svært viktig materiale, da nettopp slike tekster har sterk påvirkning på folks holdninger og oppfatning av hva som er moderne og i den forstand retningsgivende, og dermed kan gi grunnlag for normendring.

Man har antatt at delingen kommer dels av indrespråklige hensyn, f.eks. at noen komposita er mindre leksikalisert enn andre, at de har komposisjonsfuge eller ikke, at forleddet er flerleddet eller at ordet rett og slett er langt. Det forekommer også ytrespråklige forklaringer, særlig at det er påvirkning fra engelsk, men også at sjeldne komposita oftere deles, at formskrift gjør det vanskelig å se hvor et ord begynner og slutter, og altså at tekstutseendet veier tyngre enn rettskrivningsreglene.

Det norrøne språket hadde ikke klare regler for sær- eller sammenskriving av komposita, men sakte og sikkert utviklet det seg en konvensjon for sammenskriving. Bakken (1995) har vist at dette skjedde samtidig som det var en endring i trykk- og tonemønster for sammensetninger, slik at sammensetningene følt mer som enheter og hadde ett hovedtrykk. Det virker som en rimelig forklaring, og man skulle dermed tro at særskrivingstendensen i dag avspeiler en ny endring i uttalen, men det er lite som tyder på det. Det pedagogiske rådet om å lytte til uttalen og å skrive sammen uttrykk som er trykkhelheter, er ikke vanskelig å få folk med på, alle hører forskjell på *eplekake* og *eple kake*, og her kan det heller ikke oppstå mulige feiltolkninger.

På den ene siden er reglene for sær- og sammenskriving svært uklare, fordi det sannsynligvis ligger små semantiske nyanseforskjeller mellom en del sær- og sammenskrevne former, særlig av kombinasjonen adjektiv og substantiv. Det er ikke alltid lett å vite om det er betydningsforskjell mellom *venstre ben* og *venstreben* i utsagn som

- Spesielt gjelder vel dette hans **venstreben**
- Protese ønskes kjøpt, helst høyreben. **Venstreben** med "chaplín-fot" kan godtas til halv pris. Haster!
- Mitt **venstre ben** føles tungt – Begge mine ben føles tunge – Både mine armer og mine ben føles tunge

Det er ikke tilstrekkelig forklaring at det ligger en terminologisering til grunn for sammenskrivingen, slik som mellom *en halv liter* og *en halvliter*. I tekster fra fotball og medisin er det mange belegg på høyre- og venstreben, mens det i karate er snakk om høyre og venstre ben. Vi må derfor lete etter andre forklaringer.

3.3. Komposisjonstyper og særskriving

Lars Nygaard ved Tekstlaboratoriet ved HF, Universitetet i Oslo, har samlet inn et korpus som består av totalt ca. 140 mill. løpeord fra norske diskusjonslister på Internett (såkalte Usenet-grupper) i tidsrommet 1998-2002 (<http://logos.uio.no/usenet.html>). Det interessante med et slikt korpus er at det overhodet ikke er normert eller rettet. De som skriver på slike diskusjonslister, er som regel unge voksne som skriver mye og uten å legge for stor vekt på formell korrekthet.

Korpuset er ikke ordklassetagget, men er undersøkt for alle typer ordstrenger som i *Tanums store rettskrivningsordbok* (1996) forekom som ett ord, dermed fikk vi et stort materiale med potensielt særskrevne komposita. Totalt ga søket en liste på 4381 særskrevne former. Søk på et utvalg av de samme ordstrengene i Oslo-korpuset (<http://www.tekstlab.uio.no/norsk/bokmaal/>), som består av normerte tekster, ga så godt som ingen treff på særskrevne komposita. De feildelte ordene skal systematiseres i grupper etter typer av sammensetninger, for om mulig å oppdage om visse typer oftere ble særskrevet enn andre. Vi har dermed et materiale der det ligger godt til rette for å studere de forskjellige feiltypene og derigjennom kanskje finne fram til noe mer nyansert om årsakene til fenomenet.

Med støtte i syntaktisk struktur i norsk kan man automatisk kartlegge potensielle dekomposisjoner. Både kombinasjonen verb + substantiv og substantiv + verb er svært sannsynlige fraser, da det er vanlig inventar i strengen subjekt og verbal eller verbal og objekt. Automatisk sortering av fraser og særskrevne komposita krever kjennskap til seleksjonsrestriksjonene mellom komponentene, noe som forutsetter et maskinlesbart leksikon med semantisk informasjon, som f.eks. SIMPLE (jf. Pedersen 2005). Annerledes er det ved kombinasjonen substantiv og substantiv, dersom det første ikke står i genitiv: *barnas far*, er det sannsynligvis dekomponert kompositum, i norsk som i svensk.

Vi antar at følgende typer kan være interessante å studere nærmere:

Hjemlige sammensetninger:

- 1a) Ubøyd substantiv + ubøyd substantiv: *kule spill, familie rett, dag driver, bjørn unge*
- 1b) Substantiv i gen. + substantiv: *verdens erfaring, verdens rike, søndags morgen*
- 1c) Substantiv + komp.fuge + substantiv: *hjorte hund, saue melk*
- 2) Adjektiv + substantiv: *stor gård*
- 3) Preposisjon + substantiv: *etter utdannelse*

Lånte sammensetninger:

- 1) Engelsk forledd + engelsk etterledd: *land rover, best seller, western film, body lotion, live show, brass band, cover charge*

Hybride sammensetninger:

- 1) Norsk forledd + engelsk etterledd: *norges cup*
- 2) Engelsk forledd + norsk etterledd: *country musikk, fleece jakke, service næring*
- 3) Internasjonalismer: *design studio, video studio*

Dersom vi skal få mer kunnskap om motiveringen for særskriving av komposita, trenger vi grundige morfologiske, etymologiske analyser av feiltypene. Først må vi finne en måte å identifisere flest mulig av de feildelte kompositaene på, slik man er langt på vei til å klare i svensk. Deretter bør det foretas statistiske analyser av feiltypene i forskjellige teksttyper. Særskrivingsproblemet – i den grad det er et problem – kan ikke løses med moraliserende eller overflatiske vitser om forståelighet. Kan hende er selve skriftbildet blitt så dagligdags at det rent grafiske virker som en utfordrende variasjonsmulighet. Reklametekster og annen grafisk formidling kan tyde på det. Det faktum at særskriving er et lite problem i moderne tysk, som har samme typer sammensetninger som de skandinaviske språkene, men har beholdt stor forbokstav i substantiv, er også interessant å studere videre. Det fins mange språk som ikke har faste konvensjoner for orddeling, kan hende er de skandinaviske språkene i ferd med å utvikle liknende konvensjoner.

Litteratur

- Bakken, Kristin 1995: *Leksikalisering av sammensetninger. En studie av leksikaliseringprosessen belyst ved et gammelnorsk diplommateriale fra 1300-tallet*. Upublisert doktoravhandling. Oslo.
- Daher, Camilla 2004: *Fram tida vardags företeelser. En granskning av hur gymnasister särskriver*. Göteborg: Institutionen för nordiska språk, Göteborgs universitet. (Otryckt kandidatoppsats.)
- Hallencreutz, Katarina 2001: Skyll på längden, inte på engelskan. I: *Språkvård* 1/2001, 4-9.
- Hallencreutz, Katarina 2003: *Särskrivningar och andra skrivningar i elevspråk*. Uppsala.
- Linder, Nils 1886: *Regler och råd angående svenska språkets behandling i tal och skrift*. 2. uppl. Stockholm.
- Pedersen, Bolette Sandford 2005: *Datamatisk leksikografi i Norden – status og visioner*. I: Fjeld, R.V. og D. Worren (red): *Nordiske studiar i leksikografi 7, Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24 mai 2003*. Oslo, 302-314.
- Sakshaug, Laila 1992: *Særskriving av samansette ord i norsk*. Hovedoppgåve ved Universitetet i Trondheim.
- Tanums store rettskrivningsordbok*. 8. rev. utgave. Oslo: Kunnskapsforlaget 1996.
- Walmsness, Roar 2000: *Særskriving av sammensette ord. En studie av feiltypen hos grunnkurselever i Videregående skole*. I: *Norskraft* 6-2000.

Adverbiers syntaks i ordbøger

The lexical syntactic properties of Danish adverbs. This paper investigates the individual syntactic properties of Danish adverbs on the basis of recent Scandinavian grammar descriptions as well as extensive corpus investigations. The main goal is to clarify which types of syntactic information to describe in the lexical entries of adverbs, in a computational lexicon as well as in a learners' dictionary of Danish. On the basis of the investigation, a lexical description model for the word class is proposed and afterwards applied for the description of 217 monosemous adverbs. Five main syntactic groups of Danish adverbs are presented on the basis of the lexicon. But still we find lots of syntactic differences between lemmas belonging to the same group, even between synonymous lemmas which one might expect to have the same syntactic behaviour – this underlines that adverbs have very individual lexical properties.

1. Indledning

Hvilke individuelle leksikalske egenskaber har adverbier hvad angår syntaks? Hvilke positionsmuligheder i en sætning eller muligheder for at indgå i forskellige slags syntagmer har de for eksempel? Og hvad må nødvendigvis oplyses om det enkelte adverbium i et sprogteknologisk leksikon for at et grammatikprogram kan håndtere adverbiet syntaktisk korrekt? Adverbialer og dermed også adverbier er nemlig meget ofte årsag til syntaktisk tvetydighed i automatisk analyse af dansk. Ligeledes kan det være svært at placere dem korrekt ved automatisk produktion af danske sætninger. Dette problem opstår fx nemt i maskinoversættelse fra engelsk til dansk da placeringen af adverbier er forskellig i de to sprog. Sætningen *Mark rode a bicycle on the day of the transit strike willingly* (Ernst 2003) skal således ikke oversættes til **Mark cyklede på dagen for transport-strejken gerne*, men derimod til *Mark cyklede gerne den dag der var transportstrejke*. Også en person der ikke har dansk som modersmål, har brug for grammatiske oplysninger om adverbier for at kunne håndtere dem korrekt – men hvilke er leksikalsk betinget og derfor nødvendige at angive i en ordbog? Dette er det hovedspørgsmål der er søgt besvaret i et ph.d.-projekt finansieret af Nordisk Sprogteknologisk Forskningsprogram (NorFA) og afsluttet i foråret 2005. Ideen til emnet for afhandlingen opstod i opstartsfasen af STO-projektet (Braasch 2006) hvor en stor sprogteknologisk ordbase for dansk skulle opbygges. Det viste sig nemlig at man var helt på bar bund med hensyn til det nødvendige teoretiske grundlag for at beskrive adverbiers syntaks.

I danske ordbøger finder man indimellem eksempler på at leksikografer har observeret og ønsket at give informationer om syntaktiske særtræk ved det adverbium der beskrives. Fx kan man i *Politikens Nudansk Ordbog med etymologi* finde følgende om end noget upræcise beskrivelse af adverbiet *nok's* syntaks i visse

konstruktioner (understreget):

Nok ...

• {sjovt/..} **nok** udtryk for undrende eftertanke m.m.; '**nok**' står efter visse adverbier og gør at de kan stå andre steder i sætningen end ellers • underligt nok har jeg ikke hørt fra ham siden • jeg har sjovt nok også haft sådan en

Også i *Den Danske Ordbog* kan man finde syntaktiske oplysninger om et adverbium, fx for *alene*, bet. 1 (understreget):

alene² adv.

1 udtrykker en begrænsning som udelukker andre muligheder = udelukkende, kun [ofte efter nægtelse]; typisk: ~ af den grund, ~ fordi, ene og ~ □ *Konkurrencen gælder ikke alene æren men også på længere sigt en masse penge* AarhSt85

• udtrykker at der er tale om én og kun én mulighed = ingen anden end, intet andet end [NGN/NGT alene] □ *han alene havde besiddet den fornødne sjælsstyrke til at vriste sig fri* PRønJen89.

I *Macmillan English Dictionary for advanced learners* finder man også syntaktiske angivelser ved hørfrekvente adverbier, fx som her ved *enough* og *hardly*:

Enough can be used .. as an adverb (after an adjective, adverb or verb): *The rope isn't long enough* **She didn't move quickly enough* **You haven't practised enough*.

Hardly comes before the main verb of a sentence, but when there is a modal or auxiliary verb, **hardly** usually comes after it: *I can hardly breathe* **You have hardly done any work*. In stories and in formal English **hardly** is sometimes used at the beginning of a sentence before an auxiliary verb: *Hardly had she begun to speak, when there was a shout from the back of the hall*.

I ph.d.-projektet har jeg ønsket at få helt præcist overblik over hvad en leksikograf egentlig bør undersøge og beskrive for det enkelte danske adverbium når han/hun skriver sin ordbogsartikel eller opbygger en datamatisk ordbase. Samtidig har jeg ønsket at undersøge en stor gruppe adverbier for de egenskaber jeg konkluderede, var relevante, for at se om adverbier fordeler sig i veldefinerede syntaktiske klasser, eller om de snarere opfører sig meget individuelt. Endelig mente jeg også at det var interessant at undersøge om oplysninger i leksikonnet kan fjerne en del af problemerne med syntaktisk tvetydighed ved automatisk analyse af danske sætninger med adverbier.

I denne artikel gennemgås først en række væsentlige egenskaber ved adverbier, dels ud fra beskrivelser af dem i udenlandske sprogteknologiske ordbøger for adverbier og nyere skandinaviske grammatikker, dels ud fra korpusundersøgelser af

en række danske adverbiers syntaktiske opførsel. Ud fra gennemgangen opbygges en række træk med tilhørende værdier som ordklassen på en systematisk måde kan beskrives ud fra. Dernæst præsenteres nogle hovedgrupper som 217 monoseme adverbier der blev beskrevet ud fra disse træk, fordeler sig på. Samtidig vil jeg vise hvor individuelt de enkelte ord rent faktisk opfører sig.

2. Beskrivelsesmodel for adverbiers syntaktiske leksikalske egenskaber

Modellen til beskrivelsen af adverbiernes syntaks følger STO-leksikon-modellen (se Braasch 2006). Det vil sige at lemmaselektionen er foretaget ud fra Den Danske Ordbogs (2003) lemmakandidatliste¹ (1202 adverbier) med et krav om en frekvens på mindst 20 i Den Danske Ordbogs korpus på 40 millioner løbende ord (se www.dsl.dk). Dette betød at i alt 910 adverbier på lemmakandidatlisten skulle medtages i STO. Blandt disse 910 adverbier finder man selvfølgelig de ægte adverbier (fx *måske*, *nøje* og *altid*), men også visse præpositionsfraser (fx *i dag*, *i går* og *for øvrigt*) samt visse regelmæssige (-t) afledninger af adjektiver (fx *forskelligt*). Desuden finder man adverbier der samtidig er præpositioner (fx *på* og *i*). Det var også et krav i STO-projektet at den pågældende syntaktiske egenskab – fx at et bestemt præpositionssyntaxme var komplement til lemmaet – skulle være rimeligt repræsenteret i de korpusser som STO-projektet benyttede i det leksikalske arbejde (fortrinsvis Berlingske Aviskorpus (1999), men også Korpus 2000 blev konsulteret). Dette princip gælder også for beskrivelsen af adverbiers syntaktiske egenskaber.

I STO-leksikonnet udtrykkes et lemmas syntaktiske egenskaber som attribut-værdi-par i tabeller i en database. Leksikograferne anvender i det praktiske arbejde tekniske strengværdier, såkaldte deskriptorer. Et eksempel på en deskriptor er 'Dn2GP-af' for substantiver der syntaktisk opfører sig som fx lemmaet 'navngivning' (NGN's navngivning af NGN/NGT). D står for 'Deskriptor', 'n' står for 'noun', '2' står for antallet af komplement (subjekt og objekt), 'G' betyder at subjektet (eller objektet) for verbalsubstantivet kan være en genitiv, og endelig står 'P-af' for at der er tale om et objekt i form af en præposition 'af' og en styrelse i form af et substantiv. Deskriptorerne til beskrivelsen af adverbierne skal opbygges på lignende måde – der skulle altså laves et inventar af forskellige træk der kunne sættes sammen til forskellige deskriptorer efter behov. Leksikografen skal så tildele lemmaet netop den deskriptor der dækker det pågældende adverbiums syntaktiske egenskaber (eller udforme en ny hvis adverbiet opfører sig anderledes end de hidtil beskrevne lemmaer). Den information der ligger i hver type deskriptor, kan så efterfølgende omsættes til attribut-værdi-par i den databasestruktur som en sprogteknologisk ordbog er opbygget af – eller for den sags skyld omsættes til

¹ Lemmakandidatlisten blev stillet til rådighed af Det Danske Sprog- og Litteraturselskab og indeholder lemmaer markeret med ordklassen adverbium fra bl.a. *Gyldendals Store Dansk-Engelsk Ordbog* og *Retskrivningsordbogen*.

konstruktionsoplysninger i en almindelig ordbog. Men hvilke syntaktiske egenskaber skal disse adverbiumsdeskriptorer så udtrykke?

2.1. Syntaktiske egenskaber: Hvilket led modificerer adverbiet, og hvor står det i forhold til denne kerne?

Hovedegenskaben ved adverbier er hvilken type led det modificerer. Der skal derfor som udgangspunkt være en ny deskriptor for hver type syntagme der er mulig som kerne, og deskriptoren skal selvfølgelig angive hvilken type det drejer sig om. Svenska Akademiens Grammatik, herefter SAG (Teleman et al. bind 2 (1999:625)), nævner at adverbier kan modificere verbalfraaser, sætninger, adjektiver, adjektivfraser, participiumsfraser, adverbiumsfraser og substantivfraser. Norsk Referanse-Grammatikk, herefter NRG (Faarlund et al. (1997:451)), nævner desuden præpositionsfraser (*de bor næsten i centrum; Dette tilbudet gjelder bare i januar*) samt determinativer (*også den andre mannen, næsten alle nålevende mennesker*). I leksikonmodellen besluttede jeg, dels ud fra de grammatiske beskrivelser, dels ud fra korpusundersøgelser (udtræk på 100-120 tilfældige konkordanslinjer i Berlingske aviskorpus 1999) af 49 adverbier fra alle betydningsområder, at det i hvert fald var nødvendigt at beskrive de typer af modifikation der er nævnt i tabel 1.

Syntagmetype der modificeres af adverbiet	Lemma-eksempel	Eksempel: <i>adverbium modificeret led</i>
adverbium	<i>meget</i>	<i>Meget ofte</i> gik han i byen
‘nøgent’ adjektiv	<i>ganske</i>	<i>ganske stort</i> var huset
adjektivsyntagme, herunder participum	<i>sikkert</i>	Den <i>sikkert meget trætte</i> mand satte sig ned <i>Netop ankommet</i> gik han over til døren
præpositionssyntagme	<i>lige</i>	<i>Lige ved indgangsdøren</i> stod en stol
substantivsyntagme, herunder determinativ (da det ofte er syntaktisk tvetydigt)	<i>kun</i>	<i>Kun oplysning og information</i> hjælper <i>Også de</i> gamle mænd var trætte

Tabel 1.

Hvad angår modifikation på verbums- eller sætningsniveau, var der visse vanskeligheder. Togeby (2003) skelner fx her mellem fire niveauer. Han nævner at mådesadverbier udgør en del af de valensbundne led til verbet på samme måde som fx et objekt, og at det modificerer verbet inklusive dets øvrige valensbundne led (s. 50 og figurerne s. 51 samt 268). Retningsadverbier udgør derimod en inkorporeret del af selve verbet ifølge Togeby (2003:71). Derudover modificerer den gruppe som han kalder ‘udsagnsadverbialer’ (fx tids- og stedsadverbier), hele propositionen, hvorimod ‘ytringsadverbialer’ samt negationer og logiske operatører modificerer hvad

han kalder 'udsigelsens realitetsstatus' (2003:50). I den leksikalske beskrivelsesmodel besluttede jeg kun at skelne mellem tre niveauer, nemlig modifikation af selve det leksikalske verbum, hele verbalet eller hele sætningen, idet der til gengæld blev opstillet præcise syntaktiske kriterier for hvornår adverbiet modificerer hvilket af de tre typer syntagme. Adverbiet modificerer således selve verbet når det er i stand til at fungere som partikel i en partikelverbumbeskrivelse (*eleverne tog bøgerne frem; blomsten gik ud*). Det gør det også hvis det er i stand til at stå på en plads mellem det infinitive verbum og en sådan partikel (*han havde slået ordet omhyggeligt op i ordbogen*). Også adverbier der kan erstatte et præpositionsobjekt eller et bundet adverbial, modificerer selve verbet (*han havde selv tænkt derpå, han tog derhen*). Adverbiet modificerer derimod hele verbalet, dvs. verbet inklusive dets komplement, hvis det ikke opfylder betingelserne ovenfor, men stadig godt kan stå i en infinitiv-sætning (som fx *nu* kan: *at opgive projektet nu ville være ærgerligt*). Endelig modificerer det hele sætningen hvis det ikke kan stå i en infinitivsætning, som fx *faktisk* og *absolut* ikke kan (**at opgive projektet faktisk ville være ærgerligt* over for *projektet blev faktisk opgivet*; **at ville absolut i biografen er irriterende* over for *han ville absolut i biografen*).

2.2. Adverbets position i forhold til dets kerne

Ud over at man skal angive i ordbogen hvilket eller hvilke led adverbiet kan modificere, er det også nødvendigt at informere om adverbets positionsmuligheder i forhold til det/de pågældende led. Her er det igen nødvendigt at skelne mellem adverbier på sætningsniveau og adverbier der modificerer øvrige syntagmetyper. I de sidste tilfælde står adverbiet altid enten før eller efter sin kerne. Hvis man ser på tilfælde hvor det står før, finder vi nemt eksempler i dansk når kernen er et adjektiv eller participium (*meget glad; vel ankommet*), et adverbium (*næsten aldrig*), et substantivsyntagma (*kun mændene var til stede*) og et præpositionssyntagma (*lige ved indgangsdøren*). SAG (bind 4:154) nævner derudover at visse typer adverbier der denoterer approksimation, kan præmodificere finite verber: *Han {nästan/så gott som/likesom/mer eller mindre/praktisk taget} somnade under föreläsningen*. Dette gælder også for dansk (*man nærmest freder etaten i de år rammen omfatter; begejstringen for Lis Sørensen næsten væltede huset; De [: cafépersonalet] bare lissom ved med sig selv, at dette er kun et job, de har, til de om senest trekvart år står foran eller bagved kameraet* (alle Berl99)).

SAG (bind 4:160) nævner også at visse bisætningstyper, nemlig de nominelle og de adverbelle (men fx ikke de relative), kan præmodificeres af et adverbium. Dette gælder også for dansk: *Denne proces er meget omstændelig, især når der er flere lande indblandet; Tilgang af nye officerer er nødvendig, også hvis forsvaret skal beskæres, for at sikre en jævn aldersfordeling* (begge Berl99). Endelig nævner SAG at adverbier kan modificere infinitive verbalsyntagmer. Dette gælder også for dansk: *Alene at*

sende en helikopter i luften koster omkring 20.000 kr.; Han fortsatte arbejdet med især at værne om fællesskabets værdier (begge Berl99).

Det er altså nødvendigt også at kunne beskrive i ordbogen om et adverbium kan præmodificere et finit verbum, en bisætning (med forskellige bisætningsindledere) samt et infinit verbalsyntagne. De såkaldte 'fokuserende' adverbier som de netop nævnte adverbier er eksempler på, bliver i det hele taget behandlet indgående i SAG (se fx bind 4:152). Betydningsmæssigt fremhæver denne type af adverbier selve det led der modificeres, idet der sammenlignes med en anden mængde i konteksten eller diskursen. Selvom fokuserende adverbiers prototypiske placering i svensk er foran det led der modificeres, skriver SAG (bind 4:156) at '*De flesta fokuserenda adverbial kan stå postfokalt, dvs. omedelbart efter det fokuserade led*'. I dansk er dette ikke nær så udbredt som i svensk, og jeg vil derfor ikke komme nærmere ind på alle de svenske tilfælde her, men blot liste de forskellige typer af dansk postmodifikation, både mht. fokuserende adverbier og andre typer af adverbier:

Adjektiv: *at være træt nok*

Adverbium: *at løbe hurtigt nok*

Præpositionssyntagma: *I Asien derimod er der udsigt til stabilisering*

Substantivsyntagma: *Nyheden alene vil ikke kunne skabe et blad; Alle borde udenfor var optaget*

Det der i visse tilfælde opfattes som postmodifikation på sætningsniveau i SAG, vil jeg opfatte som et topologisk fænomen der bør beskrives ud fra positionsmuligheder i et sætningsskema. Både SAG, NRG og Togeby (2003) beskriver positionsmuligheder på sætningsniveau og er enige om at et adverbium kan stå både i initialfeltet (fundamentfeltet), i midtfeltet (neksusfeltet) samt i slutfeltet i sætningen, men ingen af grammatikkerne er helt udtømmende mht. positionsmulighederne. Togeby opererer godt nok med en særlig plads til mådesadverbier samt med en fælles plads for retningsadverbier og præpositionelle objekter, men angiver i øvrigt ikke i detaljer hvordan adverbierne i slutfeltet fordeler sig på forskellige pladser hvis der både er et adverbium som mådesadverbial, et som partikel, et som retningsadverbial og et som frit adverbial i slutfeltet. NRG opfatter eventuelle adverbier på en plads foran objekter som en flytning fra den 'rigtige' plads i slutfeltet fordi man i norsk sjældent har adverbier på denne plads. På linje med SAG betragter NRG partikler i partikelverbalskonstruktioner som præpositioner og ikke adverbier. I svensk har man slet ikke den plads der i dansk svarer til mådesadverbialpladsen.

Hansen og Heltoft (endnu ikke udkommet) beskriver derimod meget detaljeret hvorledes adverbierne fordeler sig topologisk i en dansk sætning hvis alle muligheder udnyttes i en såkaldt maksimalsætning. Tabel 2 illustrerer pladserne ud fra min eksempelsætning: *Egentlig havde du vist endnu ikke slået dig ordentligt ned derovre dengang.*

Egent- lig	havde	du	vist endnu	ikke	rigtig	slået	dig	ordentligt	ned	derovre	dengang
funda- ment- felt			sætningsadv. og/eller frit adv. i tema- position	negation	adver- bium i fokus- position			mådes- adv.plads	parti- kel- plads	bundet adv. plads	slutfelt
neksusfelt											

Tabel 2.

Disse positionsmuligheder skal altså angives for hvert V, VP- eller S-modificerende adverbium i ordbogen:

1. Fundamentfeltet: *Her har jeg placeret et bord; Desværre gik han hjem.*
2. Temaposition i neksusfeltet (positionen til venstre for negationen): *Han har igen ikke afleveret brevene til tiden* (jeg skelner i modsætning til Hansen og Heltoft ikke mellem sætningsadverbium og frit adverbium da det ikke er muligt at lave et syntaktisk defineret skel).
3. Fokuspositionen i neksusfeltet (positionen til højre for negationen): *Der har ikke her været plads til at gentage den skildring, jeg tidligere har givet* (Ber199).
4. Mådesadverbialpladsen mellem objekt og partikel: *Han havde ryddet værelset omhyggeligt op.*
5. Partikelpladsen, efter mådesadverbialpladsen og før bundne adverbialer: *Han sendte brevene hurtigt ud til kunderne; Han læste omhyggeligt op af Bibelen.*
6. Fællespladsen for bundne adverbialer (inklusive præpositionsobjekter), efter en eventuel partikel i partikelverbsumkonstruktioner: *Han læste omhyggeligt højt derfra; han tænkte længe derover; han gik hurtigt hjem derfra.*
7. Slutfeltet, efter alle bundne adverbialer: *Han var gået hurtigt hjem derfra forleden.*

2.3. Øvrige syntaktiske egenskaber

Ud over modifikations- og positionsmulighederne er der også andre syntaktiske egenskaber for adverbier som er relevante at beskrive i ordbogsindgangen. Hvis man ser på de informationer der er medtaget for adverbier i det spanske PAROLE²-leksikon (som ud over det italienske var det eneste af PAROLE-leksikaene med detaljerede syntaktiske informationer for adverbier (Internettet (2005, a og b)), er mange af de samme oplysningstyper også relevante for danske adverbier. Fx er det for de adverbier der kan modificere hele sætningen, relevant at beskrive om det kan stå i spørgende sætninger eller ej, eller om det måske ligefrem kræver en spørgende

² PAROLE var et morfologisk og syntaktisk ordbaseprojekt for 12 EU-sprog.

sætning. Adverbierne *hellere* og *ligeledes* kan således godt optræde i spørgsmål (*vil du hellere komme senere?; er den anden væg ligeledes dækket af reoler?* (Berl99)), hvorimod adverbierne *desværre* og *sandelig* ikke kan (**kom han værre for sent?; *har han sandelig ret?*).

Det spanske PAROLE-leksikon indeholder også information om hvorvidt et adverbium kan topikaliseres i en sætningskløvning. Dette fænomen beskrives også i Togeby (2003:269) samt i SAG (bind 4:151) og NRG (1090). Fx optræder adverbierne *derfor* og *dengang* ofte i sætningskløvninger (*det er derfor han gør det; det var dengang han opdagede det*) hvorimod adverbierne *desværre* og *eksempelvis* ikke kan indgå i en sådan konstruktionstype.

Om adverbiet selv kan modificeres af et andet adverbium, dvs. om der syntaktisk set er mulighed for et adled til adverbiet, skal også angives i leksikonindgangen (se fx SAG (bind 2:637)). Således kan *nøje* modificeres af fx *meget*, hvorimod man ved korpusundersøgelser kan konstatere at et adverbium som *ekspres* ikke forekommer med noget adled.

Zinsmeister og Heid (2003) beskriver hvordan det i tysk er leksikalsk betinget om et adverbium kan udgøre prædikativet i en prædikativkonstruktion. Det er det også i dansk, fx har adverbierne *hjemme* og *undervejs* denne evne (*han er hjemme; han er undervejs*), hvorimod et adverbium som *ihjel* ikke har den (*han frøs ihjel og han kedede sig ihjel*, men ikke **han var ihjel*).

De førnævnte korpusundersøgelser viste desuden tydeligt at nogle adverbier stiller særlige krav mht. negativ/spørgende eller positiv sætningskontekst. Fx kan *nogensinde* kun optræde i negativ/spørgende sætningskontekst, hvorimod *bestandig* kun optræder i positiv sætningskontekst (negativ sætningskontekst defineres her som tilstedeværelsen af *ikke, ingen, ingenting, intet* eller *aldrig* i sætningen, men kan dog også forekomme med andre begrænsende udtryk, fx *næppe, uden* etc.). Selv når et adverbium modificerer fx et adjektiv, kan det stille et sådant krav. Fx kan et adjektivsyntagme med adledet *synderlig* kun optræde i negativ sætningskontekst, som fx i sætningen *han var ikke synderlig imponeret* (Berl99).

Det spanske PAROLE-leksikon indeholder information om eventuelle komplementer til adverbiet. Også SAG (bind 2:637; bind 3:635) beskriver hvordan adverbier kan have en efterstillet bestemmer i form af et bundet adverbial (*tillsammans med dem, mindre end den där*). Denne type information skal også med i leksikonnet da dette også gælder danske adverbier, idet jeg også regner ikke-bundne led for komplementer hvis dette led ikke kan stå alene. Adverbiet *desværre* tager således et komplement i form af et fakultativt præpositionssyntagme indledt med *for* (*desværre for ham var hun gået*), og adverbiet *væk* tager ligeledes et fakultativt præpositionssyntagme, i dette tilfælde indledt med *fra* (*han gik væk fra vejen*).

Når et adverbium efterfølges af et andet tæt relateret adverbium, nævner SAG at der normalt er to mulige analyser: Enten opfattes det sidste adverbium som værende

komplement til det første, eller også kan de sammen analyseres som en koordineret adverbiumsfrase. Konstruktionen *Här där jag står* er et eksempel på dette (SAG, bind 3:636). På dansk har man et eksempel som *han står her udenfor*. I leksikonnet vil det sidste adverbium blive beskrevet som et komplement, men kun hvis det er hyppigt forekommende i korpus. Ifølge SAG (bind 3:638) kan adverbier i svensk tage følgende tre typer af komplementer:

- a) underordnet syntagme indledt med *än* eller *som*: *mer än, så (bra) som, lika (bra) som; hon sprang fort som vinden* (svarende til dansk *mere end; lige så god som; hun løb hurtigt som vinden*);
- b) præpositionssyntagme: *Vi har för höga skatter för att dom ska trivas här* (svarende til dansk *vi har for høje skatter til at de kan trives her*) og
- c) underordnet sætning: *Vi har äntligen kommit därhän att vi kan börja förhandla* (svarende til dansk *vi er endelig kommet derhen hvor vi kan begynde at forhandle*).

I tabel 3 gives en samlet oversigt over hvilke komplementtyper danske adverbier, i lighed med de svenske, kan tage:

Komplementtype	Indledt af	Eksempel
adverbium	%	<i>lidt væk derfra lå et andet hus</i>
underordnet frase	end, så X som, som	<i>Han løb hurtigere end før, huset var lige så stort som det andet,</i>
underordnet sætning	hvor, når..	<i>berinde hvor jeg står er der varmt nu når du siger det er her faktisk varmt</i>
præpositionssyntagme med substantiv som styrelse præpositionssyntagme med infinitivsætning som styrelse	til, med, af etc.	<i>han gik sammen med hende mest af alt er jeg træt heldigvis for hende gik han der er penge nok til at købe pc'en</i>

Tabel 3.

Tabel 4 viser en opsummering af alle de typer af syntaktiske egenskaber som bør medtages i et syntaktisk leksikon for adverbier. I højre kolonne ses den forkortelse for egenskaben som skal indgå i deskriptorernes opbygning, jf. kapitel 2.

Informationstype	Muligheder	Værdi
Hvilken type syntagme modificeres?	adjektiv adjektivsyntagme adverbium præpositionssyntagme substantivsyntagme verbum verbalsyntagme sætning	_ADJ _ADJP _ADV _PP _NP _V _VP _S
I hvilken position står adverbiet i forhold til dette syntagme?	Hvis syntagmet er ADJ, ADV, PP el. NP: præ- eller postmodifikation (eller begge muligheder, indikeret med +Post) Hvis syntagmet er V, VP el. S: præmodifikation af finit verbum præmodifikation af bisætning indledt med X præmodifikation af infinitivsætning både præmodifikation samt mulig i slutfeltet af infinitivsætning Pladser på sætningsniveau: fundament neksusfelt både før og efter negation kun temaplads i neksusfelt kun fokusplads i neksusfelt mådesadverbialplads partikelplads bundet adverbialplads slutfelt	Præ Post +Post Præfinit PræsubCl-når/ hvis.. finf +finf F N Te Fo Må P BA S
Kan adverbiet modificeres af et andet adverbium, fx af 'meget', 'næsten' eller 'lidt'?	ja nej	m %
Er adverbiet i stand til at optræde i kløvning?	ja nej	ExC %
Er det i stand til at udgøre prædikativet i en prædikativkonstruktion?	ja nej	pred %
Stiller adverbiet særlige krav til konteksten?	positiv negativ nej	pos neg %

Tager adverbiet et (evt. obligatorisk) komplement?	nej komplementet er obligatorisk komplementet er - et andet adverbium - en underordnet frase indledt af X, fx 'end' - en underordnet sætning indledt af X, fx 'når' - en præposition (fx 'med') + substantiv - en præposition (fx 'til') + at-inf	% o -adv -sp-end -sc-når -P-med -Pi-til
Kan et S-modificerende adverbium optræde i spørgsmål?	ja, også i spørgsmål ja, kun i spørgsmål nej, ikke i spørgsmål	% oblQ notQ

Tabel 4.

2.4. Leksikalsk beskrivelsesmodel

Listen over egenskaber samt de tilhørende mulige værdier i tabel 4's højre kolonne var udgangspunktet for en leksikalsk beskrivelsesmodel som 217 lemmer i ph.d.-projektet blev beskrevet ud fra. Hvert adverbium fik tildelt de egenskaber, dvs. de værdier fra højre kolonne i tabel 4, der var relevante for netop det lemma (i en bestemt rækkefølge, som jeg ikke vil komme nærmere ind på her). Således blev de nødvendige deskriptorer som tidligere nævnt faktisk opbygget undervejs i arbejdet.

Et eksempel på en deskriptor er 'Dd1FTe_S'. 'Dd1' angiver at der er tale om en Deskriptor for et adverbium med aritet 1 (dvs. at det ikke tager noget komplement), 'F' angiver at adverbiet kan stå i fundamentfeltet forrest i sætningen. 'Te' angiver at det kan stå før nægtelsen i neksusfeltet, og endelig angiver '_S' at adverbiet knytter sig til hele sætningen. Deskriptoren 'Dd1FTe_S' dækker fx adverbierne *anstændigvis* og *endvidere*'s egenskaber: Ingen af dem kan optræde i et infinitivsyntagme, derfor knytter de sig til hele sætningen S (og ikke til VP). De er begge mulige i fundamentfeltet (*Endvidere har orkestret optrådt på en række af de store jazz-festivaler; Anstændigvis må vi [kræve, at forsvarer] må have det bedste udstyr* (begge Berl99)). De er ligeledes begge mulige før en evt. nægtelse i neksusfeltet (*Retten fandt det endvidere ikke godtgjort, at træet [medførte gener]; Venstre har anstændigvis ikke tænkt sig at pille ved [forliget]* (begge Berl99)). Hver syntaktisk indgang blev desuden forsynet med et citat. De 217 lemmer var i øvrigt alle monoseme (i *Politikens Nudansk Ordbog med etymologi*) for i første omgang at undgå at blande problemer med flertydighed ind i analyserne. I tabel 5 ses et lille udsnit af leksikonnet.

allernådigst_1	Dd1posN_S	<i>Hans arbejdsgiver havde allernådigst givet ham lov til at spise og sove lidt</i>
alligevel_1	Dd1FTeS_V P	<i>Alligevel er han imod forslaget</i>
almindeligt_1	Dd1_ADJ	<i>nøgenbadning er tilladt og almindeligt accepteret, men familien var påklædt</i>
almindeligvis_1	Dd1FN_S	<i>Vi udvikler jo almindeligvis ikke autoimmune sygdomme under en infektion</i>
almindeligvis_2	Dd1_ADJP	<i>Den almindeligvis moderate russiske politiker skriver også, at han i givet fald vil anbefale [at] standse al militær-teknologisk samarbejde med USA</i>
alternativt_1	Dd1posFN_S	<i>Alternativt blendes det hele i en køkkenmaskine Boligselskabet kan alternativt købe sig fri for hjemfaldspligten</i>

Tabel 5.

Generelt var erfaringen fra arbejdet med at beskrive adverbierne i øvrigt at introspektion ikke fungerer ret godt for adverbier. Det viser sig også helt generelt ved at adspurgte informanter har meget forskelligt syn på hvad et givet adverbium 'kan' syntaktisk set. Korpusundersøgelser har derfor været helt uundværlige. Nogle gange var jeg dog nødt til at ty til introspektion, fx når jeg skulle afgøre om adverbiet tager verbalet eller hele sætningen som sin kerne, idet der sjældent er korpusforekomster af adverbier i infinitivsyntagmer. Det var heller ikke altid nemt at beslutte om en bestemt opførsel var prototypisk for et adverbium eller for sjældent forekommende i korpus til at blive beskrevet. Fx vurderede jeg at adverbiet *overhovedet ikke* kan stå efter *ikke* i neksusfeltet, idet der kun er 3 *ikke overhovedet* over for 725 *overhovedet ikke* i Berlingske korpus 1999. Derimod får *nødvendigvis* tildelt den mulighed at kunne stå før nægtelsen i neksusfeltet i leksikonnet fordi der trods alt er 18 forekomster af *nødvendigvis ikke* (godt nok over for 986 forekomster af *ikke nødvendigvis*), fx *Det handler nødvendigvis ikke bare om at skaffe flere penge* (Berl99). Man kan dog måske i virkeligheden være i tvivl om hvorvidt disse 18 forekomster overhovedet er korrekt sprogbrug. Dette leder frem til endnu et problem med hensyn til det leksikalske arbejde med adverbierne: Selvom noget findes i et korpus, er det da overhovedet korrekt? Fx finder man følgende to eksempler på *muligvis* og *velsagtens* i Berlingske Aviskorpus 1999 som i mine øjne ikke er korrekt sprogbrug (det sidste eksempel er nok en dårlig oversættelse fra svensk):

- a. ? *Velsagtens er det kun et spørgsmål om kort tid;*
- b. ? *Og Dagens Nyheter siger om stykket: Ikke noget mesterværk, men det indeholder en række paradokser og problemstillinger til muligvis at blive provokeret af, men til endnu mere at tænke over.*

Ud over at være et skridt på vejen til et fuldt leksikon med ca. 900 lemmaer gav beskrivelsen af de 217 adverbier et godt overblik over hvilke syntaktiske grupper danske adverbier kan inddeles i, samt hvor forskellige adverbierne rent faktisk er indbyrdes. Det vil jeg komme ind på i næste afsnit.

3. Syntaktiske hovedgrupper samt leksikalske forskelle

De 217 lemmaer udmøntede sig i i alt ca. 400 syntaktiske indgange idet et lemma, som tidligere nævnt, jo skulle tildeles en deskriptor for hver type af led det kunne modificere. Men rent faktisk er der temmelig mange adverbier der kun kan modificere én enkelt ledtype – nemlig næsten 50 % af lemmaerne (101/217). Dette er meget brugbar viden i datamatisk sprogbehandling idet en enkelt oplysning i leksikonnet for en stor del af lemmaernes vedkommende således gør det helt ukompliceret at fastsætte hvilket syntagma adverbiet skal tilknyttes i den automatiske syntaksanalyse.

Hvis man ser på den halvdel af adverbier der modificerer mere end en type led (116/217), opstår der nogle forholdsvist tydelige hovedgrupper, vel at mærke når man ser på præcis hvilke typer af led de modificerer. Det skal dog samtidig understreges at der er mange smågrupper af adverbier som falder helt uden for hovedgrupperne, og som jeg ikke kommer ind på her.

En af hovedgrupperne er de adverbier der knytter sig til hele sætningen, og som samtidig kan knytte sig til en adjektivfrase. Sådanne adverbier udtrykker typisk enten talerens holdning til resten af sætningen (fx *måske*), knytter to sætninger sammen (fx *ligeledes*) eller har en tidsbetydning (der kræver tempus udtrykt i verbet, fx *hidtil*). Et eksempel er *desværre* (*Der har desværre været meget stille omkring den senere tids overgreb i Kina; Jeg nåede alt i en desværre for tidlig afsluttet karriere* (begge Ber199)). Men der er mange undtagelser fra hovedgruppens prototypiske syntaktiske modifikationspotentiale. Fx kan nogle adverbier i gruppen modificere yderligere ledtyper. Et eksempel på det er adverbiet *formentlig*.

S: *Vi har formentlig kun set toppen af isbjerget* (Ber199)

ADJP: *Den formentlig ligeledes optøede fiskefilet...* (Ber199)

NP: *Formentlig et andet forhold gør, at en række østeuropæiske mærker også trygt kan stå på gaden* (Ber199)

PP: *Formentlig for første gang herhjemme, blev en reklamespot sendt ud i æteren* (Ber199).

En anden stor gruppe er de adverbier der knytter sig til adjektiver samt adverbier – de har typisk grads betydning. Denne gruppe er i øvrigt velbeskrevet i dansk grammatik. Et eksempel er *temmelig* (*Der var temmelig aggressive opkøb af euro lige efter åbningen; Den forbruger, der gemmer sine kvitteringer, når temmelig hurtigt op til denne grænse* (begge Ber199)). Men heller ikke her opfører adverbierne sig

systematisk ens, idet der er eksempler på at adverbier med grads betydning kun kan modificere adjektiver og ikke adverbier, fx *svært* (*Til gengæld er de svært tilfredse på Københavns Kommunes parkeringskontor* (Berl99); men **Han snakkede svært hurtigt*).

En tredje hovedgruppe er de adverbier der knytter sig til selve verbet på mådesadverbialpladsen, og som samtidig modificerer et adjektiv (der ofte er afledt af et participium). Et eksempel er *bredt* (*Han kommer bredt ud med sine synspunkter; [Man ønsker at] at iværksætte en bredt anlagt aktion* (begge Berl99)). Disse adverbier udtrykker mådesbetydning, men igen er der adverbier der kun kan modificere en af ledtyperne selvom de faktisk udtrykker måde; det gælder fx *forlæns* og *surt*:

Forlæns: V: *Kun hvis jeg vil pine mig selv går jeg forlæns ned* (Berl99); men ikke ADJ: **En forlæns gående mand*;

Surt: ADJ: [*Det overgår langt*], *hvad statens kommissærer tør rage til sig af skatteydernes surt indtjente midler* (Berl99); men ikke V: **Han har tjent pengene surt ind*).

En fjerde stor hovedgruppe er de adverbier der modificerer selve verbet og/eller hele verbalet, og som samtidig kan stå efter et substantivsyntagme samt optræde i en sætningskløvning. De har steds- eller tidsbetydning. *Derhjemme* er et godt eksempel:

V: *Visaet var allerede udstedt og lå derhjemme; Mine forældre vil helst have, at jeg er derhjemme* (begge Berl99).

VP: *Derhjemme sover konen og børnene stadig* (Berl99)

NP: *Konen og de to døtre derhjemme plager livet af ham* (Berl99)

Kløvning: *Det er heller ikke derhjemme, man giver plads til sin begejstring* (Berl99).

Men det er langt fra alle V- og VP-tilknyttede adverbier der følger dette prototypiske mønster, således fx ikke *nogensinde* og *tit*. *Nogensinde* kan godt optræde efter en NP, men kan ikke optræde i kløvning, hvorimod det omvendte gælder for *tit*:

Nogensinde: NP: *det er såmænd blot en måned siden, jeg fik den tætteste klipning nogensinde* (Berl99); men ikke kløvning: **Det var (ikke) nogensinde (at) jeg havde set det*

Tit: Kløvning: *Det er tit, at vores klienter er uspendende og bare slasker rundt* (Berl99); men ikke NP: **Opmærksomheden tit er et problem for ham*.

Den sidste store hovedgruppe af adverbier der umiddelbart kan udledes af de foreløbig 217 beskrevne adverbier i leksikonnet, er de adverbier som også kan modificere mere end to typer led, men som ikke hører under den forrige gruppe af V- og/eller VP-modificerende adverbier. Af de mange led de modificerer, er det ene ofte et NP eller et VP i foranstillet position. Det drejer sig om de tidligere omtalte fokuserende adverbier (SAG, bind 4:152). *Alene* er et godt eksempel på denne type

adverbium – det modificerer både NP'er, VP'er, PP'er og andre adverbier:

NP: *Alene hans udseende kvalificerer til denne status* (Berl99)

VP: *Alene at vove forsøget virkede dumdrigt* (Berl99)

PP: *Alene i Danmark er der næsten sket en fordobling af patienter* (Berl99)

ADV: *Alene derfor kan det ikke nytte noget* (Berl99).

Inden for gruppen af fokuserende adverbier er der dog stor variation mht. præcis hvilke typer af led adverbiet kan modificere. Dette gælder også for adverbier med synonyme betydninger (ifølge *Den Danske Ordbog*), fx *udelukkende* og *alene*. Begge kan modificere NP'er, VP'er og bisætninger, men *udelukkende* kan derudover også modificere adjektiver, hvor *alene* i stedet kan modificere PP'er og adverbier:

ADJ: *En ny og udelukkende (*alene) europæisk alliance bliver født ud af den gamle*

PP: *Alene (*udelukkende) i Danmark er der næsten sket en fordobling af patienter*

ADV: *Alene (*udelukkende) her i USA ville det redde flere tusind menneskeliv om året.*

Selvom man umiddelbart kan udlede disse 5 hovedgrupper af leksikonnet, er det samtidig meget tydeligt at de enkelte ord inden for hver gruppe ofte har deres eget sæt af modifikationsmuligheder. Hvis man yderligere ser på andre træk end netop adverbets modifikationsmulighed, er der også mange forskelle fra ord til ord inden for alle grupperne – og også blandt umiddelbart synonyme adverbier. Fx er der mange særtræk med hensyn til positionsmuligheder. To adverbier med meget lignende betydninger, *primært* og *fortrinsvis*, udviser fx helt forskellig opførsel mht. om de kan stå efter negation, idet der i Berlinske Aviskorpus 1999 er 36 forekomster af *ikke primært*, hvorimod der er 0 forekomster af *ikke fortrinsvis*. Man kan derfor langt fra altid erstatte *fortrinsvis* med *primært* (fx *Men skolen er ikke primært en undervisningsanstalt* (Berl99); men ikke **Men skolen er ikke fortrinsvis en undervisningsanstalt*). Andre eksempler på forskelle i positionsmuligheder er fx adverbiet *alene*, der godt kan stå efter det led der modificeres (*3.500 skal fyres i Sverige alene* (Berl99)), hvorimod to umiddelbart synonyme adverbier ikke kan (*3.500 skal fyres i Sverige *kun/ *udelukkende*). Og de fleste stedsadverbier kan stå i neksusfeltet, men ikke alle kan. Fx kan *overalt* godt (*Man har overalt fundet spor af køkkenaffald* (Berl99)), men *udenfor* og *derovre* kan ikke (*Man har *udenfor/ *derovre fundet spor af køkkenaffald*). Ligeledes kan mange adverbier stå i fundamentfeltet, men igen er det langt fra alle: *Beständig fokuserer man på det forkerte* (Berl99), men ikke **Evindeligt fokuserer man på det forkerte*; og *Formentlig har ingen læsere lagt mærke til det* (Berl99), men ikke **nok/*sikkert har ingen læsere lagt mærke til det*.

Også mht. krav til positiv eller negativ/spørgende kontekst finder man leksikalske forskelle på adverbier med synonyme betydninger. Fx finder man som tidligere

omtalt ingen korpuseksempler på *atter* i negativ eller spørgende kontekst, men masser af eksempler på *igen* (*Vore partnere ved nu, at man igen (*atter) ikke ved, hvor man har Danmark; Bliver Helmut Kohl igen (*atter) valgt som forbunds-kansler i Tyskland?* (begge Ber199)).

Alt i alt må man, når man ser på de leksikalske indgange, konkludere at der ganske vist er visse syntaktiske hovedgrupper af adverbier, men at der altid er ord der opfører sig anderledes end hvad der i øvrigt er prototypisk for gruppen. Der er faktisk stor variation fra ord til ord inden for alle grupper – også for ord med synonyme betydninger, fx mht. præcis hvilke typer af led adverbiet kan knytte sig til, præcis hvilke positionsmuligheder det har, og om det fx kræver positiv eller negativ kontekst.

4. Konklusion

Konklusionen på det leksikalske arbejde der er udført for adverbier i ph.d.-projektet, er helt klart at selvom der syntaktisk set kan udledes visse prototypiske hovedgrupper af adverbier, udviser næsten hvert eneste adverbium særlig opførsel mht. en eller flere af de undersøgte egenskaber. Man er derfor nødt til at beskrive adverbiers syntaktiske egenskaber forholdsvis detaljeret for hvert ord, både i et sprogteknologisk leksikon og i almindelige ordbøger, hvis en computer eller en person der lærer dansk som fremmedsprog, skal kunne håndtere adverbier korrekt. Selve metoden med at studere og sammenligne de fire nyere skandinaviske grammatikkers beskrivelse af adverbier viste sig i øvrigt meget udbytterig fordi de angreb adverbier fra meget forskellige vinkler og på den måde supplerede hinanden godt. Fremgangsmåden kan derfor absolut anbefales som metode til andre sproglige undersøgelser inden for skandinaviske sprog. Til sidst vil jeg gerne endnu engang understrege at korpusundersøgelser er fuldstændig uundværlige i afdækningen af adverbiers opførsel da man ifølge min erfaring sjældent er i stand til introspektivt at afgøre hvad der er udbredt i sproget, og hvad der ikke er, med hensyn til denne ordklasses syntaktiske potentiale.

Litteratur

Ordbøger:

- Den Danske Ordbog*. Bind 1, 2, 3, 4 og 5 2003-05. Ebba Hjorth & Kjeld Kristensen (red.). København: Det Danske Sprog- og Litteraturselskab. Gyldendal.
- Gyldendals Store Dansk-Engelsk Ordbog* 2001. Hermann Vinterberg & C.A. Bodelsen. København: Gyldendal.
- Macmillan English Dictionary for advanced learners* 2002. Oxford: Macmillan Publishers Limited.
- Politikens Nudansk Ordbog med Etymologi. 1. udgave (elektronisk version) 1999. København: Politikens Forlag.
- Retskrivningsordbogen* 1986. København: Dansk Sprognævn.

Korpusser:

- Berl99 = *Berlingske Aviskorpus*, ca. 30 mio. ord, Berlingske Tidende og Weekendavisen (årgang 1999). København.
- Korpus 2000*: www.korpus2000.dk. København: Det Danske Sprog- og Litteraturselskab.

Anden litteratur:

- Braasch, Anna 2006: Den danske Sprogteknologiske Ordbase og dens anvendelse i værktøj til leksikografiske formål. I denne rapport.
- Ernst, Thomas 2003: Semantic Features and the distribution of adverbs. I: Lang et al. (eds.): *Modifying Adjuncts*. Berlin: Mouton de Gruyter.
- Hansen, Erik og Lars Heltoft (endnu ikke udkommet): *Grammatik over det Danske Sprog*. Herunder Kapitel 18: Hierarki og ledstilling. Foreløbig version.
- Internettet 2005 a: www.ub.es/gilcub/SIMPLE/reports/parole/parole_IT.rtf Italian Lexicon Documentation.
- Internettet 2005 b: www.ub.es/gilcub/SIMPLE/reports/parole/parole_es.html Spanish Lexicon Documentation.
- RG = Faarlund, Jan Terje, Svein Lie og Kjell Ivar Vannebo 1997: *Norsk Referansegrammatik*. Oslo: Universitetsforlaget.
- SAG = Teleman, U., Hellberg, S. & Andersson, E. 1999: *Svenska Akademiens Grammatik*. Stockholm: Svenska Akademien.
- Togeby, Ole 2003: *Fungerer denne sætning? Funktionel dansk sproglære*. København: Gads Forlag.
- Zinsmeister, Heike & Ulrich Heid 2003: Identifying predicatively used adverbs by means of a statistical grammar model. I: 'CL2003, Proceedings'. Lancaster.

Leksikalsk beskrivelse af adverbiers semantik i norsk, svensk og dansk – LEXADV

A lexical semantic description of adverbs in Danish, Norwegian and Swedish – LEXADV. The LEXADV-project is a Scandinavian research project with the aim of extending the three Scandinavian semantic lexica building on the SIMPLE lexicon model with the word class of adverbs. In order to do so, a classification covering the many different meanings that adverbs can have, is established as well as a set of formal relations and features. In this paper we describe in detail the ontology and how three groups of adverbs are described in the lexicon, namely degree adverbs, time adverbs and modal adverbs. The formal description of each lemma is based on the definition in the recently published Danish dictionary *Den Danske Ordbog*. Afterwards the lemma is translated into Swedish and Norwegian and copies of the Danish lexical entry are created. Finally these copies are evaluated and adjusted to the target language if necessary. Some of the translation problems that arise during this process are discussed in the paper.

1. Indledning

LEXADV er et skandinavisk forskningsprojekt der er finansieret af Nordplus Sprog (Nordens Sprogråd), og som løber i perioden 1. september 2004-31. december 2005. De tre partnere i projektet er Språkdata, Institutionen för Svenska språket ved Göteborgs Universitet, Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo, samt Center for Sprogteknologi på Københavns Universitet. Projektet har som mål at udvide en fælles dansk-norsk-svensk sprogteknologisk ordbog (SIMPLE) med semantiske oplysninger for ordklassen adverbier, organiseret i en databasestruktur (ordbase). Det er udsprunget dels af samarbejdet i Skandinavien om SIMPLE-ordbaserne i netværket SPINN (Pedersen et al. 2005), dels af et ph.d.-projekt om danske adverbier (Nimb 2006). SIMPLE-ordbaserne for dansk og svensk blev opbygget i et EU-projekt (1998-2000) og indeholder for hvert sprog 10.000 semantiske indgange med formaliserede betydningsbeskrivelser for substantiver, verber og adjektiver – men ikke for adverbier. Universitetet i Oslo har efterfølgende udviklet en norsk SIMPLE-ordbog på basis af det etablerede danske og samtidig sørget for at linke en stor del af indgangene til den svenske SIMPLE-ordbog. På denne måde har man opbygget en sammenkædet sprogteknologisk ordbase for dansk, norsk og svensk, hvor på nuværende tidspunkt ca. 5000 indgange er linket til indgangen med samme betydning på de to andre skandinaviske sprog. Gode sprogteknologiske ordbøger med oversættelser mellem de tre skandinaviske sprog kan fx forbedre tværspørgsøgning på internettet, de kan anvendes i sprogteknologiske applikationer som fx et maskinoversættelsessystem og er også en fremragende

ressource når man ønsker at foretage systematiske leksikalske undersøgelser af sprogene. I LEXADV-projektet undersøger vi dels hvorledes SIMPLE-leksikonmodellen kan udvides til også at dække ordklassen adverbier, dels i hvor høj grad det er fornuftigt og givtigt at arbejde med tre beslægtede sprog samtidig når man udvikler semantiske leksika. Det er i sig selv en stor udfordring at formalisere adverbiers semantik da ord fra denne ordklasse altid har ikke-konkrete betydninger der vanskeligt kan beskrives ud fra samme metoder som fx for substantiver (hvor man både i almen og datamatisk leksikografi anvender overbegrebsrelationer, meronymirelationer osv. i beskrivelsen). Derudover er der selvfølgelig masser af oversættelsesmæssige udfordringer når man arbejder med tre sprog og skal finde præcise ækvivalenter mellem sprogene.

1.1. Arbejdsgang

I udarbejdelsen af ordbasen tager vi udgangspunkt i en dansk lemmaliste for adverbier som efterfølgende bliver oversat til svensk og norsk. Lemmalisten indeholder ca. 600-700 adverbier der stammer fra lemmakandidatlisten til ordbogsprojektet *Den Danske Ordbog* (herefter DDO) idet de mindst frekvente er sorteret fra (der var oprindeligt 1202 adverbier på lemmakandidatlisten – i den færdige DDO er der 989 selvstændige adverbiumsindgange og desuden en del adverbier beskrevet under adjektivindgange, fx et adverbium som *sikkert* – disse finder vi frem til vha. synonymangivelser fra adverbiumsindgangene). Til oversættelsesarbejdet benyttes for svensk *Nationalencyklopedins ordbok* (1999) samt for norsk *Bokmålsordboka* (2004). Til selve den formaliserede leksikalske beskrivelse af hvert enkelt adverbium tager vi udgangspunkt i betydningsdefinitionerne i DDO. Til alle tre ordbøger har vi elektronisk adgang idet vi har fået en tekstfil stillet til rådighed af Det Danske Sprog- og Litteraturselskab, der indeholder alle Den Danske Ordbogs selvstændige adverbiumsindgange. Ud over at vi udarbejder en formaliseret beskrivelse (se næste afsnit), indsættes også definitionen og citatet fra DDO i den danske indgang så man altid kan relatere den formelle beskrivelse til den almindelige ordbogsbeskrivelse. De formelle beskrivelser af de danske lemmaer er foregået ‘bottom-up’ idet vi løbende har foretaget ændringer i beskrivelsesmodellen. Når de danske indgange er endeligt færdige, skal de kopieres til en norsk og en svensk ækvivalent. Efterfølgende skal disse så gennemtjekkes af en modersmålstalende med henblik på at sikre at de forskellige formaliserede træk der er benyttet i beskrivelsen af det danske adverbium, også passer på det svenske og norske adverbium.

2. Formaliseret beskrivelse af ords betydninger

En indgang i de allerede eksisterende SIMPLE-ordbøger indeholder som nævnt formaliserede informationer om et ords betydning som et computerprogram kan benytte til informationssøgning, maskinoversættelse m.m. Hvis man ser på be-

skrivelsen af verber og substantiver i SIMPLE-modellen, udtrykkes deres betydning på tre måder. For det første med en oplysning om lemmaets tilhørsforhold i en ontologi. Hvis man tager eksemplet *brød*, hvor definitionen i *Nudansk ordbog med etymologi* er ‘en madvare som er lavet af mel, vand, gær m.m., og som er bagt; skæres fx i skiver og spises med smør og pålæg’, tildeles *brød* for det første den ontologiske type ‘Food’ fra SIMPLE-ontologien (‘Food’ er en ‘Artifact’ som igen er et ‘Concrete_Object’ som igen er en ‘Entity’ under ‘Top’-knuden i ontologien). For det andet beskrives betydningen af *brød* et andet sted i indgangen med en formaliseret oplysning om hvad det sproglige genus proximum er (taget direkte fra ordbogsindgangen i *Nudansk Ordbog*): *brød* is_a ‘madvare’. For det tredje beskrives betydningen af *brød* med en række relationer og træk, fx med relationen ‘used_for’ (‘brød’, ‘spise’), der relaterer *brød* til *spise*. For betydningen *barn* tildeles fx trækket ‘Age’ med værdien ‘young’. Desuden kan man angive selektionsrestriktioner på et lemmas argumenter. Fx kan man for et verbum som *vaske* fortælle at argument 1 skal være af den ontologiske type ‘Human’, og at argument 2 skal være af den ontologiske type ‘Concrete_Object’.

Et af de punkter hvor adverbier er vanskelige at indpasse i denne beskrivelsesmodel, er med hensyn til at angive et sprogligt overbegreb. Vi vil derfor i leksikonmodellen for adverbier nøjes med at 1) tildele adverbiet et ontologisk tilhørsforhold i en ontologi som vi opstiller til formålet, samt 2) beskrive det vha. passende relationer og træk. Her kan nogle af de allerede eksisterende relationer og træk i SIMPLE-modellen anvendes, men der skal også etableres nogle nye.

3. Ontologi for adverbier samt relevante relationer og træk

På baggrund af de syntaktiske grupper som adverbier kan inddeles i ifølge Nimb 2006, samt en semantisk inddeling som *Svenska Akademiens grammatik* (Teleman et al. 1999, herefter SAG) opstiller (bind 2:627), har vi identificeret tre semantiske hovedgrupper for adverbier. De skal alle tre indgå i SIMPLEs top-ontologi, dvs. på niveau med knuderne ‘Entity’, ‘Event’, ‘Intensional’, ‘Extensional’, ‘Telic’, ‘Agentive’ og ‘Constitutive’ der alle hører direkte under topknuden i ontologien. De tre grupper er

1. ‘Functional’ der omfatter adverbier som er funktionsord. Denne gruppe svarer til SAG’s gruppe 5 ‘valensbunden aktant’ (fx *herom*, *derpå*) og gruppe 6 ‘logisk relation’ (fx *kun* og *præcis*, hos Quirk et al. (1972:431) ‘focusing adjuncts’). SAG’s gruppe 7 ‘negation’ og gruppe 2c ‘avsedd funktion’ hører også til ‘Functional’ (fx *alligevel* og *ligeledes* – de kaldes ‘conjuncts’ hos Quirk et al. (1972:520)).
2. ‘SpeakerAttitude’ der omfatter adverbier som udtrykker talerens holdning til resten af sætningen. Gruppen svarer til SAG’s gruppe 8 ‘talarattityd’ og til ‘disjuncts’ hos Quirk et al. (1972:507).

3. ‘Situational’ der omfatter adverbier som beskriver omstændigheder som fx tid, sted og måde. Gruppen omfatter de resterende grupper hos SAG: ‘grad’, ‘sätt och andra aktionella omständigheter’, ‘tidslig omständighet’ og ‘rumslig omständighet’. Disse kaldes af Quirk et al. for ‘adjuncts’ (1972:426), dog fraregnet ‘focusing adjuncts’ som vi regner for ‘Functional’-typen.

Den første gruppe adverbier, ‘Functional’, er ikke mulig at beskrive formelt vha. traditionel leksikalsk semantik, her må man ty til logikkens beskrivelsesapparat (uden at komme nærmere ind på det i denne artikel kan vi dog nævne at vi også udtrykker dette vha. træk, fx trækket ‘AbutB’ for adverbiet *alligevel*). ‘Functional’ under-inddeles i følgende tre grupper: ‘Conjunct’ (med seks undertyper – disse adverbier forbinder sætninger eller optræder som anaforiske udtryk inden for sætningsled), ‘Focus’ (med syv undertyper – disse adverbier fremhæver et sætningsled idet der ofte sammenlignes med noget andet i diskursen) samt den tredje gruppe ‘Negation’. I figur 1 ses taksonomien for ‘Functional’-gruppen.

FUNCTIONAL [Top]

Conjunct [Functional]

- Additive [Conjunct] *desuden* (A and B)
- Adversative [Conjunct] *derimod* (A but B)
- Conclusive [Conjunct] *således* (A then B)
- Explanative [Conjunct] *nemlig* (B because of A)
- Disjunctive [Conjunct] *ellers* (A or B)
- InternalReference [Conjunct] *heraf*

Focus [Functional]

- OnlyX [Focus] *kun*
- PreciselyX [Focus] *præcis*
- EspeciallyX [Focus] *især*
- AtLeastX [Focus] *i hvert fald*
- EvenX [Focus] *endog*
- TimeFocus [Focus] *allerede*
- SuchAsX [Focus] *eksempelvis*

Negation [Functional] *ikke*

Figur 1.

‘SpeakerAttitude’-gruppen kan, modsat ‘Functional’-gruppen, godt beskrives vha. leksikalske semantiske metoder, dvs. med forskellige relationer der skaber forbindelser til andre betydninger, samt med træk. Den underinddeles i tre typer, de deontiske adverbier (‘Deontic’) og de epistemiske adverbier (‘Epistemic’) – for begge

disse grupper gælder det at taleren ikke præsupponerer at sætningen er sand. Den tredje gruppe er de aletiske adverbier ('Aletic') – her præsupponerer taleren modsat at sætningen er sand. 'Aletic'-gruppen underinddeles i de forstærkende adverbier ('Emphasize'), de adverbier der inddrager en synsvinkel ('Viewpoint'), samt de adverbier der udtrykker en evaluering ('Evaluate'), jf. i øvrigt Skaftø Jensen (1997) og Tøgeby (2003) som diskuterer samme emne. Klassifikationen for 'Speaker-Attitude-adverbierne' ses i figur 2 med et ordeksempel til hver type.

SpeakerAttitude

Deontic [SpeakerAttitude] *forbåbenlig*

Epistemic [SpeakerAttitude] *måske*

Aletic [SpeakerAttitude]

Emphasize [Aletic] *simpelthen*

Viewpoint [Aletic] *anstændigvis*

Evaluation [Aletic] *desværre*

Figur 2.

Den tredje hovedgruppe af adverbier er 'Situational'-gruppen. Den er underinddelt i typerne 'Spatial' (rumlige betydninger, underinddelt i sted og retning), 'Manner' (måde), 'Time' (tid – underinddelt i tidspunkt, frekvens og varighed, jf. Pustejovsky et al. (2004), se nedenfor) og 'Degree' (grad). De ses i figur 3.

SITUATIONAL [Top]

Spatial [Situational]

Place [Spatial] *derude, bagi*

Direction [Spatial] *bort, hjemmefra*

Manner [Situational] *forsigtigt*

Time [Situational]

PointofTime [Time] *nu*

Date [PointofTime] *forleden, i går*

Set [Time] *tit*

Durative [Time] *længe*

Degree [Situational] *aldeles, fuldt*

Figur 3.

Hermed er der opstillet en taksonomi der i vid udstrækning dækker danske adverbiers mange betydninger ('manner'-gruppen, altså gruppen af mådesadverbier (1/6 af vores lemmaliste), har vi dog ikke endeligt undersøgt, og vi når ikke at medtage dem i ordbasen i indeværende projekt. Den skal sandsynligvis inddeles i yderligere

undergrupper.) I den leksikalske indgang skal der for hver adverbiums betydning være en oplysning om hvilken af de 30 semantiske typer betydningen hører under.

Med hensyn til hvilke relationer og træk der er nødvendige for yderligere at kunne beskrive adverbiernes betydning, vil vi i de næste afsnit give nogle eksempler i selve de leksikalske indgange vi præsenterer. Indledningsvis kan vi her nævne at man fx for adverbier med tidspunktsbetydning kan beskrive om tidspunktet er fortidigt, synkront eller fremtidigt (enten som iboende egenskab uanset adverbiets kontekst eller også set i forhold til tekstens nu) vha. trækket ‘Temporality_type’ med de tre mulige værdier ‘anterior’ (fx *forleden*), ‘synchronic’ (fx *nu*) og ‘posterior’ (fx *senere*).

4. Selve beskrivelsen af gradsadverbier, tidsadverbier og talerattitudeadverbier

Vi vil nu gå over til at beskrive hvordan vi helt konkret har behandlet gradsadverbier, tidsadverbier og talerattitudeadverbier (modale adverbier) i leksikonnet – i alt ca. 235 adverbier.

67 danske gradsadverbier er i første omgang blevet oversat til svensk og beskrevet. Dels er de blevet tildelt den relevante semantiske type fra ontologien (‘degree’ [‘situational’[Top]]), dels er ‘niveaue’ af graden blevet beskrevet vha. et træk som vi kalder ‘degree_level’. Dette træk har (efter flere afprøvninger mht. at ramme et rimeligt niveau) fået fire værdimuligheder i vores beskrivelsesmodel, nemlig ‘totality’, ‘high’, ‘middle’ og ‘low’. Den største gruppe udgøres af de adverbier der har fået tildelt værdien ‘high’; det har nemlig i alt 45 af gradsadverbierne. Et eksempel er *dybt*: *Det er en dybt pinlig og besynderlig måde at føre politik på* (Berl99). Den næststørste gruppe er de adverbier der har fået tildelt værdien ‘totality’ – i alt 13. Et eksempel er *fuldkommen*: *Hun kan ikke få børn, selv om hun bliver fuldkommen rask* (Berl99). Både gruppen af de gradsadverbier der udtrykker en middelstor grad (tildelt værdien ‘middle’, fx *forholdsvis: hans revolutionære sindelag [trådte] først i karakter i en forholdsvis sen alder* (Berl99)), og de der udtrykker en lav grad (tildelt værdien ‘low’, fx *lidt*: *Det er lidt uhyggeligt at tænke på* (Berl99)), er til gengæld ret små, nemlig på henholdsvis 5 og 4.

I figur 4 ses et eksempel på leksikonindgangen for *dybt*. Det første leksikalske træk som adverbiet har fået tildelt, er for ‘ReferentialStatus’ hvilket er et generelt træk der anvendes på alle typer af adverbier. For *dybt* får det værdien ‘descriptive’. Deskriptive adverbier er ifølge SAG (bind 2:679) “mindre beroende än § 14-48 av kontext och talsituation för sin tolkning och har ofta en fylligare egen betydelse”. § 14-48 beskriver det som SAG betegner henholdsvis de definitive, de interrogative, de kvantitative, de relationelle og de perspektive adverbier. De henter deres betydning ved reference til noget andet i konteksten – dem vil vi tildele andre

værdier ved trækket ‘ReferentialStatus’, afhængig af hvilken type reference der er tale om. ‘Degree_level’ er tidligere omtalt, og ‘style=neutral’ angiver at adverbiet tilhører neutral sprogbrug. I synonymfeltet angives det danske synonym *meget*. Endelig angives de svenske og norske ækvivalenter.

Selektionsrestriktioner har vi desværre endnu ikke udviklet en metode til at angive i ordbogen selvom vi faktisk mener at det er meget relevant at informere om disse. Især er der mange gradsadverbier som anvendes inden for afgrænsede betydningsområder. Fx modificerer *dybt* som gradsadverbium i dansk typisk adjektiver der udtrykker følelser, nemlig ord som *bekymrende*, *beklageligt*, *betænkeligt*, *taknemmelig*, *ulykkelig*, *fascineret*, *forundret*, *chokeret*, *engageret*, *involveret*, *ulykkelig*, *rystet*, *krænket* (Korpus 2000, naboordsundersøgelse).

Semantisk indgang	K_SEMU_dybt_DGR_1
Definition	Meget; i høj grad (DDO)
Korpuseksempel	Det er en dybt pinlig og besynderlig måde at føre politik på (DDO)
Semantisk type	Degree / Situational / Top
Leksikalske træk	ReferentialStatus= descriptive Degree_level = high Style = neutral
Synonym	DK_SEMU_meget_DGR_1
Oversættelse svensk	SE_SEMU_djupt_DGR_1
Oversættelse norsk	O_SEMU_dypt_DGR_1
Selektionsrestriktioner	%

Figur 4.

Underinddelingen af tidsadverbier i taksonomien har sit udgangspunkt i Sauri et al. (2005) *TimeML Annotation Guidelines*. Disse guidelines er udarbejdet med henblik på opmærkning/tagging af tid i løbende tekst, og ‘TimeML’ står for Time Markup Language. De opmærkninger man her arbejder med for tidsadverbialer (og dermed for adverbier), er:

- a. ‘Point_of_time’ til adverbialer der denoterer et tidspunkt,
- b. ‘Date’ til adverbialer der denoterer eller kan svare til et kalendertidspunkt (dvs. dato, måned og/eller årstal),
- c. ‘Set’ til adverbialer der denoterer en mængde af tidspunkter, og
- d. ‘Duration’ til adverbialer der denoterer en varighed.

Disse fire semantiske typer er alle ‘knuder’ i ontologien for adverbier, se ovenfor. I

projektet har vi undersøgt 79 danske tidsadverbier og tildelt dem en af de fire semantiske typer. De fordeler sig således at 52 er af typen 'Point_of_time' (fx *førhen, senere, straks*), seks er af typen 'Date' (fx *forleden, i dag*), andre seks er af typen 'Duration' (fx *længe, årevis*), og endelig er 15 af typen 'Set' (fx *ofte, undertiden*). Ud over at tildele adverbier med tidsbetydning en af disse fire semantiske typer beskriver vi også deres tidsbetydning vha. forskellige træk. To vigtige træk er 'Punctual' og 'Iterative', begge med værdimulighederne 'yes', 'no' eller 'underspecificed'. Nogle tidsadverbiers betydning spiller tæt sammen med sætningens aspekt og verbets Aktionsart. Derfor har vi brug for værdien 'underspecified' – den vil blive tildelt når det er verbet der afgør om adverbiet har en punktuell, henholdsvis iterativ betydning. I SIMPLE-leksikonnet er oplysninger om Aktionsart en del af beskrivelsen af et verbs betydning. Dette udtrykkes i leksikonindgangen for et verbum vha. trækket 'Event_type' med tre værdimuligheder: 'state' (tilstand, fx *være, blive, bo*), 'process' (aktivitet, fx *læse, save, rydde op*) eller 'transition' (overgang, fx *give, spise op, afslutte*).

For de to undertyper af tidsadverbier 'Set' og 'Duration' spiller verbets Aktionsart dog ingen rolle. Adverbier af typen 'Set' har altid betydning af flere tidspunkter fordelt over en periode af en vis varighed: transition-verbum (overgang): *Han åbner tit en øl*; state-verbum (tilstand): *Han er tit i udlandet*; og endelig process-verbum (aktivitet): *Hun snakker tit om sit arbejde*. Dette udtrykkes i leksikonindgangen ved at give trækket 'Punctual' værdien 'no' samt trækket 'Iterative' værdien 'yes'. Følgende adverbier er af denne type: *indimellem, løbende, almindeligvis, årligt, glimtvis, kvartalsvis, lejlighedsvis, månedsvis, normalt, ofte, tit, regelmæssigt, sædvanligvis, undertiden* og *sommetider*.

'Duration'-adverbierne har ligeledes altid en bestemt betydning, nemlig varighed, uanset verbets Aktionsart, og de kan aldrig have betydning af et kort tidspunkt – men de har mulighed for at udtrykke noget iterativt afhængigt af verbet. De kan aldrig kombineres med et 'transition'-verbum (**Han åbner længe en øl*). Sammen med et state-verbum (med tilstandsbetydning) får de betydning af én hændelse af en vis varighed (*Han boede længe i London*). Sammen med process-verb, dvs. aktivitets-verb, bliver sætningen derimod tvetydig: Enten drejer det sig om én enkelt hændelse af en vis varighed, eller også er der tale om flere gentagne hændelser inden for en periode af en vis varighed (*Han snakkede længe om sit arbejde; han drak i ugevis*). Disse adverbier bliver beskrevet i ordbogen med trækket 'Punctual' sat til værdien 'no' og trækket 'Iterative' sat til værdien 'underspecified'. Der er kun seks danske adverbier (hvoraf flere kun anvendes i faste udtryk med en præposition) med disse leksikalske egenskaber, nemlig: *længe, længst, i ugevis, i årevis, i dagevis, i månedsvis*.

Den mest komplicerede gruppe af tidsadverbier, og også den største, er 'Point_of_time'-adverbierne. Disse adverbiers betydning afhænger altid af verbets

Aktionsart. Adverbierne har enten betydning af ét kortvarigt tidspunkt eller ét tidspunkt af en vis varighed, eller (mest almindeligt) begge dele, afhængigt af verbets Aktionsart (Sauri et al. understreger i øvrigt at langvarige tidspunkter (som fx et år som '1985') stadig er af typen 'Point_of_time' og ikke af typen 'Duration'). Eventuelt kan 'Point_of_time'-adverbier også få betydning af noget iterativt, altså flere hændelser der foregår inden for det pågældende tidspunkt, igen afhængigt af verbets Aktionsart. Sammen med et overgangsverbum får det således betydning af én hændelse på ét kort tidspunkt (*han gav mig bogen forleden*). Sammen med et tilstandsverbum får det igen betydning af én hændelse, men nu i et længerevarende tidspunkt (*han var i skole forleden / han har altid boet i England*). Sammen med et aktivitetsverbum bliver det for nogle 'Point_of_time'-adverbiers vedkommende tvetydigt om der er tale om én hændelse på ét (længerevarende) tidspunkt eller om flere hændelser fordelt på flere tidspunkter (*Jeg læste i bogen forleden; hun læste altid i bogen*). Det enkelte adverbiums muligheder udtrykkes vha. de to førnævnte træk med tilhørende værdier, idet trækket 'Punctual' enten får værdien 'yes' eller 'underspecified', mens trækket 'Iterative' enten får værdien 'no' eller 'underspecified'.

Når man beskriver 52 danske point_of_time-adverbier ud fra disse to træk, får man tre grupper. For det første en lille gruppe af adverbier der altid har betydning af ét kort tidspunkt. De er beskrevet med 'Punctual' = 'yes' og iterative = 'no'. De kan aldrig have betydningen længerevarende tidspunkt eller gentagelse, og de optræder som regel kun sammen med transition- eller process-verber, og evt. kun i perfektive tider: *snart, straks, øjeblikkelig, netop, lige, just, omgående, prompte, fluks*. For det andet får man en gruppe der er beskrevet med 'Punctual' = 'no', 'Iterative' = 'underspecified'. Denne gruppe adverbier har enten betydning af et længerevarende tidspunkt, nemlig sammen med state-verber (*han har altid boet i England; han skal fremover bo i England; han har hidtil boet i England*), eller de har betydning af gentagelse, nemlig sammen med transition- og process-verber (*han spiser altid sin mad op; hun snakkede altid meget med ham; han vil fremover spise sin mad op; hun vil snakke mere med ham fremover; han har hidtil spist sin mad op; han har hidtil snakket meget med ham; tidligere spiste han sin mad op; han har tidligere snakket meget med ham*). De kan til gengæld aldrig have betydning af et kort tidspunkt. Nogle eksempler er *siden hen, siden, stedse, fremover, indtil videre, evindeligt, førhen* og *tidligere*. En sidste ret stor gruppe har både trækket 'Punctual' og trækket 'Iterative' udfyldt med værdien 'underspecified'. De kan altså betyde både et kort eller længerevarende tidspunkt og eventuelt også have iterativ betydning – det afhænger helt af verbet og den øvrige kontekst. De får betydningen ét kort tidspunkt sammen med transition-verber (*senere tog han bogen frem*), men betydning af et længerevarende tidspunkt sammen med process- eller state-verber (*senere spiste han; senere læste han i bogen; senere boede han i huset*) – og en iterativ læsning er ikke umulig i den rigtige kontekst selvom den ikke er den umiddelbart

mest oplagte. Nogle eksempler er *derimellem*, *da*, *dengang*, *før*, *forinden*, *forleden*, *i dag*, *i forvejen*, *i går*, *senere*, *nogensinde*, *bestandig*, *herefter*, *herpå*, *i år*, *i morgen*, *midtvejs*, *sent* og *tidligt*.

I figur 5 ses den leksikalske indgang for tidsadverbiet *førhen*. Ud over de to omtalte træk 'Punctual' og 'Iterative' der her har værdierne 'no' henholdsvis 'underspecified', får 'ReferentialStatus'- trækket værdien 'deictic' (*førhen* refererer til tekstens nu, modsat fx *forinden* der refererer til et tidspunkt beskrevet i konteksten). *Temporality_type* har værdien 'anterior' da *førhen* denoterer et tidspunkt før sit deiktiske referencetidspunkt.

For tidsadverbierne har vi valgt at beskrive selektionsrestriktioner mht. hvilke(n) Aktionsart og hvilken verbaltid (Tempus) de typisk foretrækker. *Førhen* kan kombineres med alle typer Aktionsart, men kan ikke kombineres med verber i fremtid.

Semantisk indgang	DA_førhen_POT_1
Definition	(DDO) tidligere; før i tiden = forhen
Korpuseksempel	(DDO) <i>Førhen kunne folk godt leve med et lille handicap, det kan de ikke i dag</i>
Semantisk type	PointofTime/Time/Situational/Top
Leksikalske træk	ReferentialStatus = deictic Temporality_type=anterior Punctual=no Iterative = underspecified
Synonym	tidligere; før_i_tiden
Svensk oversættelse	SE_tidigare_POT_1
Norsk oversættelse	NO_før_POT_1
Selektionsrestriktioner	Preferable_Actionsart=process/state/transition Preferable_Tempus= imp/perf

Figur 5.

Den sidste gruppe af adverbier vi her vil beskrive mere detaljeret, er talerattitude-adverbierne, altså de adverbier der udtrykker talerens synspunkt på resten af sætningen. Leksikonmodellen for denne gruppe bygger på SAG (bind 4, kapitlerne 30 *Satsadverbial* og 32 *Modala hjælpeverb*). De tre undertyper i ontologien er 'Deontic', 'Epistemic' og 'Aletic'. Med et adverbium af typen 'Deontic' udtrykker taleren ønske, betingelse, indrømmelse eller forventning fra egen eller andres side idet han/hun ikke præsupponerer at sætningen er sand. Nogle eksempler er *forhåbentlig* (ønske), *absolut* (betingelse), *faktisk* (indrømmelse), *endelig* (forvent-

ning). I alt er 21 af de 87 talerattitudeadverbier vi har arbejdet med, af denne type. Med et adverbium af typen 'Epistemic' vurderer taleren i hvor høj grad sætningen er sand. Her præsupponer han/hun altså heller ikke at den er sand. Nogle eksempler er *formodentlig* og *åbenbart*. 33 af de 87 talerattitudeadverbier er af denne type. Med den sidste type talerattitudeadverbier, 'Aletic', præsupponerer den talende derimod at sætningen er sand. De underinddeles i tre undertyper: For det første de der udtrykker talerens forstærkning eller nedtoning af sætningens udsagn ('Emphasize'): *såmænd*, *simpelthen*. 23 af de 87 adverbier er af denne type, så den er klart den mest almindelige af disse adverbier. Dernæst gruppen af evaluerende adverbier – de udtrykker hvad taleren synes om sætningens udsagn. Nogle eksempler er *heldigvis* og *desværre*. Det er en lille gruppe på kun seks adverbier. Endelig er der 'View-point'-adverbierne: Her inddrager taleren eksterne faktorer i vurderingen af sætningens udsagn. Et eksempel er *anstændigvis* – der er kun fire af denne type adverbier.

De modale adverbiers betydninger er svære at definere, også i almindelige ordbøger. I den formelle beskrivelse forsøger vi at holde os nøje til definitionerne i DDO. Vendinger der går igen i definitionerne for mange af de modale adverbier, er 'bruges for at forstærke', 'bruges for at nedtone'; 'bruges forstærkende' etc. En definition af denne type tyder meget på at adverbiet i ordbasen skal have tildelt den semantiske type 'Emphasize'. Vi uddyber betydningsbeskrivelsen ved at tilføje trækket 'Emphasize_level' med følgende værdimuligheder: 'high' (*han har minsandten gjort det igen*); 'middle' (*han har endda gjort det igen*) eller 'low' (*han har blot gjort det igen, han har nu gjort det igen*).

Hvis DDO's definition indeholder en beskrivelse af hvilken følelse der ligger til baggrund for den talendes brug af adverbiet, beskriver vi dette vha. relationen Speakerfeeling. For eksempel står der følgende om *minsandten* i DDO:

bruges forstærkende for at understrege undren, en personlig opfattelse, forsikring e.l. = sandelig, søreme jf. minsæl □ *Ude til havs mødte vi minsandten en stor havskildpadde, der lå og drev i overfladen, Jeg er så glad for, at vi har truffet hinanden igen. – Det er jeg minsandten også, Viveca*

• bruges som udtryk for overraskelse, undren e.l. □ *da Ditte strakte hals og kikkede ud ad vinduet, gav hun et overrasket udbrud fra sig. – Minsandten! Det er Viktor, der kommer hjem. Allerede.*

I SIMPLE-leksikonet relaterer vi derfor *minsandten* til *overraskelse* og *undren*: Vi har også ønsket at formalisere den del af definitionerne i DDO der beskriver hvad den talende regner for allerede eksisterende viden hos modtageren. Fx beskriver DDO adverbierne *da* og *jo* med følgende definitioner: *da*: bruges for at angive hvad der regnes for givet el. indlysende; *jo*: bruges for at angive at noget er bekendt el. indlysende = som bekendt. I SIMPLE-leksikonet gengiver vi dette vha. trækket

'PresupposedAddresseeKnows' som kan have værdierne 'yes' eller 'no'. I figur 6 ses et eksempel på en leksikalsk indgang for et talerattitudeadverbium, nemlig *dog*:

Semantisk indgang	DA_dog_EMP_1
Definition	2 bruges forstærkende for at udtrykke at noget gør stort indtryk el. påvirker den talende følelsesmæssigt og fx giver anledning til overraskelse, forundring, begejstring el. bestyrtelse
Korpuseksempel	" <i>Hvor ser du dog affældig ud,</i> " sagde Ea (DDO), <i>Det lille pus, hvor er hun dog dejlig,</i> sagde Bo ømt (DDO)
Semantisk Type	Emphasize/ Aletic / SpeakerAttitude /Top
Semantiske træk	Lexical_Reference= descriptive Emphasize_level = middle Style= neutral Speakerfeeling (DA_dog_EMP_1, SEMU_overraskelse_N_1) Speakerfeeling (DA_dog_EMP_1, SEMU_forundring_N_1) Speakerfeeling (DA_dog_EMP_1, SEMU_begejstring_N_1) Speakerfeeling (DA_dog_EMP_1, SEMU_bestyrtelse_N_1)
Synonymy	%
Svensk oversættelse	SE_verkligen_EMP_1 , SE_sannerligen_EMP_1
Norsk oversættelse	NO_så_EMP_1, NO_jamen_EMP_1 (<i>du ser jamen gammeldags ut!</i> , <i>Hun er jamen deilig!</i>)

Figur 6.

5. Oversættelse mellem dansk, svensk og norsk

Det er selvfølgelig ikke helt uproblematisk at finde præcise ækvivalenter mellem de tre skandinaviske sprog selvom de er tæt beslægtet. For det første kan ordklasse gå på tværs af sprogene. Fx er *à jour* et adjektiv på svensk, men et adverbium i dansk. Et dansk adverbium kan svare til et svensk præfiks, fx *enormt* (*enormt godt*) der bedst oversættes til *jätte-* (*jättebra*). Der er også tilfælde hvor et dansk adverbium svarer til et svensk adverbial (i SAG kaldet et 'flerordsadverbium' eller et idiom). Et eksempel er dansk *endsige* (*Ingen sælgere ønskede af naturlige årsager at kommentere deres salg endsige at få deres navn offentliggjort*) der i svensk må udtrykkes med *långt mindre / än mindre än*. Denne slags problemer af mere 'teknisk' art kan dog løses ved at tillade flerordsadverbier samt præfikser i ordbasen. Adjektiver der anvendes som adverbium, bør i et sprogteknologisk leksikon have egen indgang, og på denne måde løses også ækvivalensproblemet med forskellig ordklassebetegnelse. Det er derimod langt sværere at løse problemerne med hensyn til manglende ækvivalenser sprogene imellem, eller delvist overlappende betydninger. Selvom vi fx endnu ikke har beskrevet alle stedsadverbierne i ordbasen, kan vi allerede nu se at der er semantiske komponenter i nogle af de danske stedsadverbier som ikke kan

gengives med et svensk adverbium. Et eksempel er dansk *herovre* (*Ideen med at have medarbejdere herovre på skift er vældig god*) som på svensk må oversættes til *på denna sidan*, og dansk *dernedefra* (*han må selv skrive til dem der-nedefra*) der på svensk skal oversættes til *därifrån* hvor man ikke kan udtrykke *nede* ("nere" går ej här, som oversætteren bemærkede). Der er ligeledes problemer med de adverbier der i dansk kan erstatte et præpositionsobjekt, fx dansk *derudfra* (*derudfra skal forskningen bygges op*) – på svensk må man oversætte til *utgående från detta*, og på moderne norsk anvendes disse adverbier heller ikke. Ligeledes er der mange udfordringer i at oversætte de meget polyseme talerattitudeadverbier hvor man nemt kan blive snydt af at der findes et lignende lemma på de to andre skandinaviske sprog, men hvor der faktisk er sket store betydningskift sprogene imellem (jf. modalverberne der også har meget forskelligt betydningsindhold på de tre sprog). Hvis vi fx betragter den leksikalske indgang for svensk *visst* i NEO, kan vi se at de enkelte betydninger ofte skal oversættes til helt andre modale adverbier i dansk og norsk:

SV: visst

1 med visshet ofta förstärkande, särsk. efter det att ngn uttryckt tvivel: ja ~; nej ~ inte; jo, jag har ~ ont i huvudet (DA: ja faktisk, nej, faktisk ikke, jeg har faktisk ondt i hovedet, NO: ja visst, nej, faktisk ikke, jeg har visst vondt i hodet).

2 efter vad det verkar → troligen: vi har inte sett honom på ett tag men han har ~ varit sjuk (= DA: vist, NO visst).

3 enligt vad alla bör inse särsk. i uttr. för vädjan e.d. → nog2: ~ är det så, det måste du medge; ~ finns det många som trivs på jobbet (DA: jo vist er det så, det må du give mig ret i; der findes helt sikkert mange som trives på deres arbejde, NO: jovisst er det slik, det er (helt) sikkert mange som trives på jobben).

Bet.nyans: försvagat visserligen: ~ är jag trött men jag skall försöka komma ändå (DA: ganske vist, godt nok, NO: riktignok).

Også hvad angår gradsadverbierne, er der problemer pga. de leksikalske restriktioner som mange adverbier af denne type udøver. Fx kan man på dansk være både *lynende hurtig*, og *lynende intelligent*, samt have *lynende travlt*. På svensk kan man være *blixtrande snabb/intelligent*, men man bruger et andet adverbium sammen med *upptagen* (*strängt upptagent*). Dette problem har vi besluttet at løse ved at angive et hvad angår selektionsrestriktioner, så neutralt gradsadverbium som muligt som oversættelsesækvivalent. Derfor oversættes dansk *lynende* til svensk *mycket* idet vi dog er opmærksomme på at oversættelsen så er mindre god. Det skal til sidst nævnes at der trods alt er mange tilfælde hvor det er nemt at finde en passende ækvivalent på de to andre sprog. Fx mener vi at dansk og norsk *formodentlig* (DDO: *Efter alt at dømme; efter al sandsynlighed*) helt uproblematisk begge kan oversættes til svensk *förmodligen* (NEO: *med rätt hög sannolikhet = antagligen, troligen*).

6. Konklusion

Det er en stor udfordring at forsøge at formalisere adverbiernes semantik, men vi mener på baggrund af det hidtidige arbejde i LEXADV-projektet at en del godt kan lade sig gøre. Med hensyn til oversættelsesarbejdet mellem dansk, svensk og norsk er det som regel muligt at finde dækkende ækvivalenter fordi sprogene er tæt beslægtede. Der er dog helt klart nogle typer af adverbier der er svære at oversætte, og hvor betydningerne er overlappende.

Ud over de tre typer adverbier der er gennemgået i denne artikel, planlægger vi at arbejde detaljeret med både stedsadverbier, de fokuserende adverbier og de konjunktionelle adverbier. Vi forventer at især stedsadverbier, som udgør en stor gruppe – ca. 100 –, bliver spændende at beskrive med formelle træk, og som vi har nævnt, kan vi allerede nu se at det for denne gruppe bliver en udfordring at finde oversættelsesækvivalenter mellem de tre sprog.

Litteratur

- Berl99: Berlingske Aviskorpus*, ca. 30 mio. ord, Berlingske Tidende og Weekend-avisen. Årgang 1999. København.
- Bokmålsordboka 2004* 2. udgave, 4. oplag. Oslo: Kunnskapsforlaget.
- DDO = *Den Danske Ordbog*. Bind 1, 2, 3, 4 og 5. 2003-05. Ebba Hjorth & Kjeld Kristensen (red.). Det Danske Sprog- og Litteraturselskab. København: Gyldendal.
- Korpus 2000*: www.korpus2000.dk. København: Det Danske Sprog- og Litteraturselskab.
- Nationalencyklopædiens Ordbok 1999*. Udarbejdet af Språkdata. Göteborgs Universitet. CD-ROM.v.2.0. Göteborg.
- Nimb, Sanni 2006: Adverbiers syntaks i ordbøger. I denne rapport.
- Pedersen, B., Ruth V. Fjeld og Maria Toporowska Gronostaj 2005: Sprogteknologiske ordbaser for de nordiske sprog – rapport fra et forsknings-netværk. I: *Nordiske studier i leksikografi 7*, Nordisk forening for leksikografi, Oslo.
- Politikens Nudansk Ordbog med Etymologi*. 1999. 1. udgave (elektronisk version). København: Politikens Forlag.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech og Jan Svartvik 1972: *A Grammar of Contemporary English*, Essex: Longman.
- SAG = Telemann, U., Hellberg, S. & Andersson, E. 1999: *Svenska Akademiens Grammatik*. Stockholm: Svenska Akademien.
- Sauri, R., J. Littman, B. Knippen, R. Gaizauskas, A. Setzer og J. Pustejovsky 2005: *TimeML Annotation Guidelines*: <http://www.timeml.org/timeMLdocs>.
- Skaftø Jensen, Eva 1997: Modalitet og dansk. I: *NyS 23: Nydanske studier og almen kommunikationsteori*. København: Dansk lærerforeningen.
- Togeby, Ole 2003: *Fungerer denne sætning? Funktionel dansk sproglære*. København: Gads Forlag.

Digital sats eller digital satsing?

Presentation or content? The Unit for Digital Documentation at the University of Oslo has been making dictionary writing systems (DWS) for more than ten years. In this paper we will present some of our ideas on how a modern DWS should deal with the source material, other dictionaries and cross references inside the dictionary, and what implications this will have on the implementation of such a system.

1. Innleiing

1.1. Om Eining for digital dokumentasjon

Eining for digital dokumentasjon (EDD) ved Universitetet i Oslo er direkte underlagt det Humanistiske fakultet (HF) og har som mål å hjelpe fageiningane ved HF med å lage digitale dokumentasjonssamlingar, tekstsamlingar, korpus og andre databasar.

EDD er eit produkt av Dokumentasjonsprosjektet som var eit stort samarbeidsprosjekt mellom dei humanistiske fakulteta ved dei fire universiteta i Noreg. Oppgåva til Dokumentasjonsprosjektet var å digitalisere dei store samlingsarkiva ved dei ulike HF-einingane.

Etter at Dokumentasjonsprosjektet vart avslutta, vart det naudsynt å etablere ein driftsorganisasjon som skulle vedlikehalde og vidareutvikle databasane som vart bygd opp i prosjektperioden, samt å utvikle nye verktøy for brukarane. EDD vart oppretta som ei permanent eining under HF. EDD har i tillegg til dei faste oppgåvene for HF også ansvaret for gjennomføringa av Museumsprosjektet, eit nytt stort nasjonalt digitaliseringsprosjekt som skulle bygge på dei røynslene som var gjort under Dokumentasjonsprosjektet.

Som ei følgje av dette har EDD ein forholdsvis liten fast tilsett stab som har hovudansvaret for oppgåvene som ligg til HF, og ein stor prosjektstab knytt til Museumsprosjektet. Trass i dette skiljet, utviklar EDD bevisst løysingar som er felles for HF og Museumsprosjektet, og ein prøver å utnytte kompetansen mellom dei to stabane til beste for begge partar.

1.2. Leksikografi

Arbeidet med dei leksikalske arkiva har vore ein sentral del av EDD sine oppgåver heilt frå tida i Dokumentasjonsprosjektet. Setelarkivet til Norsk Ordbok var mellom dei første arkiva som vart digitalisert tidleg på 1990-talet, og skanning og registrering av 3 millionar setlar tok mykje ressursar.

EDD var tidleg ute med å lage løysingar for å bruke dei leksikalske arkiva, og

utvikla også tidleg applikasjonar for redigering av ordbøker. Taggarprosjektet som var eit samarbeid mellom Tekstlaboratoriet ved HF og EDD, utvikla tidleg ein taggar basert på data frå Nynorskordboka og Bokmålsordboka. EDD laga også verktøy for redigering av dei første einspråklege ordbøkene i Zimbabwe (ALLEX-prosjektet), og har også stått bak redigeringsystema for Nynorskordboka og Bokmålsordboka. I dag jobbar EDD mellom anna med redigeringsystemet til Norsk Ordbok, og eininga har utvikla integrerte korpussystem knyta til redigeringsapplikasjonane basert på CQP/Corpus Workbench.

1.3. EDD sin filosofi

Gjennom arbeidet med samlingane har EDD fleire gonger påvist små og store manglar både ved måten arkiv er organisert på, og ved innhaldet i arkiva. Arkiva har ofte vore utforma på ein slik måte at dei berre kan brukast av ei lita gruppe menneske, og ikkje sjeldan er det store problem med oppdatering av arkiva. Ofte kjem ein over arkiv der tidlegare tolkingar av eit materiale vert kasta når det vert gjort nye tolkingar. Andre arkiv kan vere organisert slik at det ikkje er rom for nytolkingar. Eit døme på det siste er setelarkivet ved Norsk Ordbok som i papirutgåve var sortert etter ei normalisert skriftform basert på 1938-rettskrivinga av nynorsk.

Dei første digitale arkiva EDD laga, var nok på mange måtar prega av dei same problema, til dømes i høve til det å kunne ha rom for nytolking av materialet. Etter kvart har EDD lagt stor vinn på å forbetre dette, og ein har i fleire år vore aktiv deltakar både i nasjonale og internasjonale fora der ein prøver å lage standardar for datautveksling, samt diskutere løysingar på ulike problem knytt til forskingsdatabasar.

EDD har på bakgrunn av dette lagt vinn på å utvikle arkivløysingar som på alle vis kan karakteriserast som vitskaplege. Med dette meiner vi mellom anna at ein må lagre grunnlagsmaterialet i så lite bearbeidd form som mogleg. Dei mest atomære storleikane i eit arkiv bør vere så fri for tolking som mogleg. Derneft må alle hypotesar som er basert på dette grunnlagsmaterialet handsamast med like stor respekt. Arkivsystema må ta omsyn til at nye tider og nye forskarar vil ha nye tolkingar av kva som ligg i eit materiale. Og sist, men kanskje viktigast, arkivsystema må kople saman grunnlagsmaterialet og hypotesane slik at det er enkelt for nye forskarar å etterprøve resultat.

Ut over dette har EDD ønskje om å kunne knytte saman informasjon i ulike arkiv slik at forskarane kan dra nytte av kvarande sine resultat i langt større grad enn i dag.

2. Produksjon av tradisjonelle ordbøker

Produksjon av ordbøker vart tidleg ei viktig oppgåve for EDD, og EDD har vore med på å flytte fleire store prosjekt inn i framtida. Norsk Ordbok er absolutt det største, og vi vil bruke denne ordboka som eit døme på korleis dette har gått til.

2.1. Fokus på utsjånad

Før EDD kom inn i biletet, vart artiklane i Norsk Ordbok skrivne i WordPerfect og Word. Ein nytta eit sjølvutvikla kodesystem for å merke opp teksten, og all oppmerking hadde som mål å styre utsjånaden på trykk.

Problemet med denne framgangsmåten er at ordboka sjølv ikkje gjer det lett for ein brukar, eller ein forskar, å etterprøve informasjonen. Rett nok har ein i alle vitskaplege ordbøker eit referansesystem til litteratur eller arkivmateriale som underbygg påstandane, men reelt sett er det ikkje mogleg for ein brukar å sjekke desse kjeldene med mindre brukaren sit i arkivet til ordboksredaksjonen. Sjølv om ein skulle sitte i arkivet, ville det vere omtrent umogleg å finne fram til dei setlane som har danna grunnlaget for ein artikkel, og ikkje minst finne ut kva setlar redaktøren ikkje brukte. Ordboka ville med andre ord formelt sett tilfredstille vitskaplege krav, men reelt sett ville ho i avgrensa grad bli utsett for vitskapleg kritikk fordi det ville vere eit alt for stort arbeid å gå til kjeldene.

Frå redaksjonen sin synsstad har også denne gamle produksjonsmåten mange ulemper. Mellom anna er handteringa av krysstilvisingar mellom artiklar eit evig problem. Det er ingenting som garanterer at ein definisjon som ein viser til, faktisk eksisterer. Stringens i form og innhald vert også meir komplisert enn det ein gjerne vil. Det einaste som styrer dette er redaktørane sin godvilje, og redaksjonen sin vilje til å overprøve einskildredaktørar sin måte å skriva artiklar på. Handtering av mange redaktørar vert også vanskeleg. Dess fleire redaktørar, dess vanskelegare vert det å halde styr på dei to føregåande punkta.

Gjenbruk av data frå ordboka er heller ikkje særleg lett med denne måten å redigere artiklar på. Det er svært lite struktur i innhaldet, og den som måtte ønskje å bruke ordboka til noko anna enn å slå opp definisjonar, måtte bruke mykje tid på å finne det ein leita etter, trass i at teksten var tilgjengeleg digitalt.

2.2. Fokus på struktur

Ut over 1990-talet vart det opplagt for dei fleste at dei gamle måtane å redigere ordboka på, ikkje gav samfunnet nok attende i høve til innsatsen. Krav til tilgjenge for ålmenta, gjenbruk av data, vitskapleg kvalitetssikring, og ikkje minst effektiv produksjon av ordboksmanuskript, sette fart i planane om å endre måten Norsk Ordbok vart redigert på. Alle krava ein stilte til den nye ordboka, peika i retning av meir fokus på struktur, mindre på utsjånad, samt at det måtte vere mogleg å kople ordboka sine artiklar mot grunnlagsmaterialet, og at ein trong administrative rutinar

bygd inn i systemet for å auke effektiviteten. I dag er det to teknologiar som peiker seg ut, strukturert tekst (SGML/XML) eller ei eller anna form for databaseløysing.

2.2.1. SGML/XML

EDD har lang erfaring i bruk av strukturert tekst. Sidan starten av 1990-talet har EDD arbeidd med SGML i oppmarkeringa av tekst. Då XML overtok for SGML, vart dette ein teknologi som vart tilgjengeleg for eit større publikum, og absolutt ein mogleg teknologi å bruke i utforminga av ordbøker.

Den største styrken til XML som format for eit ordboksmanuskript er kombinasjonen av stringens og fleksibilitet. XML gjer det mogleg å lage klare reglar for korleis strukturen skal vere, men kan tillate stor grad av valfridom i innbyrdes plassering av tekstelement. XML kan også på ein enkel måte opne for at einskilde element kan dukke opp på fleire plassar i eit dokument utan at strukturen lir under dette.

Det største problemet med XML er at dette ikkje er eit databaseformat. I utgangspunktet er eit XML-dokument ei samling tekstar som alle har fast struktur. For store dokument vil ein rein XML-struktur gjere referanseintegritet mellom delar av dokumentet tungt og tidkrevjande, og kopling til element som ligg utanfor dokumentet sjølv vil vere enno tyngre. Med XML er det ikkje mogleg å garantere at referansar til grunnlagsmaterialet i form av til dømes ordsetlar, er korrekte. XML kan heller ikkje hjelpe til med administrasjonen av ein stor redaksjon. Eit større ordboksprosjekt er avhengig av desse funksjonane.

2.2.2. Databaseløysingar

Hovudalternativet til XML er å lagre artikkelteksten i ein database av eit eller anna slag. Databasen sin største styrke er innebygde rutinar for dataintegritet. Dataintegriteten gjer det mogleg å sikre at kryssstilvisingar til ein kvar tid er korrekte, at kopling til kjeldematerialet er på plass, osv. Databasesystem utan slik integritetskontroll er til lita nytte for eit ordboksprosjekt.

Databasen tilbyr struktur, men kan tidvis også stille for store krav til struktureringa av artiklane i ei ordbok. Dette er den største innvendinga mot å bruke ein database som basis for sjølve artikkelteksten. Databasen har ikkje innebygde system for å markere at delar av ein laupande tekst har ei spesiell mening.

2.2.3. Kombinasjonsløysingar

Konklusjonen er at korkje ei rein XML-løysing eller ei rein databaseløysing fungerer for dei fleste større ordboksprosjekt. Ein må ty til kombinasjonsløysingar.

Det finst mange verktøy som i ulik grad løysar problema med XML ved å innføre ein eller annan form for databasestruktur rundt XML-dokumenta, og dette kan ofte vise seg svært fruktbart. Ofte nyttar ein XML som internt databaseformat i eit slikt

databasesystem, og dette kan gjere desse databasesystemet ope og lett å dokumentere, men dette er på langt nær opplagt. Mange av desse systema er også svært nye og lite utprøvde, dei er laga for mindre organisasjonar, og skalerer ikkje godt i store miljø.

Mange databasesystem i dag kan handtere XML eller XML-fragment i tekstfelt i databasen. Dette løyser litt av problema knytt til databasen sine ekstreme strukturkrav.

Om ein vel å basere seg på ei utvida XML-løysing, eller ei utvida databaseløysing, kjem an på kva eigenskapar ein meiner er viktigast. Eit spesialutvikla databasebasert system vil vere mest teneleg dersom ein skal ha store redaksjonar og god kontroll med referansar og tilvisingar i ordboka. Men dette krev at ein kan definere ein streng struktur for artiklane utan at dette går ut over den faglege kvaliteten.

Ei mindre ordbok med ein mindre redaksjon og mindre krav til stringens, vil tene på å bruke eit enklare, meir generelt XML-basert verktøy.

3. Den nye ordboka

I dei føregåande avsnitta presenterte vi dei krava vi meiner ein må stille til verktøy for produksjon av dei tradisjonelle ordbøkene. Men teknologien fører ikkje berre til at ein kan gjere tradisjonelle oppgåver på ein betre måte. Det kan like gjerne føre til at oppgåva eller det endeleg resultatet totalt endrar karakter. Tenkjer ein visjonært kan ein ofte med same arbeidsinnsats, få eit anna og mykje meir nyttig resultat. Dette ser ein tydeleg innanfor mange humanistiske fag i dag, og ikkje minst innanfor språkforskinga.

Teknologien kan og gje eit omgrep eit heilt nytt innhald, og innhaldet knytt til omgrepet endrar krava publikum stiller. Dette ser ein tydeleg i samband med bøker generelt, og ordbøker spesielt. I Norsk Ordbok, band 1, som vart utgjeve i 1966, er ordet "bok" definert som ei samling innbundne, prenta ark. I dag kan vi lese på web-sidene til Norsk Ordbok at ein har som mål å kome med ei elektronisk utgåve av ordboka. Både Norsk Ordbok, men også store delar av det norske folk, meiner at ei samling av ord og definisjonar som ein kan finne på internett, er ei ordbok. I daglegtale er det ikkje lenger uvanleg å høyre om elektroniske bøker, eller lydbøker. Teknologien har gjort at ei bok i dag er noko heilt anna enn det ho var for 40 år sidan, og våre krav til bøkene speglar teknologien det er mogleg å bruke for å presentere innhaldet av boka. Den beste ordboka er ikkje berre den med dei mest korrekte definisjonane, men også den som utnytter den eksisterande teknologien best til å formidle innhaldet i definisjonane. Leksikografen har ikkje gjort jobben sin dersom han ikkje har utnytta teknologien optimalt når han formidlar kunnskapen sin om eit ord.

Dette legg eit stort press på dagens leksikografar når dei skal lage vitenskaplege ordbøker. Frå Gutenberg og fram til 1990 måtte leksikografen ta stilling til

sidestorleikar, skrifttypar, spalter, osv. Dette var nokre av dei parametrane som styrte formidlinga. I tillegg hadde formidlingskanalen klare grenser for kva ein kunne ta med. Mangel på teknologiske løysingar førte til at ein sjølvstøtt ikkje kunne prente alt kjeldemateriale, og ein kunne heller ikkje legge ved lyd eller video. Dersom ein kunne gjort dette ville den vitskapelege verdien av ordboka auka.

I dag set ikkje teknologien slike grensar. Det er fullt mogleg å lenke oppslagformer og dialektformer direkte til lyd kjelder. Definisjonar og døme kan peike til tekst i eit korpus og ein kan legge inn formell syntaktisk informasjon i tillegg til definisjonar og døme. Det er mogleg å kople mot kjelder for ensyklopedisk informasjon, og ein kan legg inn informasjon om bøyning av ord, informasjon om samansetting av ord, osv. Tekst, lyd, bilete og film i uavgrensa mengde kan knytast til artikkane. Leksikografen si oppgåve blir å organisere all denne informasjonen slik at det er mogleg å finne fram i han.

Utfordringa er at brukarane av ordboka no ventar å finne denne typen informasjon. Forskarar vil døme ordboka etter kor lett det er å etterprøve informasjon ved å følgje ordboka sin eigen dokumentasjon av påstandar. Oppdragsgjeverar og andre ventar at det skal vere mogleg å trekke informasjon ut av ordboka, og bruke dette til talesyntese, maskinomsetting og dikteringssystem. Når det er teknisk mogleg å lage ei slik ordbok/leksikalsk database, så vil alle forvente at dette vert gjort.

Vi veit alle at det ikkje er mogleg å lage denne altomfattande databasen i dag, og sannsynlegvis vil ein heller aldri nå heilt i mål. Men det er ei plikt for oss å skjule til det som ligg i dagens og morgondagens teknologi når vi lagar løysingar. Ein kan ikkje forsvare å lage ordbøker i dag som ikkje kan inngå i ein større heilskap ved eit seinare høve. Det vil vere ein hån både mot fagmiljø, brukarar og dei som betalar for arbeidet.

4. Vegen mot vitskapelege leksikalske ressursar

Er det så i praksis mogleg å lage slike ressursar på eit slikt vis at alle vert nøgd? EDD har forsøkt, og trass i einskilde feilskjer undervegs, har vi kome eit godt stykke på veg.

Men utfordringa ligg ikkje primært i å lage gode system for spesifikke problem. EDD og Norsk Ordbok har i fellesskap skapt fleire ulike databasar som på ulikt nivå tilfredsstillar dei krava vi har sett til det vi kallar vitskapelege databasar. Dess enklare basane er, dess lettare er det å oppfylle desse krava. Databasane over setelarkiva til Norsk Ordbok er døme på banale basar i denne samanhengen. Basane er berre ein digital representasjon av gamle papirbaserte arkiv.

EDD har også utvikla databasane som Norsk Ordbank mellom anna brukar til å registrere bøyingsinformasjon for meir enn 100.000 grunnformar for dei to norske målformene, bokmål og nynorsk, men heller ikkje her ligg det særleg referanse til dømes til empiriske data.

EDD har også utvikla fleire ulike redigeringsapplikasjonar for spesifikke ordbøker. Dei fleste av desse er applikasjonar som ikkje gjer anna enn å registrere den teksten redaktøren vil ha på trykk, og på det viset framstår dei, etter våre krav til vitskap, ikkje som vitskaplege databasar. Det ligg lite eller ingen informasjon som gjer det mogleg å gå til kjeldematerialet.

Databasen for Norsk Ordbok har vi derimot forsøkt å bygge på ein slik måte at alle våre krav til vitskaplege metoder er oppfylte.

Metaordboka er også ein svært enkel konstruksjon, som enkelt forklart er ein indeks til alle dei andre leksikalske ressursane EDD og Norsk Ordbok rår over. Men sidan vi har brukt Metaordboka til å kople saman mange ulike arkiv, ordbøker og ordlister, står samlinga av mindre og enkle databasar i større grad fram som eit vitskapleg heile. Det er i dag mogleg å gå frå ein artikkel i Norsk Ordbok, via Metaordboka, til Norsk Ordbank, og få ut alle brukte bøyingsformar av ordet. Ein kan også gå frå handordbøkene, Nynorskordboka og Bokmålsordboka, via Metaordboka til Norsk Ordbok, og der få utfyllande informasjon om bruksmåtar, dialektformer, osb., og sjølvstilt vidare derifrå til å sjå på kjeldematerialet, anten i form av ordsetlar, eller i form av konkordanselinjer henta frå eit korpus. Samankoplinga gjer det enkelt å navigere mellom dei ulike databasane, og dermed står dei einskilde basane fram med mykje større vitskapleg tyngde enn før.

5. Samarbeid

Frå dette forholdsvise enkle systemet for samankopling av data, ser vi at det er svært mykje å hente ved å kople saman ressursar. Vi har så langt prioritert å kople saman dei ressursane vi sjølv rår over. Vi har også planar om å legge inn eit større tal dialektordbøker i dette systemet, for ytterlegare å auke nytta av databasane.

I framtida ser vi at det kunne vore nyttig å vurdere alle dei store ressursane i Norden med omsyn på ei slik samankopling. Det er svært mange tunge og viktige databasar i Norden, og ved å kople dette saman vil sannsynlegvis leksikografar i heile området få langt meir ut av også sine egne databasar.

Det er klart at dette vil krevje ein god del avklaring, både på den tekniske sida, men også når det gjeld spørsmål om opphavsrett, kommersielle rettar, osb. Vi ønskjer likevel å kome i gang med dette arbeidet, spesielt på den tekniske sida. Det hadde difor vore ønskjeleg at personar med teknisk kompetanse frå dei ulike fagmiljøa i Norden kunne samlast for å diskutere korleis ein skal gå fram for å utvikle metodar for utveksling av data. Eit mål må vere at ein slik samling blir gjennomført før neste konferanse om leksikografi i Norden.

Litteratur

- Ore, Christian-Emil Smith 1998a: Hvordan lage fagdatabaser for humanistiske fag. I: *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*. Oslo: Universitetsforlaget. ISBN 82-00-12670-6. 30 s.
- Ore, Christian-Emil Smith 1998b: Making multidisciplinary resources. I: Lou Burnard, Marilyn Deegan and Harold Short (eds.) *The Digital Demotic, A Selection of Papers from Digital Resources in the Humanities 1997*. Publication 10, Office for Humanities Communication, King's College, London, 1998, ISBN: 1-897991-12-7.
- Ore, Christian-Emil Smith 2001: Metaordboken – et elektronisk rammeverk for Norsk Ordbok? I: Martin Gellerstam et al. (eds.): *Nordiska studier i leksikografi 5. Rapport från Konferens om Lexikografi i Norden, Göteborg 26.-29. maj 1999*. Göteborg, 202-216.
- Ore, Christian-Emil Smith; Tvedt, Lars Jørgen; Bjørnstad, Tone 2002: *The Meta Dictionary ALLC/ACH 2002*; 24.07.2002-28.07.2002.
http://www.edd.uio.no/artikler/leksikografi/meta_dictionary.html
- Text Encoding Initiative Consortium, C.M. Sperberg-McQueen and L. Burnard (red.) 2002: *TEI P4: Guidelines for Electronic Text Encoding and Interchange. Print Dictionaries*, Chapter 12, XML Version. Oxford. <http://www.tei-c.org/P4X/>

Web-sider (alle sider sjekka 22. september 2005):

- ALLEX, <http://www.dokpro.uio.no/allex/allex.html>
- CIDOC, <http://www.willpowerinfo.myby.co.uk/cidoc/>
- Corpus Workbench, <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/>
- Dokumentasjonsprosjektet, <http://www.dokpro.uio.no/>
- Eining for digital dokumentasjon, <http://www.edd.uio.no/>
- Museumsprosjektet, <http://www.muspro.uio.no/>
- Norsk Ordbok, <http://no2014.uio.no/>
- Tekstlaboratoriet, UiO, <http://www.hf.uio.no/tekstlab/>

Sprogbrug og sprogsyn hos flertal og mindretal i den dansk-tyske grænseregion

Language use and language attitudes within the majorities and the minorities in the Danish-German border region. The Danish-German border region is characterised by a traditional linguistic diversity which is determined by national and regional affiliations and the status of the languages and dialects being used within the majorities and the Danish, German and North Frisian minorities. This article discusses the multidimensional linguistic pattern resulting from these groups' use of Danish, the Danish dialect of Sønderjysk, German, Low German, North Frisian and the varieties South Schleswig Danish, North Schleswig German and Flensburg Petuh. And it clarifies that the minority languages ratified as such by the states most often are the minority members' second language acquired outside the home and characterised by language contact phenomena. It also demonstrates that language nationalism has dominated the attitudes but is now being replaced by a language pluralism that corresponds to the diversity.

1. Den traditionelle sproglige mangfoldighed

Den dansk-tyske grænseregion er i dag karakteriseret ved en sproglig mangfoldighed der er bestemt af nationale og regionale tilhørsforhold og sprogenes og dialekternes status. Traditionelt omfatter den dansk, tysk, sønderjysk, plattysk og nordfrisisk samt de sproglige varieteter sydslesvigsk, nordslesvigtysk og Flensborg petuh. Til den nye mangfoldighed hører indvandrersprog som fx tyrkisk, arabisk og russisk, men denne artikel omhandler kun det traditionelle sprog mønster siden 1920.

Dette år blev det historiske Slesvig, som svarer til nutidens grænseregion, delt efter at befolkningen havde stemt om den ville tilhøre Danmark eller Tyskland. Indtil 1864 havde den boet i hertugdømmet Slesvig under det danske monarki, men derefter hørte Slesvig til det tysksprogede Prøjsen. Folkeafstemningerne resulterede i at den nordlige del blev indlemmet i Danmark, en begivenhed som danskerne kalder for Genforeningen og tyskerne for tabet af Nordslesvig. Området betegner danskerne siden hen som Sønderjylland, og her betragter det danske flertal sig som sønderjyder, og de taler sønderjysk og det officielle sprog dansk. For de i dag ca. 15.000 tysksindede er det Nordslesvig, og de anser sig for at være nordslesvigere og udgør et tysk nationalt mindretal. Det har sine egne tyske børnehaver, skoler, biblioteker, kirke og dagblad samt en række foreninger med et kulturelt, socialt og politisk virkefelt. Tysk er det officielle mindretalssprog, men medlemmerne taler også sønderjysk og dansk.

Syd for 1920-grænsen organiserede de dansksindede sig som et dansk nationalt mindretal med dansk som officielt mindretalssprog. De kalder deres hjemegn for Sydslesvig og er enten sydslesvigere eller danskere. Indbyrdes bruger de dansk, tysk,

plattysk eller sønderjysk. Dette mindretal med ca. 50.000 medlemmer har som det tyske sine egne danske børnehaver, skoler, biblioteker og dagblad og et foreningsliv som omfatter kirke og kultur, sociale og politiske forhold.

Den tyske betegnelse for området er Landesteil Schleswig, og hvis tyskerne har en regional identitet, er de slesvig-holstenere. Foruden tysk behersker nogle af dem også plattysk eller sønderjysk, mens de betragter dansk som et nabosprog og som et mindretalssprog.

Nordfrisisk er ligeledes et mindretalssprog, og det beherskes af ca. 10.000 og bliver talt på den tyske del af vestkysten med øer fra grænsen i nord til Bredstedt i syd. Nordfrisisk består i dag af 9 nordfrisiske dialekter, som udgør det nordfrisiske mindretals officielle sprog. Foruden dialekterne taler nordfriserne tysk og nogle af dem også plattysk. Sønderjysk og dansk forekommer, men det er især blandt de nordfriserne som samarbejder med det danske mindretal.

Tysk såvel som dansk og nordfrisisk er ratificeret som mindretalssprog og plattysk som regionalsprog i Sprogpagten, som er Europarådets Europæiske Pagt om regionale sprog eller mindretalssprog (Udenrigsministeriet 2001). Tyskland underskrev den i 1998 og Danmark i 2000, og herefter har ekspertgrupper fra Europarådet fulgt mindretalenes vilkår for udfoldelse og virke på mindretalssproget, og de rådgiver regeringerne om mangler og muligheder.

Det er imidlertid ikke begrænsninger fra staternes side som er årsag til at medlemmerne i de tre mindretal i stor udstrækning har samme modersmål og hjemmesprog som flertalsbefolkningen. Det bunder i sproghistoriske forhold at de fleste tyske nordslesvigere taler sønderjysk derhjemme, og at tysk er hjemmesproget hos de fleste danske sydslesvigere og hos de fleste nordfriserne. I mindretallene falder modersmål og mindretalssprog således sjældent sammen, men medlemmerne tilegner sig mindretalssproget i mindretallets institutioner, og de bruger det fortrinsvis inden for mindretallets foreningsliv. Derved signalerer de deres tilhørsforhold. For mange nordslesvigere og sydslesvigere er mindretalssproget imidlertid mere et nationalsymbol end et brugssprog. De foretrækker at tale deres modersmål. Deres nationale sindelag, dvs. identitet, er for dem det afgørende kriterium for tilhørsforholdet, ikke hjemme- og dagligsproget. For nordfriserne er nordfrisisk et udtryk for både deres identitet og for en samhørighed med egnen, dens natur, kultur og historie, men det har også en symbolværdi eftersom man godt kan være nordfriser uden at beherske eller foretrække at tale en af de nordfrisiske dialekter.

2. Sprognationalisme

En forventning om at der skulle være overensstemmelse mellem hjemmesprog og nationalt tilhørsforhold, bunder i en nationalromantisk forestilling som har været dominerende i mere end et århundrede, ikke blot i grænseregionen, men i de europæiske nationalstater. Den rummer ideen om at den nationale identitet og kultur

er lig med nationalsproget, og at kun det sprog er udtryk for befolkningens solidaritet med nationen. Derfor har det skullet være modersmålet.

En sådan sprognationalisme har været udbredt i grænseregionen og er det stadig i meget nationale kredse. Her er et sprog lig en national identitet og en national identitet lig et sprog, og en manglende overensstemmelse rejser stærk tvivl om det nationale tilhørsforhold. Sådan var det også før 1920, hvor regeringsmagterne på skift påtvang befolkningen et andet sprog, ikke for at gøre den tosproget, men etsproget i et andet nationalsprog og i forventning om at en tilsvarende national identitet fulgte med.

Det er imidlertid ikke kun inden for statusplanlægningen at sprognationalismen stikker hovedet frem. Den forekommer også på korpusplanet i forbindelse med låneord og oversættelseslån. Denne type ord fra andre sprog kan få sprognationalister til at mene at nationalsproget bliver forurenset så det ikke længere er det rene udtryk for nationens enhed og den nationale identitet.

Både på status- og korpusplanet fortsatte dette nationale sprogsyn med at gøre sig gældende i grænseregionen efter 1920. Den danske flertalsbefolkning havde en udpræget negativ holdning til det tyske sprog som den var blevet påtvunget som administrations-, kirke- og skolesprog under det tyske styre. Sprogtvungen havde stødt selvforståelsen hos danskerne og deres nationale identitet, og den førte til sprogforagtelse. Da de ikke længere blev tvunget til at bruge tysk, ville mange ikke længere tale det. Sproget fik en meget lav status. Denne negativitet medførte tillige en afstandtagen til de folk som havde tysk som deres modersmål og talte det til daglig, for sprog og identitet blev anset for at svare til hinanden på grund af den udbredte sprognationalisme. Hvis man talte tysk, var man tysk.

Selv om mange sønderjyder ikke ville tale tysk, var deres danske dialekt blevet påvirket af den tætte kontakt med tysk som officielt sprog. De havde integreret en række udtryk derfra i sønderjysk. Fx blev en dommer omtalt som *amtsrichter*, og en dåbsattest hed en *doopschien*. Sådanne låneord ønskede meget nationale sønderjyder fjernet fra dialekten fordi de var tyske, og erstattet med ord fra rigsdansk. Denne sprogrensning gjaldt ikke blot de ord som var kommet ind i forbindelse med det tyske styre, men også plattyske låneord som sønderjysk gennem århundreder havde været endnu mere rig på end andre danske dialekter (Nyberg 1988:42). Det gjaldt fx ord som *geburtsdag*, der skulle udskiftes med fødselsdag, og en *mantel* skulle hedde en frakke. Ifølge den sønderjyske digter og nationale forkæmper Nikolaj Andersen fandt denne proces især sted blandt dannede folk, hvilket sandsynligvis vil sige dem som læste danske publikationer. I artiklen *Sproglige Ændringer* skriver han således: "Det tidligere ganske almindelige *geborsda* (geburtsdag) f. eks. har nu kun hjemme i de lavere klasser i Danmark, medens det i mere dannede kredse er blevet fortrængt af det danske fødselsdag." (Andersen 1902:238).

3. Sprogdebat i mindretallene

Mens den nationale sprogkamp foregik blandt danskerne nord for grænsen, og alt tysk blev lagt for had af mange sønderjyder, så havde de dansksindede syd for grænsen i stor udstrækning tysk som modersmål og hjemmesprog. Årsagen hertil var et sprogskifte som havde fundet sted de foregående århundreder, og som havde bevæget sig fra syd og var stoppet ved en linje som næsten svarer til 1920-grænsen. Både de som følte en samhørighed med det danske monarki, og de som var orienteret mod Holsten, havde i løbet af denne periode skiftet fra at give børnene sønderjysk som modersmål til at give dem plattysk. Det skyldtes det plattyske sprogs dominerende status. Senere skete der et skifte fra plattysk til højtyisk, også pga. sprogets status, og først derefter slog den nationalromantiske forståelsesramme igennem og dermed sprognationalismen.

De familier som i mange generationer har anset sig for at være danske i Sydslesvig, har derfor stadig tysk som modersmål og hjemmesprog ligesom dem som har valgt at slutte sig til mindretallet efter anden verdenskrig (Pedersen 2005:373). De få familier bosiddende tæt ved grænsen, som sprogskiftet ikke nåede, er dog fortsat sønderjysktalende. Og tilflyttere fra tilhørslandet Danmark har også bragt dansk ind i mindretalsfamilier som modersmål og hjemmesprog (Pedersen 2000: bind 1:189, bind 2:353).

Det officielle danske mindretal har imidlertid ligesom mange danskere i Danmark en national forestilling om at det danske sprog hører til følelsen af samhørighed med den danske nation. Så kan man ikke være dansk uden at beherske det danske sprog og bruge det i dagligdagen. Da de dansksindede i Sydslesvig ikke lever op til dette ideal fra fødslen af og heller ikke opfylder det i deres tysksprogede arbejdsliv, har det ført til en århundredlang sprogdebat i mindretallets dagblad Flensborg Avis, som i øvrigt er delvis på dansk og delvis på tysk. I den første fase rummede læserbrevene krav om at dansksindede skulle give deres børn dansk som modersmål så de kunne blive rigtige danskere. I den næste fase blev det samme fremsat som en moralsk forpligtelse uden begrundelse, men med sprog lig med nationalitet som fælles dagsorden. Senere blev det erkendt at et sprogskifte til dansk hjemmesprog var urealistisk, og debatten kom til at dreje sig om elevernes sprog i den danske skole og deres forældres kundskaber i dansk, men også sprogbrugen i mindretallets foreninger blev diskuteret. Ønsket har været at dansk skulle være enerådende eller i hvert fald være dominerende, så sprogdebatten har altid afspejlet konflikten mellem sprognationalismens etsprogede ideologi og virkelighedens sproglige mangfoldighed (Pedersen 2000, bind 2:261-313).

I Sønderjylland diskuterer det tyske mindretal ligeledes forholdet mellem den nationale identitet og mindretalssproget, men det har ikke et nationalsprog som det mest udbredte modersmål og hjemmesprog som det danske mindretal, derimod har det tyske mindretal den nationalt neutrale dialekt sønderjysk. Godt nok betragtede

sønderjyderne dialekten som deres nationalsprog 'æ sproch' under tysk styre, men sønderjysk er altid blevet brugt af både dansk- og tysksindede nord og syd for grænsen.

Sprogdebatten blandt de tysksindede foregik i de tyske aviser i 1920'erne og 30'erne bl.a. i *Nordschleswigsche Zeitung*. Også her var den tidlige debat præget af ønsket om mindretalssproget som modersmål, men den gled hurtigt over til at dreje sig om at sønderjysk var af uvurderlig værdi fordi det var modersmålet og et dagligsprog der gav samhørighed med hjemegnen. Medlemmerne skulle derimod stræbe efter at forbedre deres færdigheder i tysk, som i øvrigt blev betragtet som det eneste acceptable skriftsprog. Og så var det skolens opgave at lære børnene tysk og motivere dem til at benytte det ved siden af deres modersmål (Pedersen 2005: 365-367).

Dette er et sprogluralistisk syn, hvor andre sprog ikke forhindrer det nationale ståsted, og nationaliteten godt kan være knyttet til et nationalsprog som ikke er modersmålet, men er tilegnet som andetsprog. Sprogdebatten i det tyske mindretal ser ud til at fortsætte i dette spor, og i dag bliver tysk, sønderjysk og dansk anset for at være ligeværdige sprog, men dog med hver deres funktion. Samtidig bliver det fremhævet at den nationale identitet er truet hvis tysk ikke bliver benyttet, og det tyske sprog bliver gentagne gange trukket frem som mindretallets vigtigste markør. Man aner sprognationalismens ideologi, for mindretalssproget skal ikke blot have symbolværdi, det skal være i brug så det kan markere den tyske nationalitet.

Nordfriserne har ikke en tilsvarende problematik i relation til deres mindretalssprog, for der har aldrig eksisteret nogen frisisk nationsdannelse, hverken af nordfriserne alene eller sammen med vestfriserne i provinsen Frisland i Nederlandene og østfriserne i Nedersaksen. Der har til gengæld været et nationalpolitisk modsætningsforhold i 1900-tallet mellem nordfriserne i *Nordfriesischer Verein*, som erklærede sig tysksindede, og dem som stræbte efter en status som nationalt mindretal, nemlig nordfriserne i *Foriining for nationale Friiske*, nu kaldet *Friiske Foriining*. Da den arbejder tæt sammen med det danske mindretal, har de andre nordfriserne ofte anset foreningens medlemmer for at være dansksindede, men i deres selvforståelse er de et nordfrisisk folk. I dag arbejder de to grupperinger imidlertid sammen om at fremme nordfrisisk sprog og kultur i et flersproget og flerkulturelt samfund.

4. Enhed og mangfoldighed

På tværs af statsgrænsen, sprogrænserne og de nationale, regionale eller lokale grænser i hovedet er befolkningen i grænseregionen fælles om hilsnen *mojn*, der skrives *mojn* på tysk, *plattysk* og *nordfrisisk* og bruges uanset om man taler sønderjysk, dansk, tysk, *plattysk* eller *nordfrisisk*. Imidlertid rummer ikke blot stavemåden, men også hilsnens etymologi, dens funktion og holdningen til brugen

af mojn forskelligheder. Der er ikke blot enhed i mangfoldigheden, men også mangfoldighed i enheden. Og dermed kan ordet illustrere det kulørte mønster som mødet mellem sprog og sprogsyn gennem tiderne har resulteret i.

Hilsnen mojn er af tysk oprindelse og stammer fra den dialektale udtale *morjen*, som findes i berlinsk og brandenburgsk. Det svarer til hilsnen *Morgen*. Senere er der tale om en lydovergang fra *morjen* til *mojn/moin*. Ifølge folkeetymologien som i dag flourerer i tyske reklamer og på internettet, skulle hilsnen være af frisisk oprindelse og stamme fra *mojen dag*. Det er imidlertid helt usandsynligt da adjektivet *moje* slet ikke optræder i forbindelse med hilsner i frisisk og for den sags skyld heller ikke i hollandsk. Her bruges *goed*. Og det er tillige svært at forestille sig at det svagt betonedede førsteled *mojen* skulle være bevaret, mens det stærkt betonedede *dag* skulle være faldet væk. Normalt er det omvendt som *goddag*, der bliver til *dāv* (Pedersen 1995, 2001). Tyske sprogforskere afviser således også folkeetymologien (Byl 1989).

Denne hilsen bredte sig til det nordtyske område og dermed Slesvig sandsynligvis fordi den stammede fra de kulturelle, politiske og økonomiske magtcentre, helt ligesom hilsnen *hej* i dag har spredt sig fra København til resten af Danmark. Det blev ikke mojn som blev udbredt. I daglig tale er hilsnen ikke nået ud over Sønderjylland, kun reklamefolk spreder den i dag som symbol på et regionalt tilhørsforhold. Ellers kan den fungere på alle tider af døgnet, men der er forskel på hvad mojn bliver anvendt til, for traditionelt er hilsnen kun et goddag syd for grænsen, mens den bliver brugt som et farvel og en hilsen i forbifarten i Vestsønderjylland og som både et goddag og et farvel i Østsønderjylland. Denne skelnen mellem vest og øst ser imidlertid ud til at være ved at forsvinde.

Da mojn begyndte at blive brugt i Sønderjylland af andre end tyskere og tysksindede i det tyske mindretal, fremkaldte det en reaktion fra mange sønderjyder. De kunne ikke forlige sig med dette tyske låneord. Blandt dem var lærer Mathias Bendix Ottosen fra Hostrup ved Tønder. Han optegnede i 1920'erne og 30'erne den lokale dialekt i en række hæfter med titlen Hostrup-Dansk. Efter hans død blev de udgivet som en ordbog i 5 bind i årene 1963-69 af Institut for dansk dialektforskning ved Københavns Universitet. Fra 1961 foreligger også hans lydlære.

Titlen Hostrup-Dansk viser at det har været Ottosen magtpåliggende at fremhæve at denne sønderjyske dialekt er en dansk dialekt. Tyske forskeres forsøg på at henregne den til de tyske dialekter bliver derved manet til jorden. Ordbogsartiklerne afspejler ligeledes at han finder det vigtigt at signalere hvad han mener om tyske låneord i dansk. Han sammenligner dem med udslæt, altså noget sygeligt, og det skal tørre væk. Med andre ord skal låneordene ikke bruges. Denne holdning kommer til udtryk i forbindelse med fx *æn Svæk*, tysk *Zweck*, og hilsnen *mojn*, som han staver *måjn*. Og interessant nok betragter han hilsnen både som goddag og farvel, hvilket ikke stemmer overens med den senere brug på egnen. Han skriver:

1003. æn Svæk, tysk Zweck.

Tyskerne siger meget ofte: Es hat keinen Zweck! ☹: det er unyttigt, der opnåes ikke noget dermed, det lønner sig ikke, det er ørkesløst. På Dansk siger man altså: "Dæ har 'æ-nø Svæk!" og det høres tit. Man kan ærgre sig derover, man kan gøre Forsøg på at få denne Germanisme udryddet: dæ har 'æ-nø 'Svæk! Man opnår ikke noget dermed. Sådant "Udslet" skal tørre hen af sig selv. (...) Og nu igen det sidste Skrig: "Måjn"! Rest af den smukke Hilsen: Guten Morgen! Ung og Gammel, Store og Små hilser hinanden, når de mødes, når de skilles med: "Måjn" Hu ha! Mensch, ärgere dich nicht! få 'dæ har 'æ-nø 'Svæk = det hjælper ikke.

En sådan negativ holdning til tyske låneord kom ikke blot til udtryk i artikler og ordbøger, men også i de danske dagblade. Det sønderjyske Hejmdal bragte således i 1938 et læserbrev med overskriften "*Møjn*" er "*Ugræs*". Der stod bl.a.:

Vi, som bor der, hvor to Sprog mødes, maa altid være paa Vagt imod, at Ord fra det ene ikke blandes i det andet. Vi bør passe paa, at "Frø af Ugræs" ikke fæster Rod i vort Sprog. I det store og hele har de senere Aar ogsaa virket i rigtig Retning. De fleste sønderjydske "Germanismer" er forsvundet. Men derfor er det saa meget mere paafaldende, at et enkelt Ord har kunnet trænge meget stærkt frem, og det paa et Omraade, hvor der aldeles ingen Grund var til det, da vi har Ord, der dækker bedre og lyder kønnere i vort eget Sprog. "Ugræsset" er den plattyske Hilsen "Møjn".

Læserbrevsskribenten griber til Johan Ottosens gendigtning af folkevisen "Det haver så nyligen regnet", som han skrev i anledning af de sønderjyske nationale lederes besøg hos deres støtter i København i 1890. I første strofe står de berømte ord: "frø af ugræs er føget over hegnet", som er et billede på den tyske indflydelse og det tyske styre i Sønderjylland. Mojn bliver i dette læserbrev henregnet til sådant tysk "ukrudt", og det rummer desuden en æstetisk vurdering af det lydlige indtryk, der falder ud til fordel for danske hilsner. Mojn er med andre ord uønsket og lyder heller ikke godt. Og læserbrevet slutter med et krav om at mojn skal forbydes: "Vi vil gøre 'Møjn' til Løgn. Vi vil ha' det 'forbøjn'".

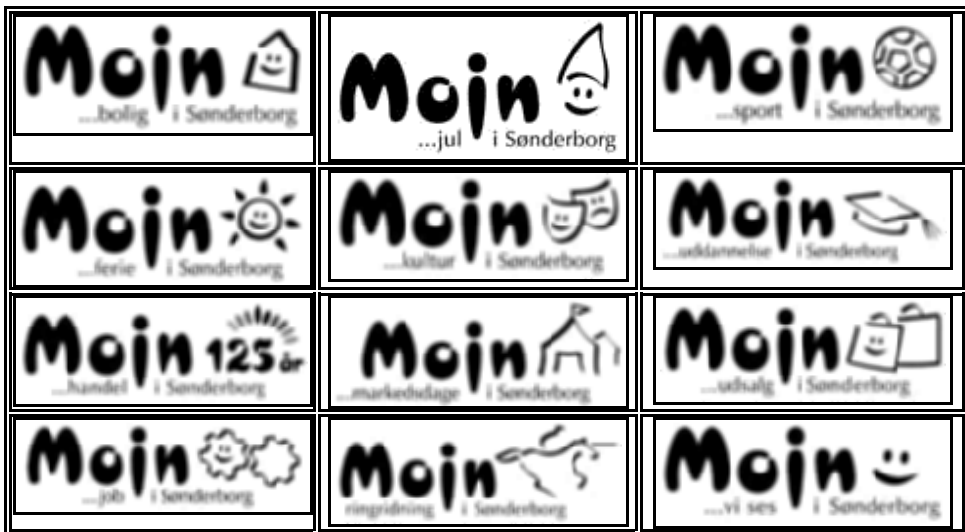
Det er stadig en udbredt holdning blandt dansk-nationale sønderjyder som voksede op omkring 1920, at "mojn er forbojn". Dette forbud har været så stærk en markør for et dansk nationalt sindelag at den efterfølgende generation også er blevet gjort bekendt med det, men børn som er født efter 1970, kender ikke forbuddet. Det viser en spørgeskemaundersøgelse fra 1994 (Pedersen 1995). De er tværtimod vokset op med mojn som en sønderjysk hilsen, en karakteristisk ordet også får i *Alsisk Ordsamling*, mens det bliver defineret som en mindre formel hilsen i *Den Danske Ordbog* med den tilføjelse at det er en almindelig hilsen i Sønderjylland.

4.1. Mojn-kampagner

Hilsnen mojn slog rigtig igennem i 1970'erne efter at Sønderborg Handel havde

markedsført byen med sloganet “Mojn, vi ses i Sønderborg”, hvor bogstavet o i mojn var en smiley med hat på. Det figurerede overalt, ikke blot i forretningerne, men også på T-shirts, krus og posthusets frankeringsstempel. Det sønderjyske bryggeri Fuglsang lavede tilmed en Sønderborg-pils med sloganet på flaskehals-etiketten. Kampagnen blev understøttet af slagsangen *Mojn* som Valdemars Orchestra spillede ved utallige lejligheder, og som opnåede at komme på Dansktoppen.

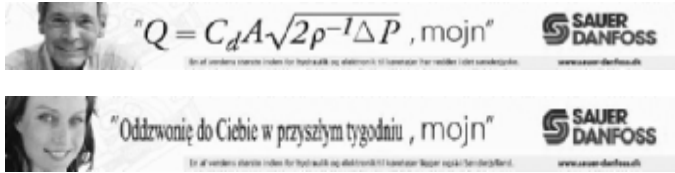
Denne lancering af mojn førte ikke til nogen national rejsning, men flere gamle medlemmer af de nationale foreninger, Sprogforeningen og Sønderjydsk Skoleforening, har over for mig udtrykt deres forargelse over kampagnen, og over at jeg i det hele taget ville beskæftige mig med dette tabu-ord. Ifølge deres opfattelse skulle “det tørre hen af sig selv”, som Ottsen udtrykte det i ordbogen. Mojn visnede imidlertid ikke, men blomstrede i de følgende årtier, og Sønderborg har haft gentagne kampagner med sloganet, hvor smileyen har haft forskellig hovedbeklædning. Da Sønderborg Handel fejrede sit 125-års jubilæum i 2004, kom der et nyt design og en række mojn-sloganer med hvert sit logo:



Fra: www.sdbg-city.dk/Jubilogo

Disse intensive kampagner i Sønderborg og Radio Mojn i Aabenraa har været med til at slå mojn fast som en sønderjysk hilsen. Hertil bidrog også en cd-rom med mojn for tilflyttere, som var vedlagt det satiriske hæfte *Æ rummelpot* 2004. Og tiltagene har nok også fået andre virksomheder til at bruge hilsnen i deres reklamer. I 2005 kører Sauer-Danfoss således en kendskabskampagne hvori mojn indgår. På plakater i læskure ved busstoppesteder og på streamere på busser bliver hilsnen brugt til at

markere det sønderjyske tilhørsforhold og til at vække opsigt både i forbindelse med en matematisk formel og en sætning på polsk. For mange bliver det uforståelige derved kombineret med det velkendte og vækker nysgerrighed.



På tilsvarende vis bliver den tyske dobbelthilsen moin-moin brugt i reklameøjemed syd for grænsen, men i endnu større udstrækning end den sønderjyske hilsen. En ugeavis hedder Moin-Moin, og det mildt sagt vrimler med hilsnen langs den slesvig-holstenske vestkyst, sandsynligvis især her pga. fejlslutningen om ordets frisiske oprindelse. Moin-Moin ses over indgangen til postkontorer og hoteller, på postkort og på souvenirs i alle tænkelige former som fx krus, kasketter, paraplyer og T-shirts. Hilsnen signalerer et regionalt tilhørsforhold, og den sælger. Og som i Sønderjylland tager store firmaer, som fx dagbladet Frankfurter Allgemeine Zeitung, hilsnen i brug i deres regionale kampagner.



Over indgangen til postkontoret i Husum var der i 1996 velkomsthilsnen Moin-Moin og en opfordring på plattysk om at kigge ind.

4.2. Moin/Moin, de unge og identiteten

En grænseoverskridende undersøgelse af unge gymnasieelevers brug og holdning til hilsnen midt i 1990'erne afslørede ligeledes en udbredt brug af moin/moin (Pedersen 1995, 2001). De sønderjyske elever gav udtryk for at moin understregede en følelse af fællesskab, men nogle få tog afstand fra hilsnen fordi den virkede bondsk. De var endvidere enige om at moin var et farvel i Vestsønderjylland og en hilsen i forbifarten, mens den fungerede som både goddag og farvel i Østsønderjylland. I de danske gymnasier ville eleverne stave hilsnen moin eller møjn, mens nogle i det tyske mindretals gymnasium foretrak den tyske stavemåde moin, men betragtede ordet som en sønderjysk hilsen.

Unge som gik på det danske gymnasium Duborg-Skolen i Sydslesvig, forbandt derimod hilsnen med tysk sprog og kultur, og i overensstemmelse hermed mente de fleste at ordet skulle staves moin og var et udtryk for et goddag. Nogle skrev dog at: "I Sønderjylland bliver moin fortrinsvis brugt som farvel, i Nordtyskland som goddag og i Sydslesvig som begge dele da vi er påvirket af begge kulturer." (Pedersen 1995:261). I de tyske gymnasier fremhævede enkelte elever også denne forskel i funktionen, men de var enige om at tyskere kun sagde goddag med hilsnen, og at de stavede den moin. De brugte den alle som enkeltord eller i den redupliserede form moin-moin.



En sydslesviger reklamerer for et løb arrangeret af Sydslesvigs Danske Ungdomsforeninger i 1997.

I Sønderjylland er mojn således gået fra at være et nationalt stridsmærke til at være et folkeligt symbol på en sønderjysk identitet. Flere medlemmer af det tyske mindretal betragter derimod hilsnen både som en sønderjysk og som en slesvigsk hilsen, der kan signalere en regional grænseoverskridende slesvigsk identitet. I Slesvig-Holsten er en sådan slesvigsk identitet i forbindelse med hilsnen mojn ikke blevet beskrevet, men derimod bliver mojn hæftet på en nordtysk eller frisisk identitet, mens den i det danske mindretal ses som udtryk for en tysk og en sønderjysk identitet samt en dansk mindretalsidentitet som sydslesviger.

5. Standardisering og sproglig variation i flertal og mindretal i Sønderjylland

5.1. *De sønderjyske dialekter og rigsdansk*

I Sønderjylland bliver hilsnen mojn brugt både i sønderjysk og rigsdansk og af indfødte og tilflyttere i Sønderjylland og er derfor ikke et barometer for hvorledes det går de sønderjyske dialekter. Der er stadig forskelle i dialekterne fra øst til vest og fra nord til syd, men de bliver mere og mere udviskede, og der er næppe tvivl om at dialekterne er vigende til fordel for et sønderjysk regionalsprog. Tilmed oplever bedste- og oldeforældregenerationen, som lærte deres børn sønderjysk, at deres børn og børnebørn ikke giver dialekten videre. De blev i modsætning til forældrene tosprogede med rigsdansk allerede i børnehaven og skolen og skulle jo i teorien nemt kunne være blevet bidialektale. Det blev nogle af dem også, men dialekter har så ringe en status i skolen og i Danmark som helhed at de for mange rangerer på linje med andre fænomener fra et uddødt bondeliv som egnsdragter og hesteplov. De finder at det er ude af trit med et moderne liv at tale sønderjysk med børnene, og måske taler deres ægtefælle også kun rigsdansk. Men der er selvfølgelig sønderjyder som vælger at give sønderjysk videre. Det er dog ofte i form af et sønderjysk regionalsprog med få dialektale træk.

De børn som vokser op med sønderjysk i dag, bliver som deres forældre bidialektale, men en undersøgelse blandt unge i Tønder foretaget af Marie Maegaard (2000) viser at de på den ene side holder af sønderjysk og på den anden side gerne vil være ligesom dem der taler rigsdansk. De ligger under for et pres fra en rigsdansk højstatusnorm. Når de blev interviewet, var de således stolte over at kunne tale sønderjysk, men når de blev udsat for en test, så vurderede de lokalsproget lavest og københavnsk sprogbrug højest.

Blandt de tysksindede i det tyske mindretal har mange stadig sønderjysk som modersmål og tysk som nationalsprog. Den sproglige standardisering inden for dansk i Danmark er imidlertid begyndt også at påvirke denne gruppe. Tidligere talte de tysksindede ikke rigsdansk i Sønderjylland, for dette sprog blev betragtet som ensbetydende med et dansk nationalt tilhørsforhold fra både tysk og dansk side, men

i dag vælger de det ofte i samtale med rigsdansktalende, og formodentlig taler nogle forældre rigsdansk med deres børn (Pedersen 2005:347-49). Så måske går min spådom om at mindretallet ville være de sønderjyske dialekters beskytter, ikke i opfyldelse (Pedersen 1992:135).

5.2. Nordslesvigtyisk og højtysk

Eftersom dansk er det officielle sprog i Sønderjylland, og mange i det tyske mindretal har sønderjysk som modersmål, er det danske sprog kilde til at mindretalssproget tysk fremtræder i en varietet som kaldes nordslesvigtyisk. Den adskiller sig fra det tyske standardsprog, højtysk, på flere punkter. Byram (1986:67,149) finder at danske prosodiske træk er det mest karakteristiske for nordslesvigtyisk, og Pedersen (2000b:213) tilføjer bl.a. leksikalske særtræk som ord til ord-oversættelser. Da denne tyske varietet bliver talt inden for mindretallet, hører den til en gruppe med en fælles etnicitet og kan derfor betegnes som en etnolekt. Dertil kommer at mindretallet fuldt ud accepterer nordslesvigtyisk, det er kun i skolen at højtysk er den gældende norm, men dette ideal lever ikke alle lærere og elever op til.

Der er dog mindretalsmedlemmer som behersker højtysk, og nogle af dem vælger denne form for tysk i kommunikation med tyskere fra Tyskland, mens de foretrækker nordslesvigtyisk i samtale med andre fra mindretallet. De markerer derved en samhørighed med dels tilhørslandet, dels mindretallet, og medlemmerne har således et mangfoldigt sprogligt repertoire som de benytter i forskellige sociale relationer, mundtligt og skriftligt.

5.3. Navne, ordbøger og retskrivning på sønderjysk

Mens de sønderjyske dialekter stadig bliver givet videre til børnene inden for det tyske mindretal, så ser de ud til at være vigende som modersmål hos flertalsbefolkningen. De kommer så til gengæld i stigende grad på tryk i enkeltstående sønderjyske ord i aviser og i navne på vejskilte som Æ. Stajsvej (Stadsvejen) og forretningsnavne som Æ' Bix. Og forretningerne kan føre produkter med dialektale navne som Æ Tynnevin (Tøndervinen), og deres reklamer kan være helt eller delvis på sønderjysk (Pedersen 1993, 2003). Som i eksemplerne bliver dialekten ofte signaleret ved hjælp af det foransatte bestemte kendeord 'æ' eller 'e', men ordet volder problemer, for ikke alle ved hvad det er en størrelse. Nogen tror det er en forkortelse og sætter punktum eller en apostrof, mens andre mener at apostroffen er en trykstreg selv om den står efter bogstavet. Men ingen af disse ortografiske eller lydligge begrundelser holder stik, og apostroffen er udtryk for en ren og skær apostrophitis dialektos, dvs. syge apostroffer eller apostroffer hvor de ikke skal være (Pedersen 2002).



Æ' Bix er en lille købmandsbutik i Genner ved Aabenraa, som åbnede i 2003.

5.4. Sønderjysk dialektbevarelse

Skiltene og navnene med dialektale indslag kan betragtes som sproglige mindetavler, men også som symboler på en sønderjysk identitet. De er led i sprogplanlægning med det formål at bevare sønderjysk. Den finder også sted i Alsingergildet og Æ Synnejysk Forening, som har stor lokal mediebevågenhed. Alsingergildet blev startet i 1946 som et alsisk dialektselskab, og det er et selvsupplerende gilde. Det har bl.a. udgivet *Alsisk Ordsamling*, som er en dialektordbog redigeret af fagfolk. Der er udtale, betydning, etymologi, eksempler på hvordan ordene bliver brugt, og så lidt kulturhistorie. Æ Synnejysk Forening er derimod en åben forening, som blev stiftet 2000, og den har 1719 medlemmer i juni 2005. De mener at de kun kan blive ved med at kalde sig sønderjyder hvis de bevarer det sønderjyske sprog i tale og skrift. Foreningen har en hjemmeside og et medlemsblad på sønderjysk, og dialekten bliver også brugt ved medlemsmøder og ved foreningens gudstjenester inden for de danske og tyske menigheder i Folkekirken. Det er således en forening for både danske og tysksindede.

Af foreningens andre aktiviteter kan nævnes udgivelsen af *Min føøst tusin' oe å synnejysk* (Amery 2002), som er en oversættelse til sønderjysk af Elin Fredsted af billedbogen *The first thousand words* (Amery 1995). Denne sønderjyske udgave er i øvrigt anmeldt i LexicoNordica (Nielsen & Zethsen 2004). Desuden har foreningens hjemmeside to ordbøger fra sønderjysk til rigsdansk, Egon Maarup: *synnejysk*

sproch og Yrsa Lønne Hansen: *Så'en snakker vi i Ballum* (www.synnejysk.dk). Et par andre medlemmer har også udgivet ordbøger. Det er Else Christensen med *Oespråg å Oebogh o synnejysk. Ordsprog og ordbog på sønderjysk* og Herbert Matzen, der har udgivet *Sønderjyske ord på Højeregnens dialekt* på opfordring af og med støtte fra Fonden til udgivelse af egnsbeskrivelser fra Højeregnen. Den er fra sønderjysk til rigsdansk, men er suppleret med et privat særtryk med ordene fra rigsdansk til sønderjysk.

Det sønderjyske bryggeri Fuglsang har også bidraget til bevaring af sønderjysk. I 2002 producerede det omkring 3 mio. øl som havde en flaskehalsetiket med en sønderjysk vending fra Else Christensens samling. Dette bliver omtalt i forordet til hendes ordbog af formanden for Æ Synnejysk Forening, hvor han skriver at: "Disse vendinger har været med til at få gang i snakken omkring det sønderjyske sprog. I takt med disse tiltag vil den nødvendige bevidsthed omkring vores sønderjyske sprog blive styrket."

5.5. Sønderjysk retstavning

Formandens definition af dialekten som et sønderjysk sprog stammer fra tiden under tysk styre og viser at han ønsker at sønderjysk skal rangere på lige fod med andre sprog. Derfor er det også magtpåliggende for foreningen at udvikle et skriftsprog. Dens første bidrag til en standardisering af de sønderjyske dialekter skete med *Min føøst tusin' oe å synnejysk*, som ifølge forordet er "et forsøg på at skabe en nogenlunde konsekvent sønderjysk retskrivning". Alsingergildet har efterfølgende publiceret en standardiseret retskrivning for de alsiske dialekter, som adskiller sig fra foreningens. Et eksempel er det foransatte bestemte kendeord som bliver skrevet 'e' fx *e huss* (huset) (Alsingergildet 2004:2) i modsætning til Æ Synnejysk Forenings brug af 'æ' i dens navn og i dens ordbog selvom forordet angiver *e/æ*, fx *æ huus* (huset) (Amery 2002:49). En sammenligning af de to eksempler viser endvidere at ordet *hus* staves forskelligt. På alsisk er der en kort vokal med accent 1, hvilket bliver markeret med dobbeltskrivning af efterfølgende konsonant, mens vokalen kan være lang på et regionalt sønderjysk, hvilket er angivet med dobbeltskrivning af vokalen.

Disse eksempler viser hvor svært det vil blive at nå frem til en autoriseret retskrivning for sønderjysk som alle sønderjyder kan være enige om. Der har således også i årevis været strid om hvorledes det foransatte bestemte kendeord skulle staves, men den lagde dialektologen Magda Nyberg låg over ved at give en begrundelse for brugen af 'æ' og 'e' i Sønderjysk Månedsskrift i 1993. Heri skriver hun at 'æ' svarer til den æ-agtige udtale af kendeordet vest for Hovedvej 10, mens den mere e-agtige udtale øst for hovedvejen har resulteret i 'e'. Dette forhold bliver afspejlet på morsom vis på vejskiltet Æ. Stajsvej, som står i Bolderslev vest for hovedvejen. Herpå er 'E' blevet rettet til 'Æ', men det har ikke været muligt at få oplyst hvornår det har fundet sted.



Æ. Stajsvej (Stadsvejen) er vejen ud af Bolderslev og mod købstaden Aabenraa. Skiltet er fotograferet i 1992.

5.6. Sprogpluralisme

Nogle af ordbøgerne og cd-rommen med mojn er med til at vise en ændret holdning til brugen af sønderjysk. Det er ikke længere forbeholdt de indfødte nu hvor de fleste taler et sønderjysk regionalsprog. Sønderjyderne har øjensynlig ikke længere en dialektal målestok. Den blev tidligere anvendt over for tilflyttere som forsøgte at tale sønderjysk, og de kom som regel under mindstemålet. I sønderjydernes ører kunne de ikke tale det *ret*, dvs. rigtigt, og derfor skulle de lade være selv om de måske havde boet på egnen det meste af deres liv. I dag beretter flere sønderjyder derimod med en vis stolthed i stemmen om tilflyttere og indvandrere som lærer sønderjysk på arbejdspladsen og i lokalmiljøet. Den regionale identitet og det regionale sprog er nu for alle, måske fordi den ellers ville dø ud.

Sprogbrugen og sprogsynet hos både flertallet og mindretallet i Sønderjylland er dermed i stigende grad præget af sprogpluralisme. Tysk har høj status som mindretalssprog og nabosprog, og sønderjysk er for alle. En stærk indikator herfor er at Sprogforeningen, som blev oprettet for at bevare dansk sprog under tysk styre, i 2004 udgav Laurids Koks vise om Danevirke "Danmark dejligst vang og vænge" på både dansk og tysk (Karstoft 2004).

6. Sproglig mangfoldighed i Sydslesvig

6.1. Sønderjysk, sydslesvigsk og rigsdansk

I det lille område syd for grænsen, hvor sproskiftet til plattysk ikke slog igennem, er sønderjysk gået så meget tilbage at dialekten kun bliver givet videre til børnene i nogle enkelte familier, men den bliver holdt i live af bl.a. teatergruppen *Æ Amatører*. Den består af medlemmer af det danske mindretal. Så ligesom det i Sønderjylland kan blive det tyske mindretal som i længst tid bevarer sønderjysk i Danmark, kan det blive det danske mindretal som bevarer sønderjysk i Tyskland.

Dette danske mindretals officielle sprog er dansk, som findes i to varieteter. Foruden rigsdansk er der sydslesvigsk, som fortrinsvis beherskes af dem som har tilegnet sig dansk som andetsprog i mindretallets institutioner. De fonologiske, grammatiske og semantiske særtræk som karakteriserer sydslesvigsk i forhold til rigsdansk, er med få undtagelser overført fra tilsvarende træk i tysk. Især de leksikalske særtræk i dansk adskiller mindretallets medlemmer fra tyskere som har dansk som fremmedsprog. De er fuldt ud accepteret og er i høj grad medvirkende til at sydslesvigsk kan karakteriseres som en etnolekt ligesom nordslesvigstysk. Særtrækkene falder i 3 kategorier: Sydslesvigismen er oversættelseslån i stedet for ord brugt af etsprogede i Danmark. Fx kaldes en pedel eller vicevært for *husmester*, oversat fra tysk *Hausmeister*. Anden kategori er danske sydslesvigord, som også er oversættelseslån, men de betegner forhold eller ting som ikke findes i Danmark, som fx en *socialstation* fra tysk *Sozialstation*. Det er en institution hvor man kan søge hjælp hos socialrådgivere, sygeplejersker og læger. Endelig er der de gammeldags parallelord som *sørgehøjtidelighed*, en parallel til tysk *Trauerfeier*, hvor danske i Danmark nu til dags foretrækker *begravelse*. En række af sådanne sydslesvigismer og danske sydslesvigord er samlet i *Ordbog over Det Danske Sprog i Sydslesvig* (Skov 1997).

I daglig tale er sydslesvigsk værdsat af de fleste sydslesvigere, men i skolen er rigsdansk stadig højstatusnormen. Det er også tilfældet inden for skriftsproget, og flere sydslesvigske forfattere er blevet kritiseret for sprogkontaktfænomener i deres værker. De ser dem dog i lighed med andre som identitetsskabende elementer og ønsker ikke at få dem rettet. De vil betragtes som sydslesvigere med en mindretalsidentitet. Måske har Flensborg Avis et lignende ønske om at markere at den er en mindretalsavis, for den bærer også præg af sprogkontakten.

Til mindretalsidentiteten hører for manges vedkommende tysk som modersmål og en loyalitet over for dette sprog. De er tosprogede og bruger begge sprog i mindretalsrelationer. Dansk finder de fortrinsvis anvendelse for inden for mindretallets institutioner og foreninger og først og fremmest med de ansatte og lederne, og bestyrelsesposter kræver at man taler dansk. På tomandshånd i samtaler af privat karakter i mindretalsregi kan sprogvælget derimod falde ud til fordel for både dansk og tysk eller et skifte mellem sprogene. Blandt de unge i mindretallets skoler har

dette skifte ofte karakter af at være kodeskift som helt integrerer sprogene, som regel tysk, sydslesvigsk, rigsdansk, engelsk og udtryk fra indvandrersprogene. De er ved at etablere en global sprognorm som også er beskrevet i storbyer med mange indvandrere (Quist 2000).

6.2. *Flensborg petuh*

Mens de unges blandingsprog er dynamisk og levende, er en anden blandingsvarietet i grænseregionen ved at uddø. Det er Flensborg petuh, som er en blanding af højtysk, plattysk, dansk og sønderjysk. Det er tidligere blevet kaldt Petuhtantendänisch af tyskerne, mens danskerne betegnede det som petuhtantetysk, velsagtens fordi ingen af dem ville kendes ved blandingsproget. Det stred jo mod sprognationalismens standardisering. I dag bliver det kort og godt kaldt petuh eller Flensborg petuh.

Varieteten fik sit navn fra ordet *partout*, som fører os tilbage til tiden før 1920. Der kunne man købe et partoutkort til at sejle med damperne på Flensborg Fjord om sommeren. Det var en yndet fornøjelse for de velstående fruer i Flensborg, og de blev derfor kaldt petuhtanter. Damperne lagde til på både nord- og sydsiden af fjorden, og passagererne kunne gå fra borde og drikke kaffe på restauranterne uden pas og toldkontrol da der var tysk styre på begge sider. Men der var forskelle i det sproglige landskab, og petuhtanterne kom selv fra en by med blandet sprogbrug. Det højere borgerskab talte højtysk, de øvrige plattysk og specielt i byens nordlige og vestlige udkant blev der talt sønderjysk. Og der var vel nogle som stadig brugte den flensborgske, sønderjyske bydialekt. Rigsdansk var der muligvis også som et levn fra dansk som skolesprog før 1864.

Petuh er således et produkt af kontakten mellem alle disse sprog og dialekter især blandt petuhtanterne. De findes ikke mere, og sejladsen er så at sige ophørt, men de ældre indfødte flensborgere elsker stadig at bruge petuhudtryk. Derved signalerer de deres tilhørsforhold til byen, og så sker der det samme som med sønderjysk i Sønderjylland, petuh bliver skrevet ned, og der bliver udgivet ordbøger af lokalpatrioter. En af dem er *Petuh ABC* af W.L. Christiansen. Den indeholder både historier som i en læsebog, vel deraf navnet ABC, men også en ordbog fra petuh til tysk. Forfatteren præsenterer sig selv sådan her: "Sze, ich hab viele jahre as fremdnführer chearbeitet un deszwech nimm ich szi nu mit auf ein kleine stadtführung auf petuh." (Christiansen 2004:16). Stavemåden afspejler petuhudtalen, hvor 's' foran vokal er et ustemt, men meget skarpt 'sz', szollen (sollen/skulle) og 'g' i forlyd foran vokal 'ch', chehören (gehören/tilhøre). Der er endvidere et specielt ordforråd med oversættelseslån, og bøjninger og sætningsstrukturer afviger fra standardgrammatikken. Fredsted (1983:30) giver følgende eksempel: "Was'n tumpigen Schnack un kommen mit, szon tumpiges Wort" (sikken en tosset snak at komme med, sikken et tosset ord), hvor højtysk, plattysk og sønderjysk er blandet sammen.

6.3. Plattysk

Plattysk kaldes også nedertysk og er betegnelsen for en række tyske dialekter i området fra grænsen og sydpå til Aachen i vest og Frankfurt an der Oder i øst. Det var officielt sprog indtil 1650, men fik derefter status som dialekt, men mange fortsatte med at skrive på plattysk. Der bliver fortsat skrevet litteratur på plattysk og artikler og historier i lokalaviser, og der er plattyske teaterforestillinger fx på *Niederdeutsche Bühne* i Flensborg og radioudsendelser i RSH (Radio Schleswig-Holstein) og NDR (Norddeutscher Rundfunk). Tv-udsendelser på plattysk forekommer også.

I Sydslesvig er der stadig mange som kan forstå eller tale plattysk. Det fremgår af en undersøgelse fra 2004 foretaget af Zentrum für Niederdeutsch in Landesdeel Schleswig, hvor 91% af folk på landet forstår plattysk, og 57% kan stadig tale det. Der er imidlertid kun få børn som får plattysk som modersmål. Nogle børn lærer senere lidt plattysk i familien eller lokalsamfundet fx vha. den plattyske version *Mien Eerste Dusennd Wöör* af Heather Amerys ordbog (1998). De kan også lære noget i de skoler hvor plattysk indgår i undervisningen enten som et valgfag eller som en dimension i tyskundervisningen. Hertil bliver der fremstillet en del undervisningsmateriale fx madopskrifter og små fortællinger på plattysk, og lærerne kan deltage i efteruddannelseskurser i plattysk. Desuden er der hvert år en konkurrence om hvilken skoleelev som er bedst til at læse op på plattysk. Den bliver arrangeret af Schleswig-Holsteinischer Heimatbund, som er meget aktiv i bestræbelserne på at fremme og bevare plattysk.

6.4. Nordfrisisk

Der er ligeledes foreninger som støtter og aktivt er med til at bevare nordfrisisk på Slesvig-Holstens vestkyst. Sproget kom med indvandrere fra syd, som bosatte sig i området omkring 800 og igen 300 år senere. Der er ikke noget standardsprog, men nordfrisisk består i dag af de 5 fastlandsdialekter Wiedingharderfrisisk, Bökingharderfrisisk, Karrharderfrisisk, Nordergoesharderfrisisk, Mittelgoesharderfrisisk og Halligfrisisk og de 3 ødialekter Sildfrisisk, Før/Amrumfrisisk og Helgolandfrisisk. Forskellene på dialekterne er på nogle punkter så store at de giver forståelsesproblemer, fx hedder et bord både Staal, boosel, Taffel, skiuw og scheew. Dertil kommer at der ikke er en fælles nordfrisisk skriftsprogsnorm, for hver dialekt har sine ortografiske konventioner (Walker & Wilts 2001:284-313), men når der er behov for faglig assistance til at skrive på dialekterne fx til undervisningsmateriale, træder Nordfrisische Wörterbuchstelle ved Kiels Universitet hjælpende til. Her bliver der arbejdet på ordbøger over de nordfrisiske dialekter. Nordfriisk Instituut i Bredstedt bidrager også, og det har bl.a. udgivet oversættelser til 6 forskellige nordfrisiske dialekter af Heather Amerys *The first thousand words*. Udgaven på Bökingharderfrisisk, som også kaldes Moringer, hedder: *Min jarste duusennd uurde. Frasch* (Amery 1998).

Nordfrisisk bliver først og fremmest brugt inden for familien og med venner, men i forretninger og på gaden høres det sjældent. I det offentlige rum forekommer det desuden ved teaterforestillinger og i de nordfrisiske foreninger og ved særlige lejligheder i kirken. Man kan også høre nordfrisisk i Welle Nord i NDR (Norddeutscher Rundfunk) nogle minutter om ugen, men ikke i tv-udsendelser. I de skrevne medier forekommer nordfrisisk sjældent, for nordfriserne har ikke deres eget dagblad som det danske og det tyske mindretal, men indimellem bringer lokalaviserne artikler på en af dialekterne, og i de private rubrikannoncer forekommer de også. Desuden har tidsskriftet *Nordfriesland*, som udgives af Nordfriisk Instituut, artikler på nordfrisisk, og der bliver skrevet en del litteratur på dialekterne. Dette institut udgiver også undervisningsmateriale som bliver fremstillet på de enkelte dialekter af engagerede lærere og pædagoger.

De er med til at fremme udviklingen af det nordfrisiske sprog hos børnene. I 14 børnehaver indgår nordfrisisk således i det sproglige repertoire i form af sange og lege, mens 2 børnehaver på øerne har en decideret tosproget tysk-frisisk dagligdag. Inden for skolerne er det kun på Risum Skole under Dansk Skoleforening for Sydslesvig at nordfrisisk både er et undervisningssprog sammen med dansk og et obligatorisk fag. På 3 andre danske skoler og 27 offentlige skoler er nordfrisisk derimod et frivilligt fag (Petersen 2004).

Selv om nordfrisisk hverken har en fremtrædende plads i dagligtale i lokalsamfundet eller i medierne, så bliver sproget stærkt markeret i det offentlige rum via tosprogede byskilte med tysk og nordfrisisk tekst og etsprogede vejskilte med nordfrisiske navne.



I Nordfrisland blev det tilladt at sætte tosprogede byskilte op i 1997. Dette eksempel er fra Niebüll, som hedder Naibel på Bökingharderfrisisk. Denne dialekt bliver også kaldt Moringen.

7. Grænseoverskridende sprogbrug og sprogsyn – en konklusion

De som behersker en af de sønderjyske eller nordfrisiske dialekter, et plattysk eller et sønderjysk regionalsprog eller en af de sproglige varieteter sydslesvigsk, nordslesvigtyisk eller Flensborg petuh, har som regel også en kompetence i et standardnært dansk eller tysk, evt. rigsdansk eller højtysk. Derved er de bidialektale eller tosprogede. Det er imidlertid først og fremmest medlemmerne af det danske og tyske og til dels det nordfrisiske mindretal som er karakteriseret ved en dansk-tysk tosprogethed i dag. De kan kommunikere frit og ubesværet på dansk og tysk.

I flertalsbefolkningerne behersker en del dog dansk eller tysk som fremmedsprog, især sønderjyder som har tilegnet sig tysk gennem undervisning i skolen og via tysk radio og tv siden 1960'erne. De unge sønderjyders færdigheder i tysk er imidlertid vigende fordi de fravælger tysk som fag i skolen, og mange foretrækker engelsksprogede tv-programmer. Det samme mediebillede hersker i Sydslesvig, hvor ikke alle skoleelever har mulighed for at vælge dansk som fremmedsprog. Trods disse begrænsninger er der en stigende bevidsthed om nabosprogets betydning og en anerkendelse af dansk som et sprog ligeværdigt med tysk. Det har fx været muligt at oprette en grænseoverskridende dansk-tysk gymnasieklasse, som bliver undervist i dels Tønder, dels Niebüll af danske og tyske lærere på dansk og tysk. Når de krydser grænsen, ser de officielle vejskilte på nabolandets sprog i deres hjemland. Det er et resultat af Sønderjyllands Amts sprogpolitik for amtets vejskilte, som er i modstrid med Trafikministeriets regler for statens vejskilte, men som svarer til det tyske trafikministeriums retningslinjer fra 2000 for vejvisning i Tyskland. Her skal grænseoverskridende vejvisning til mål i grænselandet udelukkende skrives på originalsproget, men tosproget hvis bynavnene på de to nabolandes sprog er meget forskellige.





Sådanne positive holdninger til nabolandets sprog er med til at give den sproglige mangfoldighed mulighed for at fortsætte med at eksistere, og de giver sprognationalismen ringe kår. Dertil bidrager også mindretallenes erkendelse af at mindretals-sprogets status er lav som hjemmesprog og kun høj inden for en begrænset række områder. Deres satsning på tosprogethed og et bredt sprogligt variationsmønster kunne være et forbillede for sprogbrug og sprogsyn i såvel Danmark som Tyskland.

Litteratur

Ordbøger:

- Alsisk Ordsamling*, 1. udgave 1988, 2. udgave 2002. Augustenborg: Alsingergildet.
- Amery, Heather 1995: *First Thousand Words*. London: Usborne Books.
- Amery, Heather 1998a: *Mien Eerste Dusedn Wöör. Platdüütsch*: Karl-Peter Köpp. Bredstedt: Nordfriisk Instituut.
- Amery, Heather 1998b: *Min jarste duusend uurde. Frasch*: Alfred Boysen et al. Bredstedt: Nordfriisk Instituut.
- Amery, Heather 2002: *Min føøst tusin' oe å synnejysk*. Oversat fra First Thousand Words til sønderjysk af Elin Fredsted. Aabenraa: Æ Synnejysk forening.
- Christensen, Else 2003: *Oespråg å Oebogh o synnejysk. Ordsprog og ordbog på sønderjysk*. Skærbæk: Eget forlag.
- Christiansen, W.L. 2003: *Petuh-ABC*. Goldebek: Mohland.
- Den Danske Ordbog 2003-2005*. København: Gyldendal og Det Danske Sprog- og Litteraturselskab.
- Matzen, Herbert 2004: *Sønderjyske ord på Højeregnens dialekt*. Højer: Eget forlag.

- Ottsen, Mathias B. 1961: *Hostrup-Dansk. Lydlære*. (Udg. Ella Jensen). Udvalg for Folke-
maals Publikationer. Serie A nr. 20. København: Københavns Universitet.
- Ottsen, Mathias B. 1963-1969: *Hostrup-Dansk. II Ordbog*. Bind 1-5. (Udg. Ella Jensen og
Magda Nyberg). Udvalg for Folkemaals Publikationer. Serie A nr. 21.1-21.5. Bind 1-3
1963-65, bind 4-5 1966-69. København: Københavns Universitet.
- Skov, Christian 1997: Ordbog over Det Danske Sprog i Sydslesvig. I: SSF's Presseudvalg
1998: *Lille Klaus – en cd om Sydslesvig*. Flensborg: Sydslesvigsk Forening.

Anden litteratur:

- Alsingergildet 2004: *Alsisk skriftsprog Vejledning i skrivemåde*. Augustenborg: Alsingergil-
det.
- Andersen, Nikolaj 1902: Sproglige Ændringer II. I: *Sønderjydske Aarbøger*, 161-240.
- Byl, Jürgen 1989: MOIN! Die lange Erklärung eines kurzen Grusses. I: *Ostfriesland* 1, 10 ff.
- Byram, Michael 1986: *Minority Education and Ethnic Survival. Case Study of a German
School in Denmark*. Clevedon: Multilingual Matters.
- Fredsted, Elin 1983: Spiser man mon børn i Flensborg. I: *Mål og Mæle* 9/1, 27-31.
- Karstoft, Christian B. (red.) 2004: *Om Tyre Danebod. Laurids Koks vise om Danevirke fra
omkring 1690*. Flensborg: Sprogforeningen.
- Maegaard, Marie 2001: "Jeg er da stolt af at jeg er sønderjyde – altså sådan forholdsvis!" Om
sprogbrug og sprogholdninger hos sønderjyske unge. I: *Danske Talesprog* 2. København:
Københavns Universitet, 77-166.
- Nielsen, Martin og Karen Korning Zehtsen 2004: Vondt snakke man å synnejysk, vondt
skriwe man å synnejysk? I: *LexicoNordica* 11. Oslo, 299-307.
- Nyberg, Magda 1988: *Tyske låneord i sønderjysk*. Dialektstudier 5. København: Institut for
Dansk Dialektforskning.
- Nyberg, Magda 1993: Skal det foransatte bestemte kendeord i sønderjysk skrives med e eller
æ? I: *Sønderjysk Månedsskrift*, 117-118.
- Pedersen, Karen Margrethe 1992: Det levende sprog. Har du snakket sønderjysk med dit
barn i dag? I: *Sønderjysk Månedsskrift*, 131-135.
- Pedersen, Karen Margrethe 1993: Æ trinbræt – En studie over det dialektale skriftsprog i
Sønderjylland i 1990'erne. I: *Jyske Studier*. København: Københavns Universitet, 134-49.
- Pedersen, Karen Margrethe 1995: MOJN – en hilsen i Sønderjylland i 75 år. I: Henrik Becker
Christensen (red.) *Grænsen i 75 år*. Aabenraa: Institut for grænseregionsforskning & Told-
og Skattehistorisk Selskab, 244-268.
- Pedersen, Karen Margrethe 2000: *Dansk sprog i Sydslesvig*. Bind 1 og 2. Aabenraa: Institut
for grænseregionsforskning.
- Pedersen, Karen Margrethe 2000b: German as First Language and Minority Second
Language in Denmark. I: Gabrielle Hogan-Brun (ed.) *National Varieties of German
outside Germany*. Oxford: Peter Lang, 195-221.
- Pedersen, Karen Margrethe 2001: Mojn og hej. *Pluk* nr. 3, 1-11.
- Pedersen, Karen Margrethe 2002: Bestemt kendeord med og uden apostrophitis. I: *Mål og
Mæle* 25/4, 14-22.
- Pedersen, Karen Margrethe 2003: Navne i Sønderjylland. I: *Navnet er trofast. Modersmål-
Selskabets Årbog 2003*, København, 111-123.
- Pedersen, Karen Margrethe 2005: Realiteter og rettigheder inden for sprog og identitet – i det
danske og det tyske mindretal før og efter København-Bonn Erklæringerne i 1955. I:

- Jørgen Kühl (red.) *København-Bonn Erklæringerne 1955–2005. De danske mindretalserklæringers baggrund, tilblivelse og virkning*. Aabenraa: Institut for Grænseregionsforskning, 339-399.
- Petersen, Adeline 2004: Liir friisk! Friesischunterricht – Bilanz und Ausblick. I: *Nordfriesland* 146, 22-23.
- Quist, Pia 2000: Ny københavnsk “multietnolekt”. Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer. I: *Danske Talesprog*, bind 1, 143-211.
- Udenrigsministeriet 2001: *Den Europæiske Pagt om regionale sprog eller mindretalssprog*. København.
- Walker, Alastair & Ommo Wilts 2001: Die Nordfriesischen Mundarten. I: Horst H. Munske (Hrsg.): *Handbuch des Friesischen/Handbook of Frisian Studies*. Tübingen: Max Niemeyer, 284-305.
- Wilts, Ommo 2001: Die Verschriftung des Nordfriesischen. I: Horst H. Munske (Hrsg.): *Handbuch des Friesischen/Handbook of Frisian Studies*. Tübingen: Max Niemeyer, 305-313.
- www.sdbg-city.dk/Jubilogo
- www.synnejysk.dk
- Zentrum für Niederdeutsch 2004: www.plattdeutsches-zentrum.de
- Æ rummelpot 2004. Aabenraa.

Litteratur om sprog i grænseregionen:

- Pedersen, Karen Margrethe 2003: *Sprog og sprogkontakt i den dansk-tyske grænseregion – en bibliografi*. Adresse: www.dcbib./dk/dcb/sl_saml/bibliografi_kmp.pdf

Nyt redigeringsystem – overvejelser og valg

How to select a new dictionary writing system. Before introducing a new dictionary writing system it is important to describe the functionality of the writing system and what it requires and needs. These requirements and needs are normally described in a so-called list of requirements. The list of requirements is used to test the dictionary writing systems that are under consideration. In the Dictionary Department at Gyldendal Publishers we have prepared an extensive list of requirements listing more than 450 items. In this article I would like to pass on the experience from our tests. Firstly, I offer a discussion of the concept of a dictionary writing system. Then, I examine various dilemmas and conditions and in each case describe the prioritisation they may cause.

I de seneste år har adskillige ordbogsproducerende institutioner i de nordiske lande diskuteret ordbogsredigering. I den forbindelse er markedet for digitale værktøjer til netop redigering og produktion af ordbøger blevet undersøgt. I en årrække har der slet ikke været nogen til salg, men nu er der igen grøde: I dag kan man vælge mellem flere systemer,¹ hvis man da ikke vælger at udvikle sin egen løsning.²

I forlaget Gyldendals ordbogsafdeling har vi længe været på udkig efter et nyt system til at afløse vores to gamle systemer, GestorLEX³ og Compulexis.⁴ I den forbindelse har vi udarbejdet en omfattende kravspecifikation med flere end 450 punkter. I kravspecifikationen har vi prioriteret de krav til funktionalitet og systemarkitektur mv., som vi har fundet passende at stille til et system, der skal opfylde lige netop Gyldendals behov nu – og forhåbentlig også i fremtiden. Det er med kravspecifikationen i hånden, at vi har kunnet afprøve forskellige systemer på en så grundig og systematisk måde, at vi har følt os overbeviste om at være kommet “langt ud i krogene” og “hele vejen rundt”, før vi har sagt ja eller nej til det enkelte redigeringsystem.

Det er erfaringerne fra dette arbejde, jeg gerne vil bringe videre i denne artikel. Det er derfor ikke en videnskabelig artikel – det er snarere en opsummering af, hvad vi lærte af det konkrete projekt. Alle ordbogsprojekter er forskellige, og det er institutioner og projekters organisation også. Derfor har jeg prøvet at generalisere mine overvejelser så meget, at de fremstår så relevante som muligt for andre, der står i samme situation, som vi gjorde.

¹ Jeg har i skrivende stund kendskab til *Dictionary Publishing System (DPS)* fra franske IDM, *Dictionary Workbench* fra tyske Acolada, *iLEX* fra danske EMP og *TshwaneLex* fra det belgisk/sydafrikanske TshwaneDJe.

² Som man har valgt at gøre det hos bl.a. Norsk Ordbok.

³ Udviklet af danske TEXTware A/S.

⁴ Udviklet af engelske Compulexis Ltd.

1. Hvorfor indfører man et nyt redigeringsystem?

Der kan være mange grunde til, at man ønsker at indføre et nyt redigeringsystem, men det er mit indtryk, at de fleste af disse kan sammenfattes i følgende:

1. For at opnå økonomiske fordele, fx fordi redigering og produktion kan effektiviseres
2. For at kunne genbruge og udveksle ordbogsdata i forbindelse med nye ordbøger
3. Fordi systemet giver mulighed for bedre redaktionel kontrol med ordbogens indhold
4. Fordi systemet kan yde redaktionel støtte til ordbogsforfattere og -redaktører
5. For at kunne søge på kryds og tværs i alle ordbogsdata

2. Hvad er et ordbogsredigeringsystem?

Et ordbogsredigeringsystem er et stykke software, der er udviklet med henblik på at redigere og producere ordbøger. I programmet til den 3. internationale workshop om ordbogsredigeringsystemer, der afholdtes på Mazaryk Universitet i Brno i september 2004, defineres termen *Dictionary Writing System* (kaldet *DWS*) således:

A dictionary writing system is a piece of software for writing and producing a dictionary. It might include an editor, a database, a web interface and various management tools (for allocating work etc.). It operates with a dictionary grammar, which specifies the structure of the dictionary. (DWS 2004)

Denne definition giver et udmærket indtryk af, hvilke komponenter der kan være integrerede i et ordbogsredigeringsystem. Men definitionen giver ingen beskrivelse af systemets funktionalitet bortset fra kravet om, at der skal indgå en ordbogsgrammatik, der fastlægger reglerne for ordbogens struktur.

Bo Svensén går anderledes konkret til værks i 2. udgave af *Handbok i lexikografi* (Svensén 2004:498). Her opremser han en række konkrete krav til den funktionalitet, et redigeringsystem skal have, bl.a. alfabetisk sortering, håndtering af homografi, søge- og statistikfunktion og kontrol af krydshenvisninger.

3. Hvad er et *generelt* ordbogsredigeringsystem?

Oftest hører man ønsket om et 'generelt system' fremsat. Det er dog ikke altid ganske klart, hvad der menes med det. Man kan fortolke *generelt* på mindst tre måder, når det gælder redigeringsystemer:

1. Systemet kan være bygget op af standardiseret software. Og data kan produceres eller (som minimum) eksporteres i et internationalt standardiseret

dataformat. Det generelle består da i, at både redigeringsystem og data kan tilgås i andre omgivelser end de specifikke.

2. Systemet kan indeholde netop én ‘generel’ ordbogsgrammatik. Det underforstås da, at alle ordbøger kan beskrives i én og samme grammatik. Her er det selve ordbogsbegrebet, der opfattes generelt.
3. Systemet kan være generelt i den forstand, at det tillader brugeren selv at definere ordbogsgrammatikker, præsentations- og produktionsrutiner til de enkelte ordbøger, der således ikke postuleres at være ens. Her består generaliteten i at kunne håndtere forskellige ordbøger.

3.1. Systemet er bygget op af standardsoftware, og data lagres i standardformat

Disse meget tekniske krav tilgodeser de ydre processer ved ordbogsarbejdet. Drift og vedligeholdelse gøres sandsynligvis både lettere og billigere ved brug af standardsoftwarekomponenter, og man er mindre afhængig af en bestemt leverandør. I nogle institutioner vil det være et krav, at alt bygges op af open source-software, og at systemet kan køre på uafhængige (og måske gratis) platforme som fx Linux.

At ordbogsdata lagres (eller i det mindste kan eksporteres) i et standardiseret, platformsuafhængigt dataformat, er et vigtigt krav, hvis man ønsker at kunne producere sine ordbøger billigere og mere effektivt. Også her bliver man mindre afhængig af bestemte leverandører. På samme måde bliver udveksling af data med andre ordbogsproducerende institutioner naturligvis lettere ved brug af et standardformat.

Disse krav er i sig selv relevante og efter min mening vigtige. Men de er samtidig utilstrækkelige, idet man slet ikke fokuserer på selve redigeringsprocessen eller på håndteringen af den leksikografiske struktur i en ordbog. I et ordbogsprojekt fylder både formulering af de redaktionelle principper og især selve redigeringsfasen naturligt nok langt det meste af forløbet.

3.2. Systemet kan håndtere netop én ordbogsgrammatik

Hvis redigeringsystemet kun opererer med én ordbogsgrammatik, som alle ordbøger i systemet skal følge, kan man effektivisere redigeringsproces og produktion ved at bruge prædefinerede rutiner til præsentation af data på skærmen (både til indtastning og preview) og ikke mindst ved at have faste rutiner til satsgenerering og anden publicering (fx transformation af data til html-format). En anden fordel er, at alle brugere altid skal lære nøjagtig denne ene ordbogsgrammatik at kende.

At udforme en sådan fælles grammatik kan ikke være nemt, hvis den skal gælde både mono- og bilingvale ordbøger (med forskellige målgrupper og ordbogsfunktioner), historiske ordbøger, begrebsordbøger m.fl. – måske også mere encyklopædiske opslagsværker.

En fælles ordbogsgrammatik kan være ‘maksimal’ eller ‘minimal’. En maksimal

fællesgrammatik vil indeholde så mange forskellige informationstyper som muligt i alle tænkelige strukturelle kontekster. Grunden til at vælge denne løsning vil være, at grammatikken rummer alle de informationstyper og kombinationer af disse, som man kan tænkes at få brug for i bare én af de ordbøger, der redigeres i systemet. Man vil helst ikke risikere at hæmme forfatterens udtryksmuligheder på nogen måde. Brug af en maksimal grammatik vil typisk bevirke, at systemet i en given strukturel kontekst tilbyder brugeren alt for mange (og for den enkelte ordbogstype ofte irrelevante) informationstyper. Der er kort sagt ikke megen redaktionel kontrol eller støtte at hente i denne model, og brugeren risikerer at drukne i de mange muligheder.

Arbejder man derimod med en minimal fællesgrammatik, går øvelsen ud på at beskrive den mindst mulige fællesmængde. Minimalgrammatikken vil ofte opleves som for restriktiv i den enkelte ordbog. Brugeren har jo behov for redaktionel støtte i form af tilbud om netop de relevante informationstyper i en given kontekst. Der er stor risiko for, at man overser nogle af de nødvendige strukturelle særtilfælde, der findes i enhver ordbog. En mulig konsekvens er, at den samme informationstype bruges på forskellig vis i to forskellige ordbøger. Det går ud over den redaktionelle kontrol og den nok så vigtige fælles forståelse. Desuden kan det give problemer i forbindelse med præsentationen af data.

Særlig grelt vil det være, hvis den fælles 'generelle' ordbogsgrammatik er givet på forhånd af redigeringsystemet. Så har man ikke mulighed for selv at fastlægge den leksikografiske struktur. Ordbogsgrammatikken skal nemlig være fælles for mange forskellige ordbøger.

3.3. Systemet kan operere med forskellige ordbogsgrammatikker til forskellige ordbøger

Hvis redigeringsystemet er generelt i den forstand, at det tillader brugeren selv at formulere ordbogsgrammatikken og andre definitioner til hvert enkelt projekt, kan man opnå en meget høj grad af redaktionel kontrol og støtte. Man kan definere det enkelte projekt uden at skulle tage hensyn til andre værkers opbygning. Man kan koncentrere sig om at tilpasse den enkelte ordbogsgrammatik til den ønskede mikrostruktur og til redigeringsprocessen for det enkelte ordbogsprojekt. Det er en fordel, hvis man arbejder med flere ordbogsprojekter i huset. Det er også en god forudsætning for, at systemet kan få en vis udbredelse, dvs. en større kundekreds. Det sidste kan vise sig at være meget vigtigt også for den enkelte kunde, middellevetiden for ordbogsredigeringsystemer taget i betragtning.

Med denne type system skal man bruge flere ressourcer på oplæring. Ud over at lære selve systemet at kende skal brugerne ikke længere kun lære at bruge én enkelt fællesstruktur og ét fællesinventar af informationstyper. For hver ordbog, de skal arbejde med, skal de sætte sig ind i et specifikt inventar og en specifik struktur. Hvis

ordbogsgrammatikken er tilrettelagt rigtigt, får brugeren til gengæld hjælp i form af støtte og stimulation til at følge de redaktionelle retningslinjer. Desuden vil det være nyttigt, hvis man kan tilpasse skærmpresentationen til indtastning til det enkelte projekt – måske endda til den enkelte bruger.

Det kræver ressourcer, når man selv skal udarbejde ordbogsgrammatik og præsentationsdefinitioner (til skærm og egentlig produktion). Men hvis man dermed både kan spare tid og øge datakvaliteten, er det værd at overveje.

Stordriftsfordele i form af genbrug af fx præsentations- og produktionsrutiner (hvor det giver mening) bliver ved valg af denne type system et redaktionelt ansvar i stedet for en indbygget egenskab ved systemet. Gevinsten er, at man får mulighed for selv at træffe leksikografiske, redaktionelle, pædagogiske og æstetisk-grafiske valg.

3.4. *Alternativ: at udvikle sit eget system*

Daniel Ridings fra enheden for digital dokumentation ved Oslo Universitet taler i Ridings 2003 for den stik modsatte løsning, nemlig at man i den enkelte institution udvikler et specifikt redigeringsystem pr. ordbogsprojekt. Han argumenterer for, at ethvert ordbogsprojekt bør have en softwareudvikler tilknyttet:

It is extremely unlikely that a project will find a commercial product that will meet its needs off the shelf. For software to be beneficial to lexicographers, it must be linked to the style manual of the relevant project. This makes the software user-friendly, but not commercially viable since it is specific, not generic. It is difficult, if not impossible, to find two dictionaries from two independent groups that have the same format. The software developer is a crucial member of the team of a dictionary project. (Ridings 2003:214)

Termerne *style manual* og *format* synes begge at dække over både ordbogsgrammatik og regler for ordbogens præsentation. Desuden antages det, at alle kommercielle systemer er af den type, der opererer med én fælles ordbogsgrammatik for alle værker. Endelig skelnes der ikke klart mellem programmering af selve softwaren (som typisk skal udføres af en programmør) og konfigurering af et givent ordbogsprojekt i form af udarbejdelse af ordbogsgrammatik, præsentationsdefinition mv.

Som det siges i artiklen, er specialudviklede systemer som regel kun meget lidt generelle. Det er selvfølgelig en fordel i den forstand, at man kan tilpasse systemet nøjagtigt til det enkelte ordbogsprojekt. Men man skal også kunne ændre systemet, hvis man midt i et projektføreløb ønsker at lave om i strukturen. Hvis man har flere ordbogsprojekter, kan det være svært at genbruge systemet projekterne imellem.

Det er ofte dyrt at udvikle sit eget system. Det er da også kun i forbindelse med store og langvarige projekter, at det overhovedet giver økonomisk mening at udvikle et redigeringsystem til en bestemt bog. I et forlag med et stort ordbogsprogram ville

det være helt uholdbart at udvikle og vedligeholde et redigeringsystem til hver ordbog, selv om de måtte være nok så forskellige.

Har man udviklet sit eget system, er der ikke andre kunder, der kan medvirke til finansiering af vedligehold og videre udvikling af redigeringsystemet. Og selv om man ikke skal have udviklet nye funktioner, har man med jævne mellemrum brug for at få tilpasset systemet til nye versioner af operativsystem e.l. Man har ikke en ekstern leverandør at holde økonomisk ansvarlig for, at systemet faktisk bringes til at virke – slet ikke inden for den aftalte tid. Projektorganisationen bliver sårbar; fx er det svært at erstatte en programmeringskyndig medarbejder, der forlader sit arbejde midt i processen. Det kan også være svært at afsætte tid til at få skrevet brugbar dokumentation. I det hele taget er de fleste ordbogsproducerende institutioner sjældent gearede til at fungere som softwareudviklere.

4. Krav og prioriteringer

I dette afsnit vil jeg gennemgå, hvilke funktioner der med fordel kan prioriteres højt i forskellige situationer og ved forskellige projekt- og organisationstyper.

4.1. Når man arbejder med flere ordbøger

Når der i en institution udarbejdes flere forskellige ordbøger, bliver det naturligvis vigtigt, at redigeringsystemet kan håndtere mere end én ordbog ad gangen. Man skal kunne skifte mellem projekter, så en medarbejder ikke af tekniske grunde forhindres i at være tilknyttet flere projekter.

Det bliver også vigtigt, at de enkelte ordbøger kan have deres egne ordbogsgrammatikker. Hvis redigeringsystemet skal kunne yde redaktionel støtte (dvs. præsentere brugeren for netop de relevante strukturelle valg i en given kontekst) og stå for den strukturelle kontrol, giver det ikke mening at formulere en fælles grammatik for to potentielt meget forskellige værker.

4.2. Når man skriver nye ordbøger

Når man skriver en helt ny ordbog, vil man have stor glæde af, at redigeringsystemet kan tilbyde prædefinerede skabeloner – Bo Svensén kalder dem for ‘artikelmallar’ (Svensén 2004:499). En skabelon er et udvalg af informationstyper – typisk i form af ‘elementer’ eller ‘felter’ – der kan være tomme eller helt eller delvis udfyldt med tekst. Man kan forestille sig, at man har skabeloner for ordbogsartikler for ord, der tilhører en bestemt ordklasse og altid skal have præsenteret et bøjningsparadigme på en særlig måde. Eller man kan forestille sig skabeloner for forskellige slags henvisningsstrukturer, der kan forekomme i forskellige sammenhænge i ordbogsartiklerne.

Ved nyskrivning er det særlig vigtigt, at redigeringsystemet kan gemme artikler, der er ufærdige ifølge ordbogens grammatik. Det kan dels gælde en enkelt lang

artikel, der udarbejdes over flere dage, dels også det helt normale tilfælde, hvor man tilføjer én enkelt informationstype ad gangen til alle eller mange artikler på én gang. Ofte starter man med at oprette artikler, der blot indeholder opslagsord og evt. ordklasse, morfologi, etymologi eller anden information, der adskiller de enkelte homografer. Disse ufærdige artikler skal naturligvis kunne gemmes, uden at systemet bliver ustabil.

4.3. Når man reviderer eksisterende ordbøger

Arbejder man med nye udgaver af eksisterende ordbøger, er det andre forhold, der gør sig gældende. I den proces bliver avanceret søgning og statistikfunktionalitet ofte meget vigtig. Kan man kombinere søgning i struktur og indhold, er man efter min mening godt hjulpet. Man bør i den forbindelse vurdere, om søgesprogets syntaks lever op til det kompromis mellem brugervenlighed på den ene side og styrke og præcision på den anden, der passer bedst til brugerne. Jo mere avancerede søgninger man ønsker sig, jo mere synes syntaksen at komme til at ligne matematiske udtryk.

Hurtigt artikelopslag er uundværligt. Man har hele tiden brug for at kunne finde og åbne en hvilken som helst ordbogsartikel – ofte blot for at se, hvordan den er bygget op eller for at kopiere et strukturfragment. Moderne brugeres tålmodighed er ikke stor, så et opslag skal ganske enkelt være lynhurtigt, hvis ikke de skal blive utilfredse.

Det skal være nemt at orientere sig og bevæge sig rundt i den enkelte ordbogsartikels mikrostruktur. Der skal være et klart og pædagogisk indtastningsbillede, og der skal være relevante funktioner til at hoppe mellem eller flytte rundt på afsnit, niveauer, felter, elementer og andre komponenter, som strukturen måtte være bygget op af.

Hvis den udgave af ordbogen, man arbejder med, er den første, der redigeres i systemet, er der ofte en del strukturfejl i materialet. Det kan være, at man har konverteret fra et typografisk format (fx et tekstbehandlings- eller satsformat). Så vil der normalt være strukturer, der enten er fortolket forkert eller slet ikke er genkendt. Det er ønskeligt, at systemet fungerer stabilt alligevel og kan gemme artikler, selv om de indeholder mange strukturelle fejl.

Fejl kan naturligvis undgås ved, at man udarbejder ordbogsgrammatikken, så den bliver meget tolerant over for alle mulige strukturkombinationer. En sådan overtolerant grammatik kan ikke bruges til at støtte redigeringsprocessen på en sådan måde, at man kun får tilbudt relevante informationstyper i en given kontekst. Derfor vil det være en meget nyttig egenskab ved redigeringsystemet, hvis man kan knytte flere forskellige ordbogsgrammatikker til ét og samme ordbogsværk. Det kan fx være en overtolerant (eller 'deskriptiv') grammatik, der tillader alle de strukturer, der blev genkendt i ordbogen under konverteringen, og en ideel (eller 'normativ') grammatik, der beskriver den mikrostruktur, man faktisk ønsker, at alle artikler skal

følge. Undervejs i processen vil man nok have brug for at validere mod grammatikker, der beskriver forskellige mellemstadier.

4.4. Når man hyppigt udgiver nye udgaver af en ordbog

Udgiver man nye udgaver af en ordbog med forholdsvis korte intervaller – hvilket nok oftere er tilfældet med kommercielle ordbøger end med mere videnskabelige ordbøger – vil man for hver udgave prioritere, hvilke opgaver der skal løses netop denne gang. De fleste ressourcer går som regel til at indføre nyt indhold og revidere eksisterende indhold i ordbogen. Derfor vil nogle af de strukturelle opgaver ofte løses over en række udgaver. Man kan sige, at ordbogen sjældent gøres helt færdig, før den udgives.

Det kan til mange formål være nyttigt at registrere, hvornår et givent ord kommer med i en ordbog. Også i forbindelse med markedsføring har dette værdi. Til dette formål vil man have glæde af en form for registreringsfunktion i redigeringsystemet.

4.5. Når flere forfattere arbejder på den samme ordbog

Hvis flere skal skrive i den samme ordbogsbase, bliver de klassiske databasefunktioner vigtige: Systemet skal sikre, at der tildeles udelelig adgang til artiklerne. Og man skal kunne tildele forskellige brugere forskellige rettigheder. Nogle brugere skal have læse- og skriveret, nogle skal kun have læseret, og nogle værker må måske kun ses af udvalgte brugere.

Selv om man som forfatter kun arbejder med en del af bogens artikler (fx alle adjektiver, alle juridiske termer eller alle opslagsord, der starter med 'k'), har man også her behov for at kunne se alle artikler, når man har brug for det. Bl.a. skal man jo kunne lave krydshenvisninger på sikker vis. Man kan også forestille sig, at de udvalg af ordbogsartikler, som er tildelt de enkelte forfattere, overlapper hinanden.

Til denne form for projektorganisation bliver det nødvendigt med en veludbygget workflow-håndtering. Man skal kunne følge projektets fremadskriden, og man skal kunne se, hvem der har skrevet og rettet hvor. Man kan også forestille sig, at man har brug for at kende omfanget af den enkelte forfatters artikelproduktion for at kunne udbetale ham eller hende det korrekte honorar. Det siger sig selv, at administration og bogholderi, der føres i hånden – dvs. uden for systemet – bliver både mindre præcist og mere tidskrævende.

4.6. Brugerkompetence

Der er stor forskel på, hvilke kompetencer de forskellige medarbejdere på en ordbog har. Nogle har store it-kundskaber, andre næsten ingen. Det kan skyldes alder, baggrund eller helt andre forhold. Den vigtigste kompetence hos en ordbogsforfatter er så afgjort den leksikografisk-sproglige kompetence.

Der er ligeledes forskel på, hvordan forskellige personer opfatter en ordbogsarti-

kel: Ser de den som et hierarki af indlejrede informationstyper, ser de den som en lineær sekvens af informationstyper, eller tænker de ganske enkelt på en ordbogsartikel som et stykke (typograferet) tekst? I figur 1 ses et strukturelt eksempel på alle tre opfattelser af en ordbogsartikel fra en tænkt dansk-fransk ordbog.

Hierarkisk præsentation	Sekventiel præsentation	Tekstpræsentation
<artikel> <hoved> <opslagsord> ordforråd <klasse> sb. <krop> <oversættelse> vocabulaire	opslagsord: ordforråd klasse: sb. oversættelse: vocabulaire	ordforråd <i>sb.</i> vocabulaire.

Figur 1.

At forfatterne har forskellige opfattelser af en ordbogsartikels natur, betyder ikke, at de ikke kan forstå og følge de samme redaktionelle retningslinjer. Jeg synes, det er vigtigt, at man i redigeringsystemet kan komme de enkelte forfattere så meget i møde, at de hver især stimuleres og støttes i deres artikelproduktion. Hvis redigeringsystemet tillader fleksibel præsentation af den samme artikel, kan man undgå at skulle udpege én konceptuel opfattelse som den 'rigtige'. Det vil være meget ærgerligt at måtte sige nej tak til en knaldgod forfatter, fordi redigeringsystemet er for svært at bruge for ham eller hende.

4.7. Når én ordbog publiceres på flere måder

At kunne vise ordbogsdata på flere måder kan også være en fordel, hvis man vil publicere sin ordbog på flere forskellige måder. Ofte vil man have flere faste linjebud i en elektronisk ordbog end i en trykt bog. I den situation er der behov for at kunne se flere forskellige previews af ordbogsteksten. Det vil også være fint, om man kan definere, hvilke output-formater man vil have data eksporteret i. Det kan fx være html, xsl:FO, pdf eller et proprietært satsformat.

4.8. Når man udveksler ordbogsdata med andre

Hvis man gerne vil kunne udveksle data med andre, skal kravet om eksport af data i et standardiseret dataformat prioriteres. Med til en ordbog hører – ud over selve data – ordbogsgrammatikken, alfabetiseringsnøglen og de forskellige outputdefinitioner. Hvis alle disse definitioner også kan eksporteres i et standardformat, kan man udveksle hele pakken. Andre, der kan håndtere samme format, kan så indlæse data og definitioner meget hurtigt i deres eget system og komme i gang med redigeringsarbejdet. Et oplagt og meget udbredt dataformat er XML⁵ med de tilhørende

⁵ eXtensible Markup Language

standarder for datas grammatik, DTD⁶ og xml:schema, og standarden for XML-transformation XSLT.⁷

5. Konklusion

Det er sædvanligvis en stor investering at anskaffe sig et nyt redigeringsystem. Det kan derfor betale sig at bruge tid på at kortlægge sine behov og forventninger for at kunne vælge den løsning, der passer en bedst. Man har stor fordel af at formulere sine krav og ønsker og at opstille dem i prioriteret orden, så godt det nu kan lade sig gøre.

For mig at se er fleksibilitet en meget vigtig parameter. De vilkår, man arbejder under, har det nemlig med at ændre sig. Derfor er det meget vigtigt, at man kan ændre både ordbogsgrammatik og præsentations- og publiceringsdefinitioner undervejs. Da selve redigeringsprocessen uden sammenligning udgør den længste periode i et ordbogsprojekt, synes jeg, man skal prioritere redigeringsfaciliteterne meget højt. Det er disse faciliteter, der skal bruges i det daglige arbejde – dag ud og dag ind. Det er afgørende for datakvaliteten, at man indarbejder sine redaktionelle principper så godt som muligt i en ordbogsgrammatik for at kunne yde forfatterne mest mulig redaktionel støtte og samtidig holde en stram kontrol med datastrukturen.

I mine øjne er det vigtigt, at de valg, der kræver leksikografisk, redaktionel, pædagogisk eller æstetisk-grafisk kompetence, faktisk foretages af medarbejderne. Det kan være både svært og ærgerligt at skulle indrette sine ordbøger efter et bestemt stykke softwares (tilfældige) udformning. De vigtigste opgaver, der forudsætter sådanne valg, er:

- udformning af ordbogsgrammatik i overensstemmelse med de vedtagne redaktionelle retningslinjer
- fastlæggelse af principper for alfabetisering
- design af skærmpresentation til indtastning og preview
- fastlæggelse af typografi til forskellige udgivelser af ordbogsdata (fx som trykt bog, online-ordbog eller ordbog på mobiltelefon)

Redigeringsystemet skal tage sig af de mere tekniske, entydige og gentagne operationer og ikke mindst kontrol af data ifølge de regler, man har fastlagt.

Det er min påstand, at det er ulige meget lettere for en leksikografisk skolet person at lære sig en formel syntaks til beskrivelse af ordbogsdata, end det er for en programmør at forstå og træffe egentlige leksikografiske valg.

⁶ Document Type Definition

⁷ eXtensible Stylesheet Language Transformations

Litteratur

Anvendt litteratur:

- DWS 2004: *Third International Workshop on DICTIONARY WRITING SYSTEMS*
<http://nlp.fi.muni.cz/dws2004/>.
- Ridings, Daniel 2003: Lexicographic workbench: A case history. I: Sterkenburg, Piet van (red.): *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 204-214.
- Svensén, Bo 2004: *Handbok i lexikografi*, 2. udgave. Stockholm: Norstedts Akademiska Förlag.

Anden litteratur:

- van den Broek, Thijs 2005: *Choosing an XML editor* <http://ahds.ac.uk/creating/information-papers/xml-editors/>.
- Brooks, Kenneth P. 1988: *A Two-view Document Editor with User-definable Document Structure*. Palo Alto (California): Digital Systems Research Center.
- Harold, Elliotte Rusty 2002: *XML in a Nutshell – A Desktop Quick Reference*, 2. udgave. Cambridge (Massachusetts): O'Reilly.
- Joffe, David og De Schryver, Gilles-Maurice 2004: TshwaneLex, a state-of-the-art dictionary compilation program. I: Williams, Geoffrey and Vessier, Sandra (red.): *Proceedings of the Eleventh EURALEX International Congress*. Université de Bretagne Sud, Lorient, 99-104.
- The DANLEX-GROUP (Hjorth, Ebba/Nistrup, Bodil/Norling-Christensen, Ole/Jacobsen, Jane Rosenkilde/Ruus, Hanne) 1987: *Descriptive Tools for Electronic Processing of Dictionary Data*, Lexicographica Series maior 20. Tübingen: Max Niemeyer Verlag.

Idiom och semantisk variation – i autentiskt språkbruk och i ordböcker¹

Idiom and semantic variation in authentic usage and dictionaries. This paper explores the meanings of Swedish idioms in authentic texts and how they are reflected in dictionaries. More specifically, I recount the relationship between the meanings of 36 idioms expressed in a large collection of press material and the idiomatic meanings given in four Swedish monolingual dictionaries of partially divergent type. The study shows that a large number of idioms are defined in a variety of ways under different lemmas in the same dictionary, although the definitions must be regarded as correct in the absolute majority of cases. This applies primarily to expressions that provide a uniform impression in the press material from the semantic perspective. Several idioms whose semantic variation is more apparent in authentic texts could, on the other hand, be more perfectly described. Specimens of speech and usage notes could also be used more extensively to illustrate the meanings of idioms.

1. Inledning

Idiom eller idiomatiska uttryck har undersökts ur många perspektiv och litteraturen på området är omfattande (se vidare i Sköldbberg 2004:14-18 med där anförda referenser).

Intressantast i detta sammanhang är dock de studier där man undersökt idiom och andra fasta fraser ur ett lexikologiskt/lexikografiskt perspektiv. Forskare som t.ex. Zgusta (1971), Benson (1985), Schemann (1989), Dobrovolskij (1994) och Svensén (2004) har behandlat hur fasta fraser på bästa sätt skall presenteras i ordböcker. Bl.a. har man diskuterat under vilket lemma ett idiom ska presenteras eller om idiomatet skall utgöra ett lemma i sig. Andra forskare som Clausén (1993) och Clausén & Lyly (1994) har bl.a. presenterat en frastypologi som kan användas i ordböcker. Forskare som också kan nämnas i detta sammanhang är t.ex. Kjellmer (1996) och Farø (2002), som har undersökt om fasta fraser faktiskt nämns i existerande ordböcker – och i så fall hur de behandlas. En undersökning, som presenterades i Sköldbberg 2001, visade att det förhållandevis ofta råkade en markant skillnad mellan den form som 40 välkända svenska idiom hade i ett större nutida pressmaterial respektive i fem moderna ordböcker. Som exempel kan jag nämna *få kalla handen* med betydelsen 'helt bli avvisad'. Uttrycket uppvisar flexibilitet vad gäller verbet, och i autentiska texter förekommer bl.a. varianter som *ge någon kalla handen*, *mötas av kalla handen* och *då blev det kalla handen*. I de tre ordböcker där uttrycket var upptaget stod det

¹ Tack till Kungliga Vetenskaps- och Vitterhets-Samhället i Göteborg som genom ett generöst ekonomiskt bidrag möjliggjorde denna undersökning.

dock *vinka med kalla handen*, dvs. en form som mycket sällan används i faktiskt språkbruk. I Sköldberg 2004 koncentrerade jag mig sålunda på idiomens form, men jag blev även intresserad av att undersöka överensstämmelsen mellan den betydelse som idiom har i autentiska texter och den idiombetydelse som anges i svenska ordböcker. Det är detta senare ämne som behandlas i den här artikeln.

2. Material och teoretisk utgångspunkt

I den här undersökningen har jag utgått ifrån den del av min doktorsavhandling, Sköldberg 2004, som behandlar semantisk variation hos idiom. Jag har närmare bestämt undersökt 36 svenska idiomatiska uttryck som samtliga består av ett verb jämte komplement. Med tanke på att termen *idiom* används om relativt skiftande typer av uttryck (se t.ex. Allén et al. 1975:XXIII; Svensén 2004:239-240), återger jag här vad jag avser med termen idiom:

Ett prototypiskt idiom är en institutionaliserad sammanställning av lexikaliska enheter och bildar som helhet en syntaktisk konstruktion, som utan att ha en huvudsakligen grammatisk eller textbindande funktion fungerar som en del av en mening eller sats, har relativt fixerad form samt en figurativ betydelse som inte är härledbar ur de ingående ordens betydelse. (Sköldberg 2004:275)

Bland de uttryck som jag undersökt närmare återfinns t.ex. *rinna ut i sanden* med betydelsen 'inte leda till något'. I avhandlingen har jag granskat hur de aktuella uttrycken används i ett nutida pressmaterial som ingår i Språkbanken vid Institutionen för svenska språket vid Göteborgs universitet. Närmare bestämt har jag undersökt Press 95, Press 96, Press 97 och Press 98, som tillsammans innehåller drygt 33 miljoner löpande ord. De undersökta uttrycken är förhållandevis frekventa; alla förekommer en gång eller mer per en miljon löpande ord i materialet. Med andra ord har jag undersökt minst 33 användningar av respektive uttryck.

Jag har sålunda utgått ifrån mina semantiska beskrivningar av idiomerna och jämfört dessa med hur uttrycken redovisas i fyra svenska enspråkiga ordböcker av delvis olika slag. Dessa är *Nationalencyklopedins ordbok* (1995-96), *Natur och Kulturs Svenska Ordbok* (2001), *Bonniers svenska ordbok* (2002) samt *Svenskt språkbruk. Ordbok över konstruktioner och fraser* (2003). (Hädanefter kommer förkortningarna *NEO*, *NoKSO*, *BSO* respektive *SSB* att användas). Dessa verk valdes för att de är de mest heltäckande och kompletta ordböckerna som finns vad gäller svenskan idag.

De aspekter som framför allt intresserat mig, dvs. för det första förekomst av semantisk variation hos idiom och för det andra hur detta avspeglas i ordböcker, är inte särskilt väl utforskade. Man kan dock jämföra med Malmgren (2003) som behandlar svenska kollokationer och deras förändringsbenägenhet. Generellt kan man emellertid säga att de flesta forskare som skrivit något om semantisk variation

hos idiom yttrat sig i relativt svepande ordalag. De är heller inte överens i denna fråga. En del, såsom t.ex. Lindfors Viklund (1991:2), konstaterar att idiom har obetydlig semantisk töjbarhet, medan andra istället betonar att många idiom har en viss vaghet och oskärpa i sin betydelse (se t.ex. Koller 1977:140). Det råder heller inte konsensus om kontextens påverkan. En del forskare, däribland Nuccorini (1988:153) menar att alla idiom betyder detsamma i samtliga kontexter, medan andra, såsom t.ex. Clausén (1996:37) istället poängterar att ett idioms fulla innebörd kommer fram i just en specifik kontext (se vidare i Sköldbberg 2004:75-76). I min avhandling visar jag dock att det råder stor skillnad mellan olika idiom vad gäller semantisk variation. Lite mer än hälften av de undersökta idiomerna uppvisar en starkt begränsad semantisk variation i materialet, medan återstående uttryck däremot varierar desto mer. I min avhandling klarlägger jag också vissa återkommande typer av semantisk variation hos idiomerna. Jag visar även att dessa typer i hög grad liknar de typer av semantisk variation som återfinns hos vanliga lexem (se vidare i Sköldbberg 2004 och Sköldbberg 2005).

3. Resultat

Resultatredovisningen är disponerad på följande sätt. Inledningsvis diskuterar jag hur idiomerna presenteras i ordböckerna, om de är upptagna under ett eller flera lemman etc. Vidare behandlar jag överensstämmelsen mellan den eller de betydelse(r) som uttrycken har i pressmaterialet, som ligger till grund för Sköldbberg 2004, respektive i ordböckerna. Slutligen tar jag upp de metoder som ordboksförfattarna använt för att markera att ett idioms betydelse varierar. Artikeln avslutas med en kortfattad sammanfattning och diskussion.

3.1. Överensstämmelse inom och mellan verken

En jämförelse mellan ordböckerna visar att författarna använt delvis olika strategier när de valt att presentera idiomerna och det har givetvis påverkat resultatet. För det första kan man i samtliga ordböcker finna ett och samma idiom under flera lemman. Det råder då ibland överensstämmelse mellan angivelserna. I exempelvis *NEO* upptas uttrycket *dra sitt strå till stacken* under två lemman, dels *strå*, dels *stack*, och i båda fallen anges idiomets betyda 'bidra efter bästa förmåga till ngt gemensamt'. I samtliga ordböcker förekommer det dock att ett och samma idiom har olika definitioner under olika lemman. Som exempel kan idiomerna *ta bladet från munnen* anföras. I tabell A framkommer de angivelser som finns i de olika arbetena.

Som synes upptas idiomerna under två lemman i tre av de undersökta verken, närmare bestämt *NEO*, *NoKSO* och *BSO*. Definitionerna inom respektive ordbok är inte identiska med varandra, även om de i viss utsträckning är synonyma. I några fall uppvisar den ena definitionen mer betydelsevariation än den andra. Som exempel kan jag nämna de två definitionerna i *NoKSO*, dvs. 'säga något som har varit hemligt

eller som man inte har velat säga tidigare’ respektive ’berätta något som har varit hemligt’, där den första definitionen innehåller två alternativa betydelsenyanser men däremot inte den andra. Att ett och samma idiom definieras på olika sätt är dock vanligast i *NEO*. Detta kan kopplas till att *NEO* i mycket högre grad än de andra ordböckerna upptar ett och samma uttryck under olika lemman. I andra ordböcker, som t.ex. *SSB*, anges idiomerna i regel under ett lemma, medan det under andra möjliga lemman används hänvisningar. En fördel med *NEO* är givetvis att det är enklare för användaren att kontrollera vad uttrycket betyder. Ibland är det ju inte helt enkelt att räkna ut vilket lemma ett idiom anges under. Å andra sidan är det delvis slumpartat vad ordboksanvändaren får för uppgifter – det beror på vilket lemma han eller hon har slagit upp. Jag kan dock tillägga att idiomerna åtminstone i *NEO* nu kontrolleras så att det inte skall finnas sådana här typer av inkonsekvenser bland idiomdefinitionerna i nästa upplaga (som kommer att ges ut under namnet *Svensk ordbok utgiven av Svenska Akademien*).

verk	lemma	form	definition (+ ev. tillägg)
NEO	blad	ta bladet från munnen	säga rent ut vad man vet el. tycker
	mun	ta bladet från mun(nen)	tala rent ut
NoKSO	blad	ta bladet från munnen	säga något som har varit hemligt eller som man inte har velat säga tidigare
	mun	ta bladet från munnen	berätta något som har varit hemligt
BSO	blad	ta bladet från munnen	tala ut, säga sin mening
	mun	ta bladet från munnen	tala ut
SSB	blad	ta bladet från munnen	om att säga rent ut vad man tänker el. tycker hon tog bladet från munnen och sa att olämpliga lärare borde avskedas
Sköldberg	–	ta bladet från munnen	’(efter att ha hållit tyst) säga rent ut vad man tycker eller vet’ /om person; ofta med avs. på ngt som andra håller tyst om/

Tabell A. *Angivelser av ta bladet från munnen i de aktuella verken*

De olika ordboksförfattarna förefaller vara förhållandevis överens om betydelsen hos idiomerna *ta bladet från munnen*, även om man som sagt i vissa fall angivit vissa betydelsevarianter. Av uppställningen framgår också tillämpningen av s.k. ”förtydligande förklaringar” i *SSB* som i det här fallet inleds med ett ”om” (och i andra fall med ett ”när”) (se vidare i *SSB* 2003:11, 25). Dessa har redan ifrågasatts av bl.a. Malmgren (2004:288). I ett fall som detta kan man undra varför författarna helt enkelt inte skrivit ”säga rent ut vad man tänker el. tycker”. Å andra sidan anges även ett språkexempel som illustrerar hur idiomerna typiskt används, vilket är mycket positivt (jfr Malmgren 2004:284). I Sköldberg 2004 har jag däremot valt att använda

definitionstillägg (inom snedstreck), och det bidrar till en precisering av hur uttrycket brukar användas. Men dessa definitionstillägg återkommer jag till senare.

3.2. Överensstämmelse mellan idiombetydelse(r) i autentiska texter och i ordböckerna

Nästa fråga gäller graden av överensstämmelse mellan den/de betydelse(r) som de aktuella idiomerna har i pressmaterialet och de idiomdefinitioner eller förklaringar som är aktuella.

3.2.1. Exempel med god överensstämmelse

En genomgång av uttrycken visar att det råder en mycket god samstämmighet mellan autentiskt bruk och ordboksangivelser i en absolut majoritet av fallen. Ett första exempel är *inte sticka under stol med något* som uppvisar starkt begränsad innehållsmässig variation i pressmaterialet. I exempel (1) framkommer hur uttrycket typiskt används:

- (1) Försvarsledningen *sticker inte under stol med* att ekonomin är i fritt fall och att man som det ser ut inte har råd att förverkliga försvarsbeslutet. (Press 97)

I exemplet omtalas alltså att Försvarsledningen inte försöker dölja att ekonomin är mycket instabil. Det är också denna betydelse hos uttrycket som anges i böckerna. Detta framgår av tabell B.

ordbok	lemma	form	definition (+ ev. tillägg)
NEO	sticka	inte sticka under stol med (ngt)	inte dölja (ngt)
	stol	inte sticka under stol med	inte dölja
NoKSO	stol	inte sticka under stol med något	inte dölja något <i>Han stack inte under stol med att han var missnöjd.</i>
BSO	sticka	inte sticka under stol med	inte dölja
SSB	stol	inte sticka under stol med ngt	om att inte försöka dölja ngt, ofta negativt författaren ville inte sticka under stol med att boken hade krävt många omskrivningar
Sköldberg	–	inte sticka under stol med ngt	'inte försöka dölja ngt' /om person; med avs. på (ofta problematiskt, provocerande, diskutabelt e.d.) förhållande, faktum eller uppfattning/

Tabell B. *Angivelser av inte sticka under stol med något i de aktuella verken*

De olika definitionerna ger här ett mycket enhetligt intryck. Som synes återkommer verbet *dölja* i samtliga betydelsebeskrivningar. Av *SSB* framgår det också att det ofta är något negativt som man inte sticker under stol med. I min avhandling har jag alltså använt definitionstillägg, och dessa måste betraktas som användbara. I det typiska fallet är det alltså en person som inte sticker under stol med ett problematiskt, provocerande eller diskutabelt förhållande, faktum eller uppfattning.

Ett andra exempel är idiomet *bära frukt*. Också detta uppvisar starkt begränsad innehållsmässig variation i pressmaterialet. En genomgång visar att idiomet i samtliga belägg har innebörden 'ge önskat resultat'. Vidare representeras uttryckets externa argument, dvs. det som bär frukt, i princip alltid av ett abstrakt substantiv. Oftast betecknas någon form av arbete eller satsning. Så är också fallet i exempel (2):

- (2) Nye Agachefen Lennart Selander hoppas att företagets miljardinvesteringar skall *bära frukt* snart. Det kan behövas, resultatet för 1996 infriade inte förväntningarna. (Press 97)

I (2) är det således en större investering som förutspås ge ett positivt resultat. Hur ser det då ut i undersökningsmaterialet? Detta framkommer i tabell C.

ordbok	lemma	form	definition (+ ev. tillägg)
NEO	bära	(ngt) bär frukt	(ngt) ger resultat <urspr. bibl.>
	frukt	hans ansträngningar bar frukt	(lyckat,) märkbart resultat
NoKSO	bära	bära frukt	1 (om träd eller buskar) få frukter 2 ge ett bra resultat <i>Vi får hoppas att arbetet kommer att bära frukt.</i>
	frukt	bära frukt	lyckas bra <i>Satsningen har burit frukt.</i>
BSO	frukt	bära frukt	äv. bildl.: ge resultat
SSB	frukt	bära/ge frukt	hallonbuskar bär frukt på fjolårsskotten [...]
	frukt	◆ bära/ge frukt	alla är överens om att totalt engagemang är nödvändigt för att projektet skall bära frukt [...]
Sköldberg	–	bära frukt	'ge önskat resultat' /om ngt (vanl. ett arbete el. en satsning)/

Tabell C. *Angivelser av bära frukt i de aktuella verken*

I just detta fall aktualiseras bl.a. det faktum att idiomet *bära frukt* till sin form sammanfaller med kollokationen *bära frukt* vilken används om träd och buskar. Författarna förhåller sig till detta på delvis olika sätt. I *NoKSO* har man under lemmat *bära* valt att placera in de två betydelserna under olika lexemnummer. Under

lemmat *frukt* anförs dock bara betydelsen hos idiomet. I *BSO* återges betydelsen hos idiomet, men formuleringen ”äv. bildl.” indikerar att konstruktionen också kan tolkas bokstavligt. I *SSB* har man för det första angivit att *bära frukt* är en grammatisk konstruktion som avser träd. För det andra är *bära frukt* en konkret fras, vilken föregås av symbolen ◆, och då avses den idiomatiska betydelsen hos uttrycket. Betydelseskillnaden mellan den grammatiska konstruktionen och den konkreta frasen illustreras med hjälp av språkexempel. Däremot förklaras inte uttrycket. Uppenbarligen kan man välja olika lösningar i lägen som dessa. Men ordboksförfattarna måste ta ställning till hur man ska gå tillväga i dessa inte helt ovanliga fall. Flera av de undersökta idiomerna sammanfaller nämligen med kollokationer med mer bokstavliga betydelser. Bland dessa återfinns exempelvis (*gå*) *stick i stäv* och *stå högt i kurs*.

Beskrivningarna av vad idiomet *bära frukt* betyder måste betraktas som rättvisande. Ännu en gång vill jag dock lyfta fram fördelen med definitionstillägg och illustrativa språkprov såsom de som finns i *NoKSO* och *SSB*.

3.2.2. Exempel med mindre god överensstämmelse

Hittills har jag diskuterat uttryck som är väl beskrivna i ordböckerna. De utgör som sagt en klar majoritet av de undersökta fallen. Framför allt är det de uttryck som ger ett enhetligt intryck i materialet som är väl beskrivna. Kännetecknande för flera av de uttryck vars definitioner inte är fullt så rättvisande är att de uppvisar mer påtaglig semantisk variation. Som exempel väljer jag *dra växlar på något*. I tabell D framkommer hur det definieras i de olika arbetena.

verk	lemma	form	definition (+ ev. tillägg)
NEO	dra	dra växlar (på ngt)	utnyttja (ngt) för mycket
	växel	dra (stora) växlar på (ngt)	dra (stora) fördelar av (ngt) för framtiden
NoKSO	växel	dra växlar på något	vara alldeles för säker på att man ska få något och därför ta risker <i>Man ska inte dra för stora växlar på framtiden.</i>
BSO	växel	dra växlar på	dra fördel av; räkna med för framtiden
SSB	växel	dra växlar på ngt	om att utnyttja och dra fördelar av ngt [...] tränaren drar inte alltför stora växlar på premiärsegern, eftersom motståndet var lite för svagt; [...]
Sköldberg	–	dra växlar på ngt	'dra slutsatser (som indikerar fortsatt positiv utveckling) av ngt' /främst om person; av (ofta positiv) situation, utveckling el. händelse såsom en seger/

Tabell D. Angivelser av dra växlar på något i de aktuella verken

Det är framför allt två uttryckssätt som förekommer i definitionerna av idiomerna och det är ”utnyttja” respektive ”dra fördelar av”. I Sköldberg 2004 har jag däremot valt formuleringen ’dra slutsatser (som indikerar fortsatt positiv utveckling) av något’. Som framgår av definitionen kan de aktuella slutsatserna antyda att framtiden i något avseende är positiv. Så är också fallet i exempel (3):

- (3) Löf aktar sig dock för att *dra* för stora *växlar på* storesegern. – Vi ska passa oss för att sticka ut hakan, så här stora ytor kommer vi inte få i nästa match. (Press 97)

I (3) handlar det om att ett lag har vunnit överlägset i fotboll, vilket bådär gott inför framtiden. Den omtalade ”Löf” vill dock inte ta ut några segrar i förskott. Den semantiska faktorn ’som indikerar fortsatt positiv utveckling’ är dock inte alltid närvarande och av detta skäl står den inom parentes i definitionen. Denna del av betydelsen realiseras exempelvis inte i exempel (4):

- (4) Vilka framtidsdrömmar hyser då dagens barn? Utan att *dra* alltför stora *växlar på* denna artikel, vilken bygger på intervjuer med blott sex pojkar och flickor i låg- och mellanstadieåldern, vågar man nog ändå dra slutsatsen att dessa barn är tidstypiska i den bemärkelsen att flera av dem drömmer om att som vuxna arbeta inom mediaindustrin [...]. (Press 96)

I exempel (4) handlar det om att dra en slutsats utifrån innehållet i en viss artikel. Denna slutsats är dock varken utpräglat positiv eller negativ utan förhållandevis neutral. Man kan därmed säga att bortfallet av ’som indikerar fortsatt positiv utveckling’ leder till en generalisering av uttryckets idiomatiska betydelse. Denna återkommande typ av betydelsevariation är dock inte särskilt väl återgiven i de aktuella ordböckerna. Variationen hos just detta uttryck illustreras dock till viss del i *SSB* med hjälp av språkprov. Även en av de betydelser som anges i *BSO*, nämligen ’räkna med för framtiden’, speglar förhållandevis väl idiomets betydelse, åtminstone då den en fortsatt positiv utveckling är aktuell.

Detta gällde alltså idiomerna *dra växlar på något*. Andra uttryck som också varierar semantiskt i materialet men som beskrivs på ett väl knapphändigt sätt i flera av ordböckerna är t.ex. *kasta in handduken* och *sopa något under mattan*.

3.3. Angivelse av betydelsevariation

Hittills har jag behandlat i vilken mån definitionerna speglar den betydelse som uttrycken har i mitt material. Jag kommer nu att koncentrera mig på hur betydelsevariation hos idiomerna anges i de aktuella verken. Som utgångspunkt för detta resonemang vill jag nämna bruket av *se dagens ljus*. I materialet finns framför allt två typer av användningar. Dessa illustreras i exempel (5-6):

- (5) Lille Bryan såg *dagens ljus* den 5 november 1959, i Kingston, Ontario, Kanada. (Press 97)
- (6) En ny jätte på energiområdet *ser dagens ljus* med en omsättning på 80 miljarder franc och 90 000 anställda. (Press 95)

Uttrycket *se dagens ljus* används således dels om människor (och djur) och kan då sägas betyda 'födas', dels om inanimata ting och företeelser såsom produkter, företag och politiska partier. Uttrycket har då betydelsen 'skapas' eller 'bli till'. Av tabell E framgår vad uttrycket anges betyda i de olika verken.

verk	lemma	form	definition (+ ev. tillägg)
NEO	dag	se/skåda dagens ljus	födas, komma till
	ljus	se/skåda dagens ljus	födas <högt.>
NoKSO	dag	se dagens ljus	födas eller bli till <i>Datorn såg dagens ljus redan på femtiotalet.</i>
	ljus	se dagens ljus	födas eller bli till
BSO	dag	se dagens ljus	födas
	ljus	se dagens ljus	födas
SSB	ljus	se/skåda dagens ljus	om att födas, poet. det var först på senhösten året därpå som första barnet såg dagens ljus; äv. om företeelser el. likn. som skapas en ny symfoniorkester har sett dagens ljus i den idylliska fiskestaden
Sköldberg	–	se dagens ljus	'börja existera' /om levande varelse, föremål, organisation e.d./

Tabell E. *Angivelser av se dagens ljus i de aktuella verken*

Som synes i tabell E anger samtliga ordböcker att uttrycket kan betyda 'födas'. Under vissa lemmorna anges även att uttrycket kan betyda 'komma till', 'bli till', 'skapas', dvs. att idiomets betydelse kan vara mer utvidgad. I en majoritet av beläggen i pressmaterialet är det också denna mer utvidgade betydelse som är aktuell. Det råder sålunda inget tvivel om att också denna betydelsenyans bör återges på något sätt. I Sköldberg 2004 valde jag istället att skriva en mer generell definition som är tänkt att inbegripa båda betydelserna. Definitionstillägget är avsett att vägleda användaren om vad det är som kan *se dagens ljus*.

Detta uttryck illustrerar också hur ordboks författarna valt att markera befintlig semantisk variation i materialet. I *NEO* har man under lemmat *dag* valt att placera

ett kommatecken mellan de båda betydelserna. I *NoKSO* har man istället använt ett ”eller” och i *SSB* en kombination av ett semikolon och ett ”även”.

Ett sista exempel är *sätta käppar i hjulet för något* som i mitt material dels betyder ’vålla ett besvärande hinder’, dels ’vålla ett oöverstigligt hinder’. I tabell F framkommer hur detta beskrivs i de aktuella arbetena.

verk	lemma	form	definition (+ ev. tillägg)
NEO	hjul	sätta en käpp i hjulet (för ngn/ngt)	hindra (ngn/ngt)
	käpp	sätta en käpp i hjulet (för ngn eller ngt)	åstadkomma ett direkt hinder (för ngn eller ngt) ofta på otillbörligt sätt
NoKSO	käpp	sätta en käpp/käppar i hjulet för någon/något	hindra eller stoppa någon/något
BSO	käpp	sätta en käpp i hjulet för	effektivt hindra
SSB	käpp	sätta käppar i hjulet/hjulen för ngn/ngt	om att hindra ngt el. förstöra för ngn föräldrade system får inte sätta käppar i hjulen för arbetet; [...]
Sköldberg	–	sätta käppar i hjulet för ngt	’(avsiktligt) vålla (oöverstigligt) hinder för ngn el. ngt’ /om ngn el. ngt; ofta med avs. på verksamhet, utveckling etc./.

Tabell F. *Angivelser av sätta käppar i hjulet för något i de aktuella verken*

Av tabell F framgår att någon form av betydelsevariation markeras i tre av de aktuella böckerna, närmare bestämt i *NoKSO*, *SSB* och Sköldberg. I såväl *NoKSO* som *SSB* används ett ”eller”. Båda betydelsebeskrivningarna kan sägas spegla innehållet i många av användningarna i materialet. Som redan framgått har jag i Sköldberg 2004 valt att sätta vissa led inom parentes för att på så vis markera att vissa faktorer aktualiseras ibland, men inte alltid. Vilken metod man än använder, måste det ändå betraktas som positivt med en mer nyanserad bild av vad idiommet kan betyda.

4. Sammanfattande diskussion

Nu har jag givit smakprov på hur det kan se ut i de olika arbetena. Min undersökning visar att de idiomdefinitioner som finns i ett antal svenska ordböcker i hög grad påminner om varandra. Inom samtliga ordböcker förekommer det dock att ett och samma idiom definieras på varierande sätt under olika lemmen. I flera av ordböckerna är det till och med mycket vanligt förekommande. Detta kan vara en konsekvens av den variation som idiomerna trots allt uppvisar i autentiska texter. Under ett lemma

anges vad uttrycket betyder i vissa belägg och under ett annat lemma vad det betyder i andra. Man skulle kunna säga att det faktum att det råder en så stor inkonsekvens mellan idiomdefinitioner inom ett och samma verk tillsammans med resultat från tidigare undersökningar säger något om hur lite genomarbetade idiom är i svenska ordböcker – åtminstone hade fram till publiceringen av *Svenskt språkbruk* (2003).

Vidare visar undersökningen att de flesta av idiomdefinitionerna är kortfattade men att de ändå kan sägas täcka in idiomets betydelse(r). Framför allt gäller detta de idiom som uppvisar begränsad semantisk variation i materialet. Med andra ord verkar det som om idiomens betydelser i svenska ordböcker är mer rättvisande än deras form.

De gånger semantisk variation förekommer i autentiska texter är denna dock inte lika väl återgiven i böckerna. När variation skall markeras använder ordboksförfattarna ett antal olika redskap. För det första används lexemnummer dels för att ange att ett uttryck kan tolkas både bokstavligt och idiomatiskt, dels att ett idiom kan ha två olika idiomatiska betydelser. För det andra infogas *eller* och *även*. I vissa ordböcker som t.ex. *NoKSO*, kombineras också dessa angivelser med lexemnummer, vilket kan försvåra för läsaren. För det tredje används kommatecken och semikolon. Kommatecken brukar indikera lägre grad av betydelseskilnad och semikolon en högre grad (jfr Svensén 2004:412). Men inte heller dessa tecken är särskilt konsekvent använda. Kommatecknen används också rikligt mellan närsynonymer som författarna tagit till för att så långt det är möjligt täcka idiomets betydelsenyanser i olika kontexter. Det torde vara en konsekvens av att det inte alltid är helt enkelt att förklara vad ett idiom verkligen betyder (jfr t.ex. Andersson 1999:137). För det fjärde förekommer det att vissa delar av definitionerna står inom parentes för att ange att just dessa led är fakultativa. Det förekommer även att fakultativa betydelse-tillägg anges i mindre stil.

De semantiska beskrivningarna av idiomerna kan sålunda förbättras och preciseras ytterligare. En tänkbar väg i framtiden är just att använda någon form av definitionstillägg för att på ett tydligare sätt ange hur idiomerna i fråga brukar användas. De infogade språkproven är också något mycket positivt och de skulle också kunna användas ytterligare för att belysa den variation som faktiskt förekommer hos idiomerna i autentiska texter. Det är möjligt att det inte finns tillräckligt med utrymme för detta i papperversioner av ordböcker, men man borde verkligen utnyttja denna möjlighet i elektroniska utgåvor (jfr t.ex. Bergenholtz & Vrang 2004:185-186; Lorentzen & Trap-Jensen 2004:193-198.)

Litteratur

Ordböcker:

- Bonniers svenska ordbok*. Utarbetad av Malmström, S. m.fl. 8:e rev. uppl. Stockholm: Albert Bonniers Förlag 2002.
- Nationalencyklopedins ordbok*. Del 1-3. Höganäs: Bra Böcker 1995-96.
- Natur och Kulturs Svenska Ordbok*. Stockholm: Natur och Kultur 2001.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Utarbetad av Svenska språknämnden. Stockholm: Norstedts 2003.

Övrig litteratur:

- Allén, Sture et al. 1975: *Nusvensk frekvensordbok baserad på tidningstext. 3. Ordförbindelser*. Stockholm: Almqvist & Wiksell.
- Andersson, Henrik 1999: To nye danske idiomordbøger. I: *LexicoNordica* 6, 133-150.
- Benson, Morton 1985: Collocations and idioms. I: Ison, R. (ed.): *Dictionaries, lexicography and language learning*. Oxford: Pergamon Press, 61-68.
- Bergenholtz, Henning & Vibeke Vrang 2004: Ny dansk ordbog i seks bind for sekretærer og forskere. I: *LexicoNordica* 11, 165-189.
- Clausén, Ulla 1993: Idiom och variation. I: Garde, A. & P. Jarvad (red.): *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.-14. maj 1993*. (Skrifter udgivet af Nordisk Forening for Leksikografi 2.) Oslo, 47-52.
- Clausén, Ulla 1996: Idiom i bruk. I: *Språket lever! Festskrift till Margareta Westman den 27 mars 1996*. (Skrifter utgivna av svenska språknämnden 80.) Stockholm: Norstedts, 36-42.
- Clausén, Ulla & Erika Lyly 1994: Criteria for Identifying and Representing Idioms in a Phraseological Dictionary. I: Martin, W. et al. (eds.): *Euralex 1994 Proceedings. Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*. Amsterdam, 258-262.
- Dobrovol'skij, Dmitrij 1994: Idioms in a Semantic Network: Towards a New Dictionary-Type. I: Martin, W. et al. (eds.): *Euralex 1994 Proceedings. Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*. Amsterdam, 263-270.
- Farø, Ken 2002: Somatismen als Problem der dänischen und deutschen Lexikographie. I: *Symposium on Lexicography X. Proceedings of the Tenth International Symposium on Lexicography May 4-6, 2000 at the University of Copenhagen*. (Lexicographica 109.) Tübingen, 107-124.
- Kjellmer, Göran 1996: Idiomen, kollokationerna och lexikonet. I: *LexicoNordica* 3, 79-90.
- Koller, Werner 1977: *Redensarten. Linguistische Aspekte, Vorkommensanalyse, Sprachspiel*. Tübingen: Niemeyer.
- Lindfors Viklund, Maja 1991: *Svenska idiom. Ur ett internationellt perspektiv*. (Rapporter från institutionen för nordiska språk/svenska språket vid Göteborgs universitet (Nordrapp) 7.) Göteborg.
- Lorentzen, Henrik & Trap-Jensen, Lars 2004: Kommentarer til Henning Bergenholtz og Vibeke Vrang: Ny dansk ordbog i seks bind for sekretærer og forskere. I: *LexicoNordica* 11, 191-198.
- Malmgren, Sven-Göran 2003: *Begå eller ta självmord? Om svenska kollokationer och deras*

- förändringsbenägenhet. I: *Språk och Stil. Tidskrift för svensk språkforskning* 13 (ny följd). Uppsala, 123-168.
- Malmgren, Sven-Göran 2004: En ny svensk konstruktionsordbok. I: *LexicoNordica* 11, 271-292.
- Nuccorini, Stefania 1988: The treatment of metaphorical and idiomatic expressions in learners' dictionaries. I: Snell-Hornby, M. (ed.): *ZüriLEX '86 Proceedings. Papers read at the EURALEX International Congress, University of Zürich, 9-14 September 1986*. Tübingen, 149-160.
- Schemann, Hans 1989: Das phraseologische Wörterbuch. I: Hausmann, F.J. et al. (eds.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie. Band 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1.) Berlin/New York: de Gruyter, 1019-1032.
- Sköldberg, Emma 2001: Lever idiomerna i svenska ordböcker sitt eget liv? I: *Rapport från Konferens om lexikografi i Norden, Göteborg 26-29 maj 1999*. (Nordiska studier i lexikografi 5.) (Meijerbergs arkiv för svensk ordforskning 27.) Göteborg, 344-356.
- Sköldberg, Emma 2004: *Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom*. (Meijerbergs arkiv för svensk ordforskning 31.) Göteborg.
- Sköldberg, Emma 2005: När faktorer lyser med sin frånvaro: om några typer av betydelsevariation hos svenska idiom. I: *Hermes, Journal of Linguistics* 35, 119-134.
- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm: Norstedts Akademiska Förlag.
- Zgusta, Ladislav 1971. *Manual of lexicography*. The Hague: Mouton.

Uttalslexikon för talsyntes

Pronunciation dictionaries for speech synthesis. The Swedish Library of Talking Books and Braille (TPB) is producing talking books for university students by speech synthesis. To enhance the pronunciation, TPB has started compiling a large electronic pronunciation dictionary. In the initial phase, existing printed dictionaries are imported to the lexical database. As these dictionaries were made for human reading, they are transformed to meet the quite different requirements from the speech synthesis systems.

1. Bakgrund

1.1. *Talböcker*

Talböcker är inspelningar av upplästa böcker, avsedda för läshandikappade. Till skillnad från kommersiella ljudböcker eller audioböcker är de undantagna från upphovsrätt och får därför bara lånas ut till handikappade. I Sverige har talböcker producerats sedan 1955, alltså i femtio år. Från början var de avsedda enbart för blinda och gravt synskadade som alternativ till punktskriftsböcker, vilket tidigare hade varit den enda formen av böcker för synskadade. En av nackdelarna med punktskrift är att relativt få synskadade behärskar den. Främst är det barndomsblinda som får ordenlig färdighet i punktskriftsläsning, medan personer som får synskador i vuxen ålder – vilka utgör merparten av de synskadade – sällan klarar av punktskriften tillräckligt bra för att kunna läsa böcker. Talboken nådde därför en större målgrupp, samtidigt som den innebar en bekvämare distributionsform än den mycket skrymmande punktskriften.

I början sköttes talboksutlåningen av De blindas förening. År 1980 tog staten över ansvaret genom att inrätta myndigheten Talboks- och punktskriftsbiblioteket (TPB). Målgruppen utvidgades 1982 till andra grupper av läshandikappade. Enligt WHO har 1 % av befolkningen i industriländer behov av talböcker, vilket innebär att den potentiella användarkretsen i Sverige är cirka 90.000 personer. De flesta av dessa är dyslektiker (Sköld 2005).

Från början spelades talböckerna in på rullband. År 1973 gick man över till kassetter, och numera produceras alla nya talböcker på CD-skivor. Det internationella lagringsformatet DAISY gör att man kan navigera i texten och få text och ljud presenterat parallellt. I synnerhet dyslektiker har stor nytta av det sistnämnda (Irvall 2005, Hirschfeldt 2005).

Varje år låter TPB framställa cirka 3.000 titlar med en genomsnittlig speltid på 12-14 timmar vardera. Kostnaden för att läsa in en traditionell talbok är cirka 17.000 kronor. En stor del av de talböcker som produceras är kurslitteratur för högsko-

lestudier, vilken TPB har skyldighet att tillhandahålla. Ofta är det bråttom med inläsningen för att användaren skall kunna klara studieplanen. Många av dessa kursböcker har inte mer än en eller två användare, vilket innebär att kostnaden per användare är mycket stor. För att sänka kostnaderna och minska väntetiderna kommer man därför från och med 2005 att använda talsyntes för inläsning av viss kurslitteratur. Detta innebär med dagens teknik en klar kvalitetsförsämring, men de tänkta användarna är i allmänhet vana vid att använda talsyntes i andra sammanhang och förutsätts kunna acceptera vissa brister.

1.2. Talsyntes

Svensk talsyntes började utvecklas ungefär samtidigt med talboken. Redan 1953 utvecklade Kungliga Tekniska högskolan (KTH) i Stockholm den första talsyntes-maskinen, Ove I, som bara kunde uttala vokaler. År 1961 kom Ove II, som var stor som ett klädskåp och kunde uttala både vokaler och konsonanter. Utvecklingen har sedan fortsatt både vid KTH och andra forskningsinstitutioner och inom olika företag. Numera finns flera kommersiella datorbaserade talsyntesprogram för svenska.

Talsyntes används för närvarande främst i läsprogram för handikappade och i teleupplysningstjänster. I båda fallen är det viktigt att programmen arbetar snabbt. När man markerar en text på bildskärmen för att få den uppläst upplevs en väntetid på en halv sekund som mycket lång. Det samma gäller om man har begärt till exempel en tidtabellsuppgift i telefon och väntar på svaret.

För att få tillräcklig snabbhet med begränsade datorresurser har man tidigare strävat efter system med relativt enkla regler för uttalsgenereringen och relativt små lexikon för ord med oregelbundet uttal. I takt med att processorerna har blivit snabbare och minnena större har talsynteserna blivit mer avancerade. Framför allt har den grundläggande fonetiska renderingen blivit bättre genom utvecklingen från formantsyntes till mer utrymmeskrävande difonsyntes och ännu mer utrymmeskrävande unit selection-syntes, men även de inbyggda uttalslexikonen har blivit större.

Prosodin är fortfarande långt ifrån perfekt hos de syntetiska rösterna, även om de idag inte låter sämre än vissa nyhetsuppläsare i tv. Korrekt satsbetoning kräver en avancerad grammatisk och semantisk analys. Eftersom starkt markerade felaktiga betoningar är mycket störande föredrar många användare en relativt entonig röst, vilken dock kan kännas tröttande i längden.

En förstudie gjord på tio talsyntesinlästa kursböcker visade att 21 % av de förekommande ordformerna uttalades otillfredsställande enligt en van talboksanvändares bedömning. Det var i huvudsak tre kategorier av ord som uttalades fel: sammansatta ord, egennamn och främmande citatord. Dessa tre ordkategorier förekommer ymnigt just i kurslitteratur, vilket motiverar en omfattande insats för att förbättra orduttaleet.

Typiska uttalsfel beror på felaktig ordanalys. Uppenbarligen tolkar Ingmar, den mest spridda svenska talsyntesen, ordet *kompetens* på följande sätt: *kom* är ett obetonat prefix, *en* är suffix för bestämd form och *s* är genitivsuffix, följaktligen är *pet* den betonade ordstammen – och Ingmar uttalar därför ordet som ”kompétens”. Sammansatta ord segmenteras många gånger felaktigt, särskilt när något ordled är ett egennamn eller ett lågfrekvent ord. Ingmar segmenterar exempelvis ordet *malmöbor* som *mal-möbor*, och uttalar det därför med bibetonning på andra i stället för tredje stavelsen. Ett särskilt problem utgör främmande ord och namn, som uttalas enligt vanliga svenska uttalsregler. Dessa uttalsfel åtgärdas effektivast med hjälp av lexikon.

1.3. TPB:s talsyntesprojekt

TPB håller på att utveckla ett nytt produktionssystem för framställning av talsyntesbaserade talböcker i stor skala, där böckerna kommer in i form av maskinläsbara filer från förlagen och kommer ut som färdiga talböcker med minsta möjliga mänskliga bearbetning däremellan. Detta produktionssystem kan innefatta tung databearbetning eftersom man inte har samma snabbhetskrav som för skärmläsare och dialogsystem. Det innebär bland annat att man kan göra grundlig syntaktisk analys av texten för att få korrekt prosodi.

TPB:s talsyntesprojekt består av flera delprojekt och löper under flera år. De första förundersökningarna gjordes 2003 och det fullt fungerande systemet beräknas vara klart i början av 2007. Ett av delprojekten är uppbyggnaden av ett uttalslexikon, med det preliminära namnet TPBLex, som är skraddarsytt för talboksproduktion. Målet för lexikonprojektet är att bygga upp ett fullformslexikon med grammatisk uppmärkning som täcker det ordförråd som förekommer i svenska läroböcker.

2. Befintliga lexikon

2.1. Strategi

Det kommer att krävas många års arbete att komma i närheten av målet för lexikonprojektet. Eftersom lexikonuppbyggnaden startar nästan samtidigt som man börjar använda talsyntes för talboksproduktion måste arbetet i början inriktas på att snabbt täcka in det centrala ordförrådet. För att få ett användbart lexikon inom rimlig tid måste man då utnyttja befintliga lexikon. I första hand använder man naturligtvis lexikon i maskinläsbar form, och när de inte räcker till får man komplettera med material på papper. Innan lexikonprojektet startades gjordes därför en inventering av befintliga uttalslexikon. Det visade sig finnas åtskilliga användbara lexikon, både på papper och i maskinläsbar form.

2.2. CentLex

Centrum för talteknologi (CTT) vid KTH har en uttalsdatabas, CentLex, som har

byggts upp i samband med olika talteknologiska forskningsprojekt och är tillgängligt för CTT:s intressenter. Det är ett fullformslexikon med ungefär 400.000 ingångar. Varje ordform har en eller flera uttalsangivelser samt grammatiska uppgifter. Cirka hälften av orden i databasen är egennamn.

Materialet i CentLex kommer från flera olika källor och håller ojämn kvalitet. Delar av materialet har vidarebearbetats och kvalitetskontrollerats av kommersiella aktörer. Sådant material är inte tillgängligt annat än på marknadsmässiga villkor.

En stor del av innehållet i CentLex utgörs av uttalsangivelser för personnamn och ortnamn i den svenska telefonkatalogen. Detta är resultatet av Onomastica-projektet, ett EU-projekt som genomfördes i början av 1990-talet, med KTH och Telia som svenska deltagare. Uttalet genererades utifrån skriftformerna med hjälp av algoritmer och kontrollerades i efterhand (Gustafson 1996). I den version som ingår i CentLex är namn som förekommer minst fem gånger i materialet kontrollerade. En vidarebearbetad version med totalt 800.000 namn, där även namn med lägre frekvens har kontrollerats, används av telefonkatalogföretaget Eniro för numerupplysningstjänster.

2.3. Uttalsordböcker

Det har getts ut två stora svenska uttalsordböcker i modern tid: *Norstedts svenska uttalslexikon* (NSU 1997) av Per Hedelin och *Svenska språknämndens uttalsordbok* (SUO 2003) av Claes Garlén. Båda ordböckerna anger uttalet med IPA-tecken och har producerats med datorstöd.

NSU omfattar cirka 150.000 uppslagsord, däribland många sammansättningar och egennamn. Inriktningen är snarare deskriptiv än normativ. SUO är ungefär hälften så stor – 67.000 uppslagsord. Skillnaden i omfattning beror främst på att SUO inte innehåller några egennamn och inte heller sammansatta ord med regelbundet uttal. Inriktningen hos SUO är snarare normativ än deskriptiv.

Förutom dessa uttalsordböcker för det allmänna ordförrådet har det getts ut ett par uttalsordböcker för egennamn: *Släktnamn i Norden – med uttalsuppgifter*, vilken innehåller 3.300 släktnamn som kan vålla uttalsproblem, och *Svenska ortnamn – uttal och stavning*, som innehåller 3.000 viktiga eller svåruttalade namn. Båda gavs ut före den elektroniska textproduktionens tid.

2.4. Andra ordböcker och uppslagsverk

Allmänna ordböcker och uppslagsverk har som regel uttalsangivelser för vissa uppslagsord, främst egennamn och lånord där uttalet inte direkt kan härledas ur stavningen. Ofta är uttalsangivelserna starkt förenklade, så att de innehåller enbart den betonade stavelsen eller någon annan del av ordet med oregelbundet uttal. Ofta anges uttalet bara med ett betoningstecken i själva uppslagsordet.

I *Nationalencyklopedin* (NE) finns totalt 40.000 fonetiska parenteser med

fullständiga eller ofullständiga IPA-transkriptioner av uppslagsordet och 100.000 betoningsmarkeringar av uppslagsord eller ord i texten. Webbutgåvan av NE innehåller dessutom ljudfiler med uttal av 12.000 främmande namn på 13 olika språk, i samtliga fall inlästa av modersmålstalare.

När det gäller uttal av främmande egennamn är NE den mest omfattande källan i Sverige. Tyvärr har man helt inriktat sig på att ange de främmande namnens originaluttal, inte deras normala uttal i svensk kontext. Exempelvis ges det enbart det tyska uttalet av *Hamburg* och *Nürnberg*, det franska uttalet av *Paris* och det danska uttalet av *Aalborg*, trots att det finns etablerade svenska uttal för alla dessa namn.

Nationalencyklopedins ordbok (NEO) har betoningsmarkeringar för alla flerstaviga uppslagsord och uttalsparenteser för ett mindre antal ord. I webbversionen finns även inspelat uttal av 14.000 ord.

LEXIN är en serie ordböcker med viss encyklopedisk information, avsedda för invandrare. Den innehåller 28.500 svenska uppslagsord, alla försedda med fonematiska uttalsangivelser.

2.5. Internt arbetsmaterial

De språkvårdare som sysslar med uttalsrådgivning har byggt upp samlingar av uttalsrekommendationer för eget bruk. Sveriges Radios språkvårdare har en databas, DIXI, med sammanlagt 10.000 språkliga rekommendationer av olika slag, till stor del rörande uttal av namn som förekommer i radions nyhets- och musikprogram. Svenska språknämnden har också stora samlingar av uttalsrekommendationer, både på papper och i maskinläsbar form.

3. Det nya elektroniska uttalslexikonet

Innehållet i TPBLex kommer att baseras på flera olika källor. En viktig källa är CentLex, dit också sådant nyproducerat lexikalt material som inte skyddas av upphovsrätt kommer att återföras. Detta ömsesidiga utbyte med CentLex förutsätter att fonetisk notation och grammatisk taggning är kompatibla. Fonetiska symboler för xenofoner (det vill säga främmande språkljud) har därför lagts till, bland annat franska nasalvokaler, tyska *ich-* och *ach-*ljud, engelska och franska *r-*ljud samt ett stort antal diftonger. Efter dessa tillägg kan man ange någorlunda acceptabelt uttal för namn från alla västerländska språk utom danska. Några administrativa fält har också lagts till.

En annan viktig källa är SUO, som i expanderad form kommer att utgöra själva kärnan i lexikonet. Ytterligare material, exempelvis från DIXI, kommer att läggas in efter hand som resurserna medger detta.

3.1. Lexikonets omfattning

Det nya lexikonet skall innehålla det allmänna svenska ordförrådet inklusive

egennamn och vanliga sammansättningar. Alla böjningsformer skall ingå. Tillfälliga sammansättningar måste kunna hanteras av systemet, exempelvis genom konkaterering av förled och efterled ur lexikonet. Lånord, framför allt från engelska, måste kunna uttalas någorlunda korrekt. Detta innebär att lexikonet måste bli mycket stort.

SUO innehåller 67.000 lemmar. Generering av alla böjningsformer ger sju gånger så stor ordmängd, alltså närmare en halv miljon ordformer. Till detta kommer alla egennamn, vilket innebär åtminstone en fördubbling av materialet. Inläggning av förekommande sammansatta ord och främmande ord innebär på sikt en ytterligare expansion i samma storleksordning. Sammanlagt kan alltså lexikonet beräknas omfatta minst en miljon ordformer, efter några år snarare två miljoner.

3.2. Uppmärkning av egenskaper

Alla ord i lexikonet kommer att förses med uppgifter om grammatiska egenskaper. Dessa är framför allt till för att möjliggöra avancerad textanalys för att förbättra uttalet, dels genom att satsbetoningen blir mer naturlig och dels genom att heterofona homografer kan särskiljas. Det har ju stor betydelse för talböckernas kvalitet att man uttalar hästens *hov* med o-ljud och kungens *hov* med å-ljud, eller att den grekiske fältherren *Xenofon* uttalas med betoning på den första stavelsen och facktermen *xenofon* med betoning på den sista. I databasen kommer varje post att innehålla fält för bland annat ordklass, genus, kasus, bestämdhet med mera. Informationen i dessa fält kan sedan utnyttjas vid syntaktisk parsning.

Varje post innehåller ett eller flera fält för uttalsformer. Till dessa hör ett antal fält för uttalsegenskaper, vilka skall kunna användas för variation av uttalet beroende på kontext eller talhastighet. Exempel på sådan uppmärkning är uppgifter om talstil (högtidlig, normal, ledig), betoning (betonat eller obetonat i satsen), ursprungsspråk (för främmande ord och egennamn). Dessutom finns uppgift om källan till uttalsangivelsen.

3.3. Skillnader i förhållande till tryckta ordböcker

Ett lexikon för maskinell produktion är en fundamentalt annorlunda produkt än en tryckt ordbok avsedd att läsas av människor. Även om en tryckt ordbok är producerad med datorstöd måste innehållet därför genomgå en fullständig förvandling för att fungera i det maskinella lexikonet.

Det största problemet med traditionella ordböcker är brist på utrymme, och lexikografer har därför under tidernas lopp utvecklat olika lösningar för att hålla ned textvolymen. Det handlar inte bara om att man avsiktligt begränsar antalet uppslagsord, utan framför allt om sådant som utrymmesbesparande beteckningssystem, komprimerad redovisning av böjning, parenteser för angivande av alternativa former, förkortning av ord som nämnts tidigare i samma artikel, hänvisning till andra artiklar

och utelämnning av information som bedöms vara självklar för läsaren. Dessa praktiska lösningar medför svårigheter när man skall göra en ordbok för maskinell produktion, eftersom lexikonets användare, talsyntessystemet, är för dum och obildad för att återskapa den utelämnade informationen. Å andra sidan är tillgången till minnesutrymme i princip obegränsad. All komprimerad information i de ordböcker som importeras måste därför expanderas för att fungera i sin nya miljö.

Böjning av svenska ord anges av tradition med ett begränsat antal ändelser. För substantiv anges exempelvis ändelserna för bestämd form och obestämd pluralform. Varje svenskt substantiv kan ha sammanlagt åtta böjningsformer, men alla kan enkelt härledas ur de två som anges. Informationen kan därför expanderas med hjälp av enkla regler. Även verb, adjektiv och perfektparticip kan expanderas på motsvarande sätt.

Många av de böjningsformer som genereras är bara teoretiskt möjliga former som aldrig kommer att uppträda i verkliga texter. Exempelvis kan ordet *gravid* ha alla dessa böjningsformer:

gravid (positiv, singularis, obestämd, utrum, nominativ)
 gravitt (positiv, singularis, obestämd, neutrum, nominativ)
 gravida (positiv, singularis, bestämd, nominativ)
 gravida (positiv, pluralis, nominativ)
 gravide (positiv, singularis, obestämd, maskulinum, nominativ)
 gravidare (komparativ, nominativ)
 gravidast (superlativ, obestämd, nominativ)
 gravidaste (superlativ, bestämd, nominativ)
 gravids (positiv, singularis, obestämd, utrum, genitiv)
 gravitts (positiv, singularis, obestämd, neutrum, genitiv)
 gravidas (positiv, singularis, bestämd, genitiv)
 gravidas (positiv, pluralis, genitiv)
 gravides (positiv, singularis, obestämd, maskulinum, genitiv)
 gravidares (komparativ, genitiv)
 gravidastes (superlativ, bestämd, genitiv)

De onödiga böjningsformerna är dock inte till något besvär eftersom ingen användare kommer att upptäcka att de finns. De tar visserligen upp en del utrymme i minnet, men sådant är numera billigt, och man behöver därför inte lägga ned dyr arbetstid på att rensa bort dem.

En del uppslagsord har alternativa stavningsformer som av utrymmesskäl anges med parenteser, exempelvis *farbro(de)r*, *omisskänn(e)lig* och *tid(s)plan*. I sådana fall måste artiklarna dubbleras, en med *farbroder* och en med *farbror* som uppslagsord och så vidare. På motsvarande sätt måste parenteser i uttalsangivelser ersättas av dubblerade uttalsfält.

Den stavning som man stöter på i verkliga texter är mer vildvuxen än det som

finns i ordböcker. Detta är delvis en följd av språkvårdande ambitioner från lexicografernas sida, i synnerhet om ordboken har normativt syfte. Men den språkliga variationen beror också till stor del på okunskap och slarv hos författare och redaktörer. Oavsett orsaken måste talsyntessystemet kunna uttala de felstavade orden. Ett ofta felstavat ord i svensk text är *vakuum*. Vid sökning på svenska webbsidor i Google får man följande träffbild:

vakuum	28.300
vacuum	16.900
vacum	3.200
vaccum	1.360
vakum	571
vaccuum	122
vaacum	119

samt ytterligare nio stavningar med färre än 100 träffar.

För att kunna hantera alla dessa stavningar måste TPBLex innehålla både stavningar som ordböckerna avråder från och andra vanliga felstavningar. Däremot behöver lexikonet inte innehålla några avrårda uttalsvarianter.

3.4. Överföring från befintlig ordbok

Svenska språknämndens uttalsordbok (SUO) var den första ordboken som importerades till TPBLex. Arbetet utfördes under våren 2005.

SUO är traditionellt utformad med omfattande användning av olika typer av textkomprimering. Komprimeringen är dock väl genomtänkt och ganska konsekvent genomförd. Mer än 95 % av böjningsformerna kunde därför genereras med regler. För vissa ord med oregelbunden böjning skulle det krävas en specialregel per ord, vilket bedömdes vara en alltför omständlig metod, och därför har dessa ord expanderats manuellt. Hänvisningar till andra artiklar inom ordboken, vilka ofta förekommer vid sammansatta ord och avrårda ordformer, har också behandlats manuellt.

Många ord i SUO har flera uttalsangivelser. Detta beror på att ambitionen är att ange alla uttal som är acceptabla inom svenskt riksspråk. Exempel på detta är att *arkitekt* har uttalsangivelser med både k-ljud och tje-ljud i mitten, och *image* har både franskbaserat och engelskbaserat uttal. Sådana dubbleringar är överflödiga i ett talsynteslexikon, eftersom varje förekomst av ett ord bara behöver uttalas på ett sätt. Däremot är det en fördel att behålla talstilsberoende och betoningsberoende uttalsvarianter. Dessa är försedda med kommentarer som ”högtidligt”, ”ledigt” och ”obetonat”, vilka utan svårighet har kunnat konverteras till standardiserade egenskapstaggar.

Den tryckta versionen av SUO anger uttalet med IPA-tecken. I de textfiler som

levererades från förlaget används i stället en egen kod med vanliga ANSI-tecken, vilken har konverterats till den interna uttalskoden i TPBLex. Genom särskilda konverteringstabeller som har utvecklats i samband med projektet kan den interna koden enkelt konverteras vidare till de olika talsyntessystemens interna koder.

3.5. Återstående arbete

Det fortsatta arbetet kommer under närmast att inriktas på att utöka lexikonet genom import från fler källor. Mest angeläget är att lägga in stora mängder av svenska och utländska personnamn och ortnamn. Ett ordledslexikon för fri kombination är också önskvärt för att tillfälliga sammansättningar skall kunna hanteras på ett effektivt sätt. På lite längre sikt krävs genomgång av uppmärksningen av uttalsvarianter och utveckling av regler för val av rätt uttalsvariant.

Litteratur

- Garlén, Claes; Mattisson, Ann-Christin & Nilsson, Leif O. 1991: *Svenska ortnamn – uttal och stavning*. Stockholm: Norstedts. ISBN 91-1-915051-2.
- Gustafson, Joakim 1996: *A Swedish Name Pronunciation System*. – Stockholm: Kungl. Tekniska Högskolan. – TRITA-TMH 1996:4.
- Hirschfeldt, Ingar Beckman 2005: Från bandsallad till spetsteknik – historien om DAISY. I: *Talboken 50 år!* Enskede: Talboks- och punktskriftsbiblioteket. ISBN 91-88132-39-0.
- Irvall, Birgitta 2005: Från öppna spolar till kassetter. I: *Talboken 50 år!* Enskede: Talboks- och punktskriftsbiblioteket. ISBN 91-88132-39-0.
- LEXIN = *LexIn : Språklexikon för invandrare*. Stockholm: Statens institut för läromedel 1984-1991.
- NE = *Nationalencyklopedin*. Höganäs: Bra Böcker 1989-1996. ISBN 91-7024-619-X.
- NEO = *Nationalencyklopedins ordbok*. Höganäs: Bra Böcker 1995-1996. ISBN 91-7119-970-5.
- NSU = Hedelin, Bertil: *Norstedts svenska uttalslexikon*. Stockholm: Norstedts 1997. ISBN 91-1-971122-0.
- Sköld, Beatrice Christensen 2005: Talbokspionjärer 1930-1961. I: *Talboken 50 år!* Enskede: Talboks- och punktskriftsbiblioteket. ISBN 91-88132-39-0.
- Släktnamn i Norden: med uttalsuppgifter*. Stockholm: Esselte Studium 1977. ISBN 91-24-27122-5. Skrifter utgivna av svenska språknämnden; 60.
- SUO = *Svenska språknämndens uttalsordbok*. Stockholm: Norstedts Ordbok 2003. ISBN 91-7227-309-7.

Skandinaviske lærebøker i leksikografi og leksikologi – tverrnasjonalt oversyn og vurdering

Scandinavian introductory books on lexicography and lexicology – a comparative survey and an assessment. This article gives a brief survey of Scandinavian (Danish, Norwegian and Swedish) introductory books on lexicology and lexicography. Seven books are discussed one by one, and a comparison is done as to which sectors within the field are covered in the three countries.

1. Innleiing

I *LexicoNordica* 9 hadde Bo Svensén (2002:295–311) ein jamførande omtale av fleire relativt nye engelskspråklege innføringsbøker i leksikografi. I samband med planar om ei norskspråkleg innføringsbok som Ruth Vatvedt Fjeld og eg arbeider med, har eg sett nærmare på nokre skandinaviske bøker av denne typen, og her vil eg gi eit oversyn over kva eg har funne. Det kan ikkje forundre nokon at min artikkel blir atskilleg kortare enn Svenséns.

Eg avgrensar meg til dei tre landa Danmark, Noreg og Sverige, i denne rekkjefølgja, og inkluderer både leksikografi og leksikologi slik eg forstår begrepa. Eg gjorde ei forsøksvis avgrensing av desse begrepa mot kvarandre i foredraget mitt i Volda i 2003 (Vikør 2005a:339–341), m.a. basert på definisjonane i Nordisk leksikografisk ordbok. Der er *leksikografi* knytt heilt opp til tekstsjangeren ordbøker, mens *leksikologi* er definert slik:

forskning og teorier knyttet til et språks ordforråd, særlig betydning og strukturelle egenskaper ved de leksikalske enhetene.

Med denne utdjujinga:

Leksikologi er en deldisiplin av lingvistikken og kalles også leksikalsk semantikk. Leksikologisk beskrivelse av et ordforråd danner også ofte grunnlaget for den leksikografiske kodifiseringen.

Denne forståinga sluttar eg meg til med det unntaket at eg oppfattar leksikologi som vidare enn berre leksikalsk semantikk. Leksikologi omfattar all beskriving av det leksikalske nivået i språket, både det semantiske, pragmatiske, stilistiske, historiske (endringar av ordforrådet bl.a. gjennom påverknader frå andre språk og samfunnsendringar, dvs. endringar i det vi kan kalle språkekologien), normative (purisme-problemet), morfologiske (orddanning) m.m.

Terminologifeltet held eg utanfor i denne samanhengen. Like eins ser eg bort frå bøker i meir generell språklære, grammatikk m.m., som inneheld leksikologiske kapittel.

Eg har m.a. gått ut frå det vi har i biblioteket vårt og sett på resensjonane i *LexicoNordica*. Det har gitt følgjande fangst: éi dansk bok, tre norske, og tre svenske (når Svenséns to utgåver blir rekna som éi bok). Den eldste er frå 1987, den nyaste frå 2004. Eg behandlar bøkene land for land. Eg kan ikkje gi ei grundig vurdering av bøkene i ein kort oversynsartikkel som dette, basert på ein knapp halvtimes foredrag, berre eit jamførande oversyn over kva dei dekkjer. Etter gjennomgangen av bøkene legg eg fram ei samanfatting av korleis fagfeltet er dekt i dei ymse landa. Det blir uunngåeleg overflatisk, men forhåpentlegvis kan det gi ei lita tverrskandinavisk oversikt.

2. Dei enkelte bøkene

2.1. Danmark

2.1.1. *Henning Bergholtz og Sven Tarp (o.a.): Manual i fagleksikografi. Udarbejdelse af fagordbøger. Problemer og løsningsforslag. Systime 1993. 320 s.*

Dette er den einaste danske boka i emnet eg har funne. Den er sprungen ut av miljøet rundt fagleksikografistudiet ved Handelshøjskolen i Århus, og er strengt leksikografisk. Bergholtz og Tarp er blant dei som i sterkast grad har ønskt å skilje mellom leksikografi og den lingvistiske deldisiplinen leksikologi. *Manual i fagleksikografi* (MFL) er også lagd opp som ei praktisk bok, som ordet "Manual" seier klart nok. Boka begynner sjølvsagt med å avklare visse begrep innanfor faget, særleg med å definere fagleksikografi i forhold til allmennleksikografi. Deretter får vi ein typologi over fagordbøker, delvis etter antal og art av behandla språk, delvis etter faggrupper, og grunnlagsproblem ved somme av desse typane blir drøfta. Så er det det praktiske arbeidet som blir beskrive stadium for stadium: forarbeidet i form av brukarundersøkingar, materialsamling og leksikografiske handgrep. Seleksjon av lemma og ekvivalentar, utval og framstilling av språklege opplysningar, tilsvarande om faglege opplysningar, dei ulike bestanddelane av ei fagordbok (særleg ulike typar omtekst), ulike tilgangsstrukturar, formaspektet (layout, korrektur), revisjon og ordbokskritikk, og til slutt får vi eit historisk oversyn over danske fagordbøker. Framstillinga er i stor grad normativ, i form av at det heile vegen blir gitt grunngitte råd, anbefalingar og åtvaringar. Vi har altså for oss ei handbok i eigentleg tyding. Boka kom i 1995 ut på engelsk under tittelen *Manual of Specialised Lexicography*.

2.2. Noreg

2.2.1. *Vigleik Leira: Ordlaging og ordelement i norsk. Samlaget 1992. 300 s.*

Dette er ei leksikologisk bok, som tittelen seier. Leira er mangeårig (no pensjonert)

konsulent i Norsk språkråd, og har blant anna hatt ansvaret for nyordsregistrering i Språkrådet i ei årrekke (og er forfattar av boka *Nyord i norsk* frå 1985). Også denne boka er ganske praktisk lagd opp: Forfattaren ønskjer å gi systematisk kunnskap for språkbrukarar om ordlagingsmåtar på eit synkront grunnlag, og delar av boka fungerer som eit oppslagsverk. Det er to hovuddelar. Første delen drøftar teoretisk, men likevel konkret, ulike ordlagingsmåtar: samansetjing, avleiing, og dei meir perifere: samdanning, innkorting, importord, og to viktige problem av klart normativ type: samansetjingsfuga (særleg) og fordelinga av *-ing* og *-ning* i dei to målformene (boka dekkjer både bokmål og nynorsk, men er skriven på nynorsk). Del II består primært av ei alfabetisk liste over ordlagingsselement og bøyingsendingar, og dessutan mindre lister over særlege problem: inkolentnemningar, samansetjingsmåtar (fuger), og ordpar på *-ing* vs. *-ning*. Heile boka er altså primært ei praktisk rettleiing i språkbruk på ordlagingsplanet.

2.2.2. Valerij Berkov: Norsk ordlære. *Universitetsforlaget 1997. 224 s.*

Også denne boka er (i hovudsak) leksikologisk, som tittelen viser. Boka kom først ut på russisk under tittelen *Norvezhszkaja leksikologija* – ”ordlære” er altså ei direkte omsetjing av ”leksikologi”. I sjølve teksta i boka bruker han likevel ”leksikologi” også i den norske utgåva, og han begynner med ei drøfting av begrepet og fagområdet. Dei emna han behandlar i egne kapittel, er først ordet som språkleg eining, dernest det norske ordtilfanget inndelt og analysert ut frå ulike kriterium: semantiske, stilistiske, sosiale – her drøftar han fagspråklege og dialektale ordtilfang – og historiske (importord- og purisme-problematikken). Deretter tek han for seg fraseologi og gir ein typologi over fraseologiske einingar, før han drøftar leksikalske prosessar av ulike slag (namngjeving, tydingsendring, metonymi og metaforikk, og lingvistiske og ekstralingvistiske årsaker til endringar i ordtilfanget). Så tek han opp det emnet Leira behandlar, nemleg ordlagings, men det gjer han svært knapt, så han dublerer ikkje Leiras grundige behandling. Til slutt har han eit kapittel om norsk leksikografi, dvs. eit historisk oversyn og ei inventarisering over ordbøker som dekkjer norsk, pluss korte beskrivingar av nokre av dei. Heile Berkovs bok er svært konkret, mesteparten av teksta består av eit vell med døme som blir grupperte og disponerte etter dei linene eg har skissert her ovanfor.

2.2.3. Anne Golden: Ordforråd, ordbruk og ordlæring i et andrespråksperspektiv. *Ad Notam Gyldendal 2000. 171 s.*

Også denne boka er leksikologisk. Undertittelen viser at ho fokuserer særskilt på andrespråksinnlæring, og boka kjem da også i ein serie for eit fjernundervisningsprosjekt i norsk som andrespråk. Delar av boka gir likevel ein meir allmenn leksikologisk bakgrunn. Boka blir innleidd med ei drøfting av ordbegrepet, og så blir ordforrådet behandla (inndelingskriterium, og ordtilfanget i tekster og hos individ,

der begrepet ”det mentale leksikon” blir beskrive). Leksikalsk semantikk får eit kapittel, like eins ordforståing og tilhøvet mellom språk og tanke. Deretter kjem variasjonar i ordtilfang og ordbruk i ulike språk og teksttypar. Alle desse kapitla er ganske korte; det er snakk om elementære bakgrunnsinnføringar for dei sentrale emna som deretter står på tapetet, og som får meir utførleg behandling: ordlæring, ordlæringsstrategiar, og undervisningsmetodikk i ordkunnskap. Mykje av stoffet er presentasjonar og refereringar frå ymse forskingsprosjekt, og grafar og tabellar har ein meir framstående plass her enn i dei andre bøkene. Mens dei andre bøkene er lingvistiske i trong meining, er denne meir prega av å vere tverrfagleg med ei samanflyting av lingvistik, samfunnsvitskap, psykologi og pedagogikk.

2.3. Sverige

2.3.1. *Bo Svensén*: Handbok i lexicografi. Principer och metoder i ordboksarbetet. *Esselte Studium / Tekniska nomenklaturcentralen* 1987. 279 s. / 2. utgåve: Handbok i lexicografi. Ordböcker i teori och praktik. *Norstedts Akademiska Förlag* 2004. 624 s.

Dette er, som vel kjent, klassikaren på feltet. Den er gjennomført lexicografisk, og først og fremst tenkt som det tittelen seier, ei handbok i det konkrete arbeidet med å lage ordbøker. Det er altså ordboka som tekstsjanger boka tek opp, ikkje ordet som ein lingvistisk kategori. Førsteutgåva er på knapt 300 sider. Etter eit par innleiande kapittel om ordbokstypologi og materialbehandling tek forfattere seg informasjonskategoriene ein etter ein, frå oppslagsordet til etymologi og tilvisingar. Deretter beskriv han ulike ordboksstrukturar, ordbøker som prosjekt og ordbøker i dataalderen. Materialet er heile vegen både einspråklege svenske ordbøker, og tospråklege ordbøker mellom svensk og engelsk, fransk og tysk. Boka kom ut på engelsk i 1993 under tittelen *Practical lexicography*.

Andreutgåva er fordobla i omfang, den er på over 600 sider. Men framleis er avgrensinga til lexicografi i trong tyding klar og gjennomført, og grunndisposisjonen er behalden, men med grundigare behandling av enkeltemna og nye kapittel. Vi finn altså framleis først nokre allment innleiande kapittel med begrepsbestemmingar, ordbokstypologi, datainnsamling og -utval og ordboka som tekst – den tekstteoretiske komponenten er sterkare her enn i førsteutgåva. Så får vi igjen dei ulike opplysningskategoriene behandla og gjennomdrøfta ein for ein, med hovudvekt på definisjonar i einspråklege og ekvivalentar i tospråklege ordbøker. Dei ulike strukturtypane og -nivåa kjem deretter, og så får vi til slutt ein sekvens med dei meir ”ytre” aspekta: ordboksprosjekt, juss og etikk, elektronisk lexicografi, ordboksbruk og ordbokskritikk. Det er blitt ei heildekkjande bok som kjem til å bli ståande som standardverket i mange år framover. (Sjå grundigare omtale og vurdering i Vikør 2005b.)

2.3.2. *Peter af Trampe: Språkbrukaren och orden. Studentlitteratur 1990. 126 s.*

Denne boka ligg litt på kanten av emnet mitt, sidan den ikkje berre behandlar ordnivået i språket, men delar av den gjer det. Boka er ”en introduktion till elementär ordsemantik och till några modernare teorier om vår användning av orden i kommunikationen”, til bruk for studentar på grunnivå i allmenn og nordisk språkvitskap. Etter å ha framstilt nokre semantiske grunnbegrep og ulike typar tydingsrelasjonar (synonymi, hyponymi osv.), og gått litt inn på forskingsmetodar, tek forfattere grundigare for seg ordbruk, med særleg vekt på metaforikk. Deretter kjem eit avrundande kapittel om ulike aspekt ved språkhandlingar. Både setningssemantik og språkhandlingsteori er viktige emne i boka i tillegg til ordsemantik.

2.3.3. *Sven-Göran Malmgren: Svensk lexikologi. Ord, ordbildning, ordböcker och orddatabaser. Studentlitteratur 1994. 150 s.*

Tittelen og undertittelen samanfattar godt kva denne boka tek for seg. Malmgren bruker etter svensk tradisjon *leksikologi* som eit overbegrep for det eg kallar leksikologi og leksikografi, og blir kritisert m.a. for det av Thomsen (1995:289). Men jamvel etter min og bl.a. Thomsens terminologi, ligg hovudvekta i boka på leksikologi. Vi får først ei kortfatta drøfting av ordbegrepet, og så kjem kapitlet ”Ordbildning”, som fyller femti sider, bortimot halve boka. Så eit kort kapittel om fraseologi, og eit om leksikografi, nemleg ei typologisk oversikt over svenske ordbøker. Endeleg kjem eit avslutningskapittel om datamaskinell leksikografi. Såleis kan vi seie at boka sveipar over både leksikologi og leksikografi, men altså med hovudvekta på leksikologi, og innanfor der igjen på det morfologiske aspektet – som i Sverige, liksom i Noreg, vel er det aspektet som skaper dei største praktiske problema for folk flest. Slik supplerer Malmgren kollegaen af Trampe, ved å leggje hovudvekta på dei formelle sidene i motsetning til dei semantiske som af Trampe behandlar.

3. Samanfating

Ein kan sjå av dette at fagfeltet er noko ujamt dekt i dei tre landa. I tabell 1 har eg sett opp ei oversikt over delemna innanfor faget og ført på dei bøkene der desse emna er dekte. Rett nok blir det noko tilfeldig, for det er stor forskjell på kor djupt bøkene går inn på dei enkelte emna, og der ei bok berre så vidt tek opp eit emne utan å gå særleg meir inn på det, har eg sett bort frå det. Men eit visst globalt inntrykk av feltet kan vel dette gi.

Leksikografi

<i>Allmennleksikografi</i>	Svensén
<i>Fagleksikografi</i>	Bergenholtz/Tarp
<i>Ordboksoversikter</i>	Berkov, Malmgren, Bergenholtz/Tarp

Leksikologi

<i>Ordbegrepet</i>	Berkov, Golden, Malmgren
<i>Morfologi</i>	Leira, Malmgren, Berkov
<i>Ordforrådsinndeling</i>	Berkov, Golden
<i>Diakroni</i>	Berkov
<i>Normering</i>	Berkov
<i>Semantikk</i>	Golden, af Trampe, Berkov
<i>Mentalt leksikon</i>	Golden
<i>Leksikalske prosessar (metaforikk, metonymi osv.)</i>	Berkov, af Trampe
<i>Stilistikk</i>	Berkov
<i>Fraseologi</i>	Berkov, Malmgren
<i>Kontrastivitet</i>	Golden
<i>Didaktikk / Læring</i>	Golden

Tabell 1.

Den einaste danske boka eg har registrert, handlar om leksikografi, og attpåtil eit spesialfelt innanfor leksikografien: fagleksikografi. I Noreg står derimot leksikologien i sentrum – trass i at ”leksikologi” faktisk ikkje finst som fagdisiplin ved norske universitet. Der heiter det ”leksikografi”, som blir brukt som overbegrep på same måten som ”leksikologi” i Sverige (ifølgje Malmgren). Eg har sjølv, i samarbeid med Ruth Vatvedt Fjeld, undervist i leksikografi ved Universitetet i Oslo i ei rekkje år, og den grunnboka vi til no har brukt der, er førsteutgåva av Svenséns. Av dei norske leksikologibøkene er Berkov den mest heildekkjande, men den er svært kasuistisk og lite teoretisk, for så vidt sterkt merkt av at forfattaren er leksikograf, i den forstand at registrering og opplisting av enkeltdøme har ein dominerande plass i boka. På same måten er Leiras bok konsentrert om den morfologiske sida ved ordet som språkleg storleik, dvs. avleiing, samansetning og til dels bøyning – i tråd med det tyngdepunktet morfologi har i norsk språkdebatt og språkplanlegging i det heile, og sjølv sagt sterkt basert på Leiras stilling som konsulent i Norsk språkråd. Golden på si side legg hovudvekta på andrespråksproblematikken, som er hennar fagfelt. I

Sverige har leksikografien fått eit heildekkjande verk, mens ulike aspekt av leksikologien er blitt meir kortfatta og elementært behandla i to lærebøker. Emne som ordtilfangssamansetning, ordtilfangshistorie, fraseologi, m.m. er blitt nokså stemoderleg behandla i alle landa. Slike emne kjem rett nok inn i andre samanhengar, som i grammatikkar og språkhistorier, som gjerne har med kapittel om det leksikalske nivået, men det blir likevel meir perifere innslag i verk der hovudinteressa ligg på andre felt. Eg kunne nok ha gått inn i og vurdert dei store referansegrammatikkane for norsk og svensk, men eg har valt å ikkje gjere det i denne samanhengen.

Også på det leksikografiske planet kan vi seie at at fleire emne er betre dekte enn det kan sjå ut til her, ordbokshistoriske emne, t.d., er behandla ulike stader i artikkelform, og vi har òg grundigare innføringsbøker i enkelte store ordboksverk, nemleg i SAOB ved Ekbo og Loman (1971) og Lundbladh (1992), og i ODS ved Hjorth (1970). I Noreg, og truleg i dei andre landa, finst det fleire kåserande og populærvitskaplege bøker om ord, ordbruk, ordhistorie og ordtilfangshistorie – som alltid er populære emne. Men alt dette er fragmentarisk. Såleis måtte Fjeld og eg i undervisninga vår i stor grad basere oss på eit slikt ”fragmentarisk” pensum. Svenséns bok er den overlegent beste på sitt felt og godt brukbar i alle dei tre landa, ikkje minst etter den radikale utbygginga som skjedde i samband med den nye utgåva. Den gjekk rett nok så langt at sjølve omfanget kan bli eit problem; det kan komme til å vise seg at den vil fungere meir som eit fordjupingsverk enn som ei innføringsbok. Det er ikkje utenkjeleg at førsteutgåva også i framtida vil bli mest brukt i elementærfasen, mens andreutgåva eller utvalde kapittel i den blir ei påbygging for meir avanserte brukarar. Uansett er det eit behov for ei elementær behandling av leksikografisk metodikk også frå eit dansk og eit norsk synspunkt – blant anna fordi både dei språklege problema og dei leksikografiske tradisjonane ikkje alltid er identiske med dei svenske.

Litteratur

(med unntak av dei omtalte bøkene)

- Bergenholtz, Henning o.a. 1997: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Ekbo, Sven og Bengt Loman 1971: *Vägledning till Svenska Akademiens ordbok*. Stockholm: Läromedelsförlagen.
- Hjorth, Poul Lindegård 1970: *Vejledning i brug af Ordbog over det danske Sprog*. København: Gyldendal.
- Lundbladh, Carl-Erik 1992: *Handledning till Svenska Akademiens ordbok*. Stockholm: Norstedts.
- Svensén, Bo 2002: Några nya introduktionsböcker om lexikografi. I: *LexicoNordica* 9, 295-311.
- Thomsen, Knud Troels 1995: [Melding av Sven-Göran Malmgren: Svensk lexikologi]. I: *LexicoNordica* 2, 287-299.

- Vikør, Lars S. 2005a: Språk og verd i Norsk Ordbok. I: R.V. Fjeld og D. Worren (red.): *Nordiske studiar i leksikografi* 7. Oslo: Nordisk foreining for leksikografi, 339-347.
- Vikør, Lars S. 2005b: Ei metaleksikografisk perle. I: *LexicoNordica* 12, 331-346.

Molbech som mønster for Aasen

Molbech as a Model for Aasen. The lexicographical tradition in Norway was sparse before Ivar Aasen (1813-96) published *Ordbog over det norske Folkesprog* ('Dictionary of the Norwegian dialects'; abbr. ONF) in 1850, which appeared to be a modern dictionary. *Dansk Ordbog* ('Danish dictionary'; 1833) by Christian Molbech (1783-1857) is considered, in parts, to be a model for the lexicographical practice in ONF. This article points out elements where Molbech as a model is obvious, and also where Aasen made his own contribution to Nordic lexicography, especially in applying the methods of written-language lexicography on dialect material to describe a national vocabulary.

1. Innleiing

Den etablerte leksikografiske tradisjonen i Noreg vart innleidd i 1850, då Ivar Aasen (1813-96) la fram *Ordbog over det norske Folkesprog* (ONF) (jf. Worren 2003:407). Oppspelet til den etablerte leksikografiske tradisjonen var elles svært sparsamt med berre fire prenta ordbøker over norsk ordtilfang frå 1600-talet til byrjinga av 1800-talet. Desse ordbøkene må dertil kallast dialektordbøker, ettersom dei ikkje hadde eit eige norsk standardmål til normeringsinstans (jf. Vikør 2000:119).

Den danske litteraturforskaren Hans Hauge har kalla tida då den norske leksikografiske tradisjonen etablerte seg, for postkolonial etter ein av dei nyare litterære teoriane med den grunngevinga at det heldt seg eit visst dansk hegemoni i Noreg "fra 1814 og et stykke ind i århundredet" (Hauge 2001:371), altså også etter Kielfreden då Sverige vart den nye unionspartnaren. Språkleg sett vart det danske hegemoniet faktisk forsterka av denne unionen som eit taktisk mottrekk mot svensk påverknad (Skautrup 1953:113). I vår samanheng kjem hegemoniet greitt til uttrykk i oppdraget som Ivar Aasen fekk i 1842 frå Det kongelige Norske Videnskabers Selskab i Trondheim om å samle inn (1) ordtilfang frå dei norske målføra med tanke på ein sams grammatikk og ei tilsvarande ordbok og (2) dessutan ei ordbok "hvori vort nærværende Skriftsprogs Ord" er 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. 78). Desse tre verka er hegemoniske i den forstand at to av dei største namna i dansk språkvitskap hadde hand om dei, og Aasen fekk dei også i hende.

Rasmus Rask (1787-1832) hadde også gjeve ut den islandske ordboka (eigl. Björn Halldórsson: *Lexicon Islandico-Latino-Danicum*), som kom ut i 1814. Han hadde særskilt sett til dei danske forklaringane (Kvaran 2000:85). Vi kjenner elles Rask som den filologen som utvikla den historisk-komparative metoden i språkvitskap med

utgangspunkt i dei nordiske språka (Hovdhaugen et al. 2000:159f.). Den naturvitskaplege organiske tenkjemåten syntse seg både i den historiske utviklingsdimensjonen og i den typologiske klassifiseringa som hadde til føremål å påvise samanheng og slektskap i og mellom språk (op.cit. 157).

2. Problemstilling

På denne bakgrunnen, der Noreg var i politisk union med Sverige, men framleis hadde danske kulturelle og akademiske ideal, skal det leggjast fram døme på korleis denne hegemoniske mønstergjevinga syner seg i dei leksikografiske arbeida til Ivar Aasen, då først og fremst frå ONF. Av dei tilrådde verka frå Vitskapselskapet er det særleg *Dansk Ordbog* (DO;1833) av Christian Molbech (1783-1857) som blir nytta som kjelde i denne samanlikninga.

Manifesteringar på mønstergjeving kan finnast på mange plan, såleis på det ideologiske planet (samfunns- og språksyn), på det metodiske planet (oppbygginga) og i konkret leksikografering (på makro- og mikroplanet). Manifesteringane kan òg ytre seg som indirekte eller direkte kommunikasjon mellom verka og i parallell struktur i tillegg til direkte referansar. Spørsmålet er då om Ivar Aasen var noko meir enn ein tradisjonsberande epigon.

3. Formell omtale av DO og ONF

Ytre sett er skilnaden mellom DO og ONF stor. DO er ei tobandsordbok, der kvart band har tospalta sider med eiga paginering: band 1 med 700 sider (A-L) og band 2 med 732 sider (M-Ö). Talet på oppslagsord ser ut til å vere rundt 30 000. Ordboka er trykt i fraktur, former frå andre språk er trykte i antikva (jf. figur 1).

ONF er ei eittbands tospalta ordbok på 634 sider. Her er talet på oppslagsord om lag 25 000, irekna ord på ikkjealfabetisk plass (Kruken og Aarset 2000:XXII). Også i ONF er hovudskrifta fraktur (jf. figur 2). Valet av skrifttype var eigenleg allereie lagd med Aasens grammatikk frå 1848, *Det norske Folkesprogs Grammatik*, då med den grunngevinga at boka skulle bli lettare å lese for ”norske Almuesmænd” (Aarset 1996:XXI).

I omfang kan altså ikkje ONF tevla med DO. Men skriftbiletet gjev derimot det same tradisjonelle inntrykket.

Den nye historiske kunnskapen om ”det gamle Sprog” er det filologiske beviset på at folkemålet er ”Fædrenesproget”, som skulle atterreisast.

5. Ordtilfangsvurderingar

Akademiprinsippet var opphavleg knytt til utviklinga av dei store europeiske språka på 1500-talet for mønstergyldig språkbruk (jf. MacArthur 1986:94). Molbech tilpassa prinsippet si tid og kalla DO ”en almindelig Ordbog over det danske Sprog, saaledes som det forekommer i Skrift og dannet Tale” (Molbech 1833:IX). Men han seier det i ein kontekst der utelukking (eksklusjon) av ordtilfang er like viktig som medtaking (inkludasjon). Både eksklusjon og inkludasjon krev såleis metodiske grep.

Særleg argumenterer Molbech for å ta med eit utval av ”de saakaldte forældede Ord” (Molbech 1833:VIII): Utvalet skal grunngjevast med slektskap til samtidsorda og fylle eit sagn i samtidsspråket. Molbech kombinerte dermed akademiprinsippet og den historisk komparative metoden i eit slags restitusjonsprinsipp.

For ONF med berre folkeleg talemål som materialgrunnlag var ikkje akademi-prinsippet i den klassiske forma noko eigenleg utgangspunkt. Men prinsipp for inkludasjon og eksklusjon var avgjort relevante, ikkje minst etter restitusjonsprinsippet mellom samtidsspråket og gammalnorsk.

5.1. Inkludasjon

For Molbech var autentisk materialgrunnlag frå skjønnlitteraturen eit viktig supplement til ”selvdannede Exempler” (Molbech 1833:IX) for å gje autoritet til framstillinga. Molbech brukte 10 år på materialinnsamlinga, medan den eigenlege leksikografiske gjennomarbeidinga tok berre 5 år (Borup 1954:206f.).

Skriftlege kjelder for ONF var fleire lokale ordsamlingar, som var nokre fleire enn dei fire prenta som er nemnde framanfor (jf. Kruken og Aarset 2000:5). Prinsippet om autentisk etterprøving var ein føresetnad for at ord frå slike samlingar kom med i ONF. Det vil seie at Aasen aldri tok med eit ord frå ei slik samling utan at han sjølv hadde fått stadfesta det av ein påliteleg informant, helst på det aktuelle dialektområdet. Vurderingane til Molbech i DO kan nok ha vore ein inspirasjon til ein slik vitskapleg arbeidsmåte.

5.2. Eksklusjon

Samstundes som tittelformuleringane ”*det danske Sprogs Stammeord*” og ”*efter den nuværende Sprogbrug*” i DO er innholdsforklarande, er dei sjølvsagt også eit kriterium for legitim eksklusjon. Vekta ligg på det genuint heimlege, men tillegget om samtidsspråkbruk gjer at tilhøvet til den ekskluderte delen av ordtilfanget, som vi i dag gjerne kallar importord (t.d. Sandøy 2000:19), måtte klargjerast.

I tråd med heile språkbyggingssprosjektet som ONF var ein del av, var poenget å

beskrive eit ordtilfang som synte nasjonaliteten. Ordtilfang som ikkje gjekk inn i den samanhengen, var uaktuelt.

5.2.1. *Importord*

Av importord får tyske ord ei særskild handsaming i DO, naturleg nok etter geografiske tilhøve, attåt andre framandord. Berre i dei tilfella der ”Skriftsproget” hadde gjeve slike ord ”den fornødne Borgerret”, kunne det kome på tale å ta dei med (Molbech 1833:XIII).

Tysk ordtilfang

Molbech slo særleg ned på ”Letheden i at gjøre Tydsk til Dansk” (Molbech 1833:XII). Det klassiske utestengingdømet til sitering gjeld *Geburtsdag*:

Man vilde dog ei i en dansk Ordbog fra det 19de Aarhundrede søge *Geburtsdag*, skiøndt dette i hvert dansk Øre skærende Ord, der ikke eier meer end et eneste Bogstav, som tilhører os, endnu er saa tildeles hiemme i det almindelige, men derfor i dette Tilfælde ikke desmindre barbariske Omgangssprog, selv i dannede Kredse, at det er iblandt de største Sielenheder, at høre en Dansk sige *Fødselsdag*. (Molbech 1833:XIII)

Dømet er ein attergangar i litteraturen om Molbechs kritiske haldning til slike ord, der opphav som godkjenningssgrunn vart rangert langt føre faktisk bruk (Duekilde 1998:102).

Gjennom heile nynorsksoga har importord frå dansk og tysk (gjerne kalla ”anbehetelse-orda” etter nokre typiske føre- og etterledd) vore eit stridsspørsmål, ikkje minst med ei tilsvarande grunngjeving som finst i føreordet i DO (Aasen skilde ikkje mellom lånord og framandord (Bondevik og Myhren 2000:XVII)). Det vil seie at slike importord nesten umerkeleg tek plassen frå dei heimlege orda.

Tyske importord var ikkje sameinlege med ”Nationaliteten i Sproget” (Kruken og Aarset 2000:4). Språkvitskapleg hang motstanden mot denne ordtypen saman med ein teori om at han skipla ordlagingssevna, der omlyden spela ei viktig rolle. Men det var først med *Norsk Grammatik* (1864) at Aasen gjorde greie for denne tenkjemåten (Walton 1996:418).

Med ei så klar avvising som Molbech har av slikt tysk språktilfang, er det eigenleg ikkje til å undre på at kategorien ikkje er nemnd i det heile i føreordet i ONF.

Andre importord

Andre slags importord, gjerne kalla framandord, opnar DO for når dei har blitt ”ligesaa almeenforstaaelige, som de oprindelige danske Ord” (Molbech 1833:XIII). Elles skal dei utelatast, fordi dei alltid vil vere ein uorganisk del i ei ordbok ”over det danske Tungemaal”. Dei er gjester som kjem og går utan borgarrett i ”Sprogets Hiem” (ibid.).

For denne ordtypen finst det eit parallelt synspunkt i ONF. Her heiter det at framandord som er innkomne med skriftspråket, er utelatne, ”forsaauidt som de ikke have antaget nogen egen Form eller egen Betydning” (Kruken og Aarset 2000:6).

Om dette parallelle kriteriet har ført til eit nokolunde tilsvarende utval av importord, har ikkje vore mogleg å granske innafor ramma av denne artikkelen.

Andre utestengde ord

Eksklusjonkriteriet ovanfor er knytt til opphav. Akademi prinsippet for DO stengjer sjølvsagt ut heile det tilsvarende kjeldetilfanget som ONF byggjer på, og som Molbech for dansk gav ut i ei eiga ordbok, *Dansk Dialect-Lexikon* (DDL; 1841).

Ein ordtype er det elles Aasen som spesielt nemner som eksklusjonverdig: ”Udeladte er ogsaa adskillige Eder og Skjældsord samt enkelte andre, som kun bruges i grov og obscøn Tale og som ikke synes at være til nogen Berigelse for Sprogstoffet” (Kruken og Aarset 2000:5). Disse ordkategoriane frå såkalla lågspråk er ikkje omtala i DO, men påstanden til Aasen om at slike ord ikkje gjer språket rikare, minner i nokon grad om Molbechs ”Barbarismer og slette Ord” (Molbech 1833:XIII) jamvel om hans ordlegging helst går generelt på ”uakademiske ord” (jf. Haugen 1984:19).

6. Sjølve leksikografien

6.1. DO og ONF som morsmålsordbøker

Titlane til DO og ONF gjev berre signal om kva som er utgangsspråket, høvesvis dansk og norsk. Frå hovudtittelen til DO vil ein i dag utan vidare rekne med at det er ei eittspråkleg ordbok, og slik ser det ut til at tenkjemåten var på Molbechs tid også. Når Molbech omtalar arbeidet med ordforklaringane, er ikkje nasjonaliteten på forklaringspråket emne. Berre indirekte er det sagt at det er dansk med at det er lagt vinn på ein ”oprindelig dansk Sprogtone” (Molbech 1833:XVII).

Både i føreordet til 1848-grammatikken og i ONF utviklar Aasen eit resonnement som seier at ”Skriftsproget” er ”kun sideordnet og nærbeslægtet med det Norske” (Aarset 1996:9). Men det er opplagt at forklaringane i ONF skal formast ut i ”Skriftsproget”. Etter vanleg tenkjemåte er difor ONF formelt sett ei tospråkleg ordbok. Men Aasen hadde eigenleg ikkje noko val når han ville nå flest mogleg lese-kunnige nordmenn. ”Skriftsproget” var einaste alternativet. På dette punktet er det elles første gongen Aasen nemner DO i ONF. Han drøftar problema med å gje att tydingane i eitt språk på eit anna, og nemner særleg det dilemmaet at same personen sjeldan er like god i både kjelde- og målspråk. I dette tilfellet har Aasen hatt ”en stor Hjælp i Molbechs danske Ordbog”, men legg til at ein lærer ikkje eit språk berre frå ordbøker (Kruken og Aarset 2000:13). Resonnementet er at danskar og nordmenn ikkje hadde likeverdige tilgang til skriftkulturen for dansk.

6.2. Makrostruktur (ordning av ordtilfanget)

DO er (vel) den første eittspråklege ordboka som la fram ein heilskapleg presentasjon av dansk ordtilfang med den avgrensinga som låg i akademiprinsippet. Tilsvarende er ONF den første ordboka som la fram eit fulldekkjande systematisk gjenomarbeidd norsk ordtilfang. I begge ordbøkene er hovudprinsippet for ordninga initialalfabetisk.

6.2.1. Alfabetisering (a-ø)

Jamvel om ”Skriftsproget” er sams redigeringspråk i DO og ONF, kan alfabetisering likevel vere ei utfordring med mange konvensjonar å forhalde seg til. I DO er ø siste bokstaven ettersom å-lyden enno ikkje hadde fått eit eige teikn. DO har dobbelskriving av *a* for denne lyden, dvs. *aa*.

I sjølve alfabetiseringa følgjer ONF ordninga i DO, men når det gjeld teiknet for å-lyden, går Aasen sin eigen veg. Aasen fekk laga eit eige teikn, som var ei samanskriving (ligatur) av to *a*-ar ”da dette vilde være til megen Hjælp især i de Tilfælde, hvor *aa* og *a* støde sammen” (Kruken og Aarset 2000:XXV). Ikkje berre brukte han eit eige teikn, men Aasen gav det òg ein eigen alfabetisk verdi. I rekkjefølgje gjekk det inn mellom *a* og *b*. I DO har derimot kvar av *a*-ane sjølvstendig verdi. For alfabetbolken *da-de* syner det seg med at *då*-orda i DO kjem rett etter *da*, medan dei kjem rett føre *de* i ONF (berre hovudorda er med):

DO: **Da**, Daa, Daab, Daad, Daane (v), Daane (subst.), Daare (subst.), Daare (v), Daarlig, Daase, **Daddel ... Davre, De ...**

ONF: **da** (adv), **da** (pron.), **Dabbe ... davra**, daa, daaas, daae, daaig, Daam, daama, daamlaus, daamleg, Daamp, Daankar, daar, daara, daarleg, Daarskap, Daase (m. 1) Daase (m. 2), daatt (sup.), daatt (adv), daattlege, Daavr, de ...

Dette sjølvstendige bidraget frå Aasen til norsk makrostrukturell leksikografi fekk nådestøyten då norsk skriftkultur innførte å-en som eige skrifteikn i 1917 etter at latinsk skrift hadde vore så godt som einerådande i fleire tiår. Teiknet var plassert til slutt i alfabetet etter ”ende-prinsippet” (Torp 2002:4).

Jamvel om initialalfabetisering er hovudprinsippet i begge ordbøkene, er det modifisert slik at samansetningar i DO følgjer alfabetisk etter hovudordet med førsteleddet som alfabetiseringsbase, før hovudalfabetiseringa går vidare:

Dag, *Dagarbeide ... Dagverk, Dagetal, Dagsarbeide ... Dagsverk, Dages.*

Molbech kalla ordninga for alfabetisk-etymologisk, som han såg på som brukarvenleg (Molbech 1833:XVI).

Av dei 25 000 oppslagsorda i ONF står berre 22 011 på rein alfabetisk plass

(Kruken og Aarset 2000:XXII). Resten er gjerne knytte anten til føre- eller etterledd (t.d. *Sleakar* under *Kar* 'mann' og *Brurableia* under *Bleia* 'laken' (op.cit. XXI). Med tanke på den sterke vekta både Molbech og Aasen la på å syne ordslektskap, er alfabetisering med ordflokk-innskott nesten sjølvgeven.

6.3. Mikrostruktur (den einskilde ordartikkelen)

Det er helst utforminga av mikrostrukturen som særmerkjer ei ordbok. Etter *Nordisk leksikografisk ordbok* (NLO; under *informasjonstype*) kan informasjonskategoriane i mikrostrukturen delast inn i fem hovudtypar: den formelle, den semantiske, den kombinatoriske, den ensyklopediske og den historiske.

Både av omsyn til plassen og i samsvar med at den faglege styrken til dei to leksikografane ikkje minst knyter seg til *den semantiske kategorien*, skal jamføringa av mikrostrukturen avgrensast til denne kategorien.

6.3.1. Ordforklaringane

Molbech formulerte berre dei allmenne prinsippa som han brukte i ordforklaringane: Med bakgrunn i historisk-kritisk kunnskap om språkbruken skal ein på ein nøyaktig måte sameine "logisk Bestemthed ... med Klarhed og Korthed". Sjølve grunnsetningane om ordforklaring "maa Ordbogen selv lære dem, der vil bruge den" (Molbech 1833:XVI f.).

Når DO er ei eittspråkleg ordbok med vekt på stam- eller grunnorda i dansk, vil det seie at DO må ha mange omgrepsanalysar ordlagde i fleirordige forklaringar eller definisjonar, som t.d. ved *Ko* (sjå figur 3). Det er nettopp framlegginga i definisjonsform av slike omgrepsanalysar, som har gjeve Molbech varig ry i dansk leksikografi. Det Aasen og seinare leksikografar kunne ta mønster av, var kva språklege verkemiddel han brukte i "en verbal beskrivning av begreppet", som det heiter hos Svensén (2004:273).

For ONF som tospråkleg ordbok mellom norsk folkemål og dansk vil det seie at ONF skulle ekvivalere tydingsverda i norsk folkemål mot den tydingsverda som "Skriftsproget" representerte. Mål språket er, som det er nemnt ovanfor, ikkje uttrykt i tittelen til ONF, noko det derimot er gjort i 1873-utgåva som har tittelen *Norsk Ordbog med dansk Forklaring*. I dei tilfella der det er ekvivalens mellom norsk folkemål og dansk, kunne ordforklaringa såleis berre vere ein eittords ekvivalent. Men jamvel mellom to så nærskyldige språk som dansk og norsk har det eine språket "en Mængde Udtryk for Begreber, som det andet ikke har" (Kruken og Aarset 2000:14). Aasen utdjuar dette med å nemne både ikkjespråklege og språklege moment frå ulike nivå som spelar inn i tydingsskildringa av omgrepsverdene mellom norsk og dansk:

- særeigen natur
- særeigen levemåte
- ulike uttrykk for lyd, rørsle og utsjånad
- ulikt omgrepssomfang

Konklusjonen er difor at han har ”anseet det nødvendig at bruke mange Ord i Forklaringen og at adskille de forskjellige Betydninger, om endog Ordet skulde synes at være bekjendt af Skriftsproget” (ibid.).

Utfordringane

Som tospråksleksikograf hadde då Aasen to hovudutfordringar. *Den første* galdt å fastslå om

1. same nemning fanst med same tydingsomfang
2. same nemning fanst, men med skilnad i tydingsomfang
3. omgrepet fanst, men med ei anna nemning i dansk
4. omgrepet ikkje fanst og dermed heller ikkje nemning

Den andre utfordringa låg i typane 2 og 4. Der måtte han på eiga hand gje ”en verbal beskrivning av begreppet”. Mønsteret for beskrivinga hadde han i type 1 frå Molbechs DO og eventuelt type 3. Her følgjer nokre døme på dei ulike typane:

Til 1: Same nemning og same omgrep med ’ku’ som døme (halvfeit i ordforklaringane i dette dømet og nedanfor er sett til her):

DO	ONF
Ko, en. pl. Køer. [Isl. <i>Ku</i> . N. S. <i>ko</i> .]	Ku, f. (Fl. Kyr),
et Huusdyr, som er Tyrens ell. Oxens Hun (<i>Bos taurus</i>), og faar dette Navn, naar den første Gang har kælvet. (s. <i>Kalv, Qvie</i>). <i>Koen er med Kalv, er drægtig. En gold Ko. Denne Ko staaer længe gold. En Malkeko.</i>	en Ko.
	– Formen <i>Ku</i> bruges kun i en Deel af Landet (Stav. Amt, Tr. og Ag. Stift), hvorimod man paa andre Steder siger <i>Kyr</i> baade om een og flere; saaledes i B. Stift, Rbg. Tell. og flere. (G. N. <i>kú</i> , pl. <i>kýr</i> ; Isl. <i>kýr</i> , i Eental). Fleertal paa nogle Steder <i>Ky</i> (<i>Kjy</i>); – bestemt Form: <i>Kyn-n'e</i> (for <i>Kyrne</i>), <i>Kydna</i> , <i>Kynn</i> ; – Dativ i Nordre Berg.: <i>Ku'naa</i> (for <i>Kunom</i>). I Sammensætning er Formen <i>Ku</i> mest almindelig, f. Ex. Kuband, Kubaas ; men i Søndre Berg. hedder det <i>Kyra</i> (<i>Kyraband, Kyrafot</i>).

Figur 3. Ko og Ku

Til 2: Tilsvarande nemningar, men ulikt tydingsomfang, med 'kalle' som dømme:

DO	ONF
Kalde , v.a. 2 og 1. (<i>kaldte</i> og <i>kaldt</i> er den i Talen allene brugelige Form.) [Isl. <i>kalla</i> .]	kalla , v. a. (<i>a - a</i>),
1. bede eller byde een at komme. <i>At kalde een til Bords. At kalde ad een, p. een.</i> "Jeg kaldte ad Eder, og I svarede ikke." Jer.7.13, "Om jeg end skriger og vaander mig og kalder nok saa høit, er det, som han slet ikke hørte mig." Rahb. <i>At kalde</i> (indkalde, indstævne) <i>een for Retten. At kalde een frem, kalde een for sig, kalde een ind fra Gaden. At kalde Folket sammen.</i> "Kongen kaldte Jøderne t. Hielp." 1. Macc.11. – <i>han lod kalde Præsten t. sig. Hvorfor lod du kalde mig?</i> (Ved en Egenhed i Sprogets Ordføining siger man ogsaa, og sædvanligen: <i>Han lod Præsten kalde. Hvorfor lod du mig kalde?</i>) <i>At kalde een tilside, kalde een ud. At kalde noget i Tvivl</i> ∴ tvivle derom.	1) kalde, benævne. G. N. <i>kalla</i> .
2. nævne, give Navn. (jvf. <i>bedde, nævne</i> .) "En Ting <i>hedder</i> m. det Navn, som den ifølge sin Natur har, den kaldes m. den Benævnelse, som Villie, Vane, Omstændigheder have givet den." Sporon. <i>At kalde en Ting v. sit rette Navn. Hvorfor vil du kalde ham hvad han ikke hedder?</i> "Der ere Ting, som – ikke allevegne kunde kaldes, som de hedde; derfor komme de omsider til at hedde, som man kalder dem." Sporon. <i>Hvad kaldes (hvad Navn har) den Plante? – At kalde op, give Navn efter een. Pigen blev kaldt op (opkaldt) efter Moderen. Han lod sin Fader kalde op</i> ∴ <i>gav Sønnesønnen dennes Navn.</i>	2) lokke, kalde til sig. Sædvanlig kun om at kalde Dyr til sig; hvori- mod man siger <i>ropa</i> om at kalde et Menne- ske. <i>kalle Hunden</i> : lok Hunden hid! –
3. udvælge, beskikke, udnævne (især, og nes- ten allene om Geistlige Embeder.) <i>At kaldes til Præst. Greven har kaldet ham og Kongen bekræftet Kaldet.</i>	3) udskjælde, give fornærmelige Navne. Ofte som v. n. for Ex. <i>Han baa' kalla aa banna</i> (skjendte og bandede).
	– <i>D'æ so te kalla</i> : man maa kalde det saaledes (om noget som ikke svarer ganske til Nav- net). <i>Dæ va færigt, te kalle</i> , ∴ det var saa godt som færdigt. (Nordre Berg.). – <i>kalla aat seg</i> :

	lokke til sig. <i>kalla heim Kynn'e</i> : kalde Kørerne hjem. <i>kalla upp-atte</i> : opkalde; f. Ex. <i>Ho kalla uppatte Far sin</i> , o: hun gav Barnet Navn efter hendes Fader.
--	--

Figur 4. Kalde og kalla

Til 3, same omgrep men ulike nemningar, med 'kveite' som dømme:

DO	ONF
Helleflynder , en. pl. <i>-flyndere</i> .	Kveita , f.
en meget stor Saltvands-Fisk af Flynder-Slægten, som spises, og af hvis Finner, med en Deel af Huden og Fedtet, i Norge tilberedes det saakaldte <i>Rav</i> og <i>Rækling</i> . <i>Pleuronectes hippoglossus</i> ell. <i>maximus</i> .	Helleflynder (Hippoglossus). Overalt vest- og nordenfjelds. Derimod <i>Hellefisk</i> i de søndenfjeldske Kystegne. [Hellefisk, m. Helleflynder (= <i>Kveita</i>). Kr. Stift. (Isl. <i>heilafiski</i> , og <i>heilagfiski</i> ; Sv. <i>helgeflundra</i> .)]

Figur 5. Helleflynder og Kveita

Til 4, omgrep og nemning er ukjende i dansk,

a) med 'kakse' som dømme:

DO	ONF
Manglar	Kakse , m.
	en mægtig Mand; A. Stift, om Bønder som have store Jorder og meget at raade i Bygden. (Sv. <i>kaxe</i>). Ogsaa <i>Bikse</i> og <i>Bause</i> .

Figur 6. Kakse

b) med 'knyste' som dømme, der nemninga og omgrevet finst i DDL:

DO	DDL	ONF
Manglar	knyste, knystre , v. n.	knysta , v. n. (<i>e - e</i>),
	(J.) give en sagte Lyd fra sig, smaa-hoste, mukke, smaalee, fnise, o.d. (Mors; beslægtet med det mere almindelige: at <i>knye</i> .) Ligeledes fra Ribe Stift, hvor det mest bruges om at smaa-hoste, "og en saa-	knyste, mukke. (Tildeels i B. Stift). <i>Lat ikke dæ knyste 'ti deg</i> : vær stille, giv ikke en Lyd. Ellers: <i>knyste</i> , <i>gnyste</i> . (Sv. <i>knysta</i>).

	<p>dan Hoste kaldes <i>Faareknyste</i>." S. ogsaa Aagaards Tørningehn. S. 226. – Paa Falster: <i>Knyste</i>, smaa-knurre, pukke; naar man ikke rigtig vil eller kan komme frem med sin Vrede. (jvf. det T. <i>knitern</i>, om Lyden.)</p>	
--	---	--

Figur 7. knyste, knystre og knysta

Mønstertaking

Når det gjeld mønstertaking, ser vi frå type 2 at det finst ein klar parallell i definisjonssyntaksen for *kalle* (figur 4) der både DO og ONF har verbet *gi* som hovudord og *namn* som kjerne i utfyllinga (differentia specifica):

Kalde 2: give Navn [i tillegg til *nævne*]

kalla 3: give fornærmelige Navne (men *benævne* under *kalla 1*)

Også type 4, der Aasen verkeleg måtte stå på egne bein, bør definisjonsutforminga av *Kakse* få godkjent. Valet av overomgrep ('Mand') og av særdrag ('mægtig') og bruksområdepresiseringa ('om Bønder som have store Jorder og meget at raade i Bygden') toler parallellen til den tilsvarende utforminga i DO under *Ko*.

På andre punkt har Aasen tilsynelatande gått sin eigen veg. Særleg i utlegginga av tyding 2 av *kalla* er Aasen noko lausare i uttrykksmåten enn Molbech, når bruksmåten blir utlagd. Om dette draget likevel skal knytast til Molbech, må det vere frå DDL, jf. tillegget om 'småhoste' under *knyste* (figur 7).

Frå *Dansk-norsk Ordbog av Ivar Aasen* (DNO) syner Bondevik og Myhren (2000:XII) kanskje endå klarare korleis Aasen lærte definisjonsordlegging direkte frå "Ordbogen". Oppslagsordet er *dræve*. I DO er forklaringa "tale meget langsomt", trække p. Ordene". I 1843-manuset til DNO har *dræve* fått forklaringa "drage paa Maale[t]". Det synest svært rimeleg at Aasen på bakgrunn av dei to ordleggingane til Molbech har funne fram til det idiomatisk norske motsvaret (trække/drage, p. Ordene/ paa Maale[t]).

Desse tilfeldige døma skulle samla sett vitne greitt om at Aasen verkeleg lét "Ordbogen selv lære" seg definisjonspraksisen, ikkje minst med vekt på "Klarhed og Korthed", som var Molbechs tilråding.

7. Sluttord

7.1. Konklusjon

7.1.1. Ytleg sett

Om oppgåva berre hadde vore å etterspore referansar til DO i ONF er resultatet

magert. Den einaste direkte referansen gjeld DO som kjelde for god dansk til ordforklaringane i ONF. Om vi også tek med DDL, viser Aasen i ONF direkte til denne ordboka i alle fall under oppslaga *dan* adj 'beskaffen, skikket', *soleie* f 'en Blomst' og *tveit* m (og f) 'en Spaan, Splint, Fliis'. Opplysningane i DDL støt forklaringane til Aasen om desse orda.

7.1.2. Etter nære gransking

Etter nærare gransking kan det greitt peikast på desse *fellespunkta* mellom DO og ONF frå ulike felt som konstituerer eit ordboksprosjekt:

- historisk-komparativ språkvitskapleg tenkjemåte (etymologi, analogi)
- internt syn på nasjonalt ordtilfang
- dokumentasjonskrav (etterprøving)
- skrifttype (fraktur)
- hovudprinsipp i alfabetisering og normering
- hovudprinsipp i ordforklaring

På desse felta har Aasen aktivt meir eller mindre direkte stødd seg på DO.

Skilnader mellom DO og ONF kjem til uttrykk i

- synet på materialgrunnlag for morsmålsordbøker
- statusen for norsk som nordisk språk
- alfabetisk verdi for teiknet for å-lyden
- skrivemåte for palatallyden framfor vokal (t.d. omtala av Bondevik og Myhren 2000:XI)
- heimfesting ('diatopisk markering' som "J." og "Tildeels i B. Stift" i figur 7)

På desse punkta, som delvis har synt seg som påviselege motsvar mellom DO og ONF, har Aasen gått sjølvstendig fram og på den måten vore med og utvikla nordisk leksikografi med å syne at det ikkje er nokon prinsipiell skilnad på skriftmålsleksikografi og dialektleksikografi.

Walton (1996:421) meiner jamvel at DDL fekk vel så mykje å seie og viser m.a. til at framstillingsmåten i DDL er mest samanfallande med ONF. Her tenkjer vel Walton på eit særmerke for DDL og ONF i høve til DO, nemleg heimfesting. For det einskilde ordet er det merkt av funnstad (med mange unntak i ONF) (jf. figur 7). Her skal det likevel leggjast til at å heimfeste ord og ordtydingar var Aasen van med frå Laurents Hallager: *Norsk Ordsamling eller Prøve af Norske Ord og Talemaader* (1802), som han også hadde med seg på den store innsamlingsferda (Venås 1996:28).

Etter Molbechs eigen vurdering var DO det rette valet (rett nok skrive før han hadde gjeve ut DDL): Han våga å kalle DO ”det bedste lexicalske Værk over Modersmaalet, som hidtil er fremkommet” (Molbech 1833:XXVI). Som den pålitelege dokumentaristen Molbech var, nemner han også kven han står i gjeld til. Han har gjort seg nytte av det gode, rette og brukelege tilfanget frå føregjengarane i dansk ordbokarbeid. Også Samuel Johnson (1709-84), som er sedd på opphavsmann til den moderne europeiske morsmålsordboka med *A Dictionary of the English Language* (1755) (jf. MacArthur 1986:96) er på lista over Molbechs kvalitetssikrarar. ”[M]in engelske Mester” er karakteristikken Johnson får. Aasen er på lag med dei beste.

7.2. 'The Empire Writes Back'

Med bakgrunn i den postkoloniale litteraturteorien vart det innleiingsvis peikt på det danske hegemoniet i skriftkulturen i Noreg jamvel etter 1814. Då høyrer det også med å avslutte med hjelp av forklaringskrafta i teorien. Eit hovudpoeng er at dei som har vore under hegemoniet, eller imperiet, svarar, eller som det endå meir talande heiter i den engelske opphavsformuleringa: ”the empire writes back” (Hauge 2001:376). Imperiet skal her forståast som periferien, og etterlikning (eng. *mimicry*) er det verknadsfulle uttrykksmidlet. Vidareutvikling må vere resultatet. I vår samanheng tok Ivar Aasen utgangspunkt i den hegemoniske *Dansk Ordbog*, tileigna seg akademikonseptet for morsmålsleksikografi, utvikla på det grunnlaget ein eigen norsk talemålsbasert morsmålsleksikografi i *Ordbog over det norske Folkesprog*, som han seinare perfektjonerte i *Norsk Ordbog med dansk Forklaring* (1873). Den endelege fullføringa av denne prosessen er *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* (1966-). Som tittelen seier, er dette ordboksverket syntesen av ONF og DO. Det vil seie at orda i talemålet og i skriftmålet får likeverdig behandling slik at heile ordtilfanget blir sett i samanheng.

Litteratur

- Bondevik, Jarle og Myhren, Magne 2000: Innleiing. I: Bondevik, J. et al. (red.): *Dansk-norsk Ordbog av Ivar Aasen*. Bergen: Norsk Bokreidingslag, s. IX-XXVI.
- Borup, Morten 1954: *Christian Molbech*. København: Rosenkilde & Bagger.
- DDL = Molbech 1841.
- DNO sjå Bondevik og Myhren 2000.
- DO = Molbech 1833.
- Duekilde, Anne 1998. Christian Molbech 1783-1857. I: Akselberg, G. og Bondevik, J. (red.): *Ord etter ord. Heidersskrift til Oddvar Nes*. Bergen: Norsk Bokreidingslag, 92-104.
- Hauge, Hans 2001: Postkolonialisme. I: Johannes Fibiger et al. (red.): *Litteraturens tilgange – metodiske angrepsvinkler*. København: Gads Forlag, 369-395.

- Haugen, Einar 1984: Introduction. I: Haugen, Eva: *A Bibliography of Scandinavian Dictionaries*. New York: Kraus International Publications, 1-61.
- Hovdhaugen, Even et al. 2000: *The History of Linguistics in the Nordic Countries*. Helsinki: Societas Scientiarum Fennica.
- Kruken, Kristoffer og Aarset, Terje 2000: *Ivar Aasen. Ordbog over det norske Folkesprog*. Ny utgåve. Oslo: Det Norske Samlaget.
- Kvaran, Guðrún 2000: Det attende århundredes islandske leksikograf. I: *LexicoNordica* 7, 75-90.
- MacArthur, Tom 1986: *Worlds of reference*. Cambridge: Cambridge University Press.
- Molbech, Christian 1833: *Dansk Ordbog indeholdende det danske Sprogs Stammeord, tilligemed afledede og sammensatte Ord, efter den nuværende Sprogbrug forklarede i deres forskellige Betydninger, og ved Talemaader og Exempler oplyste*. Kiøbenhavn: Gyldendalske Boghandlings Forlag.
- Molbech, Christian 1841: *Dansk Dialekt-Lexikon, indeholdende Ord, Udtryk og Talemaader af den danske Almues Tungemaal i Rigets forskellige Landskaber og Egne, forsaavidt som de ere fremmede for Skriftsproget og almindelig Sprogbrug, med Forklaring og Oplysninger*. Kiøbenhavn: Gyldendalske Boghandlings Forlag.
- NLO = Bergenholtz, H. et al. (red.) 1997: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- ODS = *Ordbog over det danske Sprog. Treogtyvende bind*. 1946. København: Gyldendalske Boghandel, Nordisk Forlag.
- ONF = Kruken og Aarset 2000.
- Sandøy, Helge 2000: *Lånte fjører eller bunad? Om importord i norsk*. Oslo: J.W. Cappelens akademisk Forlag.
- Skautrup, Peter 1953: *Det danske sprogs historie. 3*. København: Gyldendalske Boghandel, Nordisk Forlag.
- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker og ordboksarbete i teori och praktik*. Stockholm: Nordstedts Akademiska Förlag.
- Torp, Arne 2002: Skandinaviske "særbokstaver". I: *Språknytt* 1, 1-4.
- Torp, Arne og Vikør, Lars S. 2003: *Hovuddrag i norsk språkhistorie*. Oslo: Gyldendal Norsk Forlag.
- Venås, Kjell 1996: *Då tida var fullkomen. Ivar Aasen*. Oslo: Novus forlag.
- Vikør, Lars S. 2000: Ordbøker over norske bygdemål frå før 1814. I: *LexicoNordica* 7, 119-132.
- Walton, Stephen 1996: *Ivar Aasens kropp*. Oslo: Det Norske Samlaget.
- Worren, Dagfinn 2003: Ordartikkelen i nokre norske ordbøker i historisk perspektiv. I: Zakaris Svabo Hansen og Anfinnur Johansen (red.): *Nordiske studier i leksikografi 6. Rapport fra Konferanse om leksikografi i Norden. Tórshavn 21.-25. august 2001*. Tórshavn, 405-416.
- Aarset, Terje 1996: *Ivar Aasen. Det norske Folkesprogs Grammatik*. Volda: Høgskulen i Volda.
- Aasen 1848 sjå Aarset 1996.
- Aasen 1850 sjå Kruken og Aarset 2000.

Forfatterne

Bakken, Kristin

prosjektdirektør, f. 1959
Norsk Ordbok 2014, Universitetet i Oslo,
Postboks 1021 Blindern, NO-0315 Oslo
kristin.bakken@iln.uio.no

Braasch, Anna

seniorforsker, f. 1946
Center for Sprogteknologi (CST), Køben-
havns Universitet, Njalsgade 80, DK-2300
København S
anna@cst.dk

Brink, Lars

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

Djärv, Ulrika

doktorand, f. 1969
Institutionen för nordiska språk,
Stockholms universitet,
SE-106 91 Stockholm
ulrika.djarv@nordiska.su.se

Eek, Øystein

redaksjonssjef, f. 1952
Kunnskapsforlaget, Postboks 4432
Nydalen, NO-0403 Oslo
oystein.eek@kunnskapsforlaget.no

Farø, Ken

ph.d.-stipendiat, cand.mag., f. 1972
Københavns Universitet, Institut for
Engelsk, Germansk og Romansk,
Njalsgade 120, DK-2300 København S
kenfaroe@hum.ku.dk

Fjeld, Ruth Vatvedt

professor, f. 1948
Institutt for lingvistiske og nordiske
studier, Universitetet i Oslo,
Postboks 1001 Blindern, NO-0315 Oslo
r.e.v.fjeld@iln.uio.no

Gawinski, Birthe

lektor, cand.phil., ED og master i IT,
Sprog og Læring, f. 1945
Syddansk Universitet, Niels Bohrs Vej 9,
DK-6700 Esbjerg
birthe_gawinski@mac.com

Geeb, Franziskus

Prof. f. Computerlinguistik und Informa-
tionstechnologie, f. 1964
Hochschule für Angewandte Wissen-
schaften Hamburg, Fakultät Design,
Medien und Information,
Berliner Tor 5, D-20099 Hamburg
franziskus.geeb@bui.haw-hamburg.de

Granström, Alexandra

studerande, f. 1980
Kantorsgatan 32-100, SE-754 24 Uppsala
algr7156@student.uu.se

Gronostaj, Maria Toporowska

forskare, f. 1945
Institutionen för svenska språket,
Box 200, SE-405 30 Göteborg
maria.gronostaj@svenska.gu.se

Grønvik, Oddrun

forskar, hovudredaktør, f. 1947
Norsk Ordbok 2014, Universitetet i Oslo,
Postboks 1021 Blindern, NO-0315 Oslo
oddrun.gronvik@iln.uio.no

Gudiksen, Asgerd

lektor, f. 1955
Ømålsordbogen, Nordisk Forsknings-
institut, Afdeling for Dialektforskning,
Københavns Universitet, Njalsgade 136,
DK-2300 København S
Asgerd.Gudiksen@hum.ku.dk

Hallén, Karin

fil. dr., forskningsarkivarie, f. 1943
Språk- och folkminnesinstitutet, Dag

Hammarskjölds väg 19, Box 135,
SE-751 04 Uppsala
karin.hallen@sofi.se

Hallsteinsdóttir, Erla

dr.phil., f. 1970
Vægtens Kvarter 336,
DK-5220 Odense SØ
erlahall@yahoo.dk

Jacobsen, Bent Christian

ordbogsleder, mag.art., f. 1944
Ordbog over det norrøne prosasprog,
Københavns Universitet, Amager,
Njalsgade 136, DK-2300 København S
jacobsen@hum.ku.dk

Karlsen, Knut E.

forskar, redaktør, f. 1969
Norsk Ordbok 2014, Universitetet i Oslo,
Postboks 1021 Blindern, NO-0315 Oslo
k.e.karlsen@iln.uio.no

Kvaran, Guðrún

prof. dr.phil., f. 1943
Orðabók Háskólans, Neshaga 16,
IS-107 Reykjavík
gkvaran@lexis.hi.is

Larsson, Lennart

fil. dr, redaktör SAOB, univ.lektor, f. 1952
Institutionen för nordiska språk, Uppsala
universitet, Box 527, SE-751 20 Uppsala
Lennart.Larsson@nordiska.uu.se

Lenstrup, Rita

lektor, f. 1940
Handelshøjskolen i København, Dalgas
Have 15, DK-2000 Frederiksberg F
rl.eng@cbs.dk

Leroyer, Patrick

adjunkt, ph.d., f. 1955
Center for leksikografi, Handelshøjskolen
i Århus, Fuglesangs Allé 4,
DK-8210 Århus V
pl@asb.dk

Lorentzen, Henrik

seniorredaktør, f. 1960
Det Danske Sprog- og Litteraturselskab,
Christians Brygge 1,
DK-1219 København K
hl@dsl.dk

Lyly, Erika

språkvårdare, f. 1945
Språkrådet, Box 20057,
SE-104 60 Stockholm
Erika.Lyly@spraknamnden.se

Malmgren, Sven-Göran

professor, f. 1943
Institutionen för svenska språket,
Box 200, SE-405 30 Göteborg
malmgren@svenska.gu.se

Nimb, Sanni

postdoc, ph.d., f. 1961
Center for Sprogteknologi, Københavns
Universitet, Njalsgade 80,
DK-2300 København S
sanni@cst.dk

Ore, Christian-Emil Smith

førsteamanuensis, f. 1957
Eining for digital dokumentasjon, Univer-
sitetet i Oslo, Postboks 1123 Blindern,
NO-0317 Oslo
c.e.s.ore@edd.uio.no

Pedersen, Bolette Sandford

seniorforsker, f. 1962
Center for Sprogteknologi, Københavns
Universitet, Njalsgade 80,
DK-2300 København S
bolette@cst.dk

Pedersen, Karen Margrethe

lektor, dr.phil., f. 1949
Institut for Grænseregionsforskning,
Syddansk Universitet, Persillegade 6,
DK-6200 Aabenraa
kam@sam.sdu.dk

Rasmussen, Marie Bilde

forlagsredaktør, cand.mag., f. 1967
Gyldendal, Ordbogsafdelingen,
Klareboderne 3, DK-1001 København K
marie_bilde@gyldendal.dk

Sköldberg, Emma

fil. dr., f. 1968
Institutionen för svenska språket,
Göteborgs universitet,
Box 200, SE-405 30 Göteborg
emma.skoeldberg@svenska.gu.se

Thelin, Eva

fil. kand., forskningsarkivarie, f. 1966
Språk- och folkminnesinstitutet,
Dag Hammarskjölds väg 19, Box 135,
SE-751 04 Uppsala
eva.thelin@sofi.se

Trap-Jensen, Lars

ledende redaktør, f. 1960
Det Danske Sprog- og Litteraturselskab,
Christians Brygge 1,
DK-1219 København K
ltj@dsl.dk

Tvedt, Lars Jørgen

overingeniør, f. 1965
Eining for digital dokumentasjon, Univer-
sitetet i Oslo, Postboks 1123 Blindern,
NO-0317 Oslo
l.j.tvedt@edd.uio.no

Törnqvist, Lars

konsult, f. 1952
Älvkarleövägen 13, SE-11543 Stockholm
lars.tornqvist@telia.com

Vikør, Lars S.

professor, hovudredaktør, f. 1946
Norsk Ordbok 2014, Universitetet i Oslo,
Postboks 1021 Blindern, NO-0315 Oslo
l.s.vikor@iln.uio.no

Worren, Dagfinn

professor, redaktør, f. 1944
Norsk Ordbok 2014, Universitetet i Oslo,
Postboks 1021 Blindern, NO-0315 Oslo
dagfinn.worren@iln.uio.no

Stikord

A

aðkomaorð 200
administrasjon 148, 318
adverbier 283, 301
akademiprinsippet 395
akribi 133
allusion 50
almensproglige ordbøger 103
alternativa stavningsformer 379
användarperspektiv 121
apostrophitis dialektos 334
arbetsbesparing 218
argumentstrukturer 268
artikel/artikkel 135, 146
Aschehoug og Gyldendals store norske
leksikon 65
autentisk etterprøving 395

B

begrebsorienterede søgninger 259
belgssomr 182
beregningsregler 29
beskrivningsmetode 53
betydningsdifferentiering 228, 230, 232
betydningsresuméer 255
bilingvale ordbøger 86
Bokmålsordboka 66
brugergrænseflade/grensesnitt 25, 109, 135,
233, 260
bruklighetsmarkörer 267
bymål 194
bøjningsmæssig varians 86

C

bøjningsmønstre 29
citat 50
CSS 146

D

dansk 323
Den Danske Ordbog 12, 26, 241, 253, 284,
302, 329
database 22, 137, 318
databaserte metoder 22
databasestruktur 174
dataformat 349

dataintegritet 318
datalogivstikk 21
datamatisk leksikografi 302
datautveksling 316
definition 98, 155
definitionstillægg 363
deiktisk 225, 228
dekomponering 31
deskriptiv 267
diakron/diakronisk 53, 133, 191, 260
dialektleksikon 163
dialektologi 189
dialektord 21
dialektordbog/dialektordbok 160, 185, 188,
321, 335, 391
A Dictionary of Old Norse Prose 181
differentiae specifica 255
digitalisera/digitalisere 165, 253, 315
digitalisering 129
digitaliseringsprosess 21
digitaliseringsprosjekt 315
dikteringssystem 320
dimensjonering 132, 147
divergens 223
dokumentasjonsordbok 13, 130
dokumentation 110
DWS 143, 315, 348
dynamisk indretning 224

E

ekvivalens/ækvivalens 102, 176, 312, 399
elektroniska varianter 123
elektronisk bilingval ordbog 223
elektroniske parallellutgaver 14
elektronisk ordbog 26, 88, 103, 171, 224,
253, 355
encyklopædiske informasjoner 99
encyklopædiske opplysninger 153
encyklopædisk forklaring 99
endelse 29
etableret 48
etnolekt 334
etterprøve 317
etymologisk ordbok 21, 126
excerpering 165
excerpter 169

expandera 379

F

fejlformer 86
 felstavade ord 380
 fiktionmer 86
 film 116
 filmsekvenser 114
 finansiering 15
 fingeret symptomfunktion 40
 finsk 192, 209
 fjern/tilskriv 29
 Flensborg petuh 323
 forskingsdatabasar 316
 forskingsverktøy 19
 forudsigelig betydning 49
 framandord/framandorð/fremmedord 199,
 200, 396
 framandorðabók 200
 frase 49
 fraseografi 173
 fraseologi 41, 171
 fraseologikorpora 171
 frekvensoplysninger 86
 fremmedord/framandord/framandorð 199,
 200, 396
 fremmedordbog/framandorðabók 200
 fritekstsøgning 254
 fugeelement 30
 fuldtekstsøgning 114
 funktionalistisk ordbogskoncept 86
 funktionsbegrænsede
 kildehenvisningsløsninger 248
 funktionslæren 86
 funktionsmaksimerede
 kildehenvisningsløsninger 247
 funktionsminimerede
 kildehenvisningsløsninger 248
 førehandsvisning 148

G

gastronomisk terminologi 98
 gennemsigtig 49
 genusmæssig varians 86
 genus proximum 255, 303
 geografisk utbredning 168
 gjenbruk 317
 Google 68, 267, 276, 380
 gradsadverbier 306

grensesnitt/brugergrænseflade 25, 109, 135,
 233, 260
 grundlagsmateriale 147, 316
 Gundersen, Dag 84

H

heimfesting 404
 henvisninger/hänvisningar/tilvisingar 26,
 90, 133, 137, 147, 154, 166, 181, 205, 239,
 319, 362, 378, 386
 historisk dokumentasjonsordbok 19
 historiska ordböcker 53
 historisk princip 168
 homografi 278
 hypertextbaserede ordbøger 103
 hypertextstrukturen 98
 hänvisningar/henvisninger/tilvisingar 26,
 90, 133, 137, 147, 154, 166, 181, 205, 239,
 319, 362, 378, 386
 høflighedsfrase 49

I

idiom 49, 86, 171, 359, 360
 idiomatiske uttrykk 359
 idiometydelse 360
 idiomdefinitioner 362
 idiomkategorier 268
 idiomordbog 172
 illustrationer 86
 importord 131, 396
 indtastning 261
 informationsarkitektur 114
 informationshierarki 175
 informativ funktion 268
 initialalfabetisk 398
 innføringsbøger 383
 innskrivingsprogram 134
 interferens 193

J

junggrammatikken 189

K

kebabnorsk 196
 kildehenvisning 239
 kildeprog/kjeldespråk/källspråk 61, 103,
 232, 397
 kollokation 49, 268, 364
 komplementer 290

- komposisjonsfuge 280
 komposisjonstyper 281
 komposita 279
 konsekvens 85
 konsistens 147
 konstruksjonsoplysninger 286
 konstruksjonspotentiale 234
 korpus 13, 32, 130, 155, 176, 234, 242, 258,
 276, 285, 320
 korpusfunksjoner 258
 korpusundersøgelser 259
 korrektur 86
 kravspesifikasjon 347
 kronologiske markeringer 156
 krysstilvisingar 147
 krysstilvisingssystem 137
 kulturbundne leksemer 91
 kulturbundne ord 98, 103
 kulturel ækvivalent 98
 Kunnskapsforlaget 65
 kvalitetskontroll 139
 kvalitetssikring 317
 källspråk/kildesprog/kjeldespråk 61, 103,
 232, 397
- L**
 L2-produksjonsdel 86
 latinsk-svensk-finsk ordbok 209
 legitimering 17
 leksikalisme 27
 leksikalsk-semantisk ordnet 260
 leksikalske profiler 234
 leksikografering 392
 leksikografi/leksikografi 21, 57, 85, 110, 135,
 273, 315, 383, 397
 leksikografisk radikalisme 85
 leksikografisk/leksikografisk relevans 241,
 275
 leksikologi/leksikologi 359, 383, 387
 lemma 25, 61, 90, 109, 168, 223, 285, 359,
 378, 384
 lemmakandidater 133, 260
 lemmaseleksjon/lemmaseleksjon 111, 129,
 131, 158, 173, 285
 lemmatiser 32
 lemmatisering 32, 59, 109, 111
 leksikografi/leksikografi 21, 57, 85, 110, 135,
 273, 315, 383, 397
 leksikografisk/leksikografisk relevans 241, 275
- leksikologi/leksikologi 359, 383, 387
 linjedeling 146
 ljudskrift 165
 logging 148
 lukket kollokation 49
 lyd 320
 lærnerordbog 173
 låneord 98, 199
 låneoversættelser 201
- M**
 makrostrukturelle elementer 101
 makrostruktur 90, 102, 114, 218, 398
 man 223
 maskinomsetting 320
 materialeindsamling 152
 materialsortering 137
 materialvurdering 129
 mediostruktur 116
 Menota 182
 metaforisk idiom 49
 metaleksikografisk/metaleksikografisk 62,
 110, 173, 212, 240, 250
 Metaordboka 130, 145, 146, 148, 321
 mikrostruktur 58, 115, 168, 232, 350, 353,
 399
 mindretal 323
 mindretalssprog 323
 modifikationsmulighed 297
 moin-moin 331, 332
 mojñ 327
 multimedia 116
 museumsaspekt 17
 mutual information 257
 mønstertaking 403
 mådesadverbier 288, 305
 målførekjelde 186
 målføresamlinger 196
 målsprog/målspråk 97, 223, 228, 233, 397,
 399
- N**
 nasjonale ordboksprosjekta 11
 nasjonalromantikk 189
 nasjonsbygging 189
 Nationalencyklopedins ordbok (NEO) 126,
 166, 313, 360, 362, 377
 NEO 126, 166, 313, 360, 362, 377
 nordfrisisk 323

- nordnorsk 190
 nordslavisk 323
 norm 85, 266, 271, 273
 normaliserad form 168
 norrønt språk 181
 Norsk Ordbok 12, 80, 129, 143, 181, 185,
 261, 315, 405
 Norsk riksmålsordbok 12
 nul-ækvivalente ord 98
 Nusvensk ordbok 12
 nynorsk 17, 131
 Nyord i norsk 1945-75 66
- O**
 ODS 12, 160, 194, 253, 260
 omfangsdimensjonering 129
 omgrepsanalyse 399
 online-ordbog 97, 108, 245, 356
 onomasiologisk 172, 173, 176, 254, 259
 ontologi 107, 303, 306, 310
 ontologisk type 303
 opplæring 129
 ord 49
 ordbog/ordbok/orðabók
 almensproglige ordbøger 103
 bilingvale ordbøger 86
 Bokmålsordboka 66
 Den Danske Ordbog 12, 26, 241, 253,
 284, 302, 329
 dialektordbog/dialektordbok 160, 185,
 188, 321, 335, 391
 dokumentasjonsordbok 13, 130
 elektronisk bilingval ordbog 223
 elektronisk ordbog 26, 88, 103, 171, 224,
 253, 355
 etymologisk ordbok 21, 126
 framandorðabók 200
 fremmedordbog 200
 historiska ordböcker 53
 historisk dokumentasjonsordbok 19
 hypertextbaserede ordbøger 103
 idiomordbog 172
 latinsk-svensk-finsk ordbok 209
 lærnerordbog 173
 Metaordboka 130, 145, 146, 148, 321
 nasjonale ordboksprosjekta 11
 Nationalencyklopedins ordbok (NEO)
 126, 166, 313, 360, 362, 377
 Norsk Ordbok 12, 80, 129, 143, 181, 185,
 261, 315, 405
 Norsk riksmålsordbok 12
 Nusvensk ordbok 12
 online-ordbog 97, 108, 245, 356
 sprøgteknologisk 285, 301
 språkstadietordböcker 53
 Svenska Akademiens ordbok (SAOB) 12,
 54, 163, 220, 260, 389
 Tanums store rettskrivningsordbok 66,
 281
 tematisk ordbok 212
 tosprogsordbøger 86
 trykta ordböcker/trykte ordbøger 123,
 165, 173, 256, 378
 tökuorðabók 206
 uttalsordböcker 376
 vitskaplege dokumentasjonsordbøker 21
 Ordbog over det danske Sprog (ODS) 12,
 160, 194, 253, 260
 Ordbog over det norrøne prosasprog 181
 ordbogsfunktion 92, 175, 239, 349
 ordbogsgrammatik 348
 ordbogskomponent 25
 ordbogsredigering 347
 Ordbok över Sveriges dialekter (OSD) 163
 ordboksanvändning 121
 ordboksbase 275
 ordbokskunskap 124
 ordbokstextens auktoritet 220
 orddeling 146, 279
 ordnet.dk 13, 22, 172, 253
 ordslektskap 399
 ordsprog 44, 50
 orsaker till förändring 218
 ortografi 140, 218
 ortografivarians 86
 OSD 163
 over-/underbegrebsrelation 259
 overbegreb 303
 oversættelse 98, 115, 175, 200, 223, 243, 312
 oversættelselån 42, 98, 325, 338, 339
- P**
 paradigmatvang 86
 parafrazeækvivalent 98
 parsning 32, 378
 PDF 146, 355
 persondeiktisk 228
 pidginspråk 193

pladsøkonomi 258
 plattysk 323, 340
 positionsmuligheder 287
 pragmatik 86, 90, 177, 270
 pragmatisk funktion 269
 pragmatiska fraser 268
 pressmaterial 360
 problemsökning 123
 produktdesign 113
 produktutveckling 214
 progresjon 141
 prædikativkonstruktion 290
 pålydende betydning 49

Q

quizzet 116

R

radikalisme 85
 rammestruktur 114
 redaksjonell metode 129, 133
 redaktør 139
 redaktøroplæring 139
 redigering 133, 167, 316
 redigeringsprogram 131, 141
 redigeringsprogram 143, 145, 316, 347
 referansesystem 317
 regionalsprog 324
 relasjonell database 134, 144
 relevans 20
 resultatkrav 22
 Rietz, Johan Ernst 163
 riksmål 18, 196
 rituel ytring 49
 rod 28
 Ross, Hans 188
 russenorsk 193
 rutediagram 234
 rutine-ytring 49

S

samankopling 321
 samisk 192
 sammansætninger 276, 376
 samtidssprog 19
 SAOB 12, 54, 163, 220, 260, 389
 scanning/skanning 261, 315
 SDL 164
 selektionsrestriktioner 303

semantik(k) 27, 83, 93, 138, 177, 269, 302, 383
 semantisk variation 360
 SGML 318
 signal 39
 SIMPLE 281, 301, 303
 skandinavisk 282, 298, 301, 383
 skriftmål 17, 129, 185, 405
 SOFI 163
 SOL 164
 sorterar 138
 sorteringsverktøy 147
 sociolingvistikken 191
 spesialteikn/specialtegn 133, 146, 262
 sprogdebat/språkdebatt 326, 388
 sprogkontakt/språkkontakt 191, 196, 338
 sproglig kliché 50
 sprognationalisme 325
 Sprogpagten 324
 sprogpluralisme 337
 sprogproduktion 175
 sprogpurisme 200
 sprogreception 175
 sprogrensning 325
 sprogteknologisk 25, 259, 301
 sprogteknologisk leksikon 283, 298, 312
 sprogteknologisk ordbog 285, 301
 språkdebatt/sprogdebat 326, 388
 Språk- og folkminnesinstituttet (SOFI) 163
 språkkontakt/sprogkontakt 191, 196, 338
 språkleg dokumentasjonsverksemd 18
 språkstadieordböcker 53
 SQL 112
 statistikk(k) 26, 122, 149, 218, 257, 278, 348
 statistisk 108, 275, 278, 282
 stedsadverbier 286, 297, 312
 struktur 145, 317
 strukturell fortolkning 262
 supplementsbind/-band 12, 263
 Svenska Akademiens ordbok (SAOB) 12, 54, 163, 220, 260, 389
 svensklärare 121
 svensklärarstuderande 121
 Svenskt dialektlexikon 163
 Svenskt dialektlexikon (SDL) 164
 Svenskt ortnamnslexikon (SOL) 164
 sydslesvigsk 323
 symptom 39
 synkronisk 53

syntaks/syntax 21, 31, 45, 93, 190, 223, 269,
277, 283, 353
syntaktisk opmærkning 258
syntax/syntaks 21, 31, 45, 93, 190, 223, 269,
277, 283, 353
systematisk makrostruktur 102
særdialektala ord 166, 168
særskrivning 279
særskrivning 275
sætningskløvning 290
sætningskema 288
søkjeselement 138
sønderjysk 323

T

taksonomi 305
talbøger 373
talemål/talesprog/talspråk 156, 169, 185,
202, 258, 267, 395
talerattitudeadverbier 306
talesprog/talemål/talspråk 156, 169, 185,
202, 258, 267, 395
talsyntese/talsyntes 320, 374
talspråk /talemål/talesprog 156, 169, 185,
202, 258, 267, 395
talsyntes/talesyntese 320, 374
Tanums store rettskrivningsordbok 66, 281
tekstkorpus 22, 32, 110, 263
tekstproduktion 109
teleskopløsning 259
tematisk ordbok 212
tidsadverbier 306
tidsanpassning 214
tilgangsstruktur 116, 175, 384
tilvisingar/henvisninger/hänvisningar 26,
90, 133, 137, 147, 154, 166, 181, 205, 239,
319, 362, 378, 386
tosprogsordbøger 86
tospråksleksikograf 400
transparens 86
tredigram 136
tryckta ordböcker/trykte ordbøger 123,
165, 173, 256, 378
typologisering 53
tysk 89, 98, 115, 171, 215, 225, 282, 290,
323, 377, 396
tökuorð 199, 200
tökuorðabók 206

U

udvidet leksikografisk informationsbehov
241
uetableret 48
unikal 50
usability-studier 117
utbredningskartor 168
utgangsspråk 397
utrymmesbesparing 214
uttalslexikon 375
uttalsordbøger 376
uttalsuppgifter 165

V

V2-regelen 193
varians 86
varietet 168, 185, 266, 323
vestnordisk 181
video 117, 320
vitskapleg/vetenskaplig/videnskabelig/
vitenskapelig 78, 163, 317, 354
vitskapleg digitaliseringsprosjekt 17
vitskaplege dokumentasjonsordbøger 21
vitskapleg leksikografi 21

W

wordnet 260
www.ordnet.dk 13, 22, 172, 253
www.ordnett.no 65

X

xenofoner 377
XML 108, 146, 318, 356
XML-database 112
XML/XSLT 111, 356
XSL-FO 146
XSLT 146, 356

Æ

ækvivalens/ekvivalens 102, 176, 312, 399
ækvivalentvarians 86

Å

åben kollokation 49
Aasen, Ivar 188, 391